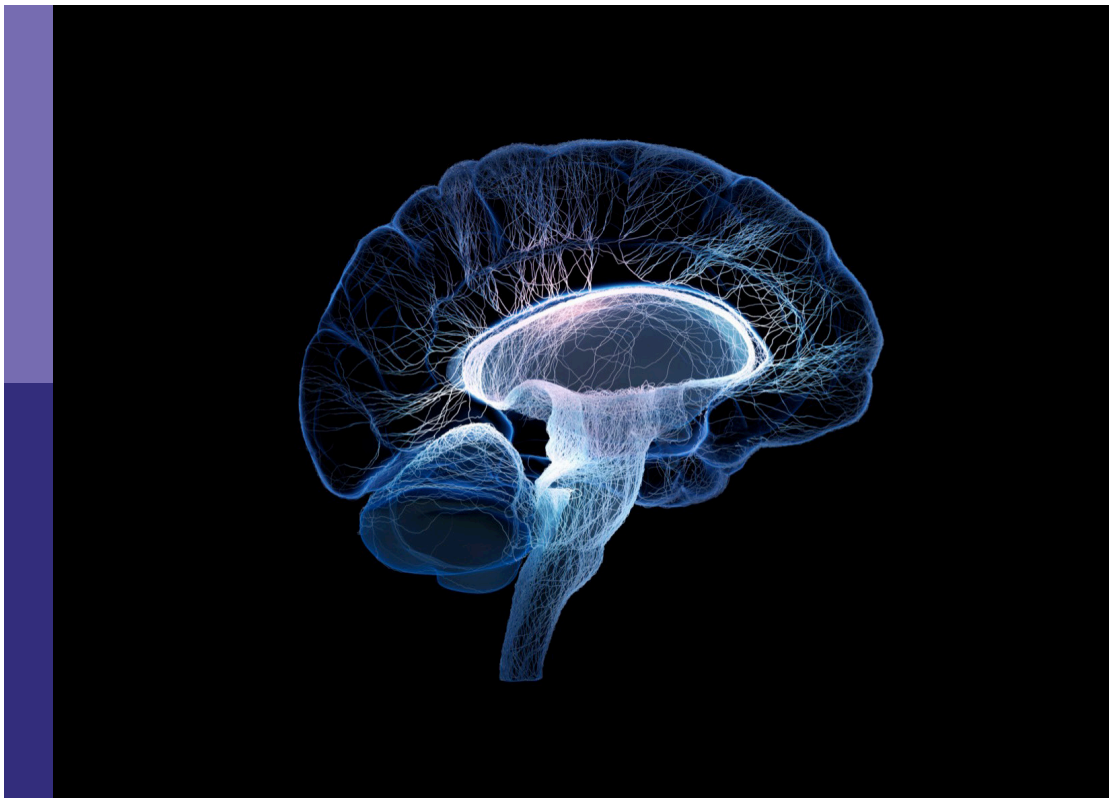
frontiers | Research Topics

Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders

Edited by
Hao Zhang, Da Ma and Lei Wang

Published in
Frontiers in Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-5550-7
DOI 10.3389/978-2-8325-5550-7

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

Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders

Topic editors

Hao Zhang — Central South University, China

Da Ma — Wake Forest University, United States

Lei Wang — Northwestern University, United States

Citation

Zhang, H., Ma, D., Wang, L., eds. (2024). *Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5550-7

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 01 October 2024
DOI 10.3389/fnins.2024.1497417

OPEN ACCESS

EDITED AND REVIEWED BY
Vince D. Calhoun,
Georgia State University, United States

*CORRESPONDENCE
Da Ma
✉ dma@wakehealth.edu
Hao Zhang
✉ hao@csu.edu.cn
Lei Wang
✉ Lei.Wang@osumc.edu

RECEIVED 17 September 2024
ACCEPTED 17 September 2024
PUBLISHED 01 October 2024

CITATION
Ma D, Zhang H and Wang L (2024) Editorial: Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders. *Front. Neurosci.* 18:1497417. doi: 10.3389/fnins.2024.1497417

Editorial: Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders

Da Ma[1]*, Hao Zhang[2]* and Lei Wang[3]*

[1]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, United States, [2]School of Electronic Information, Central South University, Changsha, Hunan, China, [3]Department of Psychiatry and Behavioral Health, Ohio State University Wexner Medical Center, Columbus, OH, United States

KEYWORDS

deep learning, artificial intelligence, brain imaging, neuroimaging, neurological disorder, psychiatric disorder

Editorial on the Research Topic
Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders

Introduction

Neuroimaging-based biomarkers have been used extensively for various neurological and psychiatric disorders, although accurate brain image-based diagnosis at the individual level remains elusive (Masdeu, 2011; Sui et al., 2020). In recent years, deep learning techniques have achieved remarkable success in fields such as computer vision and natural language processing, given their ability to learn complex patterns from large amounts of data (Zhang et al., 2020; Quaak et al., 2021). Applying deep learning to neuroimaging-assisted diagnosis, while promising, face challenges such as insufficiently labeled data, difficulty in interpretation, data heterogeneity, and multi-modal integration (Yan et al., 2022). This Research Topic highlights the development and application of cutting-edge deep learning research using neuroimaging for brain disorders, marking a collective effort to address these challenges.

The topics of the studies include differential diagnoses for brain tumors (Chen et al.; Zhang et al.) and dementia (Ma et al.) subtypes, early detection (Lang et al.; Huang et al.; Chattopadhyay et al.; Nie et al.; Liu et al.), and intervention (Yu and Fang) for neurological and neuropsychiatric disorders, as well as intracranial fluid segmentation (Puzio et al.). Various neuroimaging modalities were utilized, including structural magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), functional MRI (fMRI), Electroencephalogram (EEG), and computerized tomography (CT). A diverse range of advanced deep neural network architectures were developed and evaluated, including convolutional and graph neural networks (CNN, GNN), multi-modal neuroimaging feature fusion, vision transformers, and composited architectures.

Differential diagnosis, prognosis, and treatment response evaluation

Distinguishing different tumor types is fundamental for precision cancer treatment (Shoeibi et al., 2023; Wen et al., 2023). Chen et al. performed effective feature extraction of T1-weighted MRI by fusing multiple CNN models through pairwise feature summation, achieving an accurate classification performance of over 0.97. Zhang et al. employed a hybrid approach using EfficientNet-based feature extraction followed by a support vector machine (SVM), demonstrating comparable performance and identifying tumor regions with a Grad-CAM-based saliency map.

Identifying dementia subtypes is also crucial for personalized medicine for neurodegeneration (Ma et al., 2020; Chouliaras and O'Brien, 2023; Haller et al., 2023; Wen et al., 2023). Ma et al. introduced a multi-level, multi-type feature embedding and fusion approach to differentiate three heterogeneity clinical phenotypes of FTD: behavioral-variant (bvFTD), semantic-variant primary progressive aphasia (svPPA), and nonfluent-variant-PPA (nfvPPA), achieving a balanced accuracy of 0.84. The integrated-gradient-based explainable AI approach demonstrated more localized differential subtype patterns than groupwise statistical mapping.

Excessive accumulation of β-amyloid in the brain, a hallmark of Alzheimer's disease (AD) can be detected using PET (Jack et al., 2010; Tosun et al., 2021). Chattopadhyay et al. evaluated various machine-learning approaches to achieve this, including: (1) feature-engineered approaches, including logistic regression, XGBoost, and shallow artificial neural networks (ANN), (2) deep learning models with 2D/3D convolutional neural networks (CNN), (3) hybrid ANN-CNN models, (4) transfer learning on pretrained CNNs, and (5) Vision Transformers (MINiT). Validating a large-scale MRI/PET-paired dataset from 1,847 elderly participants, the hybrid ANN-CNN and 3D vision transformer achieved the best performance, reaching a balanced accuracy and an F1 score of around 0.8.

For neuropsychiatric disorder, Yu and Fang examined the effectiveness of exercise in Attention Deficit Hyperactivity Disorder (ADHD) patients by predicting diagnosis and intervention response through a composited approach. Random Forest was first used to select features from multi-source data. A Time Convolutional Network (TCN) was then applied to capture the behavioral and physiological signals related to motor activities over time. An Adaptive Control of Thought-Rational (ACT-R) model was used to simulate ADHD patients' cognitive processes, behavioral responses, and symptoms. Evaluation of multiple datasets demonstrated generalizable performance.

Brain network and EEG analysis

GNN has shown promising capability to analyze whole-brain connectivity to gain insight of neuropsychiatric disorders (Bessadok et al., 2023). Brain networks can be derived either from functional connectivity or structural connectivity derived from fMRI and DTI accordingly. Lang et al. introduced a novel GNN approach incorporating task-specific prior (TSP) knowledge to improve the characterization of the functional connectome patterns, demonstrating state-of-the-art performance in classifying different neuropsychiatric disorders, including ADHD, autism, and schizophrenia, as well as distinct task-specific connectivity patterns for various neuropsychiatric disorders. Huang et al. introduced a novel multi-layer brain network graph embedding to integrate multi-modal data. Complementary and unique information from structural and functional connectivity was captured through traversing nodes in each layer, with group differences computed at both the nodal and network levels, improving schizophrenia and bipolar disorder classification.

Nie et al. introduced a composited deep learning model on the electroencephalogram (EEG) data to capture the brain's electrophysiological signals for the early diagnosis of epilepsy. Fast Fourier Transform (FFT) extracted EEG signals were fed into a nested CNN-LSTM model, demonstrating state-of-the-art performance (accuracy/sensitivity/specificity = 0.96/0.93/0.96), exceeding state-of-the-art methods. Liu et al. introduced an attention-based multi-semantic dynamic graph convolutional network (AMD-GCN) to detect fatigue from EEG functional connectivity data. AMD-GCN integrates multiple modules, including channel-attention to assign weights to different input features, a multi-semantic dynamic graph convolution to capture node dependency, and a spatial-attention mechanism to remove redundant spatial node information, achieving the best classification performance on the SEED-VIG public dataset (0.90 accuracy) on fatigue detection.

Intracranial fluid segmentation in emergency settings

Image segmentation is a crucial step in clinical assessment of brain disease (Siddique et al., 2021). Puzio et al. conducted intracranial compartment (ICC) and cerebrospinal fluid (CSF) segmentation on emergency trauma head CT scans for triaging high-risk patients with traumatic brain injury for further neurosurgical treatment, achieving a dice similarity score of 0.765/0.567/0.574/556 for ICC, right/left supratentorial and infratentorial CSF regions. Comparison between automated and manual segmentation on CSF compartments demonstrated high inter-class correlation. The ICC to CSF ratio demonstrated clinical relevance in identifying patients who require surgical intervention.

Conclusions and discussions

This Research Topic presented a collection of the latest advancements in deep learning techniques on neuroimaging, demonstrating the effectiveness in diagnosing brain disorders such as neurodegeneration, neuropsychiatric symptoms, brain tumors, and traumatic brain injury. Despite these successes, challenges remain to be addressed to facilitate further clinical translation in biomedical and health applications. First, comprehensive evaluations on standard and diverse datasets will be critical for benchmarking model performance, ensuring generalizability and translatability. Second, beyond integrating multi-modal neuroimaging data, future studies would incorporate multidimensional data such as non-imaging biomarkers and

electronic health records (EHR). Finally, more advanced explainable AI approaches, such as counterfactual analysis to infer causal relationships and uncertainty measurements, are needed to ensure trustworthiness, human-in-the-loop, and successful adoption of AI models.

Author contributions

DM: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. HZ: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. LW: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. DM was supported by the Wake Forest Center for Artificial Intelligence Research Biomedical Informatics Pilot Award, Wake Forest Alzheimer's Disease Research Center Pilot Award, P30AG072947, and P30AG021332. LW received funding from R01 AG055121 and R56 AG055121.

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 VC declared a past coauthorship with the author LW.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Bessadok, A., Mahjoub, M. A., and Rekik, I. (2023). Graph neural networks in network neuroscience. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 5833–5848. doi: 10.1109/TPAMI.2022.3209686

Chouliaras, L., and O'Brien, J. T. (2023). The use of neuroimaging techniques in the early and differential diagnosis of dementia. *Mol. Psychiatry* 28, 4084–4097. doi: 10.1038/s41380-023-02215-8

Haller, S., Jäger, H. R., Vernooij, M. W., and Barkhof, F. (2023). Neuroimaging in dementia: more than typical Alzheimer disease. *Radiology* 308:e230173. doi: 10.1148/radiol.230173

Jack, C. R., Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L., Zeng, G., et al. (2010). Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain* 133, 3336–3348. doi: 10.1093/brain/awq277

Ma, D., Lu, D., Popuri, K., Wang, L., Beg, M. F., and Alzheimer's Disease Neuroimaging Initiative (2020). Differential diagnosis of frontotemporal dementia, Alzheimer's disease, and normal aging using a multi-scale multi-type feature generative adversarial deep neural network on structural magnetic resonance images. *Front. Neurosci.* 14:853. doi: 10.3389/fnins.2020.00853

Masdeu, J. C. (2011). Neuroimaging in psychiatric disorders. *Neurother. J. Am. Soc. Exp. Neurother.* 8, 93–102. doi: 10.1007/s13311-010-0006-0

Quaak, M., van de Mortel, L., Thomas, R. M., and van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. *NeuroImage Clin.* 30:102584. doi: 10.1016/j.nicl.2021.102584

Shoeibi, A., Khodatars, M., Jafari, M., Ghassemi, N., Moridian, P., Alizadehsani, R., et al. (2023). Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: a review. *Inf. Fusion* 93, 85–117. doi: 10.1016/j.inffus.2022.12.010

Siddique, N., Paheding, S., Elkin, C. P., and Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* 9, 82031–82057. doi: 10.1109/ACCESS.2021.3086020

Sui, J., Jiang, R., Bustillo, J., and Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biol. Psychiatry* 88, 818–828. doi: 10.1016/j.biopsych.2020.02.016

Tosun, D., Veitch, D., Aisen, P, Jack, C. R. Jr., Jagust, W. J., Petersen, R. C., et al. (2021). Detection of β-amyloid positivity in Alzheimer's disease neuroimaging initiative participants with demographics, cognition, MRI and plasma biomarkers. *Brain Commun.* 3:fcab008. doi: 10.1093/braincomms/fcab008

Wen, J., Varol, E., Yang, Z., Hwang, G., Dwyer, D., Kazerooni, A. F., et al. (2023). "Subtyping brain diseases from imaging data," in *Machine Learning for Brain Disorders*, ed. O. Colliot (New York, NY: Humana). Available at: http://www.ncbi.nlm.nih.gov/books/NBK597476/ (accessed September 15, 2024).

Yan, W., Qu, G., Hu, W., Abrol, A., Cai, B., Qiao, C., et al. (2022). Deep learning in neuroimaging: promises and challenges. *IEEE Signal Process. Mag.* 39, 87–98. doi: 10.1109/MSP.2021.3128348

Zhang, L., Wang, M., Liu, M., and Zhang, D. (2020). A survey on deep learning for neuroimaging-based brain disorder analysis. *Front. Neurosci.* 14:779. doi: 10.3389/fnins.2020.00779

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 29 September 2023
DOI 10.3389/fnins.2023.1269100

OPEN ACCESS

EDITED BY
Hao Zhang,
Central South University, China

REVIEWED BY
Dahua Yu,
Inner Mongolia University of Science and Technology, China
Yangding Li,
Hunan Normal University, China

*CORRESPONDENCE
Wenna Chen
✉ chenwenna0408@163.com

RECEIVED 29 July 2023
ACCEPTED 29 August 2023
PUBLISHED 29 September 2023

CITATION
Zhang J, Tan X, Chen W, Du G, Fu Q, Zhang H and Jiang H (2023) EFF_D_SVM: a robust multi-type brain tumor classification system.
Front. Neurosci. 17:1269100.
doi: 10.3389/fnins.2023.1269100

EFF_D_SVM: a robust multi-type brain tumor classification system

Jincan Zhang[1], Xinghua Tan[1], Wenna Chen[2]*, Ganqin Du[2], Qizhi Fu[2], Hongri Zhang[2] and Hongwei Jiang[2]

[1]College of Information Engineering, Henan University of Science and Technology, Luoyang, China, [2]The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China

Brain tumors are one of the most threatening diseases to human health. Accurate identification of the type of brain tumor is essential for patients and doctors. An automated brain tumor diagnosis system based on Magnetic Resonance Imaging (MRI) can help doctors to identify the type of tumor and reduce their workload, so it is vital to improve the performance of such systems. Due to the challenge of collecting sufficient data on brain tumors, utilizing pre-trained Convolutional Neural Network (CNN) models for brain tumors classification is a feasible approach. The study proposes a novel brain tumor classification system, called EFF_D_SVM, which is developed on the basic of pre-trained EfficientNetB0 model. Firstly, a new feature extraction module EFF_D was proposed, in which the classification layer of EfficientNetB0 was replaced with two dropout layers and two dense layers. Secondly, the EFF_D model was fine-tuned using Softmax, and then features of brain tumor images were extracted using the fine-tuned EFF_D. Finally, the features were classified using Support Vector Machine (SVM). In order to verify the effectiveness of the proposed brain tumor classification system, a series of comparative experiments were carried out. Moreover, to understand the extracted features of the brain tumor images, Grad-CAM technology was used to visualize the proposed model. Furthermore, cross-validation was conducted to verify the robustness of the proposed model. The evaluation metrics including accuracy, F1-score, recall, and precision were used to evaluate proposed system performance. The experimental results indicate that the proposed model is superior to other state-of-the-art models.

KEYWORDS

brain tumors, transfer learning, feature extraction, grad-CAM, robustness

1. Introduction

Brain tumors pose a serious threat to people's health and have a high fatality rate (Alyami et al., 2023). Early detection of brain tumors is crucial for patients, as they can get a greater chance of survival (Özbay and Altunbey Özbay, 2023). Medical imaging techniques have been widely used by radiologists. Among these techniques, Magnetic Resonance Imaging (MRI) is one of the most common techniques for diagnosing and evaluating brain tumors, which could provide rich brain tissue data (Gu and Li, 2021; Ayadi et al., 2022). However, the traditional MRI detection of brain tumors heavily relies on experienced doctors. Fatigue caused by prolonged working hours could affect doctor diagnosis, resulting in potential risks to patients. Therefore, it is necessary to develop an automated brain tumor classification computer-aided system to assist doctors in diagnosis (Nanda et al., 2023).

Brain tumors are commonly classified as either benign or malignant, with malignant tumors being further classified into three subtypes: glioma tumor, pituitary tumor, and meningioma tumor. Classifying brain tumors into multiple categories is more challenging than classifying them into two categories (Gu et al., 2021; Shahin et al., 2023).

Machine learning and deep learning are widely used in cancers study (Maurya et al., 2023). Typical ML classification methods encompass a series of steps: data preprocessing, feature extraction, feature selection, dimensionality reduction, and classification (Swati et al., 2019). Bi et al. (2021) and Saravanan et al. (2020) have both utilized machine learning to achieve the task of classifying skin cancers. Bi et al. (2021) utilized a combination of Support Vector Machine (SVM) and Chaotic World Cup Optimization (CWCO) optimization algorithms, whereas Saravanan et al. (2020) used SVM as a classifier and Gray-Level Co-Occurrence Matrix (GLCM) for feature extraction. Amin et al. (2020) employed SVM for brain tumors Classification. Feature extraction is a key step in achieving high performance in traditional machine learning. The accuracy of classification often depends on the features extracted with the help of experts. However, for most researchers, feature extraction is a challenging task when using traditional machine learning methods in research. The applications of machine learning and deep learning in disease classification are introduced in this paper.

In machine learning, it is necessary to perform feature extraction. Cheng et al. (2015) utilized three feature extraction techniques, namely intensity histogram, grey-scale co-occurrence matrix, and bag-of-words, achieving a model accuracy of 91.28%. Gumaei et al. (2019) employed a hybrid feature extraction approach to extract brain tumor images feature, which was combined with a regularized extreme learning machine for the classification of brain tumors, and an accuracy of 94.233% on the Chen dataset was achieved. Khan et al. (2019) used the watershed algorithm for image segmentation in a brain tumor classification system. The brain tumor classification system categorized tumors as either benign or malignant with an accuracy of 98.88%.

Since dataset features can be automatically extracted by deep learning techniques, they have got more and more attention (Bar et al., 2015). As a deep learning technique, Convolutional Neural Network (CNN) models have been widely used in the field of deep learning for tasks such as image classification, object detection, and face recognition. CNN models are mainly composed of convolutional layers, pooling layers, and fully connected layers. Convolutional layers use filters to perform convolution operations on input data and extract features of images. Pooling layers are used to downsample the features outputted by convolutional layers, reducing the number of features and parameters. The fully connected layer connects the output of the pooling layer to the final output layer for tasks such as classification or regression. Unlike traditional machine learning techniques, the CNN model can automatically learn useful features from images, eliminating the need for manual feature engineering, so it is an ideal choice for medical image processing (Yu et al., 2022; Maurya et al., 2023). Medical image datasets are generally small due to the difficulty and cost of acquisition. Therefore, as an effective small dataset processing technology, transfer learning has been widely applied in the field of medical image classification such as breast cancer, pneumonia, brain tumors, and glomerular disease (Yu et al., 2022). Talo et al. (2019) categorized brain tumors as benign or malignant using the pre-trained RestNet34. Kaur and Gandhi (2020) used pre-trained models such as Resnet50 and GoogLeNet ResnNet101 to classify brain tumors. Deepak and Ameer (2019) introduced a method using pre-trained GoogLeNet. Fine-tuned GoogLeNet was used to extract features of brain tumor images, and then SVM and KNN were employed as classifiers to complete the brain tumor classification task. EfficientNets, as lightweight models, are also extensively utilized in applications such as brain tumor classification (Tan and Le, 2019). Shah et al. (2022) used the EfficientNetB0 model to classify brain tumors as healthy and unhealthy. Nayak et al. (2022) utilized EfficientNetB0 to perform a triple classification of brain tumors, while Zulfiqar et al. (2023) utilized EfficientNetB2 for the same task. Yet, the model proposed in (Nayak et al., 2022) suffered from mild overfitting, resulting in low classification accuracy. And Zulfiqar et al. (2023) achieved a classification accuracy of only 91.35% when performing cross-validation experiments on different datasets. Additionally, Nayak et al. (2022) and Zulfiqar et al. (2023) only performed triple classification task of brain tumors.

Abiwinanda et al. (2019) created a model consisting of two convolution layers, an activation-Relu layer, and a Dense-64 layer. The model achieved an accuracy rate of 84.19%. Alanazi et al. (2022) constructed a 22-layer CNN architecture. The model was trained using a large-scale binary classification dataset, and then it was fine-tuned using a transfer learning approach. The accuracy of the model got 96.89 and 95.75% for Chen and Kaggle datasets, respectively. Kibriya et al. (2022) proposed a 13-layer CNN model and achieved 97.2 and 96.9% accuracy on Chen and Kaggle data sets. Jaspin and Selvan (2023) presented a 10-layer model using different optimizers (Adam and RMSprop) to train the model. On the Chen dataset, the accuracy of 96% was obtained using Adam and 95% was achieved using RMSprop. The studies by Swati et al. (2019) and Rehman et al. (2020) utilized the VGG19 and VGG16 models, respectively, and achieved accuracy rates of 94.82 and 98.69%. Sajjad et al. (2019) segmented the brain tumor region and used VGG19 for image classification, achieving an accuracy of 94.58%. Ghassemi et al. (2020) performed a brain tumor classification task based on a pre-trained Generative Adversarial Network (GAN) with an accuracy of 95.6%. Satyanarayana (2023) combined convolutional neural networks with a deep learning approach based on mass correlation and reported a classification accuracy of 94%. The proposed framework involved the construction of a multi-task CNN model and a 3D densely connected convolutional network. The authors combined the features extracted from a multi-task CNN and a 3D densely connected convolutional network to classify Alzheimer's disease.

Moreover, it has been proven that combining pre-trained models with machine learning is also a feasible method. Kang et al. (2021) used MobileNetV2 to extract features from brain MRI images, and adopted the SVM algorithm for classification, obtaining an accuracy of 91.58%. In reference (Sekhar et al., 2022), MobileNetV2 was used to extract features from brain tumor images. The extracted features were then classified using SVM and K-Nearest Neighbors (KNN). The best classification accuracy of 98.3% is achieved using KNN. Öksüz et al. (2022) utilized ResNet18 to extract both shallow and deep features from an enlarged Region of Interest (ROI) in brain tumors.

By integrating the shallow and deep features, a classification of the tumors was carried out using SVM and KNN classifiers. The results indicated an overall classification accuracy of 97.25% with the SVM classifier and 97.0% with the KNN classifier. Demir and

Akbulut (2022) proposed a new model, in which an R-CNN (Residual-CNN) structure was designed to extract features, using SVM as the classifier, with an accuracy of 96.6% being obtained. Deepak and Ameer (2023) used an additive loss function to train the CNN model, updating the model using different optimizers, then combined it with SVM and finally voted the classification results to derive the final classification result. The model obtained an accuracy of 95.6%. Muezzinoglu et al. (2023) built a new framework PatchResNet. Firstly, using a pre-trained ResNet50 to extract features from same-sized image blocks, feature selection was performed over Neighborhood Component Analysis (NCA), Chi2, and ReliefF. Secondly, the features were fed into the classifier KNN. Finally, majority voting was used to obtain the final prediction result with an accuracy rate of 98.1%.

Optimization algorithms have also been utilized to improve the performance of brain tumor classification systems. In reference (Kabir Anaraki et al., 2019), a Genetic Algorithm (GA) was used to optimize the CNN structure and achieved 94.2% accuracy. Kumar and Mankame (2020) combined the dolphin echolocation algorithm with the Sine Cosine Algorithm (SCA) to segment brain tumors from MRI and used the segmented images for brain tumor classification. Mehnatkesh et al. (2023) applied Improved Ant Colony Optimization (IACO) to optimize the super parameters of the ResNet architecture for brain tumor classification, achieving a classification accuracy rate of 98.694%.

The preceding discussion highlights the extensive adoption of deep learning as a prevalent technique for brain tumor classification. Nevertheless, the optimization of network structures using algorithmic approaches is time-intensive. Training the network from the ground up demands a substantial dataset and entails lengthier training compared to migration-based learning approaches. Furthermore, most of the prior studies have only employed a single dataset without conducting cross-dataset validation. However, our work utilized a pre-trained CNN model and incorporated regularization techniques to combat overfitting. The classification of brain tumors was successfully accomplished by the incorporation of machine learning techniques. Moreover, to verify the generalization performance of the proposed model, some experiments were carried out using two publicly available datasets while performing cross-data validation. And by adding Gaussian noise and salt-and-pepper (S&P) noise to the pictures of the brain tumor, the robustness of the model was further demonstrated.

We presented a novel feature extraction module based on EfficientNetB0 and employed SVM to categorize the resultant features. Specifically, we evaluated the model performance using both triple classification (glioma tumor, meningioma tumor, and pituitary tumor) and quadruple classification (glioma tumor, meningioma tumor, pituitary tumor, and healthy), providing comprehensive validation for our proposed model. In this paper, we presented an automated classification model of brain tumors, and the model was evaluated on two publicly available datasets (Chen and Kaggle). The model used a pre-trained EfficientNetB0 CNN model and combined dropout regularization and dense layers to construct a new feature extraction module EFF_D. The highest classification performance was achieved using the SVM classifier. The main research contributions of this study are as follows:

1. A new model is proposed for brain tumor classification.
2. Based on two public datasets, the proposed model has been proven to be a reliable method for brain tumor classification.
3. By using the last convolution layer of the Grad-CAM visualization model, a localized heat map was obtained, highlighting the brain tumor region.
4. The proposed model can classify brain tumors better than the available models. And the cross-data validation of the model achieves better result.

2. Materials and methods

This section focuses on our proposed approach. The base model used in this method is the pre-trained EfficientNetB0. Firstly, Relevant dropout and dense layers were introduced to construct a new model. Secondly, optimal hyperparameters were utilized to train the new model. Finally, the trained model was subsequently used to extract intricate image features, which were then classified utilizing the SVM algorithm. This approach is helpful in achieving better results in brain tumor classification tasks.

2.1. Introduction to the EfficientNetB0

EfficientNets is a series of convolutional neural network architectures developed by the Google team, making creative use of compound scaling. Of these, EfficientNetB0, as the base model, primarily consists of 16 mobile inverted bottleneck convolution (MBConv) modules (Tan and Le, 2019). In addition, the EfficientNetB0 architecture was utilized to perform 1,000 image classifications on the ImageNet dataset. According to the TensorFlow website[1], input images for the model should be represented as floating-point tensors with three color channels and pixel values ranging from 0 to 255.

2.2. Datasets and preprocessing

The experiments were performed on two publicly available brain tumor datasets. The Chen dataset is the CE-MRI dataset shared by Cheng et al. (2015), which consists of 3,064 brain MRI images from 233 patients, including three types of brain tumors, namely glioma, meningioma, and pituitary tumors. The number of images of the three types of brain tumors in the dataset is 1,426, 708, and 930. The Kaggle dataset was obtained from Kaggle (Bhuvaji et al., 2020), which is comprised of 3,264 images including four categories: glioma, meningioma, pituitary, and healthy. The number of images of the four categories in the Kaggle dataset is 926, 937, 901, and 500.

The image of the Chen dataset has a size of 512×512 and is a grayscale image. Therefore, the image of Chen needs to be resized to $224 \times 224 \times 3$. The image sizes in the Kaggle dataset are inconsistent,

1 tf.keras.applications.efficientnet.EfficientNetB0 | TensorFlow v2.12.0 (google.cn)

with some grayscale images and some RGB images. Similarly, the images should be adjusted to a uniform size of 224 × 224 × 3. In this paper, the data is randomly divided into non-overlapping training and test sets. The training set comprises 80% of the total dataset, while the remaining 20% is allocated to the test set.

2.3. Classification system

Both the datasets employed in this study, the Chen dataset, which has a total of 3,064 photos, and the Kaggle dataset, which has a total of 3,261 images—are tiny, making migration learning an effective method. The method of transfer learning is frequently used to train neural networks on a small dataset. In general, the process of training neural networks requires large dataset, but the number of brain tumor samples available is limited (Shin et al., 2016; Swati et al., 2019; Yu et al., 2022). Transfer learning offers an effective remedy for small sample size issues by enabling a transfer of knowledge from relevant tasks to new ones. Moreover, application of trained weights enhances both the efficiency and accuracy of models.

The overall architecture and method proposed in this paper are shown in Figure 1. The framework of the proposed brain tumor classification system is shown in Figure 1A. The dataset is divided into a training set and a test set, and they do not cross each other. The proposed model was trained on the training set, and the resulting trained model was saved to disk. The saved model was applied to classify the test set, and its performance was evaluated. As shown in Figure 1B, EfficientNetB0 is utilized as the foundation of our model. Table 1 describes the detailed parameters of the proposed model. The EfficientNetB0 model achieved high accuracy in classification tasks and was pre-trained on the large-scale ImageNet dataset (Tan and Le, 2019). As the dataset used in this experiment differs from the ImageNet dataset, the classification layer of the pre-trained model was removed. Then, we added two layers of Dropout to prevent overfitting, as well as two layers of Dense and one layer of Dense+Softmax to enable the model to classify our target images. The dropout ratios are 0.345 and 0.183, respectively, and the number of neurons in the Dense layer are both 69. The number of neurons in the Classification layer are either 3 or 4. When using an SVM as a classifier, the features extracted from the last Dropout layer can be used for SVM classification. The feature extraction module is called EFF_D, where the method using the Softmax classifier is called EFF_D_Softmax and the method using the SVM is called EFF_D_SVM.

2.4. Training CNNs

The training of a convolutional neural network combines forward and backward propagation. It starts at the input layer and is propagated forward. Then, the loss is back propagated to the first layer. In layer l, the i-th neuron receives the input from neuron j in layer l-1 through a computation process. Training samples xj are weighted by Eq. 1.

$$In = \sum_{j=1}^{n} W_{ij}^{l} x_j + b_i \tag{1}$$

where, W_{ij}^{l} represents weights, b_i denotes bias. After computing the weighted sum of the variables (In), the resulting values are processed through the activation functions: Swish and Relu, as represented by Eqs 2, 3, respectively.

$$S_i^l = In_i^l \times sigmoid\left(\beta In_i^l\right) \tag{2}$$

$$R_i^l = \max\left(0, In_i^l\right) \tag{3}$$

here, S_i^l is the output using Swish, and β is a constant. R_i^l is the output using Relu. The neurons in both the convolutional and fully connected layers are calculated using Eqs 1, 2 (or 3). The classification layer is calculated using the Softmax function which is shown as Eq. 4.

$$y_i = \frac{\exp\left(x_i\right)}{\sum_j^k \exp\left(x_j\right)} \tag{4}$$

where, K is the number of categories, x_i is the i-th element of the input vector x, and y_i is the i-th element of the output vector y.

The cross-entropy loss function evaluates the prediction error of the model by comparing the predicted probability distribution generated by the model with the distribution of the true labels, as represented by Eq. 5. This loss function is utilized in the backpropagation process to optimize the model's parameters and enhance the accuracy of the prediction results.

$$L = \frac{-1}{m} \sum_{i}^{m} \ln\left(p\left(\frac{y_i}{x_i}\right)\right) \tag{5}$$

here, m represents the total number of samples, x_i indicates the training sample indexed i, y_i represents the corresponding label of x_i, and P denotes the probability that x_i belongs to class y_i.

The model weights are updated according to Eq. 6.

$$\gamma^t = \gamma^{\left\lfloor \frac{tN}{m} \right\rfloor}$$

$$V^{t+1} = \mu V_l^t - \gamma^t \alpha_l \frac{\partial C}{\partial W} \tag{6}$$

$$W_i^{t+1} = W_l^t + V_t^{t+1}$$

where, α_l, γ^t and μ represent different factors affecting the current iteration of the learning algorithm. α_l corresponds to the learning rate at layer l. γ^t represents the scheduling rate which reduces the initial learning rate and μ is used to describe the influence of previously updated weights on the current iteration.

3. Results and discussion

The experiments were performed in Win11 operating system with 16 G RAM and RTX3060 graphics card of 6 G video memory.

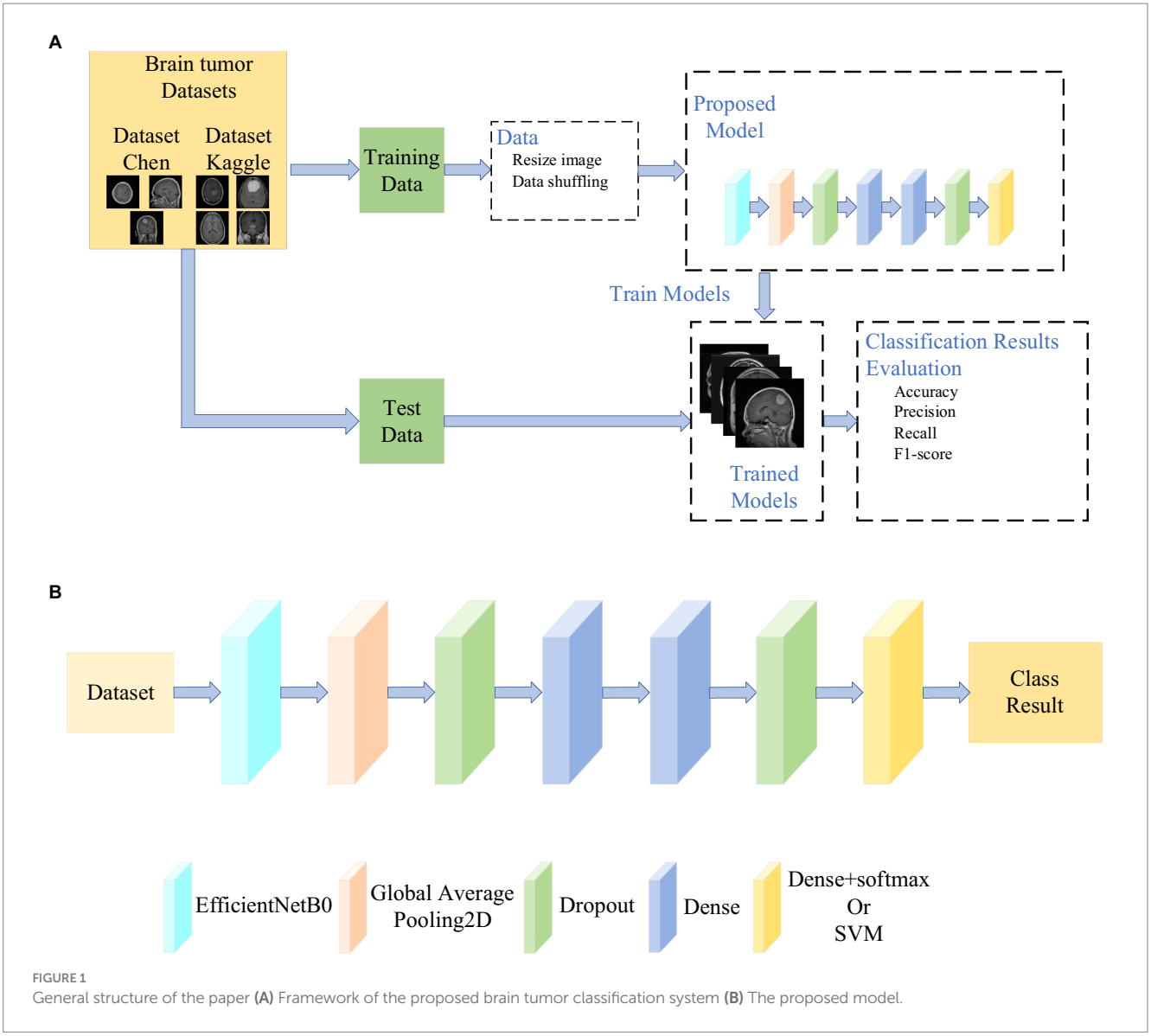


FIGURE 1
General structure of the paper **(A)** Framework of the proposed brain tumor classification system **(B)** The proposed model.

TABLE 1 Parameters of the proposed model.

Model	Parameters	Setting
EFF_D_Softmax	Dropout_1	0.345
	Dense_1	69
	Dense_2	69
	Dropout_2	0.183
	optimizer	Adam
	Learning rate	0.001
	Batch size	16
	Loss function	Cross entropy
	epoch	25
EFF_D_SVM	C	1
	kernel	linear
	probability	True

3.1. Performance evaluation

The dataset exhibits an imbalance, thus, it is insufficient to only accuracy is used to quantify model performance. Except for accuracy, precision, recall, and F1-score metrics are also utilized to evaluate the model performance (Alsaggaf et al., 2020). The calculation formulas for these metrics are expressed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

TABLE 2 Comparison of benchmark models.

Model	Training time (seconds)	Inference time (seconds)	Parameters (million)	Test accuracy(%)
VGG19	542	7	20.03	87.09
ResNet50	397	4	23.59	91.18
DenseNet121	524	3	7.04	96.57
EfficientNetB0	500	4	4.05	98.37

$$\text{F1} - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

where, TP (True Positive) is the number of correct positive predictions, TN (True Negative) is the number of correct negative predictions, FP (False Positive) is the number of wrong positive predictions, and FN (False Negative) is the number of wrong negative predictions.

3.2. The selection of the benchmark model

The paper conducts a comparative analysis of VGG19, ResNet50, DenseNet121, and EfficientNetB0 models in relation to training time, inference time, total parameters, and test set accuracy. Each model is compared using the Chen dataset, which has seen extensive used. Inference time represents the time required to predict 612 images from the test set. The outcomes of the experiments are presented in Table 2. Although fine-tuning EfficientNetB0 takes relatively more time, its inference time is also faster. Furthermore, EfficientNetB0 has the highest classification accuracy. Therefore, EfficientNetB0 is chosen as the benchmark model.

3.3. Experimental results

In order to further verify the effectiveness of the proposed model, a series of comparison models were also designed in this article. Initially, the neuron count in EfficientNetB0's classification layer is aligned with the number of categories in the dataset used for classification. The model is then subjected to fine-tuning. The model employing the Softmax classifier is referred to as EFF_Softmax, while the one employing the SVM classifier is labeled as EFF_SVM.

The training steps of the proposed EFF_D_SVM model are as follows:

Step 1: Importing the data and resizing the images to split the data into a training set and a test set.

Step 2: Loading the model and pre-trained weights, removing the Top layer, and adding the Dropout and Dense layers.

Step 3: Training EFF_D_Softmax to classify brain tumor images.

Step 4: Using EFF_D to extract features and using SVM to classify brain tumors.

Similarly, the same steps are adopted to train EFF_Softmax and EFF_SVM.

The experiments were performed using two datasets. The dataset Chen was used for testing 612 images consisting of 285 glioma tumor images, 141 meningioma tumor images and 186 pituitary tumor images. The Kaggle was used for testing 652 images including 185 glioma tumor images, 187 meningioma tumor images, 180 pituitary tumor image, and 100 no-tumor images.

Figure 2 shows the training results of the EFF_D_Softmax model and the EFF_Softmax model on the training sets of both datasets. Images in Figures 2A,B depict the training results obtained from the Chen dataset, while images in Figures 2C,D represent the training outcomes achieved using the Kaggle dataset. In relation to the Chen dataset, the EFF_D_Softmax model demonstrates an accuracy of 100 and 99.59% on the training and validation sets, respectively. Similarly, the EFF_Softmax model achieves accuracies of 100 and 99.18% on the training and validation sets, respectively. For the Kaggle dataset, the EFF_D_Softmax model achieves 99.93 and 98.21% accuracy on the training and validation sets, respectively. Similarly, the EFF_Softmax model achieves 100 and 98.51% accuracy on the training and validation sets, respectively.

The confusion matrixes for the classification results of the proposed method are shown in Figures 3, 4. Eqs 6–9 are utilized to calculate the detailed values of the model classification results from the confusion matrixes. The labels G, M, P, and NO represent different types of brain tumors: G for glioma, M for meningioma, P for pituitary tumor, and NO for the absence of a tumor. The obtained model metrics on the Chen and Kaggle are listed in Tables 3, 4, respectively. Moreover, to visually show the superiority of the adopted EFF_D_SVM model, the average metrics for classification results on the Chen dataset and Kaggle dataset are shown in Figures 5A,B, respectively. On the Chen, EFF_D_SVM showed the best classification results. On the Kaggle, as can be seen from Figure 5B, EFF_D_SVM outperformed the other models in terms of accuracy, f1-score and precision, but its recall rate was lower than that of EFF_SVM. Through the comparison in Table 4, we can see that the recall rate of EFF_D_SVM was higher than that of EFF_SVM for glioma, meningioma, and pituitary, and slightly lower than that of the EFF_SVM for no tumor. In a comprehensive analysis, the classification ability of EFF_D_SVM is still better than that of EFF_SVM. The Softmax classifier constantly strives for higher probabilities for correct classifications and lower probabilities for incorrect classifications, aiming to minimize the loss value. In contrast, the SVM classifier only needs to satisfy the boundary value and does not need to perform subtle manipulations on the concrete scores. Consequently, the Softmax classifier exhibits overfitting in brain tumor classification. Typically, Softmax is employed for large datasets, while SVM is suited for smaller datasets. In this paper, a small dataset is used, which could also contribute to the favorable performance of SVM classification.

On one hand, the model's fitting ability pertains to its capacity to accurately capture patterns and relationships within the training data. On the other hand, generalizability encompasses the model's capability to perform with data which has not encountered previously. When too much emphasis is placed on the model's ability to fit, the model may

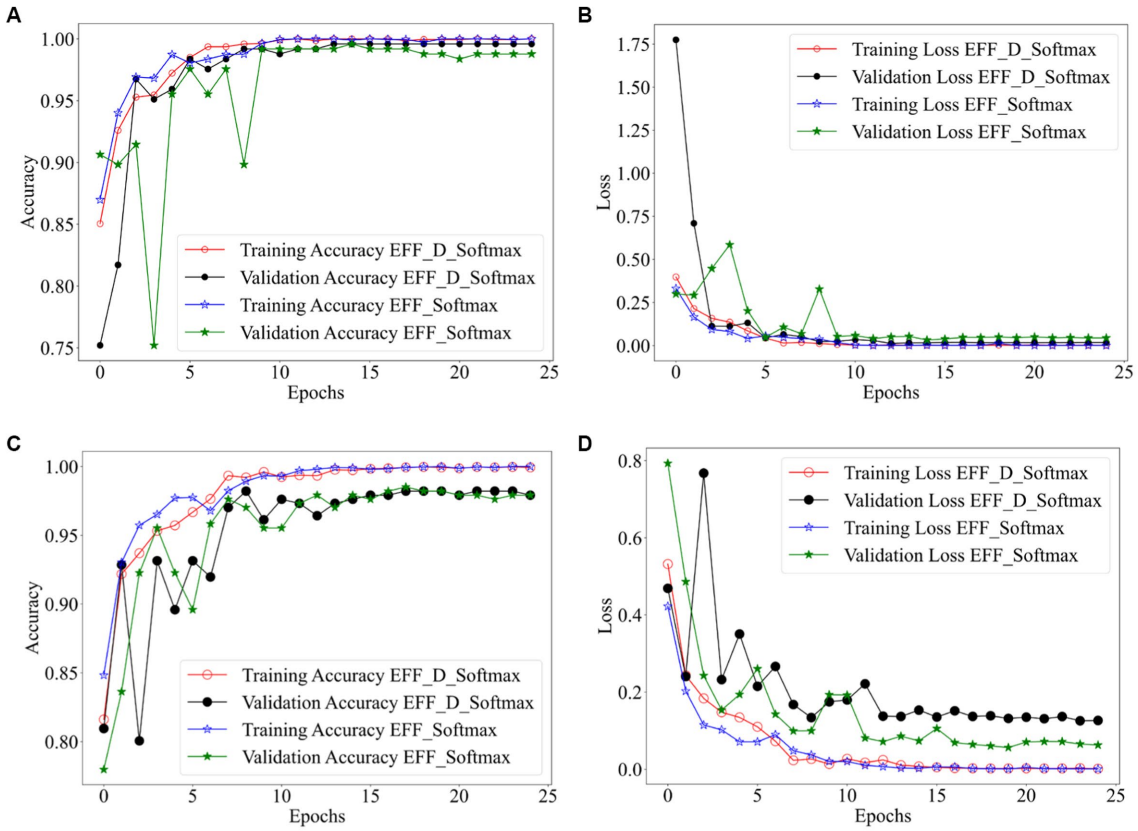


FIGURE 2
Training results for EFF_D_Softmax and EFF_SoftMax, **(A)** accuracy curve (Chen dataset) **(B)** loss curve (Chen dataset) **(C)** accuracy curve (Kaggle dataset) **(D)** loss curve (Kaggle dataset).

TABLE 3 Detailed metrics values of the proposed model on the Chen dataset.

Proposed model	Tumor type	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
EFF_D_Softmax	Glioma	98.61	99.30	98.95	98.37
	Meningioma	96.45	96.45	96.45	
	Pituitary	99.46	98.39	98.92	
	Average	**98.17**	**98.07**	**98.11**	
EFF_D_SVM	Glioma	98.95	99.30	99.12	98.86
	Meningioma	97.20	98.58	97.89	
	Pituitary	1.00	98.39	99.19	
	Average	**98.72**	**98.76**	**98.73**	
EFF_Softmax	Glioma	98.60	98.60	98.60	98.04
	Meningioma	94.48	97.16	95.80	
	Pituitary	1.00	97.85	98.91	
	Average	**97.69**	**97.87**	**97.77**	
EFF_SVM	Glioma	99.29	98.60	98.94	98.69
	Meningioma	95.86	98.58	97.20	
	Pituitary	1.00	98.92	99.46	
	Average	**98.38**	**98.70**	**98.53**	

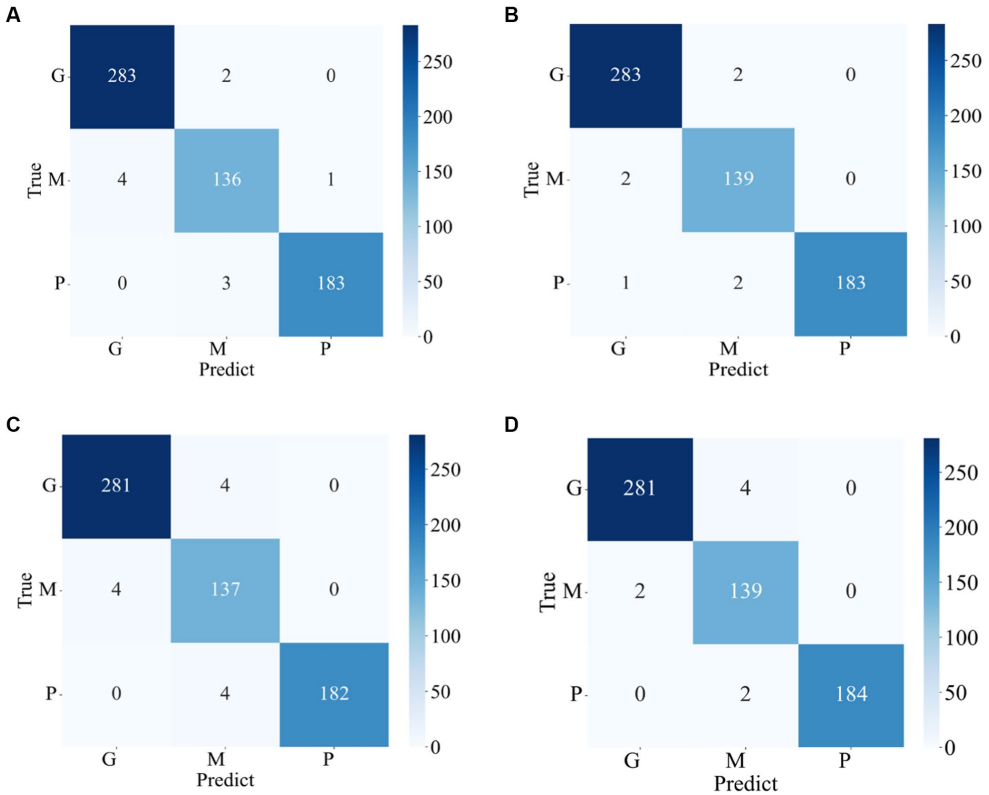


FIGURE 3
Confusion matrix of the proposed model in the Chen **(A)** EFF_D_Softmax **(B)** EFF_D_SVM **(C)** EFF_Softmax **(D)** EFF_SVM.

TABLE 4 **Detailed metrics values of the proposed model on the Kaggle dataset.**

Proposed model	Tumor type	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
EFF_D_Softmax	Glioma	96.83	98.92	97.86	97.85
	Meningioma	98.88	94.65	96.72	
	No Tumor	98.02	99	98.51	
	Pituitary	97.81	99.44	98.62	
	Average	97.89	98	97.93	
	Meningioma	98.88	94.65	96.72	
	No Tumor	98.02	99	98.51	
	Pituitary	97.81	99.44	98.62	
	Average	**97.89**	**98**	**97.93**	
EFF_D_SVM	Glioma	97.86	98.92	98.39	98.31
	Meningioma	97.33	97.33	97.33	
	No Tumor	1	97	98.48	
	Pituitary	98.9	99.44	99.17	
	Meningioma	97.33	97.33	97.33	
	No Tumor	1	97	98.48	
	Pituitary	98.9	99.44	99.17	
	Average	**98.52**	**98.17**	**98.34**	
EFF_Softmax	Glioma	95.31	98.92	97.08	97.55
	Meningioma	98.31	93.05	95.6	
	No Tumor	96.15	1	98.04	

(Continued)

TABLE 4 (Continued)

Proposed model	Tumor type	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
	Pituitary	1	99.44	99.72	
	Average	**97.88**	**98**	**97.93**	
EFF_SVM	Glioma	98.37	97.84	98.1	98
	Meningioma	97.3	96.26	96.77	
	No Tumor	95.24	1	97.56	
	Pituitary	1	98.89	99.44	
	Average	**97.73**	**98.25**	**97.97**	

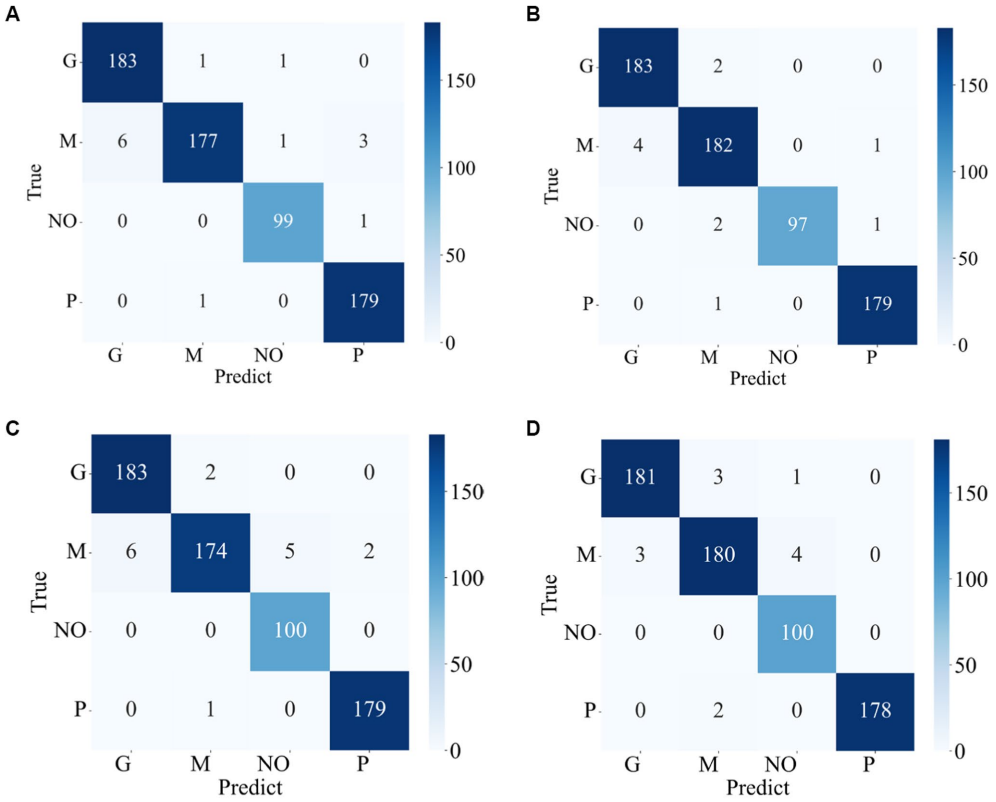


FIGURE 4
Confusion matrix of the proposed model in the Kaggle **(A)** EFF_D_Softmax **(B)** EFF_D_SVM **(C)** EFF_Softmax **(D)** EFF_SVM.

overfit on the training data set and underperform on new data not seen before. Moreover, as can be observed from Figures 2, the EFF_D_Softmax and EFF_Softmax fit well on the training sets of both datasets. However, model validation on test sets for both datasets found that the EFF_D_Softmax outperformed the EFF_Softmax. Therefore, EFF_D_Softmax has better anti-fitting and generalization ability.

The Receiver Operating Characteristic (ROC) curve offers an effective tool to assess the model classification ability by the relationship curve between the false positive rate and the true positive rate. The Area Under the Curve (AUC) provides essential information about the ability of the proposed model to differentiate between tumor types. The classifier performance is better if the AUC value is higher. The ROC curves of EFF_D_SVM for Chen and Kaggle are depicted in Figures 7A,B, respectively. These curves, which are very close to the upper-left corner, indicate that the EFF_D_SVM model has excellent classification ability. In the Chen dataset, the AUC values of EFF_D_SVM for glioma, meningioma, and pituitary are 0.9994, 0.9998, and 0.9996, respectively. And in the Kaggle dataset, the AUC values of EFF_D_SVM for glioma, meningioma, pituitary adenoma and tumor-free are 0.9937, 0.9964, 0.9999 and 0.9999, respectively.

The classification results obtained by our proposed model are compared with those obtained by previous state-of-the-art models that used the same dataset, as shown in Table 5. It can be observed that the proposed model outperforms the available state-of-the-art methods, both on Chen and Kaggle datasets. In particular, the accuracy of our proposed EFF_D_SVM model achieve 98.86 and 98.31% on Chen and Kaggle, respectively.

To understand the model's area of interest for a category, we visualized it using the Grad-CAM (Selvaraju et al., 2020) technique. This technique can help us to understand how the model

TABLE 5 Comparison of our proposed model with previous models.

Reference	Dataset	Method	Accuracy(%)	F1-score(%)	Precision(%)	Recall (%)
Swati et al. (2019)	Chen	Fine-tuned VGG19	94.82	91.73	89.52	94.25
Sekhar et al. (2022)	Chen	GoogleNet+KNN	98.3	97.24	97.24	97.23
Öksüz et al. (2022)	Chen	ResNet18 + ShallowNet+SVM	97.25	95.26	95.25	95.27
Satyanarayana (2023)	Chen	DCNN-MCN	94	–	–	–
Deepak and Ameer (2023)	Chen	Majority voting	95.6	–	–	–
Jaspin and Selvan (2023)	Chen	MCCNN	95.17	95	96	95
Mehnatkesh et al. (2023)	Chen	Optimizing ResNet	98.694	98.458	98.53	98.40
Kang et al. (2021)	Kaggle	MobileNetV2 + SVM	98.16	–	–	–
Muezzinoglu et al. (2023)	Kaggle	PatchResNet	98.1	98.1	97.91	98.15
Alanazi et al. (2022)	Chen	22-layer-CNN	96.89	–	–	–
	Kaggle		95.75	–	–	–
Kibriya et al. (2022)	Chen	13-layer CNN	97.2	-	97	96
	Kaggle		96.9	–	–	–
Proposed model	Chen	EFF_D_SVM	98.86	98.73	98.76	98.72
	Kaggle		98.31	98.34	98.52	98.17

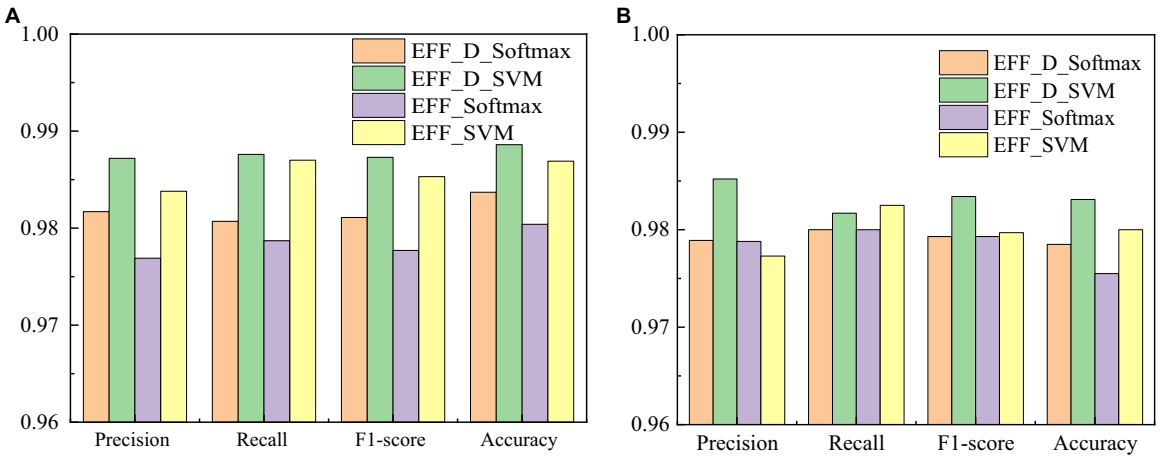


FIGURE 5
Average metrics for classification results **(A)** Chen **(B)** Kaggle.

distinguishes different types of brain tumors. In this paper, Grad-CAM is used to create a class activation heat map. The contribution of a specific part in differentiating between different brain tumors is directly proportional to the darkness of its corresponding color. Figure 6 shows a visual depiction of EFF_D_SVM for brain tumor image categorization using Grad-CAM. The heat map produced by Grad-CAM is displayed in Figure 6B, while Figure 6C exhibits the outcome of superimposing the heat map onto the original image. Figure 6C visually demonstrates the application of the grad-cam technique, where the area of the brain tumor is highlighted in red. This indicates that the tumor region serves as a prominent feature in differentiating brain tumors, although the surrounding area is also included.

3.4. Cross-dataset validation and robustness validation

To further demonstrate the robustness of our proposed model, cross-validating experiment on multiple datasets was also carried

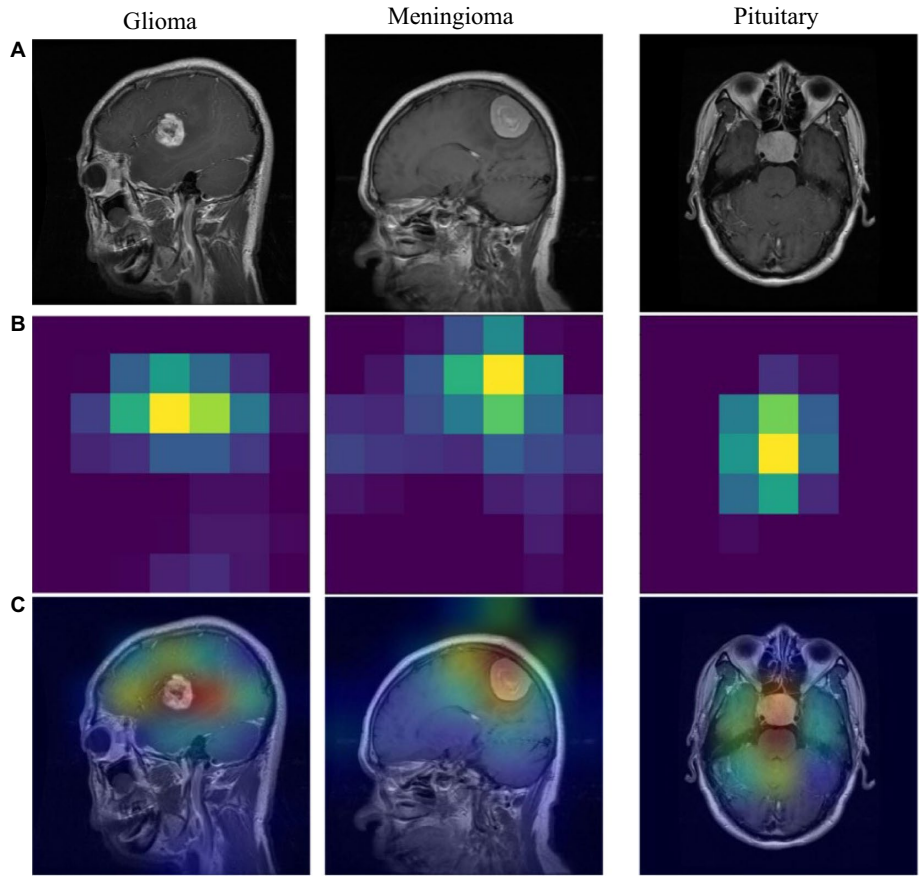


FIGURE 6
Grad-CAM visualization of different tumors. **(A)** brain tumor **(B)** heatmap **(C)** superimposed image.

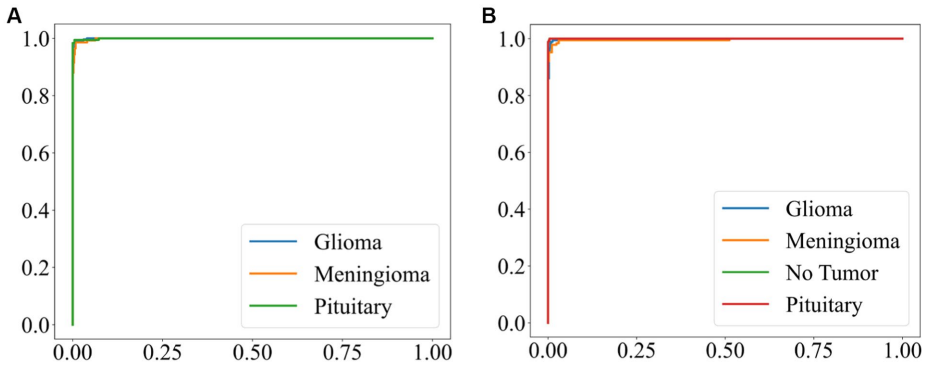


FIGURE 7
ROC curve for EFF_D_SVM **(A)** Chen **(B)** Kaggle.

out. Considering that the Chen dataset is three-class dataset while the Kaggle dataset comprises four classes, EFF_D_SVM and EFF_SVM will be evaluated on Kaggle while excluding the normal category classes. This decision was made to ensure the model reliability and validity while avoiding any potential confounding factors. Table 6 es the results of cross-dataset validation. EFF_D_SVM achieves an F1-score of 97.61% and accuracy of 97.62%, which performs better than other models. These results suggest that the proposed EFF_D_SVM model has strong robustness.

To further evaluate the robustness of the model, gaussian noise and S&P noise were added to the test sets of brain tumor images, respectively. Gaussian noise constitutes a form of noise characterized by a probability density function that adheres to a Gaussian distribution. This type of noise frequently manifests in digital images. The emergence of Gaussian noise stems from intricate interplays among circuit components, prolonged functioning of the image sensor, and various other contributing factors. S & P is often referred to as impulse noise, which randomly modifies certain pixel values to appear as sporadic

TABLE 6 Results of cross-data validation.

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
EFF_D_Softmax	97.20	97.08	97.08	97.09
EFF_D_SVM	97.67	97.61	97.61	97.62
EFF_Softmax	92.67	94.37	93.42	94.04
EFF_SVM	97.44	97.37	97.36	97.37

TABLE 7 Classification results of models after adding noise.

Dataset	Type of noise	Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Chen	Gaussian noise	EFF_D_Softmax	94.96	94.23	94.36	94.44
		EFF_D_SVM	95.76	95.17	95.34	95.42
		EFF_Softmax	93.46	89.49	90.42	92.32
		EFF_SVM	93.69	89.31	90.57	92.16
	Salt and pepper noise	EFF_D_Softmax	95.73	94.81	95.10	95.42
		EFF_D_SVM	96.67	95.22	95.83	96.24
		EFF_Softmax	95.73	94.81	95.10	95.42
		EFF_SVM	94.58	93.16	93.45	94.28
Kaggle	Gaussian noise	EFF_D_Softmax	94.14	92.89	93.25	93.10
		EFF_D_SVM	93.76	93.25	93.54	93.40
		EFF_Softmax	84.62	85.47	83.40	83.44
		EFF_SVM	86.86	88.41	86.12	86.81
	Salt and pepper noise	EFF_D_Softmax	94.38	95.04	94.45	94.33
		EFF_D_SVM	95.99	96.25	96.06	95.71
		EFF_Softmax	92.89	93.05	92.31	92.02
		EFF_SVM	91.09	91.93	91.14	90.95

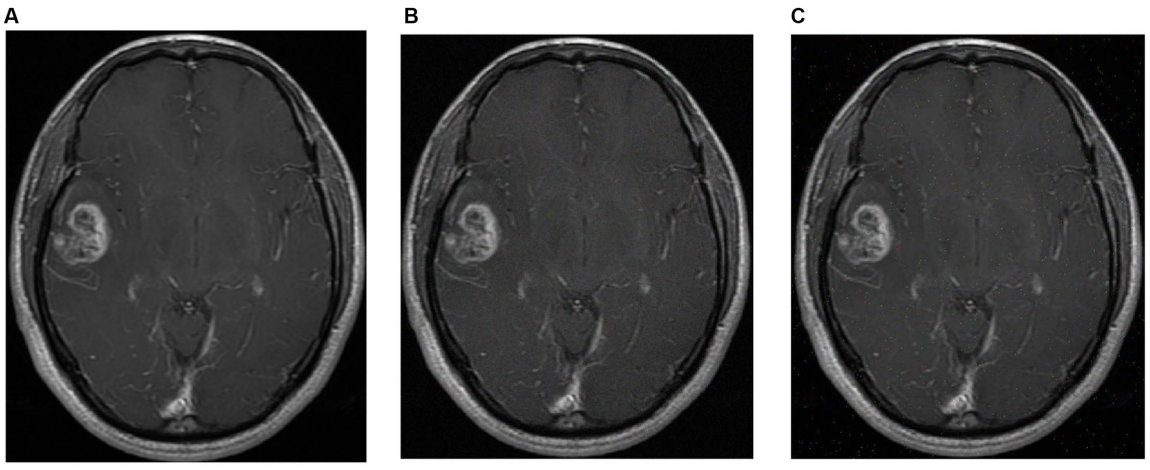


FIGURE 8
Images after adding noise **(A)** original image **(B)** Gaussian noise **(C)** S&P.

black-and-white dots in the image. This form of noise arises from the image sensor, transmission channel, decoding, and processing stages, resulting in both bright and dark dots scattered throughout the image. The robustness of models was verified by adding noise to datasets. Here, the variance of the Gaussian noise has been configured at 0.001, while the S&P noise affects 0.005 of the total pixels. Subsequently, the resulting image, which encompasses both Gaussian and S&P noise, is visually depicted in Figure 8. The Table 7 reveals that EFF_D_SVM demonstrates superior robustness compared to the other three models. Following the introduction of Gaussian and S&P noise to the images, EFF_D_SVM achieves classification accuracies of 95.42 and 96.24% for the Chen dataset, and 93.40 and 95.71% for the Kaggle dataset. Notably, for the

test set of the Kaggle dataset, both EFF_Softmax and EFF_SVM exhibit classification accuracies below 90% upon the introduction of Gaussian noise, which shows that they have weak robustness.

4. Conclusion

Early diagnosis of brain tumors is critical for selecting appropriate treatment options and saving the lives of patients. The manual examination of brain tumors is a laborious and time-consuming process, therefore, it is necessary to develop an automated detection method to aid physicians. This paper proposes a novel approach to detect multiple types of brain tumors. In this paper, a new feature extraction module EEF_D is proposed. Features are extracted from brain tumor images using EFF_D and the features are classified using SVM. To verify the effectiveness of our approach, a series of comparative experiments were also performed. The EFF_D_SVM model exhibits excellent classification ability for brain tumors with minimal Data pre-processing, as validated on both the Chen and Kaggle datasets. On the Chen dataset, EFF_D_SVM achieves a classification accuracy of 98.86% and an F1-score of 98.73%, and on the Kaggle dataset, it yields the corresponding values of 98.31 and 98.34%, respectively. Through comparison with other state-of-the-art models, the proposed model outperforms the available state-of-the-art methods. Moreover, by means of cross-validation experiments, the proposed model is proved to be very robust. In future work, samples from other types of brain disorders could be added to expand the dataset to improve the performance of the model, in turn to enhance the ability to identify other disorders.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427 https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

Author contributions

JZ: Investigation, Methodology, Project administration, Writing – original draft. XT: Software, Writing – original draft. WC: Writing – review & editing, Formal analysis, Software, Validation, Visualization. GD: Writing – review & editing, Project administration, Supervision. QF: Writing – review & editing, Validation. HZ: Investigation, Writing – review & editing. HJ: Investigation, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Major Science and Technology Projects of Henan Province (Grant No. 221100210500), the Medical and Health Research Project in Luoyang (Grant No. 2001027A), and the Construction Project of Improving Medical Service Capacity of Provincial Medical Institutions in Henan Province (Grant No. 2017–51).

Conflict of interest

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

Publisher's note

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

References

Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., and Mengko, T. R. (2019). "Brain tumor classification using convolutional neural network" in *World congress on medical physics and biomedical engineering 2018* (Singapore, FL: Springer Nature Singapore), 183–189.

Alanazi, M. F., Ali, M. U., Hussain, S. J., Zafar, A., Mohatram, M., Irfan, M., et al. (2022). Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* 22:372. doi: 10.3390/s22010372

Alsaggaf, W., Cömert, Z., Nour, M., Polat, K., Brdesee, H., and Toğaçar, M. (2020). Predicting fetal hypoxia using common spatial pattern and machine learning from cardiotocography signals. *Appl. Acoust.* 167:107429. doi: 10.1016/j.apacoust.2020.107429

Alyami, J., Rehman, A., Almutairi, F., Fayyaz, A. M., Roy, S., Saba, T., et al. (2023). Tumor localization and classification from MRI of brain using deep convolution neural network and Salp swarm algorithm. *Cogn. Comput.*, 1–11. (in press) doi: 10.1007/s12559-022-10096-2

Amin, J., Sharif, M., Yasmin, M., and Fernandes, S. L. (2020). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recogn. Lett.* 139, 118–127. doi: 10.1016/j.patrec.2017.10.036

Ayadi, W., Charfi, I., Elhamzi, W., and Atri, M. (2022). Brain tumor classification based on hybrid approach. *Vis. Comput.* 38, 107–117. doi: 10.1007/s00371-020-02005-1

Bar, Y., Diamant, I., Wolf, L., and Greenspan, H. (2015). "Deep learning with non-medical training used for chest pathology identification," in *Medical imaging 2015: computer-aided diagnosis, Orlando, FL.*

Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., and Kanchan, S. (2020). *Brain tumor classification (MRI).* Available at: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri (Accessed October 20, 2022).

Bi, D., Zhu, D., Sheykhahmad, F. R., and Qiao, M. (2021). Computer-aided skin cancer diagnosis based on a new meta-heuristic algorithm combined with support vector method. *Biomed. Signal Process. Control* 68:102631. doi: 10.1016/j.bspc.2021.102631

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

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

Deepak, S., and Ameer, P. M. (2023). Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion. *Neurocomputing* 520, 94–102. doi: 10.1016/j.neucom.2022.11.039

Demir, F., and Akbulut, Y. (2022). A new deep technique using R-CNN model and L1NSR feature selection for brain MRI classification. *Biomed. Signal Process. Control* 75:103625. doi: 10.1016/j.bspc.2022.103625

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, 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

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

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

Jaspin, K., and Selvan, S. (2023). Multiclass convolutional neural network based classification for the diagnosis of brain MRI images. *Biomed. Signal Process. Control* 82:104542. doi: 10.1016/j.bspc.2022.104542

Kabir Anaraki, A., 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

Kang, J., Ullah, Z., and Gwak, J. (2021). MRI-based brain tumor classification using Ensemble of Deep Features and Machine Learning Classifiers. *Sensors* 21:2222. doi: 10.3390/s21062222

Kaur, T., and Gandhi, T. K. (2020). Deep convolutional neural networks with transfer learning for automated brain image classification. *Mach. Vis. Appl.* 31:20. doi: 10.1007/s00138-020-01069-2

Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M., Saba, T., et al. (2019). Brain tumor detection and classification: a framework of marker-based watershed algorithm and multilevel priority features selection. *Microsc. Res. Tech.* 82, 909–922. doi: 10.1002/jemt.23238

Kibriya, H., Masood, M., Nawaz, M., and Nazir, T. (2022). Multiclass classification of brain tumors using a novel CNN architecture. *Multimed. Tools Appl.* 81, 29847–29863. doi: 10.1007/s11042-022-12977-y

Kumar, S., and Mankame, D. P. (2020). Optimization driven deep convolution neural network for brain tumor classification. *Biocybern. Biomed. Eng.* 40, 1190–1204. doi: 10.1016/j.bbe.2020.05.009

Maurya, S., Tiwari, S., Mothukuri, M. C., Tangeda, C. M., Nandigam, R. N. S., and Addagiri, D. C. (2023). A review on recent developments in cancer detection using machine learning and deep learning models. *Biomed. Signal Process. Control* 80:104398. doi: 10.1016/j.bspc.2022.104398

Mehnatkesh, H., Jalali, S. M. J., Khosravi, A., and Nahavandi, S. (2023). An intelligent driven deep residual learning framework for brain tumor classification using MRI images. *Expert Syst. Appl.* 213:119087. doi: 10.1016/j.eswa.2022.119087

Muezzinoglu, T., Baygin, N., Tuncer, I., Barua, P. D., Baygin, M., Dogan, S., et al. (2023). Patch res net: multiple patch division–based deep feature fusion framework for brain tumor classification using MRI images. *J. Digit. Imaging* 36, 973–987. doi: 10.1007/s10278-023-00789-x

Nanda, A., Barik, R. C., and Bakshi, S. (2023). SSO-RBNN driven brain tumor classification with saliency-K-means segmentation technique. *Biomed. Signal Process. Control* 81:104356. doi: 10.1016/j.bspc.2022.104356

Nayak, D. R., Padhy, N., Mallick, P. K., Zymbler, M., and Kumar, S. (2022). Brain tumor classification using dense efficient-net. *Axioms* 11:34. doi: 10.3390/axioms11010034

Öksüz, C., Urhan, O., and Güllü, M. K. (2022). Brain tumor classification using the fused features extracted from expanded tumor region. *Biomed. Signal Process. Control* 72:103356. doi: 10.1016/j.bspc.2021.103356

Özbay, E., and Altunbey Özbay, F. (2023). Interpretable features fusion with precision MRI images deep hashing for brain tumor detection. *Comput. Methods Prog. Biomed.* 231:107387. doi: 10.1016/j.cmpb.2023.107387

Rehman, A., Naz, S., Razzak, M. I., Akram, F., and Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits Syst. Signal Process.* 39, 757–775. doi: 10.1007/s00034-019-01246-3

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. Computat. Sci.* 30, 174–182. doi: 10.1016/j.jocs.2018.12.003

Saravanan, S., Heshma, B., Ashma Shanofer, A. V., and Vanithamani, R. (2020). Skin cancer detection using dermoscope images. *Mater. Today Proc.* 33, 4823–4827. doi: 10.1016/j.matpr.2020.08.388

Satyanarayana, G. (2023). A mass correlation based deep learning approach using deep convolutional neural network to classify the brain tumor. *Biomed. Signal Process. Control* 81:104395. doi: 10.1016/j.bspc.2022.104395

Sekhar, A., Biswas, S., Hazra, R., Sunaniya, A. K., Mukherjee, A., and Yang, L. (2022). Brain tumor classification using fine-tuned Goog LeNet features and machine learning algorithms: IoMT enabled CAD system. *IEEE J. Biomed. Health Inform.* 26, 983–991. doi: 10.1109/JBHI.2021.3100758

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

Shah, H. A., Saeed, F., Yun, S., Park, J.-H., Paul, A., and Kang, J.-M. (2022). A robust approach for brain tumor detection in magnetic resonance images using Finetuned efficient net. *IEEE Access* 10, 65426–65438. doi: 10.1109/ACCESS.2022.3184113

Shahin, A. I., Aly, W., and Aly, S. (2023). MBTFCN: a novel modular fully convolutional network for MRI brain tumor multi-classification. *Expert Syst. Appl.* 212:118776. doi: 10.1016/j.eswa.2022.118776

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi: 10.1109/TMI.2016.2528162

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., et al. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput. Med. Imaging Graph.* 75, 34–46. doi: 10.1016/j.compmedimag.2019.05.001

Talo, M., Baloglu, U. B., Yıldırım, Ö., and Rajendra Acharya, U. (2019). Application of deep transfer learning for automated brain abnormality classification using MR images. *Cogn. Syst. Res.* 54, 176–188. doi: 10.1016/j.cogsys.2018.12.007

Tan, M., and Le, Q. V. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks", in International Conference on Machine Learning, *vol. 97*, Long Beach, CA.

Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., and Zhang, Y.-D. (2022). Transfer learning for medical images analyses: a survey. *Neurocomputing* 489, 230–254. doi: 10.1016/j.neucom.2021.08.159

Zulfiqar, F., Ijaz Bajwa, U., and Mehmood, Y. (2023). Multi-class classification of brain tumor types from MR images using efficient nets. *Biomed. Signal Process. Control* 84:104777. doi: 10.1016/j.bspc.2023.104777

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 21 November 2023
DOI 10.3389/fnins.2023.1275065

OPEN ACCESS

EDITED BY
Lei Wang,
Northwestern University, United States

REVIEWED BY
Michael Lassi,
Institute of BioRobotics, Sant'Anna School of Advanced Studies, Italy
Arcady A. Putilov,
Federal Research Center of Fundamental and Translational Medicine, Russia

*CORRESPONDENCE
Mincheng Cai
✉ minchengcai@whut.edu.cn

RECEIVED 09 August 2023
ACCEPTED 27 October 2023
PUBLISHED 21 November 2023

CITATION
Liu H, Liu Q, Cai M, Chen K, Ma L, Meng W, Zhou Z and Ai Q (2023) Attention-based multi-semantic dynamical graph convolutional network for eeg-based fatigue detection. *Front. Neurosci.* 17:1275065. doi: 10.3389/fnins.2023.1275065

Attention-based multi-semantic dynamical graph convolutional network for eeg-based fatigue detection

Haojie Liu, Quan Liu, Mincheng Cai*, Kun Chen, Li Ma, Wei Meng, Zude Zhou and Qingsong Ai

School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei, China

Introduction: Establishing a driving fatigue monitoring system is of utmost importance as severe fatigue may lead to unimaginable consequences. Fatigue detection methods based on physiological information have the advantages of reliable and accurate. Among various physiological signals, EEG signals are considered to be the most direct and promising ones. However, most traditional methods overlook the functional connectivity of the brain and fail to meet real-time requirements.

Methods: To this end, we propose a novel detection model called Attention-Based Multi-Semantic Dynamical Graph Convolutional Network (AMD-GCN). AMD-GCN consists of a channel attention mechanism based on average pooling and max pooling (AM-CAM), a multi-semantic dynamical graph convolution (MD-GC), and a spatial attention mechanism based on average pooling and max pooling (AM-SAM). AM-CAM allocates weights to the input features, helping the model focus on the important information relevant to fatigue detection. MD-GC can construct intrinsic topological graphs under multi-semantic patterns, allowing GCN to better capture the dependency between physically connected or non-physically connected nodes. AM-SAM can remove redundant spatial node information from the output of MD-GC, thereby reducing interference in fatigue detection. Moreover, we concatenate the DE features extracted from 5 frequency bands and 25 frequency bands as the input of AMD-GCN.

Results: Finally, we conduct experiments on the public dataset SEED-VIG, and the accuracy of AMD-GCN model reached 89.94%, surpassing existing algorithms.

Discussion: The findings indicate that our proposed strategy performs more effectively for EEG-based driving fatigue detection.

KEYWORDS

EEG, driving fatigue detection, channel attention mechanism, graph convolutional network, spatial attention mechanism

1 Introduction

Drivers driving for a long time or driving at night can lead to a decline in physical and psychological abilities, seriously affecting the ability to drive safely. Fatigue while driving can impair basic skills such as attention, decision-making, and reaction time, while also affecting cognitive processes, sensory perception, and overall mental well-being. In severe cases, this may result in a decline in motor function and increase the likelihood of being involved in traffic accidents. Statistically, in 2004, the World Health Organization released the "World Report on Road Traffic Injury Prevention", which pointed out that approximately 20% ∼ 30% of traffic accidents were caused by fatigue driving. By 2030, the number of road traffic fatalities is projected to rise to about 2.4 million people annually, making road traffic

deaths the fifth leading cause of death worldwide (WHO, 2009). As the number of casualties due to fatigue driving continues to increase, it is urgent to develop reliable and effective driving fatigue detection methods.

The existing fatigue detection methods mainly include vehicle information-based, facial feature-based, and physiological signal-based approaches. The vehicle information-based detection method indirectly assess the driver's fatigue state based on the driver's manipulation of the vehicle (Li et al., 2017; Chen et al., 2020). This method utilizes on-board sensors and cameras to collect data such as steering wheel angle, grip force, vehicle speed, and driving trajectory. By analyzing the differences in driving behavior parameters between normal driving and fatigue states, it assesses the driver's fatigue condition. However, it is challenging to collect accurate and stable data using this method due to variations in driving habits and proficiency among drivers. The facial feature-based detection method infers the driver's fatigue state through analyzing eye status, mouth status, and head posture (Wu and, 2019; Quddus et al., 2021; Huang et al., 2022). This method mainly uses the camera to capture the driver's face image, and extracts the fatigue-related information through the computer vision technology. In contrast, physiological signal-based detection methods can directly reflect the driver's driving state, including electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (ECG), and electromyogram (EMG). Among various physiological signals, EEG signals contain all the information of brain operation and are closely related to mental and physical activity, with good time resolution and strong anti-interference ability (Yao and Lu, 2020), which are the result of excitatory or inhibitory postsynaptic potentials generated by the cell bodies and dendrites of pyramidal neurons (Zeng et al., 2021). Meanwhile, the EEG caps tend to be intelligent and lightweight (Lin et al., 2019), making it convenient to keep an EEG cap while driving. EEG signals are considered the most direct and promising.

EEG signals are recordings of the spontaneous or stimulus-induced electrical activity generated by specific regions of the brain's neurons during physiological processes, reflecting the brain's biological activities and carrying a wealth of information (Jia et al., 2023). From an electrophysiological perspective, every subtle brain activity induces corresponding neural cell discharges, which can be recorded by specialized instruments to analyze and decode brain function. EEG decoding is the separation of task-relevant components from the EEG signals. The main method of decoding is to describe task-related components using feature vectors, and then use classification algorithms to classify the relevant features of different tasks. The accuracy of decoding depends on how well the feature algorithm represents the relevant tasks and the discriminative precision of the classification algorithm for different tasks. The EEG signals record the electrical wave changes in brain activity, making them the most direct and effective reflection of fatigue state. Based on the amplitude and frequency of the waveforms, EEG waves are classified into five types: δ(1-3Hz), θ(4-7Hz), α(8–13Hz), β(14–30Hz), γ(31–50Hz) waves (Song et al., 2020). It is worth noting that, during the awake state, EEG signals are mainly characterized by α and β waves. As fatigue increases, the amplitude of α and β waves gradually diminishes, and they may even disappear, while δ and θ waves gradually increase, indicating significant variations in EEG signals during different stages of fatigue (Jia et al., 2023). Therefore, many scholars regard EEG signals as the gold standard for measuring the level of fatigue (Zhang et al., 2022). Lal and Craig (2001) tested non-drivers' EEG waves and analyzed the characteristics of EEG wave changes in five stages: non-fatigue, near-fatigue, moderate fatigue, drowsiness, and anti-fatigue. They concluded that EEG is the most suitable signal for evaluating fatigue. Lal and Craig (2002) collected EEG data from 35 participants in the early stage of fatigue using 19 electrodes. The experimental results indicated a decrease in the activity of α and β waves during the fatigue process, while there was a significant increase in the activity of δ and θ waves. Papadelis et al. (2006) introduced the concept of entropy in a driving fatigue experiment. The study found that under severe fatigue conditions, the number of α waves and β waves exhibited inconsistent changes, and shannon entropy and kullback-leibler entropy values decreased with the changes in β waves.

In recent years, thanks to the rapid development of sensor technology, information processing, computer science, and artificial intelligence, a large number of studies have proposed combining fatigue driving detection based on EEG signals with machine learning or deep learning methods. Paulo et al. (2021) proposed using recursive graphs and gramian angular fields to transform the raw EEG signals into image-like data, which is then input into a single-layer convolutional neural network (CNN) to achieve fatigue detection. Abidi et al. (2022) processed the raw EEG signals using a tunable Q-factor wavelet transform and extracted signal features using kernel principal component analysis (KPCA). They then used k-nearest neighbors (KNN) and support vector machine (SVM) for EEG signal classification. Song et al. (2022) proposed a method that combines convolutional neural network (CNN) and long short-term memory (LSTM) called LSDD-EEGNet. It utilizes CNN to extract fe atures and LSTM for classification. Gao et al. (2019) introduced core blocks and dense layers into CNN to extract and fuse spatial features, achieving detection. In the study (Wu et al., 2021), designed a finite impulse response (FIR) filter with chebyshev approximation to obtain four EEG frequency bands (i.e., δ, θ, α, β), and constructed a new deep sparse contracting autoencoder network to learn more local fatigue features. Cai et al. (2020) introduced a new method referred to as graph-time fusion dual-input convolutional neural network. This method transforms each EEG epoch of sleep stages into limited penetration visible graph (LPVG) and utilizes a new dual-input CNN to assess the degree sequences of LPVG and the original EEG epochs. Finally, based on the CNN analysis, the sleep stages are classified into six states. Gao et al. (2021) were the first to explore the application of complex networks and deep learning in EEG signal analysis. They introduced a fatigue driving detection network framework that combines complex networks and deep learning. The network first calculates the EEG signals for each channel and generates a feature matrix using a recursive rate. Then, this feature matrix is fed into a specially designed CNN, and the prediction results are obtained through the softmax function.

The above deep learning and convolutional neural network (CNN) methods mainly focus on the features of individual electrode EEG signals and overlook the functional connectivity of the brain, that is the correlation between EEG channels. Due to the

non-Euclidean structure of EEG signals, CNN based on Euclidean space learning is limited in handling the functional connections between different electrodes. Therefore, using CNN to process EEG signals may not be an optimal choice.

In recent years, the emergence of graph convolutional neural networks (GCN) has been proven to be the most effective method for handling non-Euclidean structured data (Jia et al., 2021; Zhu et al., 2022). Using GCN to process EEG signals allows to represent the functional connections of the brain through topological data. In this case, each EEG signal channel is treated as a node in the graph, and the connections between EEG signal channels serve as the edges of the graph. Jia et al. (2023) proposed a model called MATCN-GT for fatigue driving detection, which consists of a multi-scale attention time convolutional neural network block (MATCN) and a graph convolution-transformer (GT) block. The MATCN directly extracts features from the raw EEG signals, while the GT processes the features of EEG signals from different electrodes. Zhang et al. (2020) introduced the PDC-GCNN method for detecting driver's EEG signals, which uses partial directed coherence (PDC) to construct an adjacency matrix, and then employs graph convolutional neural network (GCN) for EEG signal classification. Song et al. (2020) proposed a multi-channel EEG emotion recognition method based on dynamic graph convolutional neural network (DGCNN). The basic idea is to use graphs to model multi-channel EEG features and then perform EEG emotion classification based on this model. Jia et al. (2020) proposed a novel deep graph neural network called GraphSleepNet to classify EEG signals. This network can dynamically learn the adjacency matrix and utilizes a spatio-temporal graph convolutional network (ST-GCN) to classify EEG signals. The method demonstrated excellent classification results on the MASS dataset. Zhang et al. (2019) designed a graph convolution broad network (GCB-net) to explore deeper-level information in graph-structured data. It utilizes graph convolutional layers to extract features from the input graph structure and stacks multiple regular convolutional layers to capture more abstract features. Additionally, a broad learning system (BLS) is employed to enhance the features and improve the performance of GCB-net.

Although GCN is proficient at learning the internal structural information of EEG signals, it relies on the connectivity between nodes provided by the adjacency matrix. Most methods obtain functional connectivity of EEG signals by using predefined fixed graphs such as PLI, PLV, PDC, or spatial relationships, which prevents the model from adaptively constructing adjacency graphs simultaneously related with subjects, fatigue states and samples, thereby overlooking the data-driven intrinsic correlations. However, constructing a suitable graph representation for the adjacency matrix of each data in advance requires time and effort. Additionally, GCN faces challenges in learning dependencies between distant nodes (long-range vertices). Increasing the depth of GCN to expand the receptive field remains difficult and may lead to over-smoothing of nodes.

To address the above problem, we propose a new fatigue driving detection network, referred to as the attention-based multi-semantic dynamical graph convolutional network (AMD-GCN). First, the network utilizes a channel attention mechanism based on average pooling and max pooling to assign weights to the fused EEG input features. This helps the model focus on the crucial information parts related to fatigue detection. Next, the adjusted EEG input features are fed into the GCN, we determine the adjacency matrix using spatial adjacency relationships, Euclidean spatial distances, and self-attention mechanism to construct data-driven intrinsic topology under multiple semantic patterns, thereby enhancing the spatial feature extraction capability of GCN. Furthermore, a spatial attention mechanism based on average pooling and max pooling is employed to calculate the weights of spatial nodes in the output of GCN, which helps in removing redundant node information and reducing interference in fatigue detection. Finally, the prediction results are output by softmax.

2 Dataset description and EEG pre-processing

2.1 Public dataset SEED-VIG

We validated the proposed method on the publicly available dataset SEED-VIG (Zheng and Lu, 2017) for driving fatigue detection researches. SEED-VIG adopt the international 10-20 electrode system standard, and the EEG signals were collected from 6 channels in the temporal region of the brain (FT7, FT8, T7, T8, TP7, TP8) and 12 channels from the posterior region (CP1, CPZ, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, O2), where CPZ channel serves as the reference electrode, and the specific electrode placement is shown in Figure 1. The experiment simulated a driving environment by creating a virtual reality scenario, in which 23 participants engaged in approximately 2 hours of simulated driving during either a fatigue-prone midday or evening session. The subjects comprised 12 females and 11 males, with an average age of 23.3 years and a standard deviation of 1.4. All subjects had normal or corrected vision.

The SEED-VIG dataset was vigilantly annotated using eye-tracking methods, capturing participants' eye movements with the assistance of SMI eye-tracking glasses. These glasses categorized eye states into fixation, blink, and saccade, and recorded their respective durations. The "CLOS" state, referring to slow or long-duration eye closure, is undetectable by the SMI eye-tracking glasses. In such cases, fixation and saccade represent normal states, while blink or CLOS indicates fatigue in participants. Therefore, PERCLOS represents the percentage of time in a specific period when participants were in a fatigued state (Dinges and Grace, 1998). The calculation of PERCLOS is as follows:

$$\begin{aligned} PERCLOS &= \frac{blink + close}{interval}, \\ interval &= blink + fixation + saccade + close \end{aligned} \tag{1}$$

Where blink, close, fixation, and saccade denote the duration of eye states (blink, close, gaze, and sweep, respectively) recorded by the eye tracker within the 8-second intervals. PERCLOS is a continuous value between 0 and 1, with smaller values indicating higher vigilance. The standard procedure for using this publicly available dataset for research is to set two thresholds (0.35 and 0.7) in order to classify the samples into three types:

- Awake class: $PERCLOS < 0.35$;
- Tired class: $0.35 \leq PERCLOS < 0.7$;

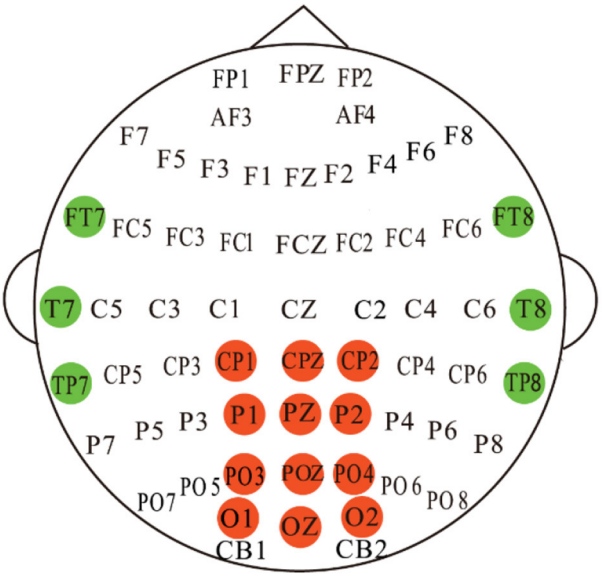


FIGURE 1
Electrode placements for the EEG setups. 12-channel and 6-channel EEG signals were recorded from the posterior region (red color) and the temporal region (green color), respectively.

- Drowsy class: $PERCLOS \geq 0.7$.

In addition, we validated our proposed method on the SEED-VIG dataset, dividing each subject's 885 samples into 708 samples for training and 177 samples for testing by a way that preserves the temporal order, then we trained the model separately on each subject and evaluated it on the testing samples of the same subject. Finally, in order to mitigate the impact of data imbalance within one subject on the model performance evaluation as much as possible, the average classification accuracy and individual variation of 23 subjects were computed as evaluation metrics. It is worth noting that SEED-VIG adopts an 8-second non-overlapping sliding window to sample data, and we split the dataset by preserving the temporal order. Therefore, training is based on past data, and testing is based on future data. This ensures that the model is evaluated on unseen data, thereby alleviating the risk of data leakage (Saeb et al., 2017).

2.2 EEG pre-processing

The signal preprocessing method is consistent with other works (Zheng and Lu, 2017; Ko et al., 2021; Peng et al., 2023; Shi and Wang, 2023), we directly used the clean EEG signals provided by the study (Zheng and Lu, 2017), which has removed eye blinks, and the raw EEG data was downsampled from 1000 Hz to 200 Hz to reduce computational burden. Subsequently, it is bandpass filtered between 1-50 Hz to remove irrelevant components and power line interference. For SEED-VIG, there are two different methods to segment the frequency range into different bands. One widely used approach is to divide the frequency range into bands as follows: δ(1-3Hz), θ(4-7Hz), α(8-13Hz), β(14-30Hz), γ(31-50Hz). The other method is to uniformly divide the range into 25 bands with a 2-Hz resolution.

TABLE 1 Summary of the overall properties of SEED-VIG.

Dataset	Samples	Channels	Frequency bands
SEED-VIG-5band	885	17	5
SEED-VIG-2Hz	885	17	25
PERCLOS-labels	885	N / A	N / A

NA, Not Applicable.

For each frequency band, the computation of the extracted differential entropy (DE) feature is as follows:

$$h(X) = -\int_X f(x)\ln f(x)dx \tag{2}$$

Here, X is a random variable whose probability density function is defined by $f(x)$. Assuming that the probability density function $f(x)$ of the EEG signal follows the Gaussian distribution $N(\mu, \delta^2)$, the DE feature can then be computed as:

$$\begin{aligned} h(X) &= -\int f(x)(-\frac{1}{2}\ln(2\pi\delta^2) - \frac{(x-\mu)^2}{2\delta^2}) \\ &= \frac{1}{2}\ln(2\pi\delta^2) + \frac{Var(X)}{2\delta^2} = \frac{1}{2}\ln(2\pi e\delta^2) \end{aligned} \tag{3}$$

Here, we used the facts that $\int f(x)dx = 1$ and $Var(x) = \int f(x)(x-\mu)^2 dx = \delta^2$. DE features were extracted by short-term Fourier transform with an 8-second non-overlapping time window.

The overall properties of SEED-VIG are summarized in Table 1. In our study, we concatenate the DE features extracted based on 5 frequency bands and the DE features extracted based on 25 frequency bands within the same time window as one sample input to the neural network. This allows us to fully utilize the information contained in the original EEG signal and thereby enhance the effect of fatigue driving detection. The overall data form of one subject can be expressed as $R^{885\times17\times30}$.

3 Method

Our proposed AMD-GCN model consists of three functional modules: channel attention mechanism based on average pooling and max pooling (AM-CAM), multi-semantic dynamical graph convolution (MD-GC), and spatial attention mechanism based on average pooling and Max pooling (AM-SAM). The AMD-GCN model enables end-to-end fatigue state assessment of drivers based on the extracted DE features from EEG signals. The AMD-GCN model retains crucial input features through AM-CAM, performs multi-semantic spatial feature learning through MD-GC, and eliminates redundant spatial nodes information through AM-SAM. The overall architecture of fatigue driving detection based on AMD-GCN is illustrated in Figure 2.

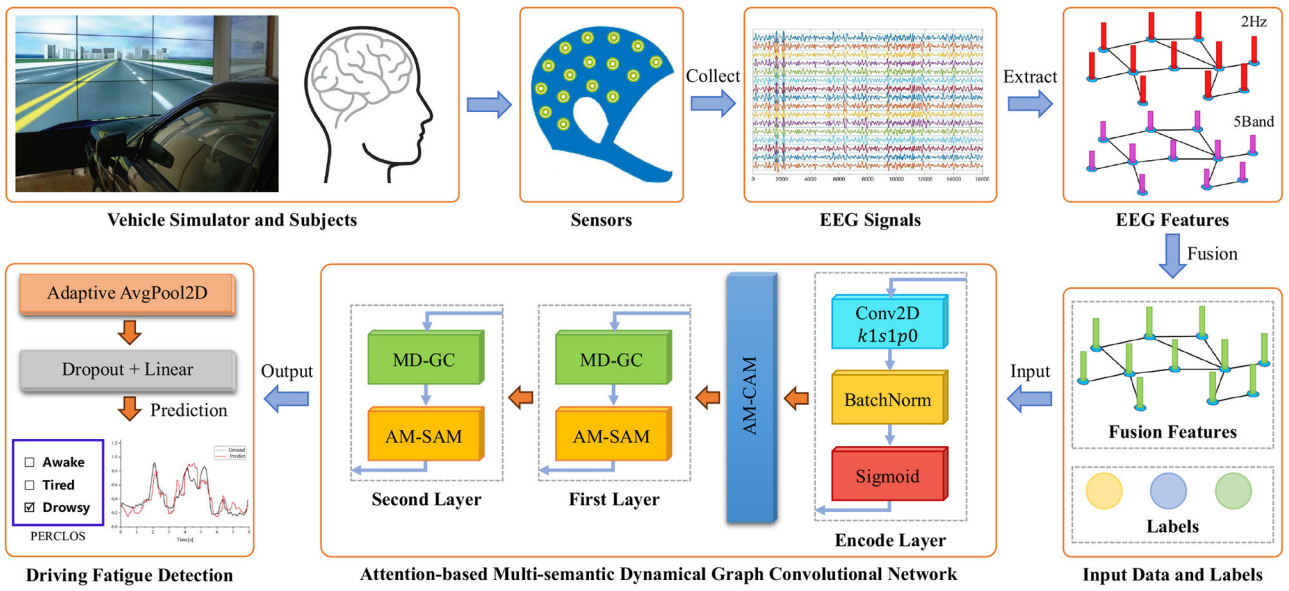


FIGURE 2
Overall schematic diagram of fatigue driving detection based on AMD-GCN. AMD-GCN consists of three modules: AM-CAM module, MD-GC module, and AM-SAM module. The input to the model is the fused feature of DE features extracted based on 5 frequency bands and DE features extracted based on 25 frequency bands. The output is the predicted label with probabilities.

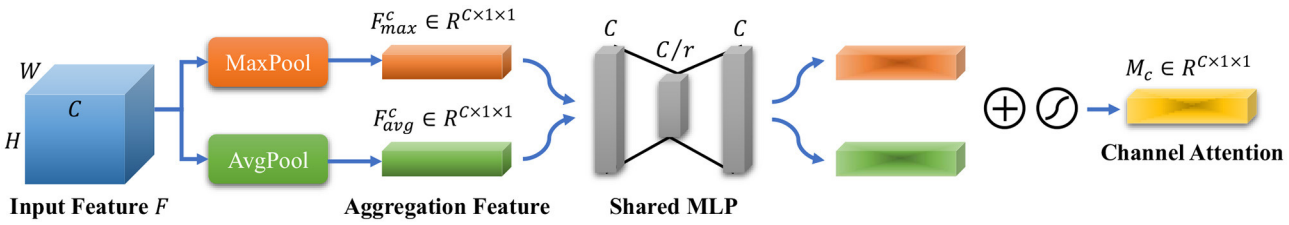


FIGURE 3
Schematic diagram of AM-CAM. As illustrated, the channel attention sub-module utilizes both the max pooling output and average pooling output with a shared network.

3.1 Preliminary

In our paper, we designed the AMD-GCN model adopting graph convolutional neural networks to process spatial features. To facilitate reader comprehension, we first elucidate the fundamental concepts and relevant content of GCN before introducing AMD-GCN.

Consider a graph $G = (V, \varepsilon, A)$, which represents a collection of all nodes and edges. Here, $V = (v_1, v_2, ..., v_n)$ signifies that the graph has N nodes, v_n denotes the n-th node, and E is a set of edges representing relationships between nodes. $A \in R^{N\times N}$ stands for the adjacency matrix of graph G, denoting connections between two nodes. It's worth noting that GCN (Kipf and Welling, 2016) employs graph spectral theory for convolutional operations on topological graphs. It primarily explores the properties of the graph through the eigenvalues and eigenvectors of the graph's Laplacian matrix. The Laplacian matrix of a graph is defined as follows:

$$L = D - A \tag{4}$$

where $D \in R^{N\times N}$ is the degree matrix of the vertices (diagonal matrix), that is, the elements on the diagonal are the degrees of each vertex in turn. L denotes the Laplacian matrix, whose normalized form can be expressed as:

$$L = I_n - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = UAU^T \tag{5}$$

Where I_n is the identity matrix. UAU^T represents the orthogonal decomposition of the Laplacian matrix, where $U = [u_0, u_1, ..., u_{n-1}] \in R^{n\times n}$ is the orthogonal matrix of eigenvectors obtained through the singular value decomposition (SVD) of the graph Laplacian matrix, and $\Lambda = [\lambda_0, \lambda_1, ..., \lambda_{n-1}] \in R^{n\times n}$ is the diagonal matrix of corresponding eigenvalues. For a given input feature matrix X, its graph Fourier transform is:

$$\hat{X} = U^T X, X = U\hat{X}(inverse) \tag{6}$$

The convolution of the graph for input X and filter K can be expressed as:

$$Y = X * GK = U((U^T X) \odot (U^T G)) = U\hat{K}U^T X \tag{7}$$

Here, $\odot$ denotes the element-wise Hadamard product. However, directly computing the Eq.7 would require a substantial amount of computational resources. To mitigate energy consumption, Kipf and Welling (2016) proposed an efficient variant of convolutional neural networks that directly operate on graphs, approximating the graph convolution operation through a first-order Chebyshev polynomial. Supposing a graph G with N nodes, each node possessing its own features, let these node features form a matrix $X \in R^{N\times D}$. With an input feature matrix X and an adjacency matrix A, we can obtain the output:

$$Y = \sigma(\hat{D}^{-\frac{1}{2}}A\hat{D}^{-\frac{1}{2}}XW) \tag{8}$$

Where σ represents the nonlinear activation function.

3.2 Channel attention mechanism based on average pooling and max pooling

Firstly, we employ an autoencoder layer to perform re-representation of the input data, creating inputs with richer semantic information, as depicted in Figure 2, where the input channels are 30 and the output channels are 128. Then, in order to focus the model on crucial parts of the input related to the fatigue detection category, we generate channel attention maps by exploiting inter-channel relationships of features. This is achieved through the design of a channel attention mechanism based on average pooling and max pooling (AM-CAM) layer. The channel attention mechanism focuses on determining "what" in the input is meaningful, treating each channel of the feature map as a feature detector (Zeiler and Fergus, 2014). To compute channel attention effectively, we compress the spatial dimensions of the input feature maps. To gather spatial information, we employ an average pooling layer to gain insights into the extent of the target object effectively, utilizing it in the attention module to compute spatial statistics. Additionally, we use a max pooling layer to collect salient information about different object features, enabling the inference of finer channel attention. Figure 3 illustrates the computation process of channel attention maps, and the detailed operations are described as follows.

Given an intermediate feature map $F \in R^{C\times H\times W}$ as input, we first utilize average pooling and max pooling operations to aggregate spatial information from the feature map, generating two distinct spatial context descriptors: F^c_{avg} and F^c_{max}, representing average-pooled features and max-pooled features, respectively. Subsequently, both of these descriptors are fed into a multilayer perceptron (MLP) with a hidden layer to generate the channel attention map $M_c \in R^{C\times 1\times 1}$. To reduce parameter overhead, the hidden activation size is set to $R^{\frac{C}{r}\times 1\times 1}$, where r is the reduction ratio and is set to 16 in our study. After applying the shared network to each descriptor, we merge the output feature vectors using element-wise summation. In short, the channel attention is computed as:

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \end{aligned} \tag{9}$$

Where σ denotes sigmoid function, $W_0 \in R^{\frac{C}{r}\times C}$ and $W_1 \in R^{C\times\frac{C}{r}}$, Note that the MLP weights, W_0 and W_1, are shared for both inputs and the ReLU activation function is followed by W_0. The output F_{out} of AM-CAM can be formulated as:

$$F_{out} = M_c(F) \odot F \tag{10}$$

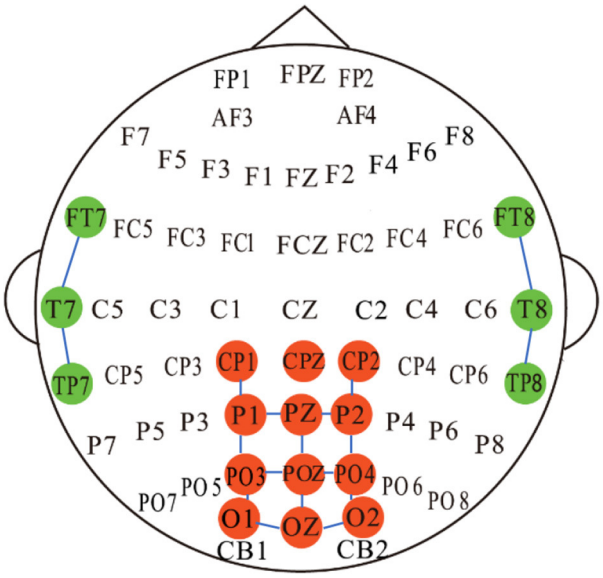


FIGURE 4
A schematic diagram illustrating the connections between the 17 EEG channels based on spatial adjacency relationships is used to construct the adjacency matrix for SRGC. CPZ serves as the reference electrode and is not involved in the construction of the adjacency matrix.

3.3 Multi-semantic dynamical graph convolution

In this study, we propose a multi-semantic dynamical graph convolution (MD-GC) for extracting spatial features from the input. It determines the adjacency matrix based on spatial adjacency relationships, Euclidean spatial distance, and self-attention mechanism. Our approach constructs data-driven intrinsic topology under various semantic patterns, enhancing the spatial feature extraction capability of graph convolution. Overall, given an intermediate feature map $X \in R^{C\times V}$ as input, the output of MD-GC can be computed as:

$$\begin{aligned} MDGC(X) = \sigma(&BN(SRGC(X) + EDGC(X) \\ &+ SAGC(X))) \end{aligned} \tag{11}$$

Where σ is sigmoid function, BN is batch normalization, SRGC represents spatial relationship-based graph convolution, EDGC represents Euclidean distance-based graph convolution, and SAGC stands for self-attention-based graph convolution.

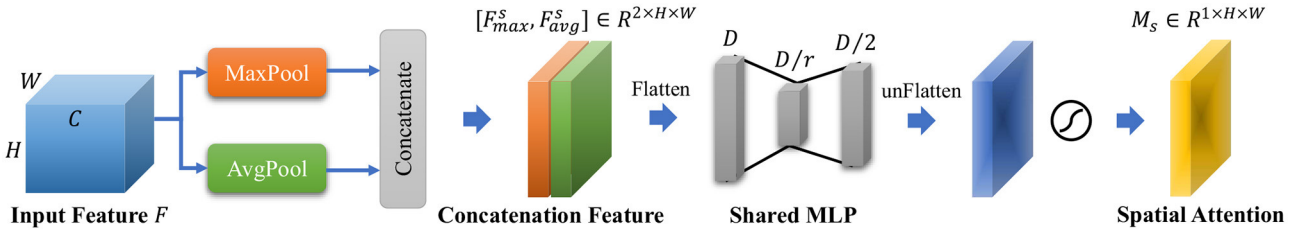


FIGURE 5
Schematic diagram of AM-SAM. As illustrated, the spatial attention sub-module utilizes both the max pooling output and average pooling output with a shared network.

3.3.1 Graph convolution based on spatial relationship

Intuitively, the correlation between EEG electrodes is constrained due to the distribution of nodes on the brain (Song et al., 2020), which represents inherent connections. To capture this relationship, we developed a spatial adjacency graph, denoted as $G_{SR}(V, A_{SR})$. A_{SR} represents the spatial adjacency matrix between brain nodes, as shown in Figure 4, where adjacent nodes are connected by solid blue lines. A_{SR} considers the adjacency relationships of 6 channels from the temporal region of the brain and 12 channels from the posterior part of the brain. We first normalize the spatial adjacency matrix A_{SR} using

$$\tilde{A}_{SR} = D_{SR}^{-1} A_{SR} \tag{12}$$

$D_{SR}^{-1} \in R^{N\times N}$ is a diagonal degree matrix of A_{SR}. $\tilde{A}_{SR}$ provides nice initialization to learn the edge weights and avoids multiplication explosion (Brin and Page, 1998; Chen et al., 2018). Given the computed $\tilde{A}_{SR}$, we propose the spatial relationship-based graph convolution (SRGC) operator. Let $X \in R^{V\times C}$ and $Y_{SRGC} \in R^{V\times C_{out}}$ be the input and output features of SRGC, respectively. The SRGC operator can be formalized as:

$$Y_{SRGC} = SRGC(X) = \tilde{A}_{SR} X W_{SR}^T \tag{13}$$

Where $W_{SR} \in R^{C_{out}\times C}$ is the trainable weight used to facilitate feature updating in the SRGC.

3.3.2 Graph convolution based on Euclidean-space distance

Considering that SRGC can only capture relationships between nodes connected by physiological connections, here we introduce a Euclidean distance-based graph convolution (EDGC) operator to capture potential relationships between physically non-connected nodes, thereby imposing higher-order positional information. Specifically, we define a Euclidean space distance adjacency matrix for the potential sample dependencies in EDGC, where the adjacency weight between nodes i and j is calculated as:

$$a_{i,j} = max(E) - e_{i,j} \tag{14}$$

where $e_{i,j}$ is an element at row i and column j in the matrix $E \in R^{V\times V}$ that represents the distance between every pair of nodes. To calculate $e_{i,j}$, we first assume the input takes the form of $X \in R^{V\times C}$. Then, we have $e_{i,j} = \|\bar{x}_i - \bar{x}_j\|_2$, where $\|\bar{x}_i - \bar{x}_j\|_2$ represents the Euclidean spatial distance between nodes i and j in X. Finally, subtracting $e_{i,j}$ from the maximum value in matrix E defines the adjacency relationship between nodes i and j, implying that nodes closer together have higher adjacency weights. Let $Y_{EDGC} \in R^{V\times C_{out}}$ be the output features of EDGC, the EDGC operator can be formulated as:

$$Y_{EDGC} = EDGC(X) = A_{ED} X W_{ED}^T \tag{15}$$

Where $W_{ED} \in R^{C_{out}\times C}$ is the trainable weight used to facilitate feature updating in the EDGC.

3.3.3 Graph convolution based on self-attention mechanism

In addition to EDGC, we also propose a novel module based on the self-attention mechanism for graph convolution (SAGC) to derive context-dependent intrinsic topology. Specifically, SAGC employs self-attention (Vaswani et al., 2017) on node features to infer intrinsic topology and uses topology as neighborhood vertex information for graph convolutions. A self-attention is an attention mechanism that relates different brain nodes. Considering all possible node relations, SAGC infers positive bounded weights, termed self-attention map, to represent the strength of relationships. For a given SAGC input $X \in R^{V\times C}$, we linearly project node representations X to the query and key of D dimensions with learnable matrices $W_Q, W_K \in R^{C\times D}$ to obtain a self-attention map, as shown in Eq.16.

$$A_{SA} = softmax\left(\frac{XW_K(XW_Q)^T}{\sqrt{D}}\right) \tag{16}$$

Where *softmax* is used to normalize the self-attention map, D is the output channel size and $D = \frac{C}{8}$. The scaling factor $\frac{1}{\sqrt{D}}$ is used to ensure even distribution of data and avoid elements with large values in the self-attention map having small gradients during backpropagation, which could hinder the training of neural network. Then, let $Y_{EDGC} \in R^{V\times C_{out}}$ be the output features of SAGC, the SAGC operator can be formalized as:

$$Y_{SAGC} = SAGC(X) = A_{SA} X W_{SA}^T \tag{17}$$

Where $W_{SA} \in R^{C_{out}\times C}$ is the trainable weight used to facilitate feature updating in the SAGC.

3.4 Spatial attention mechanism based on average pooling and max pooling

After extracting spatial features, to retain crucial spatial node information and eliminate redundancy, we generate a spatial attention map based on the inter-spatial relationships between features. We design a spatial attention mechanism based on average pooling and max pooling (AM-SAM) to achieve this. Different from the channel attention, the spatial attention focuses on "where" is an informative part, which is complementary to the channel attention. Given an intermediate feature map $F \in R^{C\times H\times W}$ as input, to compute the spatial attention map, we first apply average pooling and max pooling operations along the channel axis of F and concatenate them to generate an efficient feature descriptor. On the concatenated feature descriptors, we apply a multilayer perceptron (MLP) to generate the spatial attention map, which encodes emphasis or suppression of locations. The schematic diagram of AM-SAM is illustrated in Figure 5, and the detailed operational description of AM-SAM is as follows.

We aggregate channel information of a feature map by using two pooling operations, generating two 2D maps: $F^s_{max} \in R^{1\times H\times W}$ and $F^s_{avg} \in R^{1\times H\times W}$, which denotes average-pooled features and max-pooled features across the channel respectively. F^s_{max} and F^s_{avg} are first concatenated and flattened into $F^s_{fla} \in R^{2HW\times 1\times 1}$, which is then passed through a multilayer perceptron (MLP) with a hidden layer. To reduce computational resource consumption, the hidden layer size is set to $\frac{D}{r}$, where $D = 2 \times H \times W$ and r is a reduction factor, set to 4 in our study. After obtaining the MLP's output, we use unflatten and nonlinear activation operation to transform the output into a two-dimensional spatial attention map. In short, the spatial attention is calculated as:

$$\begin{aligned} M_s(F) &= \sigma(MLP([MaxPool(F); AvgPool(F)])) \\ &= \sigma(W_1(ReLU(W_0([MaxPool(F); AvgPool(F)])))) \end{aligned} \tag{18}$$

Where $[\cdot]$ denotes concatenation operation, σ is sigmoid function, $W_0 \in R^{\frac{D}{r}\times D}$ and $W_1 \in R^{\frac{D}{2}\times\frac{D}{r}}$. It is worth noting that $[\cdot]$ and W_1 are followed by flatten and unflatten operations, respectively. The output F_{out} of AM-SAM can be formulated as:

$$F_{out} = M_s(F) \odot F \tag{19}$$

4 Experiment

4.1 Method comparison

To better demonstrate the advancement of the AMD-GCN model, we compared it with the state-of-the-art methods on the SEED-VID dataset. Since the codes for these models was not publicly available, we followed the descriptions provided in the original papers for replication, so the final test results might differ. Here, PSD, DE, and WPCA represent different types of features extracted from the raw EEG signals. For the KNN classifier, we set the number of neighbors to 3. The SVM classifier utilized a radial basis function (RBF) kernel for training. EEGNet (Lawhern et al., 2018) is a single CNN architecture capable of accurately classifying EEG signals from various brain-machine interface paradigms.

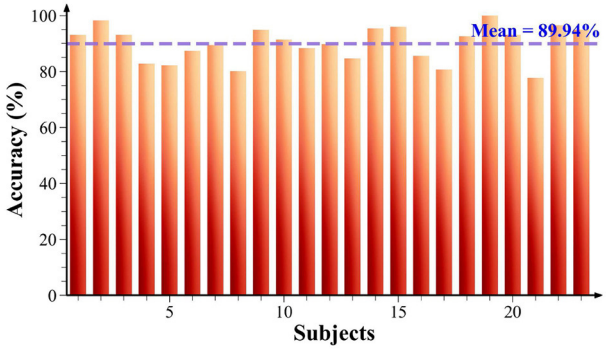


FIGURE 6
Fatigue detection accuracy of 23 subjects in the SEED-VIG dataset.

TABLE 2 Comparison with accuracy and individual variation of state-of-the-art methods on the SEED-VIG dataset.

Method	Accuracy (%)	IV (Individual variation)
DE-KNN	77.37	15.45
PSD-SVM (Barua et al., 2019)	77.64	20.41
DE-SVM (Barua et al., 2019)	78.60	19.10
WPCA-SVM (Dong et al., 2019)	79.71	17.69
EEGNet (Lawhern et al., 2018)	84.50	13.24
ESTCNN (Gao et al., 2019)	86.55	11.23
SAT-IFDM (Hwang et al., 2021)	85.28	11.50
LPCCs + R-SCM (Chen et al., 2022)	87.10	8.07
PDC-GCN (Zhang et al., 2020)	89.42	10.22
GCNN-LSTM (Yin et al., 2021)	89.31	10.45
AMD-GCN (Ours)	**89.94**	**6.14**

The bold values represent the best accuracy and individual variation.

ESTCNN (Gao et al., 2019) is a spatio-temporal CNN model that emphasizes the temporal dependencies of each electrode and enhances the ability to extract spatial information from EEG signals. SAT-IFDM (Hwang et al., 2021) is a subject-independent model for classifying driver fatigue states, aimed at mitigating individual differences among subjects. LPCCs + R-SCM (Chen et al., 2022) is a novel psychological fatigue detection algorithm based on multi-domain feature extraction and fusion. It employs linear prediction to fit the current value with a set of past samples to calculate linear predictive cepstral coefficients (LPCCs) as temporal features. PDC-GCN (Zhang et al., 2020) has been introduced in the section slowromancapi@. GCNN-LSTM (Yin et al., 2021) is a model that combines GCN and LSTM. The model uses GCN for feature extraction and processes the obtained features using LSTM, followed by classification using dense layers. The chosen models for comparison are relatively representative and reproducible. Figure 6 presents the fatigue detection accuracy of all subjects using the AMD-GCN model on the SEED-VIG dataset, and the results of model comparisons are reported in Table 2.

TABLE 3 Experimental results of ablation study on the SEED-VIG dataset, where w/o indicates the removal of specific functional module.

Method	Accuracy (%)	IV (Individual variation)
w/o SEED-VIG-5band	$87.19^{\downarrow 2.75}$	$7.06^{\uparrow 0.92}$
w/o SEED-VIG-2Hz	$86.28^{\downarrow 3.66}$	$7.61^{\uparrow 1.47}$
w/o AM-CAM	$86.47^{\downarrow 3.47}$	$7.56^{\uparrow 1.42}$
w/o MD-GC	$82.64^{\downarrow 7.30}$	$9.44^{\uparrow 3.30}$
w/o AM-SAM	$87.98^{\downarrow 1.96}$	$6.81^{\uparrow 0.67}$
w/o SRGC	$88.03^{\downarrow 1.91}$	$6.75^{\uparrow 0.61}$
w/o EDGC	$86.65^{\downarrow 3.29}$	$7.33^{\uparrow 1.19}$
w/o SAGC	$85.92^{\downarrow 4.02}$	$7.89^{\uparrow 1.75}$
AMD-GCN	**89.94**	**6.14**

The down and up arrow indicates a decrease in accuracy and an increase in individual variation after the removal of specific functional modules, respectively. The bold values represent the best accuracy and individual variation.

Obviously, Figure 6 shows that the detection accuracy is 77.74% for 21-th subject, while the detection accuracy for the remaining participants is all above 80%, and even 19-th subject achieved 100% accuracy. This indicates that the AMD-GCN model possesses great generalization capabilities and has the potential to achieve fatigue detection for a wide range of drivers. As can be seen in Table 2, our proposed AMD-GCN model has an accuracy improvement of about 10.23 $\sim$ 12.57% compared to the traditional machine learning methods (KNN, SVM). Compared to CNN-based methods, the accuracy improvement is about 2.84 $\sim$ 5.44%. Compared with the GCN-based method, the accuracy improvement is about 0.52 $\sim$ 0.63%. The experimental results prove that the performance of the AMD-GCN model outperforms existing detection methods.

4.2 Ablation study

In this section, to further validate the impact of fused features and the role of each module in AMD-GCN, we performed a series of ablation studies, and the experimental results are documented in Table 3. From rows 2, 3, 10 of Table 3, it can be observed that the detection accuracy decreases by 2.75% and 3.66% when SEED-VIG-5band or SEED-VIG-2Hz is removed from the fused features, respectively. This indicates that both SEED-VIG-5band and SEED-VIG-2Hz are indispensable for enhancing the performance of EEG-based driver fatigue detection, and their effects are complementary. Furthermore, the detection accuracy of SEED-VIG-2Hz is higher by 0.91% compared to SEED-VIG-5band, indicating that DE features extracted from 25 frequency bands can better capture the heterogeneity of different fatigue states.

Rows 4, 5, and 6 of Table 3 shows the detection accuracy of the AMD-GCN without the AM-CAM, MD-GC, and AM-SAM functional modules, respectively. Firstly, the AM-CAM module is beneficial to aid the model in focusing on important information related to fatigue detection, and removing the AM-CAM module could introduce noise and confusion to fatigue state detection. The experimental results indicate that AM-CAM contributes to a 3.47% accuracy improvement for the model. Secondly, MD-GC can establish adjacency topologies of numerous semantic patterns, enabling rich non-Euclidean spatial feature learning. Removing MD-GC would disregard functional connections and inherent relationships between EEG nodes, thus weakening the performance of AMD-GCN and reducing the model accuracy by 7.3%. Furthermore, the AM-SAM module can eliminate redundant spatial node information from the output of MD-GC, aiding in enhancing the network's capability to differentiate data from different fatigue states. The experimental results show that AM-SAM contributes to a 1.96% accuracy improvement for the model. In summary, the designed modules successfully enhance the performance of EEG-based driving fatigue detection.

To validate the effectiveness of the adjacency topologies for the three semantic patterns in MD-GC, we obtained the detection accuracy of AMD-GCN without SRGC, EDGC, and SAGC, as described in rows 7, 8, and 9 of Table 3. Apparently, AMD-GCN without SRGC, EDGC, SAGC achieve 88.03%, 86.65%, 85.92%, underperforming the vanilla one by 1.91%, 3.29%, 4.02% respectively. The intrinsic topologies of these semantic patterns are crucial for AMD-GCN to learn category-dependent and data-dependent spatial features, which enhance the performance of AMD-GCN significantly. Moreover, it is evident that the improvements brought by these graph convolutions based on different semantic patterns can be superimposed, implying their roles are complementary to each other.

4.3 Supplement experiment

To verify the reliability of our algorithm, we conducted 10 repeated experiments on the SEED-VIG dataset. In each experiment, the dataset was randomly divided into 5 folds, with one fold used for testing and the remaining four for training, the results are depicted in Figure 7. It can be found that the accuracy varies from 89.62% to 90.37%, and individual variations range from 5.94 to 6.25, this indicates the stability of our method in terms of both detection accuracy and individual variation metrics. Figure 7 presents an average accuracy of 89.94% and an average individual variation of 6.14 for the AMD-GCN, both of which surpass the state-of-the-art methods reported in Table 2. Note that the values reported in Table 2 are average accuracy and average individual variation.

Then, we visualize the channel attention map and spatial attention map of first layer for the first subject under three fatigue states, as shown in Figure 8. Obviously, AM-CAM can achieve channel filtering for inputs with richer semantic information, allowing the model to capture essential parts of the input related to fatigue detection category, and AM-SAM is able to retain crucial spatial node information associated with fatigue states to mitigate interference from redundant information. It can be summarized that our proposed AM-CAM and AM-SAM effectively enhance the feature representation ability of neural network on input data,

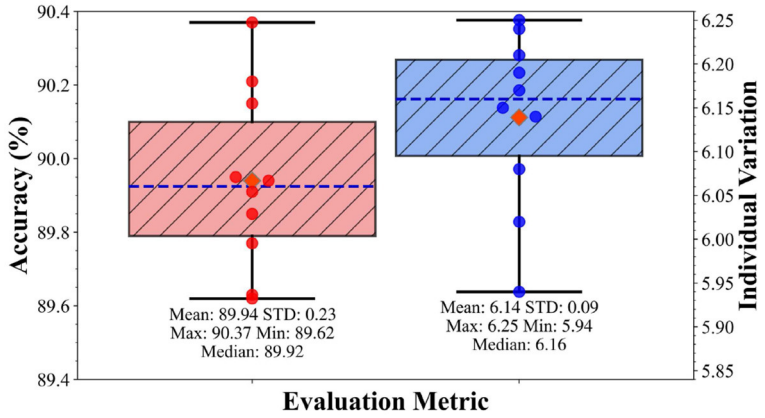


FIGURE 7
Results of 10 repeated experiments. The orange diamond points represent the mean value, the deep blue dashed lines represent the median value, the red and blue scattered points denote the accuracy and individual variations of the repeated experiments, respectively.

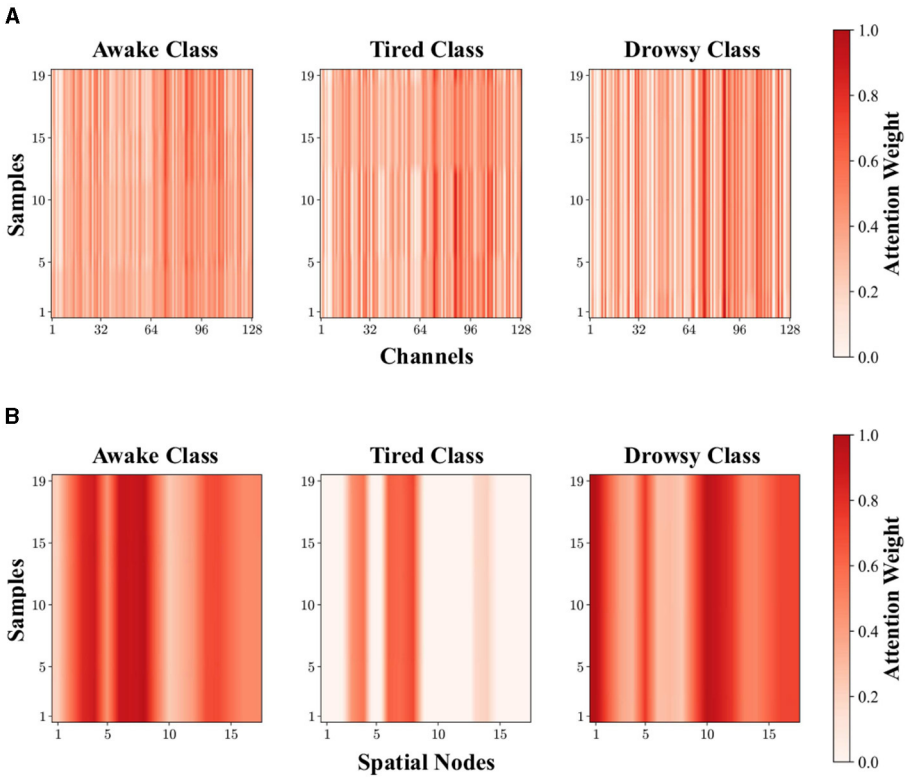


FIGURE 8
The visualization of attention map for the first subject under different fatigue states. **(A)** Channel attention map. **(B)** Spatial attention map of first layer.

thereby improving the performance of EEG-based fatigue detection task.

Furthermore, we visualize the adjacency matrices of the three semantic patterns constructed by AMD-GCN for different subjects, fatigue states, and samples, as shown in Figure 9. This can be concluded that due to SRGC containing a predetermined fixed adjacency graph, it remains consistent for all input data, thereby representing the inherent adjacency between brain nodes. In contrast, EDGC and SAGC construct intrinsic adjacency graphs based on the input data. They exhibit heterogeneity for different subjects, fatigue states, and samples, which benefits AMD-GCN in capturing potential data-dependent intrinsic adjacency relationships between brain nodes. This facilitates AMD-GCN in learning discriminative features for different fatigue states, thus enhancing the performance of driver fatigue detection. Additionally, from the adjacency matrices formed by SRGC, EDGC and SAGC, it can be observed that the adjacency weights among the 6 channels in the temporal region of the brain or the 11 channels in

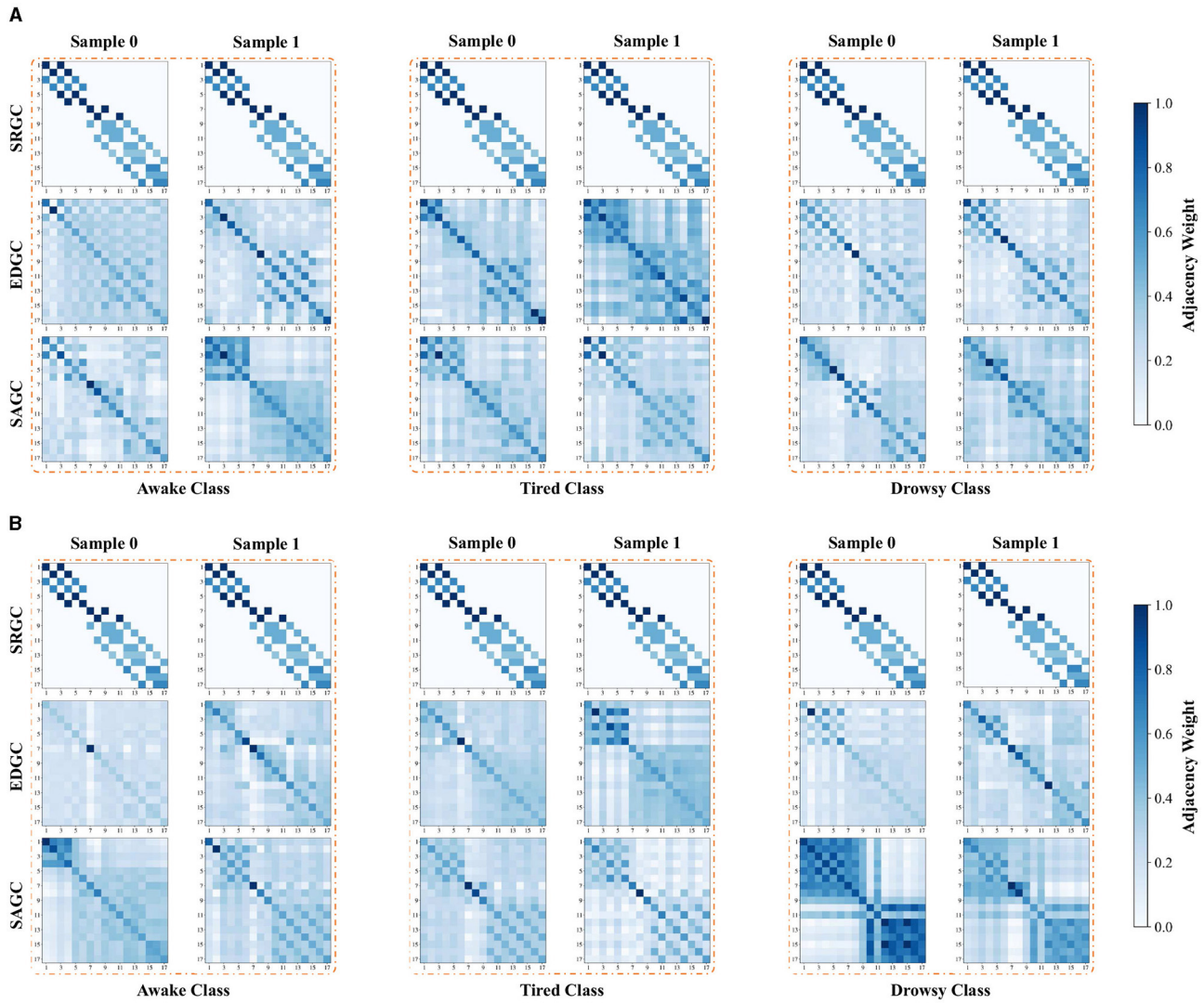


FIGURE 9
In the first layer of AMD-GCN, adjacency matrices for the three semantic patterns corresponding to samples of different fatigue states from two subjects. Sample 0 and Sample 1 represent two samples from the same category of the same subject. **(A)** First subject. **(B)** Second subject.

the posterior region of the brain are significantly stronger than the adjacency weights between the temporal and posterior regions. This consistency aligns with the brain tissue structure. Creating suitable adjacency matrices specifically for the temporal and posterior brain regions is crucial for efficient driver fatigue detection.

5 Conclusion

In this work, we have designed a driving fatigue detection neural network, referred to as the attention-based multi-semantic dynamical graph convolutional network (AMD-GCN), which integrates a channel attention mechanism, a spatial attention mechanism and a graph convolutional network. It aims to classify fused features extracted from EEG signals, where the fused features are obtained by concatenating DE features extracted from 5 frequency bands and DE features extracted from 25 frequency bands. In simple terms, we designed a channel attention mechanism based on average pooling and max pooling (AM-CAM), the mechanism helps the network retain crucial features in the input data that are relevant to driving fatigue detection. We introduced a multi-semantic dynamical graph convolution (MD-GC) that constructs intrinsic adjacency matrices for numerous semantic patterns based on input data., this enhancement improves the GCN's ability to learn non-Euclidean spatial features. We established a spatial attention mechanism (AM-SAM) based on average pooling and max pooling, enabling the network to eliminate redundant spatial node information from MD-GC outputs. Ultimately, we evaluated the performance of AMD-GCN on the SEED-VIG dataset, and the experimental results demonstrated the superiority of our algorithm, outperforming state-of-the-art methods in driving fatigue detection.

The limitations of the proposed AMD-GCN model are summarized from two aspects.

1) Although AMD-GCN model showed superior performance over existing deep learning models on the SEED-VIG dataset,

its network architecture is still a shallow one which limits its feature learning ability in characterizing the underlying properties of EEG data.
2) We find significant differences in the recognition results of different subjects, indicating the existence of individual differences in the driving fatigue detection task. This has not yet been considered by AMD-GCN.
3) The outstanding performance of AMD-GCN is only evident in the subject-dependent experiments, but its performance has not been assessed in the subject-independent experiments.

As our future work, first, we intend to extend AMD-GCN into a deeper architecture to further enhance its data representation learning capacity. Second, we will investigate knowledge transfer strategies to mitigate cross-subject discrepancies in EEG-based driving fatigue detection. Third, we will utilize the leave-one-subject-out cross-validation strategy to evaluate the performance of AMD-GCN in subject-independent experiments on the large-scale fatigue detection dataset. Moreover, we plan to collect EEG fatigue data from numerous subjects and generate simulated volume conduction effect data for each subject, which aims to construct a novel fatigue detection dataset, to examine whether the learning process of the adjacency matrix by AMD-GCN from the raw EEG signals is influenced by spurious correlations introduced by volume conduction effects. We will also apply AMD-GCN to other physiological signals and adopt a combination of multiple physiological signals to comprehensively assess the driver's fatigue state.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://bcmi.sjtu.edu.cn/home/seed/seed-vig.html.

Author contributions

HL: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing—original draft. QL: Data curation, Formal analysis, Investigation, Methodology, Writing—original draft. MC: Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing—review & editing. KC: Data curation, Investigation, Visualization, Writing—review & editing. LM: Software, Validation, Writing—review & editing. WM: Data curation, Formal analysis, Resources, Writing—review & editing. ZZ: Conceptualization, Data curation, Validation, Writing—review & editing. QA: Data curation, Validation, Visualization, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (52275029 and 52075398), Natural Science Foundation of Hubei Province (2022CFB896), State Key Laboratory of New Textile Materials and Advanced Processing Technologies (FZ2022008), and the Fundamental Research Funds for the Central Universities (2023CG0611, 2023-VB-035).

Conflict of interest

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

Publisher's note

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

References

Abidi, A., Ben Khalifa, K., Ben Cheikh, R., Valderrama Sakuyama, C. A., and Bedoui, M. H. (2022). Automatic detection of drowsiness in EEG records based on machine learning approaches. *Neural Process. Lett.* 1-25. doi: 10.1007/s11063-022-10858-x

Barua, S., Ahmed, M. U., and Ahlström, C. S. (2019). Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert Syst. Appl.* 115, 121–135. doi: 10.1016/j.eswa.2018.07.054

Brin, S., and Page, L. (1998). "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the 10th international conference on World Wide Web*, 107–117, doi: 10.1016/S0169-7552(98)00110-X

Cai, Q., Gao, Z., An, J., Gao, S., and Grebogi, C. (2020). A graph-temporal fused dual-input convolutional neural network for detecting sleep stages from EEG signals. *IEEE Trans. Circuits Syst. II*, 68, 777–781. doi: 10.1109/TCSII.2020.3014514

Chen, K., Liu, Z., Liu, Q., Ai, Q., and Ma, L. (2022). EEG-based mental fatigue detection using linear prediction cepstral coefficients and Riemann spatial covariance matrix. *J. Neural Eng.* 19, 066021. doi: 10.1088/1741-2552/aca1e2

Chen, S., Tian, D., Feng, C., and Vetro, A., Kovacčvić, J. (2018). Fast resampling of three-dimensional point clouds via graphs. *IEEE Trans. Signal Process.*, 66, 666–681. doi: 10.1109/TSP.2017.2771730

Chen, W., Wang, W., Wang, K., Li, Z., Li, H., Liu, S., et al. (2020). Lane departure warning systems and lane line detection methods based on image processing and semantic segmentation: a review. *J. Traffic Transp. Eng.* 7, 748–774. doi: 10.1016/j.jtte.2020.10.002

Dinges, D. F., and Grace, R. (1998). *PERCLOS? a valid psychophysiological measure of alertness as assessed by psychomotor vigilance. US Department of Transportation, Federal Highway Administration, Publication Number FHWA*-MCRT-98-006. US Department of Transportation; Federal Highway Administration.

Dong, N., Li, Y., Gao, Z., Ip, W. H., and Yung, K. L. A. (2019). WPCA-based method for detecting fatigue driving from EEG-based internet of vehicles system. *IEEE Access*, 7, 124702–124711. doi: 10.1109/ACCESS.2019.2937914

Gao, Z., Dang, W., Wang, X., Hong, X., Hou, L., Ma, K., et al. (2021). Complex networks and deep learning for EEG signal analysis. *Cogn. Neurodyn*. 15, 369–388. doi: 10.1007/s11571-020-09626-1

Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., et al. (2019). EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. *IEEE Trans. Neural Netw. Learn. Syst*. 30, 2755–2763. doi: 10.1109/TNNLS.2018.2886414

Huang, R., Wang, Y., Li, Z., Lei, Z., and Xu, M. (2022). Multi-granularity deep convolutional model based on feature recalibration and fusion for driver fatigue detection. *IEEE Trans. Intell. Transp. Syst*. 23, 630–640. doi: 10.1109/TITS.2020.3017513

Hwang, S., Park, S., Kim, D., Lee, J., and Byun, H. (2021). "Mitigating inter-subject brain signal variability for EEG-based driver fatigue state classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*. Toronto: IEEE, 990–994.

Jia, H., Xiao, Z., and Ji, P. (2023). End-to-end fatigue driving EEG signal detection model based on improved temporal-graph convolution network. *Comp. Biol. Med*. 152, 106431. doi: 10.1016/j.compbiomed.2022.106431

Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., et al. (2021). Multiview spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Trans. Neural Syst. Rehabil. Eng*. 29, 1977–1986. doi: 10.1109/TNSRE.2021.3110665

Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., et al. (2020). GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. *IJCAI*. 1324–1330. doi: 10.24963/ijcai.2020/184

Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv*. doi: 10.48550/arXiv.1609.02907

Ko, W., Jeon, E., Jeong, S., and Suk, -I. H. (2021). Multi-scale neural network for EEG representation learning in BCI. *IEEE Comput. Intell. Mag*. 16, 31–45. doi: 10.1109/MCI.2021.3061875

Lal, S. K., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. Biol. *Psychol*. 55, 173–194. doi: 10.1016/S0301-0511(00)00085-5

Lal, S. K., and Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* 39, 313–321. doi: 10.1017/S0048577201393095

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., Lance, B. J., et al. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng*. 15, 056013. doi: 10.1088/1741-2552/aace8c

Li, Z., Li, S. E., Li, R., Cheng, B., and Shi, J. (2017). Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors* 17, 3, 495. doi: 10.3390/s17030495

Lin, B., Huang, Y., and Lin, B. (2019). Design of smart EEG cap. *Comp. Methods Prog. Biomed*. 178, 41–46. doi: 10.1016/j.cmpb.2019.06.009

Papadelis, C., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I., Koufogiannis, D., Bekiaris, E., et al. (2006). "Indicators of sleepiness in an ambulatory EEG study of night driving," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. New York, NY: IEEE, 6201–6204.

Paulo, J. R., Pires, G., and Nunes, U. J. (2021). Cross-subject zero calibration driver's drowsiness detection: Exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification. *IEEE Trans. Neural Syst. Rehabil. Eng*. 29, 905–915. doi: 10.1109/TNSRE.2021.3079505

Peng, B., Zhang, Y., Wang, M., Chen, J., and Gao, D. (2023). T-A-MFFNet: Multi-feature fusion network for EEG analysis and driving fatigue detection based on time domain network and attention network. *Comp. Biol. Chem*. 104, 107863. doi: 10.1016/j.compbiolchem.2023.107863

Quddus, A., Shahidi Zandi, A., Prest, L., and Comeau, F. J. (2021). Using long short term memory and convolutional neural networks for driver drowsiness detection. Accid. Anal. *Prev*. 156, 106107. doi: 10.1016/j.aap.2021.106107

Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., and Kording, K. P. (2017). The need to approximate the use-case in clinical machine learning. *GigaScience* 6:gix019. doi: 10.1093/gigascience/gix019

Shi, J., and Wang, K. (2023). Fatigue driving detection method based on Time-Space-Frequency features of multimodal signals. *Biomed. Signal Proc. Control*. 84, 104744. doi: 10.1016/j.bspc.2023.104744

Song, T., Zheng, W., Song, P., and Cui, Z. (2020). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comp*. 11, 532–541. doi: 10.1109/TAFFC.2018.2817622

Song, X., Yan, D., Zhao, L., and Yang, L. (2022). LSDD-EEGNet: an efficient end-to-end framework for EEG-based depression detection. *Biomed. Signal Process. Control*, 75, 103612. doi: 10.1016/j.bspc.2022.103612

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*, 5998–6008.

WHO (2009). *Global Status Report on Road Safety: Time for Action*. Geneva: World Health Organization.

Wu, E. Q. Deng, P., Qiu, X., Tang, Z., Zhang, W., Zhu, L., et al. (2021). Detecting fatigue status of pilots based on deep learning network using EEG signals. *IEEE Trans. Cogn. Dev. Syst*. 13, 575–585. doi: 10.1109/TCDS.2019.2963476

Wu, Y., and Ji. Q. (2019). landmark detection: a literature survey. *Int. J. Comput. Vis*. 127, 115–142. doi: 10.1007/s11263-018-1097-z

Yao, C. L., and Lu, B. L. (2020). "A robust approach to estimating vigilance from EEG with neural processes," in *IEEE International Conference on Bioinformatics and Biomedicine*, Seoal: IEEE, 1202–1205.

Yin, Y., Zheng, X., Hu, B., Zhang, Y., and Cui, X. (2021). EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput*. 100, 106954. doi: 10.1016/j.asoc.2020.106954

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *Procedings of European Conference on Computer Vision (ECCV)* Computer Vision (ECCV).

Zeng, H., Li, X., Borghini, G., Zhao, Y., Aricò, P., Di Flumeri, G., et al. (2021). An EEG-based transfer learning method for cross-subject fatigue mental state prediction. *Sensors* 21, 2369. doi: 10.3390/s21072369

Zhang, T., Wang, X., Xu, X., and Chen, C. P. (2019). GCB-Net: Graph convolutional broad network and its application in emotion recognition. *IEEE Trans. Affect. Comput*. 13, 379–388. doi: 10.1109/TAFFC.2019.2937768

Zhang, W., Wang, F., Wu, S., Xu, Z., Ping, J., Jiang, Y., et al. (2020). Partial directed coherence based graph convolutional neural networks for driving fatigue detection. *Rev. Sci. Instrum*. 91, 074713. doi: 10.1063/5.0008434

Zhang, Y., Guo, R., Peng, Y., Kong, W., Nie, F., Lu, L. B., et al. (2022). An auto-weighting incremental random vector functional link network for EEG-based driving fatigue detection. *IEEE Trans. Instrument. Measure*, 71, 1–14. doi: 10.1109/TIM.2022.3216409

Zheng, W., and Lu, L. B. (2017). A. multimodal approach to estimating vigilance using EEG and forehead EOG. *J. Neural Eng*. 14, 026017. doi: 10.1088/1741-2552/aa5a98

Zhu, J., Jiang, C., Chen, J., Lin, X., Yu, R., Li, X., et al. (2022). based depression recognition using improved graph convolutional neural network. *Comput. Biol. Med*. 148, 105815. doi: 10.1016/j.compbiomed.2022.105815

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 21 December 2023
DOI 10.3389/fnins.2023.1288882

OPEN ACCESS

EDITED BY
Da Ma,
Wake Forest University, United States

REVIEWED BY
Xiaoxiao Wang,
University of Science and Technology of China, China
Yuanqiang Zhu,
Fourth Military Medical University, China

*CORRESPONDENCE
Li-Zhuang Yang
✉ lzyang@cmpt.ac.cn;
Hai Li
✉ hli@cmpt.ac.cn

RECEIVED 05 September 2023
ACCEPTED 01 December 2023
PUBLISHED 21 December 2023

CITATION
Lang J, Yang L-Z and Li H (2023) TSP-GNN: a novel neuropsychiatric disorder classification framework based on task-specific prior knowledge and graph neural network. *Front. Neurosci.* 17:1288882. doi: 10.3389/fnins.2023.1288882

TSP-GNN: a novel neuropsychiatric disorder classification framework based on task-specific prior knowledge and graph neural network

Jinwei Lang[1,2], Li-Zhuang Yang[1,3]* and Hai Li[1,3]*

[1]Anhui Province Key Laboratory of Medical Physics and Technology, Institute of Health and Medical Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China, [2]University of Science and Technology of China, Hefei, China, [3]Hefei Cancer Hospital, Chinese Academy of Sciences, Hefei, China

Neuropsychiatric disorder (ND) is often accompanied by abnormal functional connectivity (FC) patterns in specific task contexts. The distinctive task-specific FC patterns can provide valuable features for ND classification models using deep learning. However, most previous studies rely solely on the whole-brain FC matrix without considering the prior knowledge of task-specific FC patterns. Insight by the decoding studies on brain-behavior relationship, we develop TSP-GNN, which extracts task-specific prior (TSP) connectome patterns and employs graph neural network (GNN) for disease classification. TSP-GNN was validated using publicly available datasets. Our results demonstrate that different ND types show distinct task-specific connectivity patterns. Compared with the whole-brain node characteristics, utilizing task-specific nodes enhances the accuracy of ND classification. TSP-GNN comprises the first attempt to incorporate prior task-specific connectome patterns and the power of deep learning. This study elucidates the association between brain dysfunction and specific cognitive processes, offering valuable insights into the cognitive mechanism of neuropsychiatric disease.

KEYWORDS

neuropsychiatric disorders, task-specific prior knowledge, brain decoding, functional connectivity, graph neural network

1 Introduction

Neuropsychiatric disorder (ND) defines a wide range of psychiatric symptoms accompanying specific emotional, memory, social, or other cognitive impairments (Eddy, 2019; Porcelli et al., 2019; Jahn et al., 2021). Different subtypes of diseases, such as attention-deficit/hyperactivity disorder (ADHD) (Zepf et al., 2019), autism spectrum disorder (ASD) (Vaidya et al., 2020; Wadhera and Kakkar, 2020), and schizophrenia (SZ) (Ioakeimidis et al., 2022; Riedel et al., 2022) show abnormal brain activity during specific task context compared to healthy controls. Mental disorder diagnosis using neuroimaging and machine learning is thus promising (Lanillos et al., 2020; Perez et al., 2021).

Recent years have seen explosive growth in applying deep learning to facilitate ND classification (Chan et al., 2019; Canario et al., 2021; Liu et al., 2021). Previous studies often use brain functional connectivity (FC) or graph theory features (Farahani et al., 2019) and build convolutional neural networks (CNNs) for disease classification (Kim et al., 2016; Guo et al., 2017). However, brain networks are generally irregular and non-Euclidean structures, which can be better captured by graph neural networks (GNNs) than CNNs (Parisot et al., 2017; Zhang et al., 2020; Li L. et al., 2021; Li X. et al., 2021; Zhao et al., 2022). The benefit of GNN is due to the peculiarities of the message-passing mechanism on the graph (Ying et al., 2019). A pioneering study by Parisot and colleagues integrated the FC matrix and phenotype information to construct a sparse graph that captures participants' relationships (Parisot et al., 2017). Subsequently, various graph structures (Li L. et al., 2021) and graph modules, such as graph pooling (Li X. et al., 2021) and even dynamic graph strategies (Zhao et al., 2022), have been proposed, significantly enhancing GNN models for neuropsychiatric disease classification. These models utilize node pooling or edge convolution layers to selectively aggregate important node features, thereby providing insights into relevant diseases from a regional perspective within the brain. For example, default mode network (DMN) and memory-associated brain regions have been identified as biological markers of ASD (Li X. et al., 2021), while damage to the DMN associated with occipital and frontal lobes may explain ADHD (Zhao et al., 2022).

The whole-brain resting-state FC matrix contains redundant and spurious correlations because of confounding or collider effects (Sanchez-Romero and Cole, 2021). It is thus valuable to extract and define distinct connectivity patterns specific to certain cognitive contexts. Recent studies have demonstrated that task-state FC patterns play an essential role in dynamically reshaping brain networks and modulating the flow of neural activity during task performance (Cole et al., 2021; Hearne et al., 2021). These task-related changes in brain network activity provide valuable prior knowledge for understanding the mechanisms underlying brain disorders (Briend et al., 2019; Xia et al., 2019; Kofler et al., 2020; Riedel et al., 2022). However, previous research on ND classification often overlooked this valuable prior information (Gupta et al., 2022; Jiang et al., 2022).

Decoding studies on brain-behavior relationships provide an insightful framework (Jiang et al., 2020; Finn, 2021). We hypothesize that incorporating prior knowledge of task-specific connectivity patterns can improve the performance of ND classification. Motivated by the underlying association between brain decoding and disease diagnosis, the present study seeks to integrate task-specific prior (TSP) knowledge (task-specific functional connectivity) and GNN into a ground-breaking framework for detecting neuropsychiatric disease, dubbed TSP-GNN. We use the Elastic-Net regression model to decode task-specific brain connectome patterns from task-state fMRI in healthy people. Then, task-specific connectome patterns were migrated to illness classification using resting-state fMRI. Finally, we build a population-based graph convolution network to detect brain disease in two neuropsychiatric datasets. The brain decoding approach reduces the dimension of the brain network while providing interpretive information relevant to the task context. Our results demonstrate that task-specific connectome improves disease categorization compared to whole-brain nodes and sheds light on the relationship between brain pathology and specific cognitive processes.

2 Materials and methods

2.1 Participants

2.1.1 HCP dataset

The Human Connectome Project (HCP) (Van Essen et al., 2013) is a remarkable and widely available dataset aimed at defining the anatomical and functional interconnection of the human brain. This dataset contains high-resolution structural MRI, resting-state fMRI, task fMRI scans, and detailed behavioral information for over 1,000 healthy individuals. Subjects completed seven scanner tasks: motor execution, language, emotion, social cognition, working memory (WM), relational, and gambling-related processes. The seven tasks, which lasted for about 20–30 frames under different conditions during each block, and the detailed task paradigm were described in Supplementary Table S1.

2.1.2 Neuropsychiatric dataset

The present study consisted of two datasets, ADHD[1] and ABIDE,[2] for the investigation of disease classification. The ADHD dataset consists of eight cohorts of structural MRI and resting-state fMRI scans (Bellec et al., 2017). Similarly, the ABIDE dataset has the same acquisition modalities from 20 data sites (Cameron et al., 2013). To address the potential impact of heterogeneity in equipment and scanning parameters across different sites, we selected five data sites for the ADHD dataset and three for the ABIDE dataset. Demographic information for the two datasets mentioned above can be found in Table 1.

2.2 fMRI data preprocessing

To ensure the reproducibility of our investigation, we utilized preprocessed fMRI results from ConnectomeDB as a basis for our subsequent analysis. We applied restricted data usage to exclude any influence of inter-individual synchronization among participants within the same family, and finally, 473 unrelated individuals were included. Additionally, we obtained two neuropsychiatric datasets that offered a standard preprocessing workflow. These datasets were directly accessible from their respective data buckets. The preprocessing of fMRI data involves numerous steps to clean and standardize the data prior to statistical analysis. All preprocessing is conducted using fMRIPrep (Esteban et al., 2019), a best-in-breed workflow that ensures high-quality preprocessing to address the challenges of robust and reproducible fMRI data preparation. The minimal preprocessing steps defined by fMRIPrep include motion correction, field unwarping, normalization, bias field correction, and brain extraction.

Subsequently, we conducted a fist-level analysis on each task-state fMRI within HCP using the general linear model (GLM). Our study used the '3dDeconvolve' command in AFNI v20.3.02 to perform first-level GLM analysis. Specifically, the '-stim_times_FSL' parameter was used to specify the timing of stimulus events, while the '-stim_file' parameter was employed to include six head motion parameters. The

1 https://preprocessed-connectomes-project.org/adhd200/

2 https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html

TABLE 1 Demographic and clinical characteristics of ADHD and ABIDE datasets.

Clinical Phenotype	HCP	ADHD			ABIDE		
	n = 473	TD (*n* = 239)	ADHD (*n* = 220)	*P* Value	TD (*n* = 201)	ASD (*n* = 155)	*P* value
Age (years)	28.8 ± 3.69	11.2 ± 2.58	10.9 ± 2.48	0.315	15.05 ± 5.24	14.21 ± 4.32	0.110
Gender (M/F)	227/246	122/117	164/56	< 0.001	164/37	134/21	0.218
FIQ	–	–	–	–	110.67 ± 12.77	107.29 ± 15.94	0.032
PIQ	–	–	–	–	107.58 ± 12.62	103.89 ± 15.63	0.017
VIQ	–	–	–	–	109.44 ± 12.91	106.98 ± 16.32	0.126

Age value computed using two-sample Student's t-test with two tails; Gender value computed using chi-square test; FIQ, Full-scale IQ; PIQ, Performance IQ; VIQ, Verbal IQ.

'-mask' parameter was also used to specify the brain mask generated by fMRIprep. The total number of stimuli '-num_stimts' represented the sum of task conditions and head motion directions. All these parameters collectively constitute the design matrix for each task type, which consists of columns for each condition, nuisance variables, and a constant term, with rows corresponding to each time point of the fMRI data acquisition. The specifics of the design matrix vary according to the exact nature and timing of the task conditions within each of the seven tasks in the HCP dataset. After GLM analysis, we obtained the distribution of brain activation under different task conditions and the purified fMRI time series, devoid of noise signals from task events and motion parameters, which can enable us to investigate the neural correlates of the tasks accurately (Spencer et al., 2022).

We utilized the '3dNetCorr' command by AFNI v20.3.02 to calculate the FC matrixes for both HCP and neuropsychiatric datasets based on the fMRI time series residual preprocessed by GLM. The command will calculate the correlation matrix between the time series of each pair of ROIs defined by parameter '-in_rois.' The average time series and the functional connections between brain regions can be found in the destination file. The atlas adopted in our research was the Brainnetome Atlas (Fan et al., 2016), which has been extensively employed in various clinical studies (Li et al., 2020; Lee et al., 2021). The atlas consists of 246 distinct brain areas that have been carefully delineated. These brain regions can be parcellated into eight functional subsystems (Jiang et al., 2020; Lee et al., 2021). For more details on the names of brain regions in the atlas and their corresponding network allocation, please refer to Supplementary Table S2.

2.3 HCP behavioral performance

Due to the HCP dataset consisting of seven task fMRI scans covering various cognitive abilities, we employed corresponding performance measures as markers of these abilities. For the social task, we used the ratio of precious divided by the median response time (median_RT) under random mode. Working memory ability was evaluated using the accuracy (Acc) divided by the Median_RT score under the 2-back conditions. Emotion reflection performance was assessed using the Acc/Median_RT ratio. In the language task, the story condition was selected to indicate language competence, as performance under both story and math conditions showed a substantial association. However, no significant performance-related markers were detected for the gambling and motor tasks. We used the delay discounting measure to approximate the gambling task performance involving impulsive decision-making. Specifically, we calculated the difference in the area under the curve (AUC) scores between DDisc_AUC_40k and DDisc_AUC_200 as the gambling task score (Cai et al., 2020). A smaller AUC value indicates a higher degree of decision impulsivity. For the motor task, which does not quantitatively reflect participants' athletic ability, we substituted the endurance measure obtained from the NIH Toolbox 2-Minute Walk Test.

In addition to the task-based fMRI, we considered resting-state fMRI, which reflects a baseline state of cognitive ability without task requirements. We utilized general ability (intelligence) measures related to reasoning, problem-solving, abstract thinking, planning, and learning. These measures, which reflect individual cognitive skills like brain fingerprint, were combined into a general factor score using exploratory factor analysis (Dubois et al., 2018; Thiele et al., 2022). Task performance indicators and their corresponding calculations for all fMRI tasks mentioned above can be found in Supplementary Table S3.

2.4 Task-specific functional connectome decoding based on corresponding behavioral performance

Acknowledging the advantages of the task-state connectome in predicting cognitive traits, we constructed eight models to decode task-specific brain connectome patterns across various fMRI tasks. By incorporating task performance as a driving factor, we aimed to reveal the brain connectivity patterns that contribute to cognitive traits and potentially improve our understanding of the neural mechanisms underlying these traits. Considering the superior performance of classical linear regression methods in terms of computational efficiency and their ability to capture complex brain-behavior relationships (Sui et al., 2020; Kim et al., 2021), we developed a task performance-driven brain decoding model utilizing the Elastic-net algorithm:

$$\min_{\beta} \sum_{i=1}^{n} \left(f(x_i) - y_i \right)^2 + \lambda \sum_{j=1}^{p} \left(\alpha \left| \beta_j \right| + \frac{1}{2} (1-\alpha) \left\| \beta_j \right\|^2 \right) \quad (1)$$

The Elastic-net algorithm is known for handling high-dimensional data and selecting relevant features. The above formula, λ represents the weight coefficient of the linear regression and regularization terms, while α determines the balance between the L1 (Lasso regression) and L2 (Ridge regression) norms. For $\alpha = 0$, the model is equivalent to ridge regression, and for $\alpha = 1$, it becomes equivalent to lasso regression. The weight coefficients assigned to the features in the

Elastic-Net model can quantify the contribution of FC pairs between different brain regions to predicting cognitive traits. To construct our brain functional decoding models, we tailored them for each specific fMRI state (as depicted in the top half of Figure 1). Initially, we screened out edges highly correlated with connectome strength. Subsequently, we employed a 10-fold cross-validation approach to creating regression models to decipher task-specific connectivity patterns. By aggregating the non-zero coefficients obtained from each fold in the Elastic-Net model, we obtained a functional subnetwork that best reflected the specificity of the given task (Caunca et al., 2021). To assess the reliability of the prediction outputs, we combined the predictors from each fold and performed a permutation test. Specifically, we calculated the Pearson correlation coefficient between the predicted and observed (random shuffled) scores. The permutation test probability was determined by evaluating the frequency of correlation coefficients in a set of 10,000 permutations that exceeded the initial coefficient.

2.5 Graph theory measures the connectome

Changes in graph theory measures of brain connectome have been recognized as significant aspects of various brain diseases (Savanth et al., 2022). By quantifying the graph-theoretical properties, researchers can gain insights into the essential brain regions and unravel the underlying organizational principles of the brain network (Fallahi et al., 2021; Zhang T. et al., 2021; Zamani et al., 2022). Our investigation included several graph theory measures as supplementary features for disease classification. These measures, namely graph strength, clustering coefficient, local efficiency, page rank centrality, betweenness centrality, eigenvector, flow coefficient, and k-coreness centrality, were calculated based on binary or weighted graphs after implementing a sparsity threshold (Wang B. et al., 2022). The ideal sparse brain graphs were constructed by optimizing the global brain efficiency, and the graph theory features extracted from the corresponding task-specific brain nodes.

2.6 Task-specific prior-knowledge graph neural network model

The population and brain parcellation methods are two commonly used GNN frameworks for diagnosing brain diseases. The population graph methodology involves constructing a graph representation at the population level (Parisot et al., 2017, 2018), while the brain-level graph methodology focuses on building graphs based on individual brain connectivity patterns (Felouat and Oukid, 2020; Wang L. et al., 2021). In our study, we employed a population GNN for further computations after decoding task-specific brain regions (as shown in the bottom half of Figure 1). We chose the population GNN approach due to its superior classification performance demonstrated in previous studies (Pan J. et al., 2022). Neuropsychological scale score, gender, or age were considered as the set of non-imaging phenotypic features $N = (N_h)$. The adjacency weights of the population graph were defined as follows:

$$W(x,y) = Sim(A_x, A_y) \sum_{h=1}^{H} \gamma(N_h(x), N_h(y)) \quad (2)$$

where $Sim(A_x, A_y)$ is a similarity measure between subjects x and y, γ is the distance between phenotypic measures. For every category in h, we adopt a threshold θ and define γ as a unit-step function:

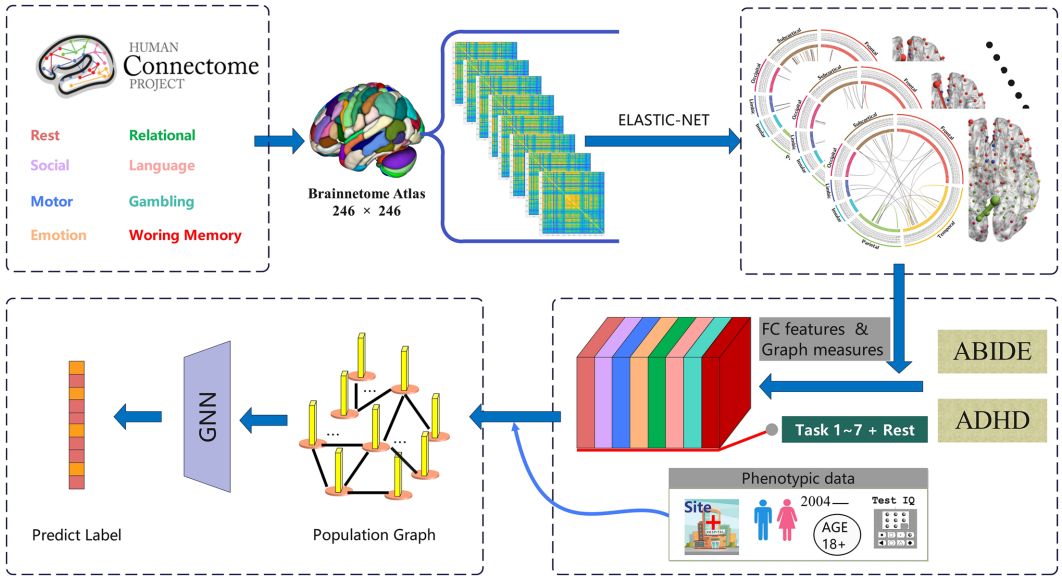


FIGURE 1
The architecture overview of the proposed TSP-GNN framework which combines task-specific patterns for disease diagnosis. The top half of the framework decodes task patterns based on cognitive performance, while the bottom extracts task-specific functional connectome and graph theoretical measures from various disease datasets. Subsequently, phenotypic information is integrated to construct the population-based graph neural network to achieve disease classification.

$$\gamma\left(N_h(x),N_h(y)\right)=\begin{cases}1, if\ \left|N_h(x)-N_h(y)\right|<\theta\\ 0, \text{otherwise}\end{cases} \quad (3)$$

The similarity of graph features was defined as:

$$Sim\left(A_x,A_y\right)=\exp\left(-\frac{\left[\rho\left(F(x),F(y)\right)\right]^2}{2\sigma^2}\right) \quad (4)$$

Where ρ is the correlation distance, and σ determines the width of the kernel. Due to network connectivity and graph theory measures based on the interconnected nodes on both sides of the edges to form subnetworks, the features remain in a relatively high dimension. We adopt a ridge classifier to perform recursive feature elimination (RFE) with a fixed number of features (Ravishankar et al., 2016). In the graph convolutional component of the TSP-GNN model, the normalized graph Laplacian function of a weighted graph $\mathcal{G}=\left(\{\mathcal{V},\mathcal{E},W\}\right)$ is defined as $\mathcal{L}=I_N-D^{-1/2}WD^{1/2}$ where I_N and D are, respectively, the identity matrix of size $N*N$ and diagonal degree matrix. The GNN architecture is derived from (Parisot et al., 2018), consists with L fully convolutional hidden layers activated using the Rectified Linear Unit (ReLU) function.

$$g_\theta * \mathrm{X}=g_\theta(\mathcal{L})\mathrm{X}=g_\theta\left(U\Lambda U^T\right)\mathrm{X}=Ug_\theta(\Lambda)U^T\mathrm{X} \quad (5)$$

The input layer encompasses the entire population graph, while a SoftMax activation function follows the output layer. To evaluate the performance of our model, we employed a five-fold cross-validation approach across all databases. During training, the training fold consisted of a subset of tagged graph nodes, the loss function was assessed, and gradients were backpropagated on this subset.

2.7 Compare with other classification methods

The current research comprehensively compared the TSP-GNN method with various machine learning techniques, deep learning models, and graph neural networks. Specifically, the comparison included support vector machine (SVM), K-nearest neighbor (KNN), and several ensemble learning methods. In addition, we included two deep neural networks (DNN) methods, namely multilayer perceptron (MLP) and convolutional neural networks (CNN). The MLP method, a supervised feedforward neural network which consists of one hidden layer, was connected to the stacked autoencoder (Parisot et al., 2018). The CNN method uses the most classical design, using dropout and linear layers to achieve reduction and forecast. As for the GNN model, we employed MAGE and EV-GNN, which have demonstrated superior performance in previous studies. It is worth noting that the original MAGE utilizes a variety of brain atlas features to improve the accuracy of disease diagnosis (Wang Y. et al., 2022). We adopted this concept in our paper to effectively integrate relevant prior information from multiple task modalities. Additionally, the EV-GNN model demonstrated the ability to automatically integrate imaging data and phenotype data within a learnable adaptive population graph (Huang and Chung, 2020).

3 Results

3.1 Functional connectivity patterns of different cognitive tasks

Our study demonstrates that brain connectivity patterns exhibit both task-specific characteristics and commonalities. We observed that the decoded edges traverse multiple functional brain regions and are distributed across various intrinsic resting-state networks (RSNs), indicating shared patterns across different tasks. The assessment metrics presented in Table 2 indicate the strength of the decoding results, with all expected correlation coefficients (r values) exceeding 0.3 and the corresponding value of ps being less than 0.05. Notably, we found that the prediction models for all tasks passed the permutation test, confirming the reliability and consistency of our decoding results (Figure 2). In addition to the permutation test, we employed several evaluation measures to assess the performance of the decoding models. These measures included the mean squared error (MSE), explained variance score (EVS), and mean absolute error (MAE). By examining these metrics, we gained further insights into the accuracy and precision of our prediction models.

3.2 Anatomical and functional localization of task-specific network edges

Significant interconnections were identified by analyzing the non-zero coefficients in the Elastic-Net model. Our analysis results revealed the most prominent interconnections associated with each task state, with the following number of edges identified: emotion (47 edges), gambling (46 edges), language (21 edges), motor task (15 edges), relational (27 edges), social (44 edges), working memory (22 edges), and rest (99 edges). Importantly, it was observed that the seven task-specific regions were widely distributed across different anatomical locations, and the number of specific edges involved in rest-state fMRI was greater than that in task fMRI. A circular diagram has depicted the distribution of the essential connected edges of social cognition and gambling tasks (Figure 3). The specific connectivity

TABLE 2 Prediction and evaluations of various cognitive abilities.

Task	*r* value	Value of *p*	R_2	MSE	EVS	MAE
W	0.400	0.017*	0.125	0.851	0.141	0.730
S	0.489	0.044*	0.205	0.741	0.228	0.685
L	0.394	0.019*	0.147	0.845	0.151	0.747
E	0.445	0.018*	0.169	0.803	0.186	0.725
R	0.383	0.016*	0.112	0.873	0.141	0.734
M	0.337	0.043*	0.097	0.887	0.108	0.713
G	0.432	0.006**	0.165	0.816	0.183	0.740
REST	0.418	0.011*	0.149	0.824	0.161	0.715

W, working memory; E, emotion processing; L, language; S, social cognitive; R, relation processing; M, motor; G, gambling.

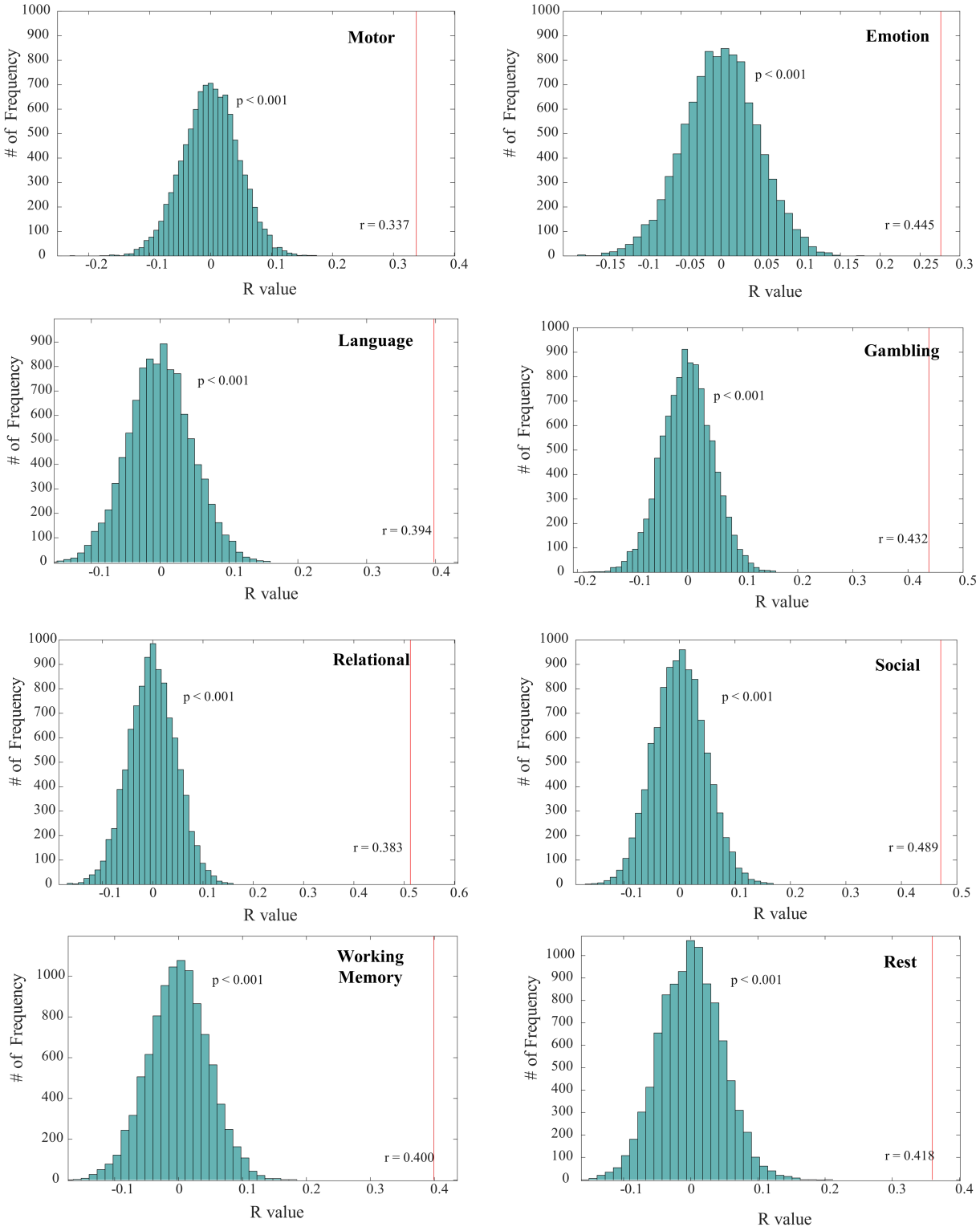


FIGURE 2
Results of permutation tests on task-state and resting-state fMRI decoding. The green histograms illustrate the correlation values' distribution between the predicted task performances and those obtained from 10,000 permutation tests. The red line marks the correlation from the predictions of the original Elastic-Net model to the actual outcomes, clearly showing that the permutation test outcomes systematically register below the baseline correlation.

distribution patterns of the brain networks for the other five tasks and resting-state fMRI are presented in Supplementary Figures S1, S2.

The social task-related FC patterns were distributed inter-LIM-VIS, LIM-SUB, VAN-SUB networks, and intra-DMN and SUB networks. In the gambling task, participants were asked to guess the number of a mystery card. Decoding results showed significant ROIs, such as inter-insular subsystem, angular gyrus (IPL_L_6_2), supramarginal gyrus (IPL_L_6_3), superior parietal lobule (SPL),

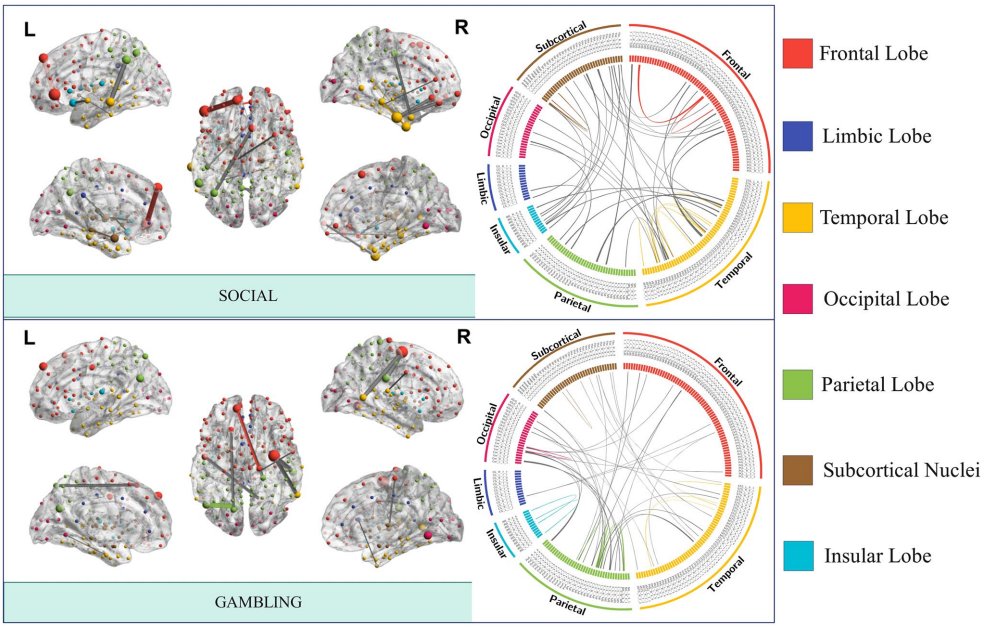


FIGURE 3
FCs with the best task performance prediction capability. The nodes and edges of the brain network are created by averaging the FC strength of a particular task across all people, and the strength determines the node size and edge thickness. Connections within a module are depicted using the same color as the module in which it is situated, whereas gray lines represent inter-module connections.

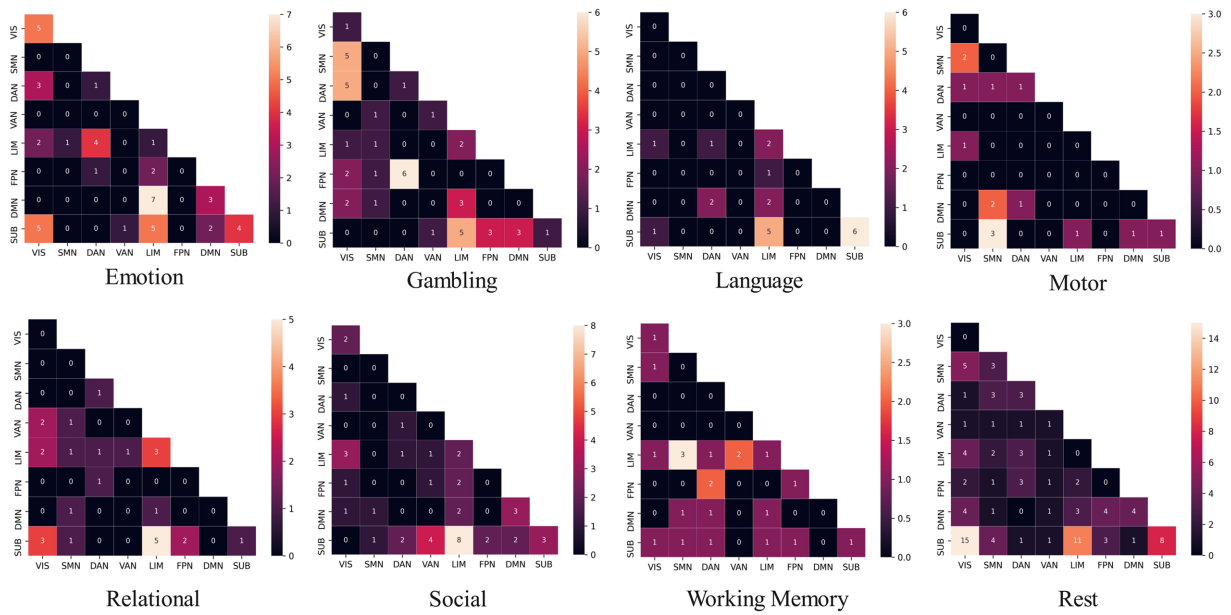


FIGURE 4
The distribution of functional brain networks associated with edges differs across decoding modes of task states. DAN, dorsal attention network; DMN, default mode network; FPN, frontoparietal network; LIM, limbic network; SMN, somatomotor network; SUB, subcortical network; VAN, ventral attention network; VIS, visual network.

precuneus (Pcun_L_4_1), right cuneus (Cun_R_5_3, Cun_R_5_4). From the RSN perspective, brain edges related to gambling or risk decision were mainly distributed inter- SMN-VIS, DAN-VIS, DAN-FPN, and LIM-SUB networks (Figure 4), indicating a broader cross-network interaction. In the resting state, the FC pattern has the highest number of brain edges and almost exhibits the highest proportion of connections within brain anatomical locations. Resting-state fMRI predicts general intelligence, which includes reasoning, problem-solving, abstract thinking, planning, and learning, in our model.

3.3 Task-specific brain connectivity for disease classification

To evaluate the impact of task-specific prior knowledge on brain disease classification, we extracted a subnetwork comprising all the nodes involved in the task-based connectome. Additionally, we incorporated graph-theoretical properties of these task-specific nodes derived from binary and weighted brain network analyzes. These steps allowed us to amalgamate FC strength with graph metrics, culminating in a refined set of input features for the GNN model. This particular methodology facilitated a comprehensive exploration of the influence of prior knowledge on disease categorization. Notably, demographics and behavioral statistics are also incorporated into the construction of the population graph. Table 3 presents the classification performance ranking by task paradigm of each dataset. The findings suggest that the classification of different types of mental illnesses exhibited a preference for the specific task prior knowledge. As two prevalent neurodevelopmental disorders, ASD and ADHD frequently co-occur. Interestingly, they exhibited distinct task preferences in classification tasks. For ADHD, task-specific features related to social and relational processing tasks can achieve higher classification accuracy. In contrast, the ABIDE dataset has shown that gambling, motor, and relational processing are the top three task-specific patterns that yielded the best classification performance.

TABLE 3 The implications of priori information decoded by different tasks on neuropsychiatric disease classification.

ADHD			ABIDE		
TASK	AUC	ACC	TASK	AUC	ACC
M	0.697	0.653	S	0.670	0.652
W	0.700	0.653	REST	0.696	0.655
L	0.705	0.632	W	0.723	0.702
G	0.705	0.636	E	0.724	0.680
REST	0.705	0.649	L	0.728	0.722
E	0.705	0.658	R	0.734	0.688
R	0.711	0.680	M	0.739	0.711
S	0.720	0.651	G	0.760	0.670

3.4 Investigate the categorization effect of various task combination models

We further conducted task-specific prior knowledge experiments on disease classification to evaluate previous task information's influence on disease classification and investigate if information complementarity between tasks may enhance diagnosis performance. We selected four, five, and six tasks from seven different task categories to create diverse combinations, C_7^4, C_7^5 and C_7^6. We presented the top three ranking AUC results for each combination of task quantities, as shown in Table 4. Our findings reveal that C_7^4 yields the best classification performance, whereas increasing the accurac y of C_7^5 and C_7^6.

From the perspective of the classification effect of the combination mode, brain diseases exhibit differential task combination preferences. Specifically, the combination of M_R_S_W achieved the best classification results on the ADHD dataset. Not exactly consistently, the combination of E_G_S_W performed best on the ABIDE dataset. Compared with single-task experiments, the classification performance is slightly improved by selecting task-specific information for combinations. Additionally, the types of tasks frequently appearing in the 4-task combination also perform well in single-task experiments.

We displayed the task-specific brain node interactions effect for best task combinations under ADHD and ABIDE datasets (Figure 5). The best task combinations for these two diseases involve working memory and social cognition. In ADHD, social cognition and working memory tasks contribute the most nodes, whereas gambling and social cognition do in ABIDE. Table 5 shows that when all ROIs are included, i.e., FC features (246 * 245/2 = 30,135) or graph theory features (15 attributes, 246 * 15 = 3,690), the classification accuracy decreases, further highlighting the superiority of task-specific nodes.

3.5 Comparison results with other baseline models

In this present investigation, various machine learning and deep learning methods were used to illustrate the superiority of the TSP-GNN model in ND diagnosis. To ensure the uniformity of input

TABLE 4 The effects of task decoding information combination patterns on neuropsychiatric disease classification.

	ADHD			ABIDE		
	TASK group	AUC	ACC	TASK group	AUC	ACC
Task_4	E_M_S_W	0.722	0.666	G_L_R_W	0.740	0.691
	E_M_R_S	0.723	0.660	G_M_R_W	0.754	0.705
	M_R_S_W	0.724	0.671	E_G_S_W	0.759	0.702
Task_5	E_G_L_R_S	0.721	0.669	G_L_M_R_W	0.738	0.716
	L_M_R_S_W	0.721	0.662	E_G_R_S_W	0.740	0.670
	E_G_M_R_S	0.721	0.656	E_G_L_M_S	0.741	0.705
Task_6	E_L_M_R_S_W	0.712	0.662	E_G_M_R_S_W	0.722	0.680
	E_G_L_M_R_S	0.715	0.656	E_G_L_M_R_S	0.725	0.677
	G_L_M_R_S_W	0.720	0.680	E_G_L_M_S_W	0.728	0.694
Task_7	G_L_M_R_S_W_E	0.712	0.659	G_L_M_R_S_W_E	0.732	0.677

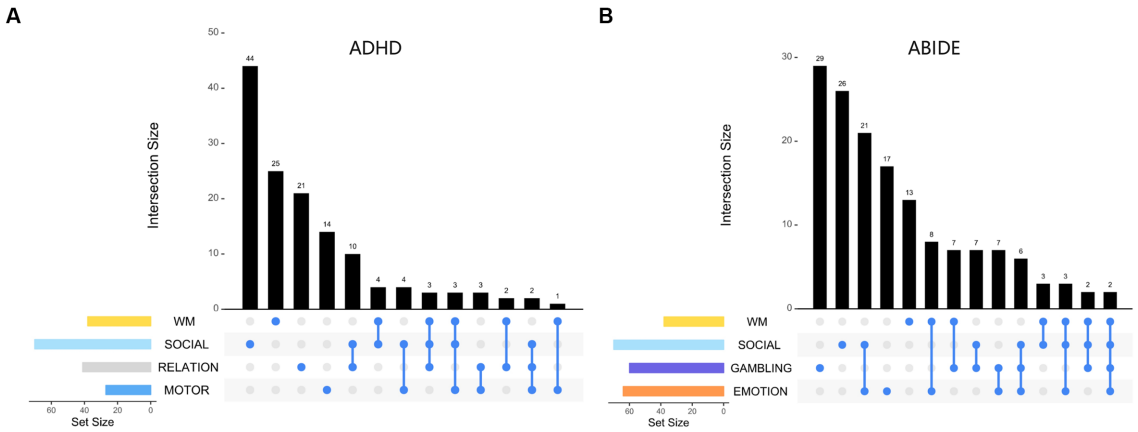


FIGURE 5
Visualizing the distribution set of nodes involved in the optimal task combination: **(A)** for ADHD dataset, **(B)** for ABIDE dataset. The findings indicate that the decoded results (node distribution) are relatively independent, with a low proportion of nodes belonging to the intersection of multiple tasks.

TABLE 5 Comparing the classification performance of task-specific features and whole brain features during two datasets.

	ADHD		ABIDE	
	AUC	ACC	AUC	ACC
All FCs	0.706	0.649	0.693	0.671
All Graph Measures	0.675	0.630	0.738	0.719
FCs + Graph Measures	0.663	0.621	0.739	0.716
Best TSP-GNN	0.724	0.671	0.759	0.702

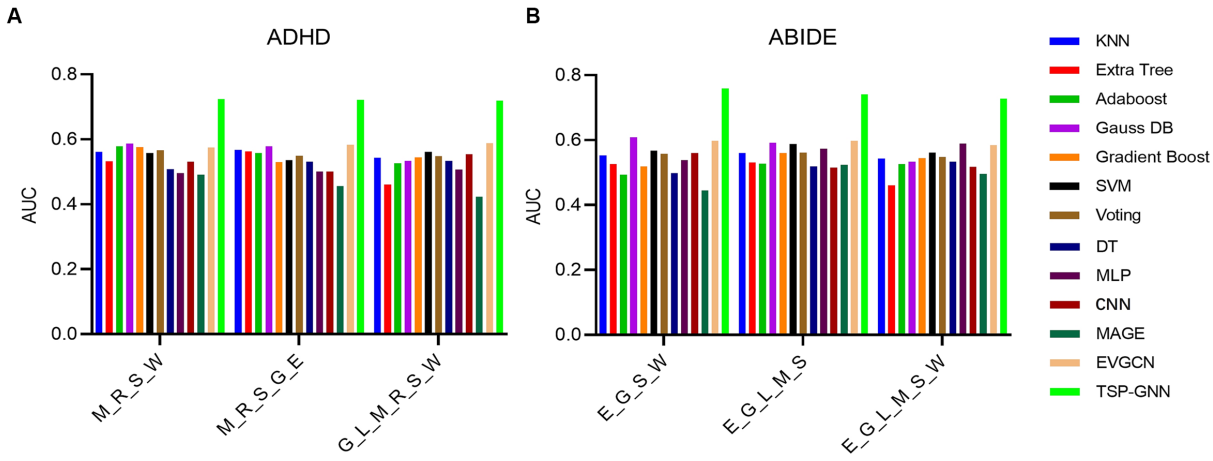


FIGURE 6
On the ADHD **(A)** and ABIDE **(B)** datasets, the classification performance of the TSP-GNN framework was compared to that of various machine learning and deep learning methods. The task priors used were combinations of the best four, five, and six combinations described in section 3.3. M_R_S_W stands for a task combination of motor, relational processing, social cognitive, and working memory tasks, and the remaining acronyms are similar.

features, we conducted experiments using the optimal task combination stated in section 3.4. In the classification experiments of ADHD and ABIDE datasets, TSP-GNN has obtained the optimal results (Figure 6), and the detailed numerical values of the classification results can be found in Supplementary Tables S2, S3. In comparison to classic machine learning approaches such as SVM (Abraham et al., 2017) and ensemble learning (Liu et al., 2020), GNN models the individual-based topologies structure (Zhou and Zhang, 2021) between subjects utilizing participant similarity, which is advantageous for enhancing classification performance. After numerous layers of graph convolution computation, highly relevant characteristics are continually aggregated (Wang L. et al., 2021). MLP

and CNN apply fully-connected and convolutional layers to achieve dimensionality reduction on brain network features, which are spatial topological graphs between brain areas and cannot be equated to the image receptive field (Dvornek et al., 2017; Khosla et al., 2018). The TSP-GNN architecture blends multi-task information from FC characteristics and graph measures to collect better and characterize the most discriminative information than typical machine learning and deep neural network models.

4 Discussions

This study represents the first investigation in brain disease classification that focuses explicitly on task-specific FC patterns. Task-based fMRI offers distinct advantages in exploring and understanding the mechanisms and brain-behavior relationships specific to cognitive impairments, which may not be evident in resting-state fMRI. Task paradigms provide structured cognitive engagement (Jiang et al., 2020; Yoo et al., 2022), allowing for a better examination of individual differences in critical neural circuits (Greene et al., 2018). Given the advantages of task fMRI, we employed the Elastic-Net regression model to explore task-specific FC patterns decoded relying on brain-behavior relationships. Additionally, we used resting-state fMRI to decode general intelligence as a baseline for comparison with task-specific FC (Dubois et al., 2018; Thiele et al., 2022; Anderson and Barbey, 2023). The decoding results for different tasks exhibited high heterogeneity, highlighting the brain regions and connectivity patterns that are more representative of the current task, in contrast to traditional supervised models based on task labels alone (Zhang et al., 2022).

In decoding brain-behavior relationships, selecting predictors and outcomes for the predictive model is a topic worthy of exploration. While predictions about various behaviors can be made based on resting-state data, our research prioritizes focus on the relationship between task-state fMRI and corresponding cognitive performance under task scenarios. The predictive modeling based on task-state fMRI is inspired by the potential of task-state FC to enhance cognitive outcome prediction (Jiang et al., 2020). Additionally, it fully explores the multiple task states within the HCP dataset. We consider utilizing the Acc/RT ratio as a behavioral index for tasks with accuracy and response speed metrics in predicting behavioral performance. Literature also conceptualizes the trade-off between speed and accuracy as 'throughput' (Thorne, 2006; Heitz, 2014). It reflects the accuracy of the response and its rapidity, thereby providing a composite measure of cognitive processing efficiency.

Our research corroborates the efficacy of integrating task-specific connectome priors into classification models for diagnosing a spectrum of psychiatric disorders across various datasets. Specifically, enhanced classification performance is observed in differentiating diseases when utilizing FC patterns associated with specific cognitive domains (Chauvin et al., 2021). Network patterns related to working memory tasks contribute significantly to both ADHD and ASD datasets. The previous study also reveals that impairments in working memory are prevalent across psychiatric conditions (Wang X. L. et al., 2021), and memory assessments are crucial for predicting and mitigating high-risk disorders (Seabury and Cannon, 2020). In classifying ADHD, leading tasks also encompass motor task, social cognition, and relational processing. Previous studies have demonstrated that severe declines in social cognition and motor speed (Haining et al., 2020) correlate with a high risk of clinical psychiatric conditions. ADHD is also associated with abnormalities in the large-scale cognitive control network that impact social attention (Fateh et al., 2022), with adolescents among the patient population exhibiting impairments in social cognition and communication abilities (Chen and Chen, 2020). Children with ADHD have a deficit in relational reasoning (Brunamonti et al., 2017), a skill subtending the acquisition of many cognitive abilities and social rules. In the classification of the ABIDE dataset, leading tasks also encompass emotion, gambling, and social cognition. Facial emotion recognition disorder is typical of people with autism. Facial emotion recognition disorder is a classic symptom of autism (Yeung, 2022). Cognitive inflexibility in people with autism appears characterized by the unwillingness to switch toward processing socio-emotional information (Latinus et al., 2019). Individuals with ASD frequently report difficulty making flexible decisions across various contexts to resolve social or moral conflicts(Tei et al., 2022). Concurrently, studies based on gambling paradigms also suggest they tend to exhibit a more cautious decision-making style (Hosozawa et al., 2021).

Integrating multi-task FC and graph theory has further enhanced classification accuracy, achieving optimal performance using four task combinations. However, the addition of features from more tasks did not continue to improve classification results, presenting an intriguing avenue for investigation. In constructing brain FC-based diagnostic models, selecting features is more critical than quantity (Du et al., 2018; Chen et al., 2020). An increased number of features may offer a richer representation of task-specific FC information, but it can also lead to the "curse of dimensionality"—a phenomenon where the introduction of noise, overfitting, and the increased difficulty of identifying meaningful patterns in high-dimensional spaces may decrease classification performance (Wee et al., 2014; Barbieri et al., 2022). Our research also validates that opting for a more suitable selection of features, rather than simply increasing their number, is the superior strategy.

The TSP-GNN system achieves a balanced trade-off between model interpretability and classification performance. In contrast to previous studies that incorporated whole-brain connectome features, our model utilizes a task-specific FC pattern, which enhances the interpretability of features by linking them to specific cognitive activities. Furthermore, the classification stage of the TSP-GNN framework employs a population graph model, simplifying the modeling of brain areas as nodes and improving classification performance. Regarding classification performance, our TSP-GNN outperforms various classical machine learning and deep network models, underscoring the superiority of our task-prioritized population graph model in detecting brain diseases. Although our classification accuracy may differ from recent studies (Chen et al., 2021; Pan J. et al., 2022), this may be due to trade-offs and parameter adjustments made during model construction. Our framework prioritizes the interpretation of cognitive processes and their extended values related to underlying disease, and task-specific prior information from brain areas can be easily transferred to other studies of cognitive brain disease and disorders. In summary, we consider the decoding model in our TSP-GNN framework as a pre-task, effectively reducing feature dimensionality and elucidating the role of task-specific prior information in the classification model for brain disease diagnosis. The model effectively bridges the gap between cognitive

behavior decoding and brain illness research, offering valuable insights and serving as a reference for task-related investigations in brain diseases.

Several considerations need to be addressed in our research. Firstly, it should be acknowledged that the ADHD and ABIDE illness cohorts in our study were not comprehensive and may not represent all available data sources. The inherent imbalance resulting from variations in data collection parameters and equipment across different locations is a significant challenge in our investigation. Constrained by the differing intended uses of data acquisition between HCP and ND, a strict age match between groups was not feasible, thus warranting further investigation into the exclusion of age-related differences in brain network impacts (Zhang et al., 2023). Secondly, the task-specific FC derived from the regression process has enhanced the efficacy of disease diagnosis and is considered, to some extent, correlative rather than causally direct. Employing causal correlation-based FC (Sanchez-Romero et al., 2023) and evidence of neural modulation (Zhou et al., 2020) based on brain networks holds promise for overcoming this limitation. Lastly, our current classification results can be further enhanced by refining the incorporation of prior information and optimizing future models to approach state-of-the-art performance. Continual efforts to improve the quality of prior knowledge and refine model development are necessary to ensure our approach remains at the forefront of research in this field.

Future research aims to develop deep learning models integrating cognitive performance and task state labels for brain decoding. Recognizing the intricate relationship between brain decoding and classification, despite their distinct objectives, we intend to explore the application of zero-shot learning and advanced transfer learning models that can achieve mutual benefits for both brain function decoding and disease classification tasks (Zhang P. et al., 2021). An exciting prospect is the collection of psychiatric disorder data using appropriate task paradigms in clinical settings (Birba et al., 2022). By incorporating task performance in actual clinical circumstances, we can investigate and evaluate the underlying causes of illnesses, expand our prior knowledge about task-based brain activity, and further optimize our models accordingly. Our future endeavors aim to bridge the gap between brain decoding and disease classification by developing advanced deep-learning models informed by clinical data and task performance. This approach has the potential to significantly contribute to the field by providing valuable insights into the underlying mechanisms of brain disorders and facilitating more accurate diagnoses.

5 Conclusion

The present study introduces a novel TSP-GNN framework to improve brain disease classification. By leveraging functional connection-based cognitive performance prediction, this study decodes task-specific FC patterns and transfers them as prior knowledge for diagnosing ND. As far as we know, this study represents the first attempt to transfer task-specific connectivity patterns as a priori knowledge in brain disease research. Our results demonstrate that integrating task-specific priors leads to improved classification accuracy compared to traditional methods. The finding highlights the informativeness of task-specific connection patterns. Besides, the optimal task combinations for each kind of ND offer valuable insights into the underlying mechanisms of that brain disease. By incorporating task-specific connectivity patterns, our framework enhances the understanding and prediction of brain diseases, opening up new avenues for future investigations in this domain.

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

JL: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. L-ZY: Conceptualization, Methodology, Supervision, Writing – review & editing. HL: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We thank all participants for their time and effort. Preliminary analyzes were sponsored by the Natural Science Fund of China (82371931), the Natural Science Fund of Anhui Province (2008085MC69), the Natural Science Fund of Hefei City (2021033), HFIPS Director's Fund (YZJJ202207-TS), the General scientific research project of Anhui Provincial Health Commission (AHWJ2021b150), Collaborative Innovation Program of Hefei Science Center, CAS (2021HSC-CIP013), Anhui Province Key Laboratory of Medical Physics and Technology (LMPT201904).

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1288882/full#supplementary-material

References

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., et al. (2017). Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *Neuroimage* 147, 736–745. doi: 10.1016/j.neuroimage.2016.10.045

Anderson, E. D., and Barbey, A. K. (2023). Investigating cognitive neuroscience theories of human intelligence: a connectome-based predictive modeling approach. *Hum. Brain Mapp.* 44, 1647–1665. doi: 10.1002/hbm.26164

Barbieri, M., Lee, P. K., Brizi, L., Giampieri, E., Solera, F., Castellani, G., et al. (2022). Circumventing the curse of dimensionality in magnetic resonance fingerprinting through a deep learning approach. *NMR Biomed.* 35:e4670. doi: 10.1002/nbm.4670

Bellec, P., Chu, C., Chouinard-Decorte, F., Benhajali, Y., Margulies, D. S., and Craddock, R. C. (2017). The neuro bureau ADHD-200 preprocessed repository. *NeuroImage* 144, 275–286. doi: 10.1016/j.neuroimage.2016.06.034

Birba, A., Fittipaldi, S., Cediel Escobar, J. C., Gonzalez Campo, C., Legaz, A., Galiani, A., et al. (2022). Multimodal neurocognitive markers of naturalistic discourse typify diverse neurodegenerative diseases. *Cereb. Cortex* 32, 3377–3391. doi: 10.1093/cercor/bhab421

Briend, F., Marzloff, V., Brazo, P., Lecardeur, L., Leroux, E., Razafimandimby, A., et al. (2019). Social cognition in schizophrenia: validation of an ecological fMRI task. *Psychiatry Res. Neuroimag.* 286, 60–68. doi: 10.1016/j.pscychresns.2019.03.004

Brunamonti, E., Costanzo, F., Mammi, A., Rufini, C., Veneziani, D., Pani, P., et al. (2017). Evaluation of relational reasoning by a transitive inference task in attention-deficit/hyperactivity disorder. *Neuropsychology* 31, 200–208. doi: 10.1037/neu0000332

Cai, H., Chen, J., Liu, S., Zhu, J., and Yu, Y. (2020). Brain functional connectome-based prediction of individual decision impulsivity. *Cortex* 125, 288–298. doi: 10.1016/j.cortex.2020.01.022

Cameron, C., Yassine, B., Carlton, C., Francois, C., Alan, E., András, J., et al. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Front. Neuroinform.* 7:41. doi: 10.3389/conf.fninf.2013.09.00041

Canario, E., Chen, D., and Biswal, B. (2021). A review of resting-state fMRI and its use to examine psychiatric disorders. *Psychoradiology* 1, 42–53. doi: 10.1093/psyrad/kkab003

Caunca, M. R., Wang, L., Cheung, Y. K., Alperin, N., Lee, S. H., Elkind, M. S. V., et al. (2021). Machine learning-based estimation of cognitive performance using regional brain MRI markers: the northern Manhattan study. *Brain Imaging Behav.* 15, 1270–1278. doi: 10.1007/s11682-020-00325-3

Chan, N. K., Kim, J., Shah, P., Brown, E. E., Plitman, E., Carravaggio, F., et al. (2019). Resting-state functional connectivity in treatment response and resistance in schizophrenia: a systematic review. *Schizophr. Res.* 211, 10–20. doi: 10.1016/j.schres.2019.07.020

Chauvin, R. J., Buitelaar, J. K., Sprooten, E., Oldehinkel, M., Franke, B., Hartman, C., et al. (2021). Task-generic and task-specific connectivity modulations in the ADHD brain: an integrated analysis across multiple tasks. *Transl. Psychiatry* 11:159. doi: 10.1038/s41398-021-01284-z

Chen, M. H., and Chen, Y. L. (2020). Functional connectivity of specific brain networks related to social and communication dysfunction in adolescents with attention-deficit hyperactivity disorder. *Psychiatry Res.* 284:112785. doi: 10.1016/j.psychres.2020.112785

Chen, Y. L., Tu, P. C., Huang, T. H., Bai, Y. M., Su, T. P., Chen, M. H., et al. (2020). Using minimal-redundant and maximal-relevant whole-brain functional connectivity to classify bipolar disorder. *Front. Neurosci.* 14:563368. doi: 10.3389/fnins.2020.563368

Chen, H., Zhuang, F., Xiao, L., Ma, L., Liu, H., Zhang, R., et al. (2021). *AMA-GCN: Adaptive multi-layer aggregation graph convolutional network for disease prediction*. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21).

Cole, M. W., Ito, T., Cocuzza, C., and Sanchez-Romero, R. (2021). The functional relevance of task-state functional connectivity. *J. Neurosci.* 41, 2684–2702. doi: 10.1523/jneurosci.1713-20.2021

Du, Y., Fu, Z., and Calhoun, V. D. (2018). Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Front. Neurosci.* 12:525. doi: 10.3389/fnins.2018.00525

Dubois, J., Galdi, P., Paul, L. K., and Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373:20170284. doi: 10.1098/rstb.2017.0284

Dvornek, N. C., Ventola, P., Pelphrey, K. A., and Duncan, J. S. (2017). *Identifying autism from resting-state fMRI using long short-term memory networks*. Berlin: Springer International Publishing, pp. 362–370.

Eddy, C. M. (2019). What do You have in mind? Measures to assess mental state reasoning in neuropsychiatric populations. *Front. Psych.* 10:425. doi: 10.3389/fpsyt.2019.00425

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116. doi: 10.1038/s41592-018-0235-4

Fallahi, A., Pooyan, M., Lotfi, N., Baniasad, F., Tapak, L., Mohammadi-Mobarakeh, N., et al. (2021). Dynamic functional connectivity in temporal lobe epilepsy: a graph theoretical and machine learning approach. *Neurol. Sci.* 42, 2379–2390. doi: 10.1007/s10072-020-04759-x

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., et al. (2016). The human Brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb Cortex* 26, 3508–3526. doi: 10.1093/cercor/bhw157

Farahani, F. V., Karwowski, W., and Lighthall, N. R. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review. *Front. Neurosci.* 13:585. doi: 10.3389/fnins.2019.00585

Fateh, A. A., Huang, W. X., Mo, T., Wang, X. Y., Luo, Y., Yang, B. R., et al. (2022). Abnormal insular dynamic functional connectivity and its relation to social Dysfunctioning in children with attention deficit/hyperactivity disorder. *Front. Neurosci.* 16:596. doi: 10.3389/fnins.2022.890596

Felouat, H., and Oukid, S. (2020). *Graph convolutional networks and functional connectivity for identification of autism Spectrum disorder*. 2020 Second International Conference on Embedded & Distributed Systems (EDiS).

Finn, E. S. (2021). Is it time to put rest to rest? *Trends Cogn. Sci.* 25, 1021–1032. doi: 10.1016/j.tics.2021.09.005

Greene, A. S., Gao, S. Y., Scheinost, D., and Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* 9:920. doi: 10.1038/s41467-018-04920-3

Guo, X., Dominick, K. C., Minai, A. A., Li, H., Erickson, C. A., and Lu, L. J. (2017). Diagnosing autism Spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. *Front. Neurosci.* 11:460. doi: 10.3389/fnins.2017.00460

Gupta, S., Lim, M., and Rajapakse, J. C. (2022). Decoding task specific and task general functional architectures of the brain. *Hum. Brain Mapp.* 43, 2801–2816. doi: 10.1002/hbm.25817

Haining, K., Matrunola, C., Mitchell, L., Gajwani, R., Gross, J., Gumley, A. I., et al. (2020). Neuropsychological deficits in participants at clinical high risk for psychosis recruited from the community: relationships to functioning and clinical symptoms. *Psychol. Med.* 50, 77–85. doi: 10.1017/s0033291718003975

Hearne, L., Mill, R., Keane, B., Repovs, G., Anticevic, A., and Cole, M. (2021). Activity flow underlying abnormalities in brain activations and cognition in schizophrenia. *Sci. Adv.* 7:eabf2513. doi: 10.1126/sciadv.abf2513

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front. Neurosci.* 8:150. doi: 10.3389/fnins.2014.00150

Hosozawa, M., Mandy, W., Cable, N., and Flouri, E. (2021). The role of decision-making in psychological wellbeing and risky Behaviours in autistic adolescents without ADHD: longitudinal evidence from the UK millennium cohort study. *J. Autism Dev. Disord.* 51, 3212–3223. doi: 10.1007/s10803-020-04783-y

Huang, Y., and Chung, A. C. S. (2020). "Edge-Variational graph convolutional networks for uncertainty-aware disease prediction" in *Medical image computing and computer assisted intervention – MICCAI 2020*. eds. A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga and S. K. Zhou (Berlin: Springer International Publishing), 562–572.

Ioakeimidis, V., Haenschel, C., Fett, A.-K., Kyriakopoulos, M., and Dima, D. (2022). Functional neurodevelopment of working memory in early-onset schizophrenia: a longitudinal FMRI study. *Schizophr. Res. Cogn.* 30:100268. doi: 10.1016/j.scog.2022.100268

Jahn, F. S., Skovbye, M., Obenhausen, K., Jespersen, A. E., and Miskowiak, K. W. (2021). Cognitive training with fully immersive virtual reality in patients with neurological and psychiatric disorders: a systematic review of randomized controlled trials. *Psychiatry Res.* 300:113928. doi: 10.1016/j.psychres.2021.113928

Jiang, Z., Wang, Y., Shi, C., Wu, Y., Hu, R., Chen, S., et al. (2022). Attention module improves both performance and interpretability of four-dimensional functional magnetic resonance imaging decoding neural network. *Hum. Brain Mapp.* 43, 2683–2692. doi: 10.1002/hbm.25813

Jiang, R., Zuo, N., Ford, J. M., Qi, S., Zhi, D., Zhuo, C., et al. (2020). Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage* 207:116370. doi: 10.1016/j.neuroimage.2019.116370

Khosla, M., Jamison, K., Kuceyeski, A., and Sabuncu, M. R. (2018). "3D convolutional neural networks for classification of functional connectomes" in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. eds. D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel and L. Maier-Hein (Berlin: Springer International Publishing), 137–145.

Kim, M., Bao, J., Liu, K., Park, B.-Y., Park, H., Baik, J. Y., et al. (2021). A structural enriched functional network: an application to predict brain cognitive performance. *Med. Image Anal.* 71:102026. doi: 10.1016/j.media.2021.102026

Kim, J., Calhoun, V. D., Shim, E., and Lee, J. H. (2016). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage* 124, 127–146. doi: 10.1016/j.neuroimage.2015.05.018

Kofler, M. J., Soto, E. F., Fosco, W. D., Irwin, L. N., Wells, E. L., and Sarver, D. E. (2020). Working memory and information processing in ADHD: evidence for directionality of effects. *Neuropsychology* 34, 127–143. doi: 10.1037/neu0000598

Lanillos, P., Oliva, D., Philippsen, A., Yamashita, Y., Nagai, Y., and Cheng, G. (2020). A review on neural network models of schizophrenia and autism spectrum disorder. *Neural Netw.* 122, 338–363. doi: 10.1016/j.neunet.2019.10.014

Latinus, M., Cléry, H., Andersson, F., Bonnet-Brilhault, F., Fonlupt, P., and Gomot, M. (2019). Inflexibility in autism Spectrum disorder: need for certainty and atypical emotion processing share the blame. *Brain Cogn.* 136:103599. doi: 10.1016/j.bandc.2019.103599

Lee, J. J., Kim, H. J., Ceko, M., Park, B. Y., Lee, S. A., Park, H., et al. (2021). A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* 27:174-+. doi: 10.1038/s41591-020-1142-7

Li, L., Jiang, H., Wen, G., Cao, P., Xu, M., Liu, X., et al. (2021). TE-HI-GCN: an Ensemble of Transfer Hierarchical Graph Convolutional Networks for disorder diagnosis. *Neuroinformatics* 20, 353–375. doi: 10.1007/s12021-021-09548-1

Li, A., Zalesky, A., Yue, W. H., Howes, O., Yan, H., Liu, Y., et al. (2020). A neuroimaging biomarker for striatal dysfunction in schizophrenia. *Nat. Med.* 26:558. doi: 10.1038/s41591-020-0793-8

Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., et al. (2021). BrainGNN: interpretable brain graph neural network for fMRI analysis. *Med. Image Anal.* 74:102233. doi: 10.1016/j.media.2021.102233

Liu, M., Li, B., and Hu, D. (2021). Autism Spectrum disorder studies using fMRI data and machine learning: a review. *Front. Neurosci.* 15:7870. doi: 10.3389/fnins.2021.697870

Liu, Y., Xu, L., Li, J., Yu, J., and Yu, X. (2020). Attentional connectivity-based prediction of autism using heterogeneous rs-fMRI data from CC200 atlas. *Exp. Neurobiol.* 29, 27–37. doi: 10.5607/en.2020.29.1.27

Pan, J., Dong, Y., and Chen, H. (2022). Review of research on auxiliary diagnosis of autism based on graph neural networks. *Comput. Eng.* 48, 1–11. doi: 10.19678/j.issn.1000-3428.0064352

Pan, J., Lin, H., Dong, Y., Wang, Y., and Ji, Y. (2022). MAMF-GCN: multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder. *Comput. Biol. Med.* 148:105823. doi: 10.1016/j.compbiomed.2022.105823

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., et al. (2018). Disease prediction using graph convolutional networks: application to autism Spectrum disorder and Alzheimer's disease. *Med. Image Anal.* 48, 117–130. doi: 10.1016/j.media.2018.06.001

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Moreno, R., Glocker, B., et al. (2017). *Spectral graph convolutions for population-based disease prediction*. In: International conference on medical image computing and computer-assisted intervention: Springer International Publishing, pp. 177–185.

Perez, D. L., Nicholson, T. R., Asadi-Pooya, A. A., Begue, I., Butler, M., Carson, A. J., et al. (2021). Neuroimaging in functional neurological disorder: state of the field and research agenda. *Neuroimage Clin.* 30:102623. doi: 10.1016/j.nicl.2021.102623

Porcelli, S., Van der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., et al. (2019). Social brain, social dysfunction and social withdrawal. *Neurosci. Biobehav. Rev.* 97, 10–33. doi: 10.1016/j.neubiorev.2018.09.012

Ravishankar, H., Madhavan, R., Mullick, R., Shetty, T., Marinelli, L., and Joel, S. E. (2016). *Recursive feature elimination for biomarker discovery in resting-state functional connectivity*. In: *38th annual international conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC)*, pp. 4071–4074.

Riedel, P., Lee, J. H., Watson, C. G., Jimenez, A. M., Reavis, E. A., and Green, M. F. (2022). Reorganization of the functional connectome from rest to a visual perception task in schizophrenia and bipolar disorder. *Psychiatr. Res. Neuroimag.* 327:111556. doi: 10.1016/j.pscychresns.2022.111556

Sanchez-Romero, R., and Cole, M. W. (2021). Combining multiple functional connectivity methods to improve causal inferences. *J. Cogn. Neurosci.* 33, 180–194. doi: 10.1162/jocn_a_01580

Sanchez-Romero, R., Ito, T., Mill, R. D., Hanson, S. J., and Cole, M. W. (2023). Causally informed activity flow models provide mechanistic insight into network-generated cognitive activations. *NeuroImage* 278:120300. doi: 10.1016/j.neuroimage.2023.120300

Savanth, A. S., Vijaya, P. A., Nair, A. K., and Kutty, B. M. (2022). Classification of Rajayoga meditators based on the duration of practice using graph theoretical measures of functional connectivity from task-based functional magnetic resonance imaging. *Int. J. Yoga* 15, 96–105. doi: 10.4103/ijoy.ijoy_17_22

Seabury, R. D., and Cannon, T. D. (2020). Memory impairments and psychosis prediction: a scoping review and theoretical overview. *Neuropsychol. Rev.* 30, 521–545. doi: 10.1007/s11065-020-09464-2

Spencer, D., Yue, Y. R., Bolin, D., Ryan, S., and Mejia, A. F. (2022). Spatial Bayesian GLM on the cortical surface produces reliable task activations in individuals and groups. *NeuroImage* 249:118908. doi: 10.1016/j.neuroimage.2022.118908

Sui, J., Jiang, R., Bustillo, J., and Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biol. Psychiatry* 88, 818–828. doi: 10.1016/j.biopsych.2020.02.016

Tei, S. S., Tanicha, M., Itahashi, T., Aoki, Y. Y., Ohta, H., Qian, C. Y., et al. (2022). Decision flexibilities in autism spectrum disorder: an fMRI study of moral dilemmas. *Soc. Cogn. Affect. Neurosci.* 17, 904–911. doi: 10.1093/scan/nsac023

Thiele, J. A., Faskowitz, J., Sporns, O., and Hilger, K. (2022). Multitask brain network reconfiguration is inversely associated with human intelligence. *Cereb. Cortex* 32, 4172–4182. doi: 10.1093/cercor/bhab473

Thorne, D. R. (2006). Throughput: a simple performance index with desirable characteristics. *Behav. Res. Methods* 38, 569–573. doi: 10.3758/bf03193886

Vaidya, C. J., You, X., Mostofsky, S., Pereira, F., Berl, M. M., and Kenworthy, L. (2020). Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders. *J. Child Psychol. Psychiatry* 61, 51–61. doi: 10.1111/jcpp.13114

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

Wadhera, T., and Kakkar, D. (2020). Multiplex temporal measures reflecting neural underpinnings of brain functional connectivity under cognitive load in autism Spectrum disorder. *Neurol. Res.* 42, 327–337. doi: 10.1080/01616412.2020.1726586

Wang, X. L., Cheng, B. C., Roberts, N., Wang, S., Luo, Y., Tian, F. F., et al. (2021). Shared and distinct brain fMRI response during performance of working memory tasks in adult patients with schizophrenia and major depressive disorder. *Hum. Brain Mapp.* 42, 5458–5476. doi: 10.1002/hbm.25618

Wang, L., Li, K., and Hu, X. P. (2021). Graph convolutional network for fMRI analysis based on connectivity neighborhood. *Netw. Neurosci.* 5, 83–95. doi: 10.1162/netn_a_00171

Wang, B., Li, L., Peng, L., Jiang, Z., Dai, K., Xie, Q., et al. (2022). Multigroup recognition of dementia patients with dynamic brain connectivity under multimodal cortex parcellation. *Biomed. Sig. Proc. Control* 76:103725. doi: 10.1016/j.bspc.2022.103725

Wang, Y., Liu, J., Xiang, Y., Wang, J., Chen, Q., and Chong, J. (2022). MAGE: automatic diagnosis of autism spectrum disorders using multi-atlas graph convolutional networks and ensemble learning. *Neurocomputing* 469, 346–353. doi: 10.1016/j.neucom.2020.06.152

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

Xia, M., Womer, F. Y., Chang, M., Zhu, Y., Zhou, Q., Edmiston, E. K., et al. (2019). Shared and distinct functional architectures of brain networks across psychiatric disorders. *Schizophr. Bull.* 45, 450–463. doi: 10.1093/schbul/sby046

Yeung, M. K. (2022). A systematic review and meta-analysis of facial emotion recognition in autism spectrum disorder: the specificity of deficits and the role of task characteristics. *Neurosci. Biobehav. Rev.* 133:104518. doi: 10.1016/j.neubiorev.2021.104518

Ying, R., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. J. A. (2019). *GNNExplainer: generating explanations for graph neural networks*. Available at: https://ui.adsabs.harvard.edu/abs/2019arXiv190303894Y (Accessed March 1, 2019).

Yoo, K., Rosenberg, M. D., Kwon, Y. H., Scheinost, D., Constable, R. T., and Chun, M. M. (2022). A cognitive state transformation model for task-general and task-specific subsystems of the brain connectome. *NeuroImage* 257:119279. doi: 10.1016/j.neuroimage.2022.119279

Zamani, J., Sadr, A., and Javadi, A.-H. (2022). Classification of early-MCI patients from healthy controls using evolutionary optimization of graph measures of resting-state fMRI, for the Alzheimer's disease neuroimaging initiative. *PLoS One* 17:608. doi: 10.1371/journal.pone.0267608

Zepf, F. D., Bubenzer-Busch, S., Runions, K. C., Rao, P., Wong, J. W. Y., Mahfouda, S., et al. (2019). Functional connectivity of the vigilant-attention network in children and adolescents with attention-deficit/hyperactivity disorder. *Brain Cogn.* 131, 56–65. doi: 10.1016/j.bandc.2017.10.005

Zhang, Z., Cui, P., and Zhu, W. (2020). Deep learning on graphs: a survey. *IEEE Trans. Knowl. Data Eng.* 34, 249–270. doi: 10.1109/TKDE.2020.2981333

Zhang, Y., Farrugia, N., and Bellec, P. (2022). Deep learning models of cognitive processes constrained by human brain connectomes. *Med. Image Anal.* 80:102507. doi: 10.1016/j.media.2022.102507

Zhang, P., Li, W., Ma, X., He, J., Huang, J., and Li, Q. (2021). Feature-selection-based transfer learning for Intracortical brain–machine Interface decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 60–73. doi: 10.1109/TNSRE.2020.3034234

Zhang, T., Liao, Q., Zhang, D., Zhang, C., Yan, J., Ngetich, R., et al. (2021). Predicting MCI to AD conversation using integrated sMRI and rs-fMRI: machine learning and graph theory approach. *Front. Aging Neurosci.* 13:688926. doi: 10.3389/fnagi.2021.688926

Zhang, B., Zhang, S., Feng, J., and Zhang, S. (2023). Age-level bias correction in brain age prediction. *Neuroimage Clin.* 37:103319. doi: 10.1016/j.nicl.2023.103319

Zhao, K., Duka, B., Xie, H., Oathes, D. J., Calhoun, V., and Zhang, Y. (2022). A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. *NeuroImage* 246:118774. doi: 10.1016/j.neuroimage.2021.118774

Zhou, T., Kang, J., Li, Z., Chen, H., and Li, X. (2020). Transcranial direct current stimulation modulates brain functional connectivity in autism. *Neuroimage Clin.* 28:102500. doi: 10.1016/j.nicl.2020.102500

Zhou, H., and Zhang, D. (2021). *Graph-in-graph convolutional networks for brain disease diagnosis*. In: 2021 IEEE International Conference on Image Processing (ICIP)), pp. 111–115.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 February 2024
DOI 10.3389/fnins.2024.1331677

OPEN ACCESS

EDITED BY
Delia Cabrera DeBuc,
University of Miami, United States

REVIEWED BY
Chao Huang,
Florida State University, United States
Dandan Li,
Taiyuan University of Technology, China

*CORRESPONDENCE
Da Ma
✉ dma@wakehealth.edu
Jane Stocks
✉ janestocks2018@u.northwestern.edu
Lei Wang
✉ lei.wang@osumc.edu

†These authors share first authorship

RECEIVED 01 November 2023
ACCEPTED 08 January 2024
PUBLISHED 07 February 2024

CITATION
Ma D, Stocks J, Rosen H, Kantarci K, Lockhart SN, Bateman JR, Craft S, Gurcan MN, Popuri K, Beg FM and Wang L and (2024) Differential diagnosis of frontotemporal dementia subtypes with explainable deep learning on structural MRI. *Front. Neurosci.* 18:1331677. doi: 10.3389/fnins.2024.1331677

Differential diagnosis of frontotemporal dementia subtypes with explainable deep learning on structural MRI

Da Ma[1*†], Jane Stocks[2*†], Howard Rosen[3], Kejal Kantarci[4], Samuel N. Lockhart[1], James R. Bateman[5], Suzanne Craft[1], Metin N. Gurcan[1], Karteek Popuri[6], Mirza Faisal Beg[7] and Lei Wang[2,8*], on behalf of the ALLFTD consortium

[1]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, United States, [2]Department of Psychiatry and Behavioral Health, Northwestern University Feinberg School of Medicine, Chicago, IL, United States, [3]Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, United States, [4]Department of Radiology, Mayo Clinic, Rochester, MN, United States, [5]Department of Neurology, Wake Forest University School of Medicine, Winston-Salem, NC, United States, [6]Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada, [7]School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, [8]Department of Psychiatry and Behavioral Health, Ohio State University Wexner Medical Center, Columbus, OH, United States

Background: Frontotemporal dementia (FTD) represents a collection of neurobehavioral and neurocognitive syndromes that are associated with a significant degree of clinical, pathological, and genetic heterogeneity. Such heterogeneity hinders the identification of effective biomarkers, preventing effective targeted recruitment of participants in clinical trials for developing potential interventions and treatments. In the present study, we aim to automatically differentiate patients with three clinical phenotypes of FTD, behavioral-variant FTD (bvFTD), semantic variant PPA (svPPA), and nonfluent variant PPA (nfvPPA), based on their structural MRI by training a deep neural network (DNN).

Methods: Data from 277 FTD patients (173 bvFTD, 63 nfvPPA, and 41 svPPA) recruited from two multi-site neuroimaging datasets: the Frontotemporal Lobar Degeneration Neuroimaging Initiative and the ARTFL-LEFFTDS Longitudinal Frontotemporal Lobar Degeneration databases. Raw T1-weighted MRI data were preprocessed and parcellated into patch-based ROIs, with cortical thickness and volume features extracted and harmonized to control the confounding effects of sex, age, total intracranial volume, cohort, and scanner difference. A multi-type parallel feature embedding framework was trained to classify three FTD subtypes with a weighted cross-entropy loss function used to account for unbalanced sample sizes. Feature visualization was achieved through post-hoc analysis using an integrated gradient approach.

Results: The proposed differential diagnosis framework achieved a mean balanced accuracy of 0.80 for bvFTD, 0.82 for nfvPPA, 0.89 for svPPA, and an overall balanced accuracy of 0.84. Feature importance maps showed more localized differential patterns among different FTD subtypes compared to groupwise statistical mapping.

Conclusion: In this study, we demonstrated the efficiency and effectiveness of using explainable deep-learning-based parallel feature embedding and visualization framework on MRI-derived multi-type structural patterns to differentiate three clinically defined subphenotypes of FTD: bvFTD, nfvPPA, and svPPA, which could help with the identification of at-risk populations for early and precise diagnosis for intervention planning.

KEYWORDS

FTD (frontotemporal dementia), differential diagnosis algorithm, explainable deep learning, multi-type features, multi-level feature fusion, bvFTD, nfvPPA, svPPA

1 Introduction

Frontotemporal dementia (FTD) is an umbrella term describing the many clinical syndromes underlain by frontotemporal lobar degeneration (FTLD) neuropathology. FTD is characterized by the progressive impairment of cognitive and behavioral functions such as executive functioning, language, social comportment, and motor functioning (Dickerson and Atri, 2014). FTLD is the third most common cause of dementia and is as common as Alzheimer's disease (AD) in individuals under the age of 65 (Erkkinen et al., 2018). Clinically, FTLD is typically associated with one of several diagnoses characterized by specific constellations of symptoms. Patients who present with early impairments in social comportment and executive dysfunction are typically diagnosed with behavioral-variant FTD (bvFTD). Primary progressive aphasia (PPA) is a clinical syndrome characterized by a selective deterioration of language functions and can be further subdivided into semantic (svPPA) and nonfluent variants (nfvPPA) (Mesulam et al., 2014). Regardless of the initial clinical syndrome, FTD syndromes eventually result in global dementia and death (Mioshi et al., 2010).

Although clinical trials of potential disease-altering therapies (e.g., anti-tau antibodies, tau aggregation inhibitors) are currently underway (Boxer et al., 2013; Tsai and Boxer, 2016; Mis et al., 2017; Logroscino et al., 2019; Panza et al., 2020; Huang et al., 2023), the significant degree of clinical, pathological and genetic heterogeneity observed in FTD hinders the development of sensitive and specific biomarkers that would allow for targeted recruitment of groups at highest risk for clinical/cognitive decline (Katzeff et al., 2022). Critically, early and accurate diagnosis of the clinical syndrome is essential for the targeted recruitment of participants in clinical trials, as treatments will only be effective if patients are accurately diagnosed. In bvFTD, patients show significant gray matter volume loss of the frontal and temporal lobes, with early and most distinctive loss of volume in the insula and anterior cingulate cortex (Seeley et al., 2008; Mandelli et al., 2016; Ranasinghe et al., 2016). Among the PPA syndromes, svPPA is associated with striking asymmetric (typically left > right) atrophy of the temporal pole, while nfvPPA shows atrophy of the left inferior frontal/insular cortex (Agosta et al., 2015). Across FTD clinical phenotypes, the spatial distribution of atrophy is consistent with the constellation of clinical symptoms.

While each FTD clinical syndrome has a typical anatomical pattern of neurodegeneration, early manifestations can vary greatly across people. Moreover, early patterns of neurodegeneration can be highly overlapping across clinical syndromes, such as in the case of anterior temporal lobe atrophy for both svPPA and bvFTD, and inferior frontal and insular atrophy in both bvFTD and nfvPPA. Indeed, Vijverberg et al. (2016) found that a visual review of a single MRI had insufficient sensitivity (70%) to identify cases with bvFTD. Researchers have therefore attempted to employ machine learning methods for pattern analysis to improve the classification and diagnosis of FTD (Ducharme, 2023). Similar research in the field of AD has achieved high accuracy levels when classifying diseased individuals compared to controls (often >90% accuracy) (Falahati et al., 2014; Rathore et al., 2017). Similarly, several studies have demonstrated that machine learning methods can aid in the reliable discrimination of AD and FTD (Ma et al., 2020, 2021). However, the use of machine learning methods for discrimination between FTD syndromes is rarer (see McCarthy et al., 2018 for review), often only covering a few subtypes (Wilson et al., 2009; Bisenius et al., 2017; Di Benedetto et al., 2022). Both Wilson et al. (2009) and Bisenius et al. (2017) classified PPA subtypes against each other using a principal component analysis approach based on gray matter volume, particularly for the comparison of svPPA from nfvPPA, finding moderately high accuracy (89.1%), sensitivity (84.44%) and specificity (93.8%), equivalent to an balanced accuracy of 89.1%. Similarly, Kim et al. (2019) classified bvFTD, nfvPPA and svPPA using principal component analysis and hierarchical classification and reached moderately high accuracy (overall balanced accuracy of 79.9% with 67.1% sensitivity and 92.6% specificity, and lower specificity when comparison between each FTD subtypes). Di Benedetto et al. (2022) compared different deep learning approaches but specifically for detecting bvFTD population only, and reported balanced accuracy ranging from 73.6 to 91.0% through independent validation.

In the present study, we trained a deep neural network classifier to differentiate bvFTD, nfvPPA, and svPPA patients using a multi-level feature embedding and fusion framework on multi-type morphological features derived from T1-weighted MRI scans drawn from two multi-site neuroimaging consortiums. To our knowledge, this is the first study using deep learning to examine the multi-class discrimination of all three FTD subtypes (bvFTD, nfvPPA, and svPPA) using multi-type MRI-based features.

2 Materials and methods

The overall schematic diagram of the proposed neuroimaging-based differential diagnosis framework is shown in Figure 1. The framework consists of four major steps: (1) feature extraction to derive patch-based multi-type features of cortical thickness and cortical/subcortical volumes; (2) W-score-based feature harmonization to control confounding factors such as scanner difference and study site bias, as well as demographic-related covariates; (3) the differential diagnosis model using multi-layer-perceptron (MLP)-based multi-level parallel feature embedding deep neural network to achieve FTD subtype classification; and (4) neuroimaging-derived feature visualization that differentiates FTD subtypes.

2.1 Experimental data

The experimental data consists of 173 bvFTD patients, 63 nfvPPA patients, and 41 svPPA patients, aggregated from the baseline visit

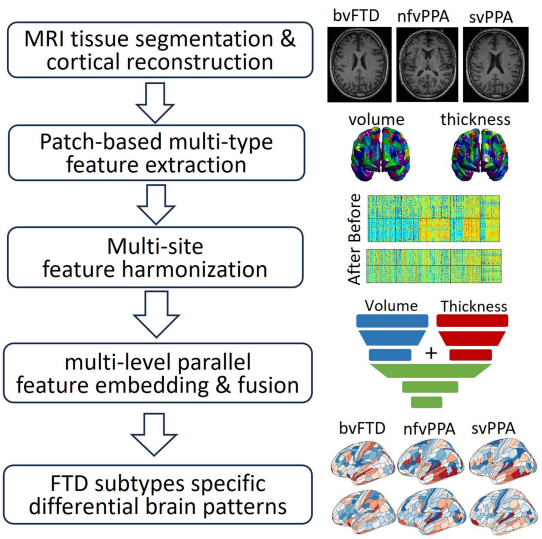


FIGURE 1
Schematic diagram of the framework in this study, which consists of: (1) feature extraction to derive patch-based multi-type features of cortical thickness and cortical/subcortical volumes; (2) W-score-based feature harmonization to control confounding factors such as scanner difference and study site bias, as well as demographic-related covariates; (3) the differential diagnosis model using multi-layer-perceptron (MLP)-based multi-level parallel feature embedding deep neural network to achieve FTD subtype classification; (4) neuroimaging-derived brain differential patterns among different FTD subtypes.

studies of two cohorts: the ARTFL-LEFFTDS Longitudinal Frontotemporal Lobar Degeneration (ALLFTD) cohort (Rosen et al., 2020) and the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI, also referred to as NIFD) cohort (Boeve et al., 2019). We excluded the cognitively normal healthy subjects in the aggregated dataset due to the limited sample size ($n = 27$). Table 1 shows patient demographic information. The clinical diagnosis of FTD subtypes was defined as the ground truth to train the proposed differential diagnosis framework, regardless of their mutation carrier status.

ALLFTD is a multi-site study consisting of data collected from 23 North American institutions, which is a combination of two previously independently initiated longitudinal neuroimaging studies, ARTFL and LEFFTDS. It aims to longitudinally follow FTLD mutation carriers to improve understanding of the FTLD disease progression based on both biological markers and clinical manifestation. Participants were primarily enrolled based on probable familial FTLD due to family history (i.e., with prior enrollment of a symptomatic proband), along with a small percentage of symptomatic and asymptomatic non-carriers enrolled. Mutation carriers of *MAPT*, *GRN*, or *C9orf72* genes were most common. Clinical consensus diagnosis for each clinical subtype was conducted by multidisciplinary teams following widely accepted published criteria (Gorno-Tempini et al., 2011; Rascovsky et al., 2011) and included comprehensive neurologic assessment, neuropsychological testing, brain MRI, and biofluid collection, as well an interview with caregiver or companion. Detailed information regarding the subject recruitment, diagnostic criteria, neuroimaging scanning protocols as well as image processing are available at [1,2].

NIFD is also a multi-site cohort with both clinical and MRI data collected at the University of California San Francisco, Mayo Clinic Rochester, and Massachusetts General Hospital. The NIFD consortium was initiated in 2010. NIFD did not collect information regarding familial mutations, and the comprehensive clinical evaluation for consensus diagnoses of FTD subtypes follows the similar criteria of ALLFTD, which includes neurologic history, neuropsychological testing, neurologic and physical examinations, structured interviews with caregiver, and neuroimaging. Detailed information regarding the subject recruitment, diagnostic criteria, neuroimaging scanning protocols, and image processing are available at [3].

2.2 Image preprocessing and patch-based multi-level multi-type feature extraction

2.2.1 Brain anatomical structural parcellation and patch segmentation

Deep learning approaches such as convolutional neural network (CNN) require large-sample data to train. However, our sample size does not lend itself to those methods. Therefore, we designed a multi-type feature extraction and multi-level feature embedding framework based on a multi-layer perceptron (MLP) architecture that is appropriate for this sample size. We employed neuroimaging-based preprocessing pipelines to extract the structural features from the raw T1 MR. Two primary structural feature types were extracted from the raw T1 structural MRI data: the regional brain structure volume and cortical mantle thickness. Each MRI scan was parcellated into small patch-based features (also called super-pixels) to reduce the dimensionality of the input data while preserving anatomically relevant MRI features.

The manifold of cerebral cortical surface data was first derived through brain tissue segmentation (gray matter, white matter, and cerebral spinal fluid – CSF), followed by cortical surface reconstruction using FreeSurfer 5.3 (Fischl, 2012). The initial vertex-based data was then further segmented into 360 patches, or regions of interest (ROIs), using the HCP-MMP1 atlas (Glasser et al., 2016) to preserve critical local discriminative features. The mean cortical measurements, both volume and thickness, were then calculated for each patch as the input features. In addition, the volumes of 15 FreeSurfer-segmented subcortical gray matter structures were also included as additional volumetric features (thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbent, both left and right hemisphere, plus brainstem). The final multi-type features resulted in a total of 735 features: 360 cortical thickness features plus 360 cortical volume features, as well as 15 subcortical volume features.

2.2.2 Feature harmonization

When combining multi-cohort data, confounding factors such as demographic variation as well as discrepancies within the data

1 https://www.allftd.org/
2 https://memory.ucsf.edu/research-trials/research/allftd
3 http://4rtni-ftldni.ini.usc.edu/

TABLE 1 Demographics information of the patients collected from multiple cohorts, in terms of sample size and age, stratified by sex, study cohort, as well as FTD subtypes.

		Overall	Grouped by Sex	
			Male	Female
Sample size (%)		277	151 (54.5%)	126 (45.5%)
Age, mean (SD)		63.7 (7.7)	63.5 (6.9)	63.9 (8.6)
Cohort, n (%)	ALLFTD	131 (47.3%)	79 (52.3%)	52 (41.3%)
	NIFD	146 (52.7%)	72 (47.7%)	74 (58.7%)
Subtype, n (%)	bvFTD	173 (62.5%)	97 (64.2%)	76 (60.3%)
	nfvPPA	63 (22.7%)	32 (21.2%)	31 (24.6%)
	svPPA	41 (14.8%)	22 (14.6%)	19 (15.1%)

acquisition devices and protocols will introduce unwanted heterogeneity within the data. Such data heterogeneity not only reduces the power of the analysis but may also introduce systematic bias. Neuroimage-derived measurements such as cortical thickness and subcortical volume will likely inherit such confounder-induced intrinsic biases. To control the confounders including cohort difference, scanner and coil difference, sex, as well as total intracranial volume (TIV), we used the generalized linear model (GLM)-based data harmonization that we have previously developed (Ma et al., 2019), using bvFTD as the reference group to calculate the reference mean and standard deviation. The resulting standard-residual term of the original feature, which is termed as w-score, is then used as the harmonized feature for the downstream tasks. It is worth noting that the GLM model used for feature harmonization was constructed using only the training data in each validation fold in the cross-validation. Details about cross-validation are described in the "model training and evaluation" section below.

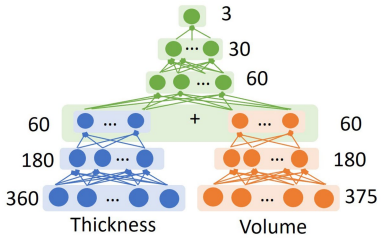


FIGURE 2
The schematic diagram of the neural network architecture of the multi-level parallel feature embedding framework used in this study to achieve accurate classification to differentiate FTD subtypes. Each block represents a multi-layer-perceptron (MLP) block. The number displayed in each of the MLP-based feature embedding blocks indicates the number of nodes in the corresponding layer.

2.3 Deep neural network (DNN)-based FTD subtype differential diagnosis model

2.3.1 Neural network architecture design

To achieve accurate differentiation between the three FTD subtypes based on neuroimaging information, we designed and trained a deep neural network (DNN) classifier through a two-level multi-type parallel feature embedding and fusion process (Figure 2). Each of the feature-embedding blocks was built using a multi-layer perceptron (MLP). Specifically, both the patch-wise cortical thickness features and cortical/subcortical volume features were fed into the two parallel input arms of the first-level network (shown in blue and red blocks) and optimized simultaneously. The embedded features from the first level were then concatenated into a fused intermediate latent feature vector and fed into the second level network (shown in green blocks), to derive the final output node of three classes of FTD subtypes.

2.3.2 Model training

A 10-fold nested cross-validation procedure was used to evaluate the robustness of the classification model, with each fold containing 80% training data, 10% validation data, and the remaining 10% of the data reserved as the independent testing set. The train/validation/test split was stratified based on the sample size ratio among FTD subtypes to ensure a comparable percentage sample for each class in each fold. The final predicted subtype classifications were derived from the probabilistic ensemble of the nine models trained in the inner folds. Weighted cross-entropy loss function was used to account for unbalanced sample size across subtypes, with weights calculated as the inverse proportion of class samples for each class. Stochastic gradient descent was used to optimize the model parameters of the DNN to minimize the loss function, with a learning rate of 1×10^{-3} and an L2 weight decay rate of 1×10^{-5}.

2.3.3 Performance evaluation and ablation study

To evaluate the classification performance of the differential diagnosis model, we measured the balanced accuracy for each FTD subtype, which was defined as the mean of sensitivity (the true positive rate) and specificity (the true negative rate), as well as the overall balanced accuracy calculated as the averaged across all FTD subtypes. We performed model comparisons to evaluate the effect of each component of the multi-type, multi-level feature embedding framework. A set of different experimental setups were included: (1) the proposed multi-level multi-type parallel feature embedding framework, in which the volume and thickness features were embedded into latent feature space independently in the first level before fusing and feeding into the second-level feature

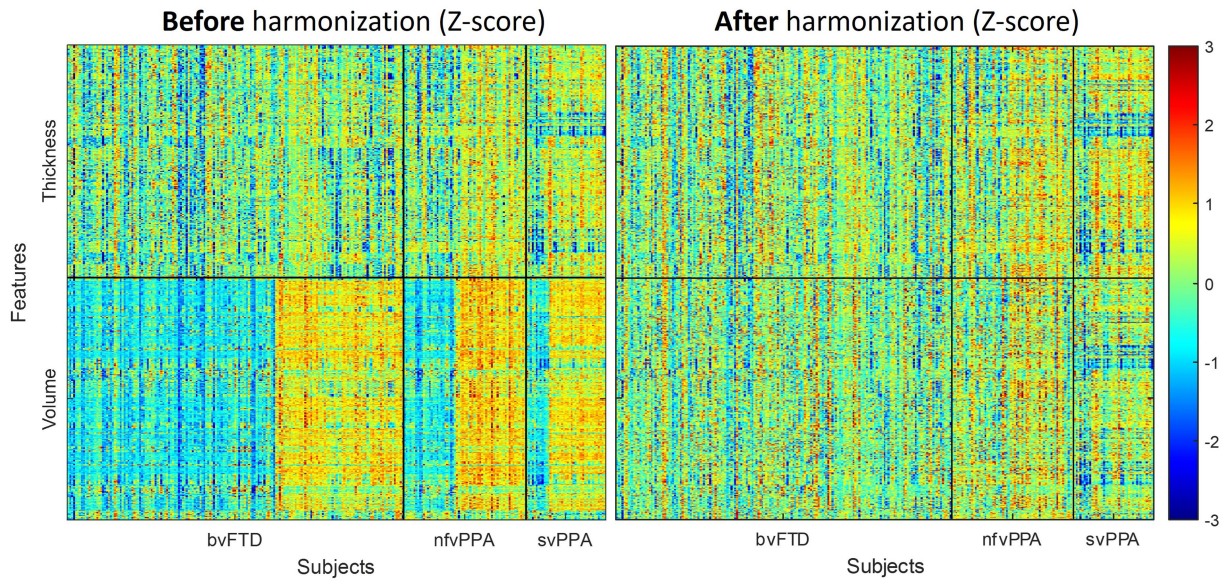


FIGURE 3
Effects of feature harmonization in preprocessings. The paranomic heatmap shows the Z-scores of thickness and volume features before (left) and after (right) the data harmonization. Z-score values for each features represents the difference between individual measurements compared to the reference group mean, standardized by the reference group standard deviation. Negative Z-scores indicate lower value than the reference mean (i.e., smaller volume, thinner cortex), while positive Z-scores represent higher value than reference mean (i.e., larger volume, thicker cortex). Cohort-dependent biases were noticeable before the harmonization (left), which were reduced after the GLM-based feature harmonization step (right).

embedding block; (2) "naïve concatenation" model that concatenated the volume and thickness input features into a long feature vector as a naïvely-fused multi-type feature and trained a conventional MLP network with the same number of nodes at each level; (3) ablation model that used only the thickness features as input; and (4) ablation model that only used volume features as input. All the model evaluations were performed on the test sets across all 10 outer folds.

2.4 Clinical explainability via local feature importance

To investigate the local distinguishable structural features that contribute more toward differentiating FTD subtypes, we used an explainable AI (XAI) approach called "Integrated Gradient" (Sundararajan et al., 2017), which assigned importance scores to each input feature (i.e., volume and thickness patches) reflecting their relevant contribution to the model's outcome prediction. This was achieved by computing the integral of the gradients of the predicted output for the given input features. The populational mean Integrated Gradient based feature importance map of each FTD subtype was then projected onto the template cortical manifold (HCP-MMP1 atlas) using the R package *ggseg* (Mowinckel and Vidal-Piñeiro, 2020). Additionally, for both volume and thickness, we conducted patch-wise linear models with the diagnostic group (vs. other groups) as the main effect and age, sex, and education as covariates. Multiple comparisons for the patch-wise cortical statistical mapping was controlled with a false discovery rate (FDR) set to 0.05.

3 Results

3.1 Multi-type structural feature extraction and harmonization

Figure 3 displays the panorama visualization of the Z-scores for each of the input features (columns) across the entire sample population of patients (rows) for all three FTD subtypes, both before and after the feature harmonization. Z-score value for each feature represents the difference between individual measurements compared to the reference group mean, standardized by the reference group standard deviation. Negative Z-scores indicate values lower than the reference mean (i.e., smaller volume, thinner cortex); while positive Z-scores represent higher than the reference mean (i.e., larger volume, thicker cortex). The raw volumetric features showed a significant cohort effect between the ALLFTD and NIFTD data compared to the thickness feature (Figure 3 left). Comparatively, no visible cohort bias was observable after the feature harmonization (Figure 3 right).

3.2 Differential diagnosis model evaluation and ablation study

The proposed differential diagnosis model showed the best classification performance among all compared models, achieving a balanced accuracy of 79.7% for bvFTD, 81.9% for nfvPPA, 89.2% for svPPA, and an overall balanced accuracy of 83.6%. Table 2 shows the results of the ablation study to evaluate the performance of the proposed FTD subtype differential diagnosis model using 10-fold class-stratified nested cross validation, in terms of balanced accuracy

for each subtype as well as the overall performance, and Figure 4 shows the corresponding box plot of the class-specific balanced accuracy as well as the overall multi-class balanced accuracy. When comparing single-type features as input, the thickness-only feature input (Table 2B; Figure 4 yellow) showed stronger discriminative power compared to the volume-only feature input (Table 2A; Figure 4 blue) for bvFTD, svPPA, as well as the overall performance. Interestingly, simply concatenating the volume and thickness feature types into a single input feature vector (Table 2C; Figure 4 green) resulted in reduced classification performance compared to the thickness-only feature input. On the contrary, the proposed multi-level parallel feature embedding approach (Table 2D; Figure 4 red) demonstrated performance improvement in terms of balanced accuracy for the classification of two out of the three FTD subtypes (bvFTD and nfvPPA), as well as the overall balanced accuracy.

3.3 FTD subtype differential patterns through explainability deep learning

The Integrated Gradient based FTD subtype feature attribute visualization patterns are shown in Figure 5 for both cortical thickness and volume features. The magnitude of the feature attributions (i.e., absolute value) represents the influence of each feature toward the output classification, while the sign of the feature attribution (i.e., positive and negative) reflect the direction of the feature influence toward the classification output. For example, for features with positive attributions (as shown in red), increasing in scalar value of the feature (i.e., structural volume or cortical thickness) will increase the likelihood of prediction for the correct FTD subtype; while for features with negative attribution (as shown in blue), decrease in scalar value of the feature will increase the likelihood of prediction for the correct

TABLE 2 The ablation study of the FTD subtype differential model.

Feature Type	bvFTD	nfvPPA	svPPA	Overall
A) Volume	0.742	0.791	0.854	0.796
B) Thickness	0.781	0.790	0.885	0.819
C) Thickness + Volume	0.760	0.796	0.867	0.808
D) Thickness + Volume (multi-level)	**0.797**	**0.819**	**0.892**	**0.836**

The classification performances were reported as the mean balanced accuracy on the test sets across all the 10 outer folds of the nested cross-validation. Different combination of input features and network architecture design are reported, including: (1) volume-only input features; (2) thickness-only input features; (3) Joint volume and thickness input feature using single-level MLP; and (4) Joint volume and thickness input feature using two-level MLP. The balanced accuracy for each individual FTD subtype (bvFTD, nfvPPA, and svPPA) as well as overall balance accuracy was reported. Bold values indicate the model with the best performance in terms of mean balanced acuracy.

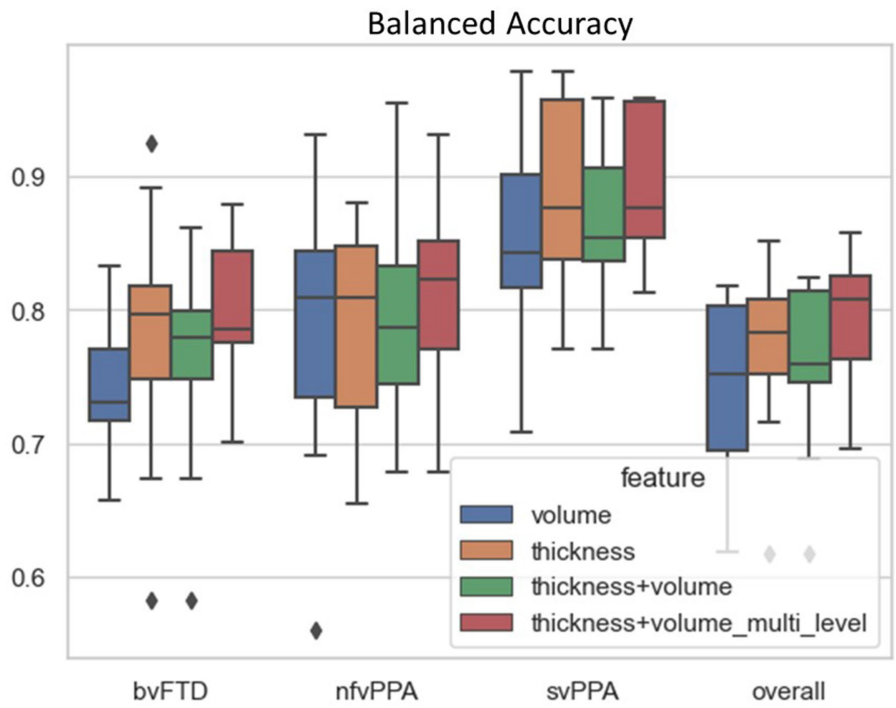


FIGURE 4
Classification performance in terms of balanced accuracy comparison among different combination of input features and network architecture design: (1) volume-only input features; (2) thickness-only input features; (3) Joint volume and thickness input feature using single-level MLP; and (4) Joint volume and thickness input feature using two-level MLP. The balanced accuracy for each individual FTD subtype (bvFTD, nfvPPA, and svPPA) as well as overall balance accuracy was reported. The black diamond box both in each subtype group as well as the overall performance represent the values of balanced accuracy that beyond two standard deviations among the 10-fold cross-validation.

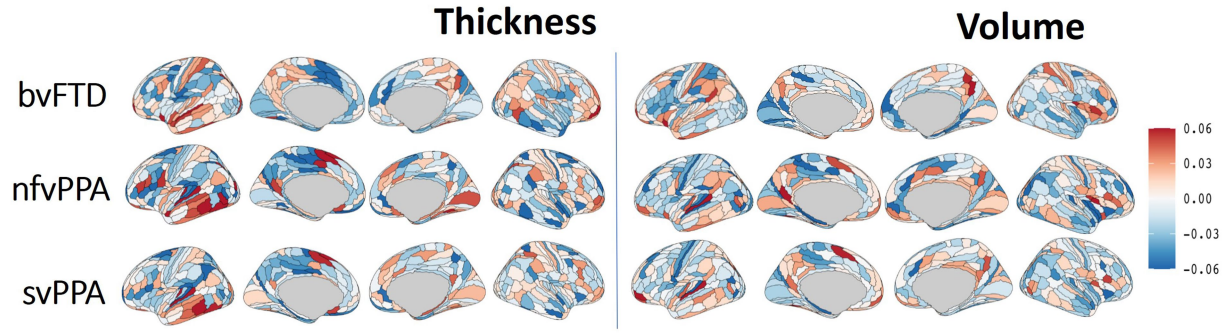


FIGURE 5
Differential cortical patterns for each of the FTD subtype. The cortical manifold plot visualizes populational average feature importance map using Integrated Gradient based feature importance analysis projected onto the template cortical manifold (HCP-MMP1 atlas) for both the cortical thickness (left) and volume (right) features. The color maps represent Integrated Gradient based feature importance scores ranging from −0.06 to 0.06. The magnitude of the feature attribution (i.e., absolute value) represent the influence of each feature towards the output classification, while the sign of the feature attribution (i.e., positive and negative) reflect the direction of the feature influence towards the classification output. Attributions that are close to zero represent features that have minimal influence in models prediction.

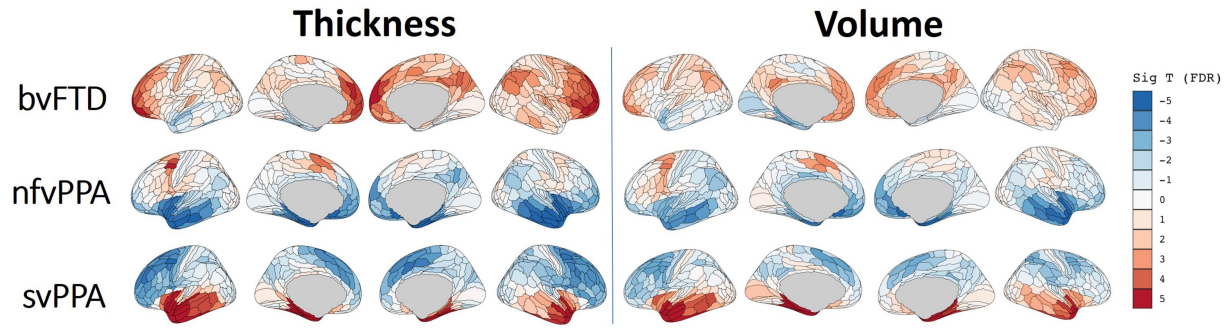


FIGURE 6
the statistical cortical mapping for each FTD subtype in which the patch-wise cortical features (both thickness and volume) were statistically compared with the combination of remaining populations that belong to the combination of the other two FTD subtypes.

FTD subtype. In other words, both the positive attributions (red) and negative attribution (blue) with the same Integrated Gradient value will have equivalent feature importance for making the correct classification, but with different direction of the influence. Attributions that are close to zero represent features that have minimal influence in the model prediction. Based on the thickness features (Figure 5, left), patches within the left temporal lobes appear to positively impact differentiation for both nfvPPA and svPPA. Regions from the inferior frontal and frontal operculum also positively influence the model for nfvPPA. For bvFTD, left-sided anterior temporal and frontal opercular/insular as well as bilateral frontal pole regions showed positive influences on the model, while cingulate and paracentral regions had negative influences (shown in blue). Volume-based features showed relatively diffuse differential patterns for both positive and negative influence than thickness features, although in generally similar overall patterns. This observation aligns with the results of the ablation study that thickness features showed stronger power to classify FTD subtypes compared to volume features. Figure 6 displays their corresponding patch-based statistical cortical mapping visualization, demonstrating canonical patterns of cortical atrophy in each subtype. Patterns of cortical atrophy in each subtype generally correspond to the Integrated Gradient based features of importance (i.e., in the temporal regions for svPPA and nfvPPA, and in frontal regions for bvFTD). However, it's worth noting that the patterns of atrophy tend to be more evenly distributed across neighboring patches, whereas Integrated Gradient based feature importance displays a more scattered distribution.

4 Discussion

In this study, we developed a deep-learning-based framework for the identification and differentiation of three subtypes of FTD (bvFTD, nfvPPA, and svPPA) based on structural MRI data drawn from two multi-site neuroimaging consortiums. We showed that the ensembled DNN classifier achieved promising differentiation power, with a balanced accuracy of 0.80 for bvFTD, 0.82 for nfvPPA, and 0.89 for svPPA. We additionally implemented a novel feature visualization tool to identify the most discriminative cortical and subcortical regions and explore their clinical relevance, which can provide insights into the underlying neuropathological processes and aid in the development of targeted interventions for different FTD subtypes.

The high balanced accuracy achieved by the DNN classifier in this study is an important step towards developing more reliable tools for

differentiating FTD subtypes using neuroimaging data. Machine learning methods have been extensively implemented in the differential diagnosis of Alzheimer's disease (AD) from cognitively normal controls (Wang et al., 2007; Lucas et al., 2011; Raamana et al., 2014; Dominic et al., 2018; Popuri et al., 2018; Bae et al., 2019; Gyujoon et al., 2022), and between AD, FTD, and cognitively normal (CN) groups (Wang et al., 2016; Kim et al., 2019; Ma et al., 2020; Hu et al., 2021). Prior work has also implemented machine learning methods for differential diagnosis of PPA subtypes, including svPPA, nfvPPA and logopenic PPA (Agosta et al., 2015; Themistocleous et al., 2021). However, the performance of these models has been inconsistent, with few studies reporting high accuracy levels but with small sample sizes (McCarthy et al., 2018).

One of the challenges in machine learning classification of FTD subtypes is the significant clinical, pathological, and genetic heterogeneity of FTD, making it difficult to develop a universal model that can accurately classify all subtypes. Additionally, the lack of large and standardized datasets, as well as the variability in imaging protocols across studies, have also limited the generalizability. DNN classifiers have shown superior performance compared to traditional machine learning methods, such as support vector machines (SVM) and random forest for accurate classification of disease groups using neuroimaging data (Schmidhuber, 2015; Eslami and Saeed, 2019; Amini et al., 2021). On the other hand, end-to-end deep learning frameworks such as CNN-based usually require a large sample size for training. In the current study, we designed a multi-type feature extraction and multi-level feature embedding framework based on the multi-layer perceptron (MLP) framework, with dimension reduction and feature extraction achieved through neuroimaging-based preprocessing pipelines to extract the structural features from the raw T1 MR. Specifically, we demonstrated that the fusion of multi-type input features in DNN is most effective through multi-level parallel feature embedding, in which each feature type was embedded into independent feature-specific low-dimensional representation before fusion together for a higher-level concurrent representation learning. Our results (Table 2; Figure 4) demonstrated the effectiveness of such a multi-type feature fusion approach as compared to the naïve feature concatenation at the input layer. Such a multi-type parallel feature embedding framework could be generalizable to other multi-modal deep learning problems such as neuroimaging genomics (Mirabnahrazam et al., 2022).

Our results showed the highest balanced accuracy of classification for svPPA at 0.89. svPPA is commonly associated with striking asymmetric atrophy of the dominant hemisphere temporal pole (Rogalski et al., 2011). This distinctive atrophy pattern is usually due to the presence of TDP-43 Type C neuropathology in these regions (Kawles et al., 2022; Keszycki et al., 2022). The high discriminative accuracy found in the present study is, therefore, unsurprising given this distinctive neuropathological profile and resultant neuroanatomical pattern of atrophy. Regions of the temporal lobes were identified as most useful in the discrimination, both for nfvPPA and svPPA, potentially driven by the semantic and linguistic variations that are identified as clinical features to define these two FTD subtypes. Moreover, subcortical regions, including the hippocampus and amygdala, were identified by the feature visualization tool as aiding in the differentiation (Supplementary Figure S3), aligning with the fact that more posterior elements of the medial temporal lobe in svPPA spared (Tan et al., 2014).

For bvFTD, our classifier achieved a balanced accuracy of 0.80. Individuals diagnosed clinically with bvFTD typically show significant gray matter volume loss of the frontal and temporal lobes, with early and most significant loss of volume in the insula and anterior cingulate cortex (Seeley et al., 2008; Mandelli et al., 2016; Ranasinghe et al., 2016). The lower classification accuracy observed in bvFTD than in svPPA may represent the greater clinical, neuroanatomical, and pathological heterogeneity of bvFTD. Indeed, bvFTD can be due to underlying FTLD-Tau, FTLD-TDP, or less commonly, AD neuropathology (Peet et al., 2021). Based on the feature visualization map, brain regions that more strongly contributed to the classification of bvFTD vs. others include the left posterior insula, superior temporal gyrus, and right prefrontal lobe for cortical thickness. For volume-based input data, the right posterior cingulate and bilateral insular and frontal opercular regions were identified as strongly contributing to the classification. This is consistent with reports showing that atrophy of the insular cortex is common in bvFTD (Mandelli et al., 2016; Fathy et al., 2020) and has even been shown to correlate with key clinical features, such as social cognition (Baez et al., 2019).

Finally, we showed that nfvPPA classification balanced accuracy was 0.82. Patients who present clinically with nfvPPA typically show atrophy of the left inferior frontal, insular and premotor cortex (Agosta et al., 2015), consistent with the pattern of motor speech deficits that are observed clinically (Rogalski et al., 2011). The lower observed classification accuracy of bvFTD and nfvPPA may be attributable to overlapping neuropathological and neuroanatomical signatures, as both syndromes are frequently associated with FTLD-Tau pathology (Mesulam et al., 2008, 2014). In the feature visualization map, regions identified as contributing to the classification included the left lateral and medial temporal lobes, left inferior frontal lobe, and left paracentral/midcingulate for the thickness inputs. In addition, regions from the volume inputs that were identified as important included the left superior temporal and right frontal operculum. Interestingly, prior work by Mandelli et al. (2016) found that nfvPPA subjects showed greater atrophy in the left posterior insula, which corresponds more to speech production, whereas bvFTD subjects showed greater atrophy in the ventral anterior insula, which corresponds to social–emotional functions. We observed similar results in our feature importance map, with regions of importance for nfvPPA being more congruent with inferior frontal motor speech areas, while bvFTD areas of importance were more apparent in the posterior insula and the anterior superior temporal lobe. Feature visualization maps also indicated that the bilateral hippocampal and right amygdala volumes were important in the classification (Supplementary Figures). Analyses of subcortical structural changes in nfvPPA are limited. However previous research has indicated possible effects on structures of the basal ganglia due to their role in hypothesized speech production pathways (Mandelli et al., 2018).

4.1 Limitations and future directions

In the current study, we considered demographic information as cofounding factors and controlled their effects on neuroimaging features through a regression-based harmonization step (Ma et al., 2019). This harmonization approach has been shown to be effective in increasing the classification power when predicting the risk of future

dementia onset (Popuri et al., 2020) and differentiating dementia subtypes (Ma et al., 2020).

Furthermore, disease subtypes might have populational prevalence among different demographic groups (Ma et al., 2022), and this information might aid discrimination. Indeed, incorporating demographic information into deep-learning frameworks has shown benefits to the deep-learning model in clinical applications such as dementia onset risk (Mirabnahrazam et al., 2022). Future directions of the current research could include investigating an alternative strategy to incorporate demographic information into the differential diagnosis framework instead of treating them as confounding factors in the harmonized preprocessing step, potentially improving the efficacy and generalizability of the differential diagnosis framework.

Additionally, our classification of interest was clinical diagnosis, as clinical syndromes are known to correspond more closely to neuroanatomical lesions as compared to neuropathology (Seeley et al., 2009). However, future research may choose to incorporate clinical, pathological, or genetic information to evaluate how this impacts classification accuracy.

In this work, we did not include cognitively normal control subjects as the healthy aging population due to insufficient samples, so we selected bvFTD as the reference group for data harmonization. Therefore, the resulting feature importance map mainly accounts for the more subtle differences among the three FTD subtypes rather than their differential atrophy patterns compared to the cognitively normal subjects. Future studies may choose to incorporate a large representative healthy aging population to be regarded as the reference group to achieve the most unbiased data harmonization (Ma et al., 2020), as well as extend to multi-syndrome dementia subtypes (Lampe et al., 2022) to capture brain patterns that include both predominant pathological factors as well as secondary subtype-driven differential patterns that are more likely to be subtle and relatively more heterogeneous.

We used structural features from T1-weighted MRI in the current study to derive differential features for detecting subtypes within FTD. Extension of current work could involve additional neuroimaging modalities such as diffusion tensor imaging (DTI) (Torso et al., 2020) or functional MRI (fMRI) (Gonzalez-Gomez et al., 2023). Another future direction for dealing with limited features would be to use a self-supervised approach as a feature extractor, to be trained on larger datasets, to extract disease-agnostic generalized neuroimaging features in lower dimensions, and then train a using the low-dimension representation space (Krishnan et al., 2022; Tang et al., 2022; Huang et al., 2023).

Finally, in terms of the model explainability, we mainly focused on using the deep-learning-based integrated gradient to derive the feature importance map. In follow-up studies, other feature importance methods, especially model-agnostic approaches such as SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) and multi-type feature permutation tests (Mirabnahrazam et al., 2022) could be incorporated to achieve more comprehensive and comparative analysis on the clinical explainability of deep-learning-based models.

5 Conclusion

In conclusion, we present here what we believe represents the first study to use a deep neural network classifier to differentiate the FTD subtypes of bvFTD, nfvPPA, and svPPA with feature visualization. We showed promising differentiation power using a combination of feature harmonization and a parallel multi-type feature embedding framework. Our approach has several potential clinical applications. For example, it could be used to identify at-risk populations for early and precise diagnosis, leading to more effective intervention planning. Further, our work may also help to advance our understanding of the underlying neurobiological mechanisms of FTD, providing important insights into the pathophysiology of the disorder.

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

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

DM: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HR: Conceptualization, Data curation, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing. KK: Data curation, Validation, Methodology, Writing – review & editing. SL: Investigation, Methodology, Resources, Validation, Writing – review & editing. JB: Investigation, Supervision, Validation, Writing – review & editing. SC: Funding acquisition, Investigation, Supervision, Validation, Writing – review & editing. MG: Investigation, Validation, Writing – review & editing. KP: Investigation, Methodology, Writing – review & editing. MB: Investigation, Methodology, Supervision, Writing – review & editing. LW: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. DM, SL, JB, and SC are funded by Wake Forest University School of Medicine Alzheimer's Disease Research Center (P30AG072947). DM received partial funding from Wake Forest Center for Biomedical Informatics

Pilot Award. JS is funded by National Institutes of Health T32 Mechanisms of Aging and Dementia Training Program, grant number: 5T32AG020506–02. LW received funding from R01 AG055121, R56 AG055121.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024.1331677/full#supplementary-material

References

Agosta, F., Ferraro, P. M., Canu, E., Copetti, M., Galantucci, S., Magnani, G., et al. (2015). Differentiation between subtypes of primary progressive aphasia by using cortical thickness and diffusion-tensor MR imaging measures. *Radiology* 276, 219–227. doi: 10.1148/radiol.15141869

Amini, M., Pedram, M., Moradi, A., and Ouchani, M. (2021). Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN). *Comput. Math. Methods Med.* 2021, 1–15. doi: 10.1155/2021/5514839

Bae, J., Stocks, J., Heywood, A., Jung, Y., Jenkins, L., Katsaggelos, A., et al. (2019). Transfer learning for predicting conversion from mild cognitive impairment to dementia of alzheimer's type based on 3d-convolutional neural network. *Neurobiol. Aging.* 99, 53–64. doi: 10.1016/j.neurobiolaging.2020.12.005

Baez, S., Pinasco, C., Roca, M., Ferrari, J., Couto, B., García-Cordero, I., et al. (2019). Brain structural correlates of executive and social cognition profiles in behavioral variant frontotemporal dementia and elderly bipolar disorder. *Neuropsychologia* 126, 159–169. doi: 10.1016/j.neuropsychologia.2017.02.012

Bisenius, S., Mueller, K., Diehl-Schmid, J., Fassbender, K., Grimmer, T., Jessen, F., et al. (2017). Predicting primary progressive aphasias with support vector machine approaches in structural MRI data. *Neuroimage* 14, 334–343. doi: 10.1016/j.nicl.2017.02.003

Boeve, B., Bove, J., Brannelly, P., Brushaber, D., Coppola, G., Dever, R., et al. (2019). The longitudinal evaluation of familial frontotemporal dementia subjects protocol: framework and methodology. *Alzheimers Dement.* 16, 22–36. doi: 10.1016/j.jalz.2019.06.4947

Boxer, A. L., Gold, M., Huey, E., Hu, W. T., Rosen, H., Kramer, J., et al. (2013). The advantages of frontotemporal degeneration drug development (part 2 of frontotemporal degeneration: the next therapeutic frontier). *Alzheimers Dement.* 9, 189–198. doi: 10.1016/j.jalz.2012.03.003

Di Benedetto, M., Carrara, F., Tafuri, B., Nigro, S., De Blasi, R., Falchi, F., et al. (2022). Deep networks for behavioral variant frontotemporal dementia identification from multiple acquisition sources. *Comput. Biol. Med.* 148:105937. doi: 10.1016/j.compbiomed.2022.105937

Dickerson, B. C., and Atri, A. (2014). *Dementia: comprehensive principles and practice*. Oxford University Press, Oxford

Dominic, J., Tom, M., Emanuele, T., Samuel, D., Susana, M.-M., Dominic, J., et al. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement.* 10, 519–535. doi: 10.1016/j.dadm.2018.07.004

Ducharme, S. (2023). Brain MRI research in neurodegenerative dementia: time to deliver on promises. *Brain* 146, 4403–4404. doi: 10.1093/brain/awad320

Erkkinen, M. G., Kim, M.-O., and Geschwind, M. D. (2018). Clinical neurology and epidemiology of the major neurodegenerative diseases. *Cold Spring Harb. Perspect. Biol.* 10:a033118. doi: 10.1101/cshperspect.a033118

Eslami, T., and Saeed, F. (2019). Auto-ASD-network: a technique based on deep learning and support vector machines for diagnosing autism spectrum disorder using fMRI data. Proceedings of the 10th ACM international conference on bioinformatics, New York, NY Association for Computing Machinery

Falahati, F., Westman, E., and Simmons, A. (2014). Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *J. Alzheimers Dis.* 41, 685–708. doi: 10.3233/JAD-131928

Fathy, Y. Y., Hoogers, S. E., Berendse, H. W., van der Werf, Y. D., Visser, P. J., de Jong, F. J., et al. (2020). Differential insular cortex sub-regional atrophy in neurodegenerative diseases: a systematic review and meta-analysis. *Brain Imaging Behav.* 14, 2799–2816. doi: 10.1007/s11682-019-00099-3

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

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178. doi: 10.1038/nature18933

Gonzalez-Gomez, R., Ibañez, A., and Moguilner, S. (2023). Multi-class characterization of frontotemporal dementia variants via multi-modal brain network computational inference. *Netw. Neurosci.* 7, 322–350. doi: 10.1162/netn_a_00285

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014. doi: 10.1212/WNL.0b013e31821103e6

Gyujoon, H., Murat, B., Yong, F., Dhivya, S., John, C. M., Marilyn, S. A., et al. (2022). Disentangling Alzheimer's disease neurodegeneration from typical brain ageing using machine learning. *Brain Commun.* 4:fcac117. doi: 10.1093/braincomms/fcac117

Hu, J., Qing, Z., Liu, R., Zhang, X., Lv, P., Wang, M., et al. (2021). Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease. *Front. Neurosci.* 14:626154. doi: 10.3389/fnins.2020.626154

Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., and Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: A systematic review and implementation guidelines. *NPJ Digit Med.* 6:1. doi: 10.1038/s41746-023-00811-0

Huang, M.-H., Zeng, B.-S., Tseng, P.-T., Hsu, C.-W., Wu, Y.-C., Tu, Y.-K., et al. (2023). Treatment efficacy of pharmacotherapies for frontotemporal dementia: A network meta-analysis of randomized controlled trials. *Am. J. Geriatr. Psychiatry* 31, 1062–1073. doi: 10.1016/j.jagp.2023.06.013

Katzeff, J. S., Bright, F., Phan, K., Kril, J. J., Ittner, L. M., Kassiou, M., et al. (2022). Biomarker discovery and development for frontotemporal dementia and amyotrophic lateral sclerosis. *Brain* 145, 1598–1609. doi: 10.1093/brain/awac077

Kawles, A., Nishihira, Y., Feldman, A., Gill, N., Minogue, G., Keszycki, R., et al. (2022). Cortical and subcortical pathological burden and neuronal loss in an autopsy series of FTLD-TDP-type C. *Brain* 145, 1069–1078. doi: 10.1093/brain/awab368

Keszycki, R., Jamshidi, P., Kawles, A., Minogue, G., Flanagan, M. E., Zaccard, C. R., et al. (2022). Propagation of TDP-43 proteinopathy in neurodegenerative disorders. *Neural Regen. Res.* 17, 1498–1500. doi: 10.4103/1673-5374.330609

Kim, J. P., Kim, J., Park, Y. H., Park, S. B., Lee, J. S., Yoo, S., et al. (2019). Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. *Neuroimage* 23:101811. doi: 10.1016/j.nicl.2019.101811

Krishnan, R., Rajpurkar, P., and Topol, E. J. (2022). Self-supervised learning in medicine and healthcare. Nature. *Biomed. Eng.* 6, 1346–1352. doi: 10.1038/s41551-022-00914-1

Lampe, L., Niehaus, S., Huppertz, H.-J., Merola, A., Reinelt, J., Mueller, K., et al. (2022). Comparative analysis of machine learning algorithms for multi-syndrome classification of neurodegenerative syndromes. *Alzheimers Res. Ther.* 14:62. doi: 10.1186/s13195-022-00983-z

Logroscino, G., Imbimbo, B. P., Lozupone, M., Sardone, R., Capozzo, R., Battista, P., et al. (2019). Promising therapies for the treatment of frontotemporal dementia clinical phenotypes: from symptomatic to disease-modifying drugs. *Expert. Opin. Pharmacother.* 20, 1091–1107. doi: 10.1080/14656566.2019.1598377

Lucas, R. T., Lorena, A. C., Fraga, F. J., Kanda, P. A., Anghinah, R., and Nitrini, R. (2011). Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin. EEG Neurosci.* 42, 160–165. doi: 10.1177/155005941104200304

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Adv. Neural Inf. Proces. Syst.* 30, 4768–4777. doi: 10.5555/3295222.3295230

Ma, D., Kumar, M., Khetan, V., Sen, P., Bhende, M., Chen, S., et al. (2022). Clinical explainable differential diagnosis of polypoidal choroidal vasculopathy and age-related macular degeneration using deep learning. *Comput. Biol. Med.* 143:105319. doi: 10.1016/j.compbiomed.2022.105319

Ma, D., Lu, D., Popuri, K., and Beg, M. F. (2021). Differential diagnosis of frontotemporal dementia and Alzheimer's disease using generative adversarial network. *arXiv* 9:5627. doi: 10.48550/arXiv.2109.05627

Ma, D., Lu, D., Popuri, K., Wang, L., and Beg, M. F.Alzheimer's Disease Neuroimaging Initiative (2020). Differential diagnosis of frontotemporal dementia, Alzheimer's disease, and Normal aging using a multi-scale multi-type feature generative adversarial deep neural network on structural magnetic resonance images. *Front. Neurosci.* 14:853. doi: 10.3389/fnins.2020.00853

Ma, D., Popuri, K., Bhalla, M., Sangha, O., Lu, D., Cao, J., et al. (2019). Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database. *Hum. Brain Mapp.* 40, 1507–1527. doi: 10.1002/hbm.24463

Mandelli, M. L., Vitali, P., Santos, M., Henry, M., Gola, K., Rosenberg, L., et al. (2016). Two insular regions are differentially involved in behavioral variant FTD and nonfluent/agrammatic variant PPA. *Cortex* 74, 149–157. doi: 10.1016/j.cortex.2015.10.012

Mandelli, M. L., Welch, A. E., Vilaplana, E., Watson, C., Battistella, G., Brown, J. A., et al. (2018). Altered topology of the functional speech production network in nonfluent/agrammatic variant of PPA. *Cortex* 108, 252–264. doi: 10.1016/j.cortex.2018.08.002

McCarthy, J., Collins, D. L., and Ducharme, S. (2018). Morphometric MRI as a diagnostic biomarker of frontotemporal dementia: A systematic review to determine clinical applicability. *Neuroimage* 20, 685–696. doi: 10.1016/j.nicl.2018.08.028

Mesulam, M.-M., Rogalski, E. J., Wieneke, C., Hurley, R. S., Geula, C., Bigio, E. H., et al. (2014). Primary progressive aphasia and the evolving neurology of the language network. *Nat. Rev. Neurol.* 10, 554–569. doi: 10.1038/nrneurol.2014.159

Mesulam, M., Wicklund, A., Johnson, N., Rogalski, E., Léger, G. C., Rademaker, A., et al. (2008). Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. *Ann. Neurol.* 63, 709–719. doi: 10.1002/ana.21388

Mioshi, E., Hsieh, S., Savage, S., Hornberger, M., and Hodges, J. R. (2010). Clinical staging and disease progression in frontotemporal dementia. *Neurology* 74, 1591–1597. doi: 10.1212/WNL.0b013e3181e04070

Mirabnahrazam, G., Ma, D., Beaulac, C., Lee, S., Popuri, K., Lee, H., et al. (2022). Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. *Neurobiol. Aging* 121, 139–156. doi: 10.1016/j.neurobiolaging.2022.10.005

Mis, M. S. C., Brajkovic, S., Tafuri, F., Bresolin, N., Comi, G. P., and Corti, S. (2017). Development of therapeutics for C9ORF72 ALS/FTD-related disorders. *Mol. Neurobiol.* 54, 4466–4476. doi: 10.1007/s12035-016-9993-0

Mowinckel, A. M., and Vidal-Piñeiro, D. (2020). Visualization of brain statistics with R packages ggseg and ggseg3d. *Adv. Methods Pract. Psychol. Sci.* 3, 466–483. doi: 10.1177/2515245920928009

Panza, F., Lozupone, M., Seripa, D., Daniele, A., Watling, M., Giannelli, G., et al. (2020). Development of disease-modifying drugs for frontotemporal dementia spectrum disorders. Nature reviews. *Neurology* 16, 213–228. doi: 10.1038/s41582-020-0330-x

Peet, B. T., Spina, S., Mundada, N., and La Joie, R. (2021). Neuroimaging in frontotemporal dementia: heterogeneity and relationships with underlying neuropathology. *Neurotherapeutics* 18, 728–752. doi: 10.1007/s13311-021-01101-x

Popuri, K., Balachandar, R., Alpert, K., Lu, D., Bhalla, M., Mackenzie, I. R., et al. (2018). Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *Neuroimage Clin.* 18, 802–813. doi: 10.1016/j.nicl.2018.03.007

Popuri, K., Ma, D., Wang, L., and Beg, M. F. (2020). Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. *Hum. Brain Mapp.* 41, 4127–4147. doi: 10.1002/hbm.25115

Raamana, P. R., Wen, W., Kochan, N. A., Brodaty, H., Sachdev, P. S., Wang, L., et al. (2014). The sub-classification of amnestic mild cognitive impairment using MRI-based cortical thickness measures. *Front. Neurol.* 5:76. doi: 10.3389/fneur.2014.00076

Ranasinghe, K. G., Rankin, K. P., Pressman, P. S., Perry, D. C., Lobach, I. V., Seeley, W. W., et al. (2016). Distinct subtypes of behavioral variant frontotemporal dementia based on patterns of network degeneration. *JAMA Neurol.* 73, 1078–1088. doi: 10.1001/jamaneurol.2016.2016

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., et al. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 134, 2456–2477. doi: 10.1093/brain/awr179

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage* 155, 530–548. doi: 10.1016/j.neuroimage.2017.03.057

Rogalski, E., Cobia, D., Harrison, T., Wieneke, C., Weintraub, S., and Mesulam, M.-M. (2011). Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology* 76, 1804–1810. doi: 10.1212/WNL.0b013e31821ccd3c

Rosen, H. J., Boeve, B. F., and Boxer, A. L. (2020). Tracking disease progression in familial and sporadic frontotemporal lobar degeneration: recent findings from ARTFL and LEFFTDS. *Alzheimers Dement.* 16, 71–78. doi: 10.1002/alz.12004

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Seeley, W. W., Crawford, R., Rascovsky, K., Kramer, J. H., Weiner, M., Miller, B. L., et al. (2008). Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. *Arch. Neurol.* 65, 249–255. doi: 10.1001/archneurol.2007.38

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., and Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron* 62, 42–52. doi: 10.1016/j.neuron.2009.03.024

Sundararajan, M., Taly, A., and Yan, Q. (2017). Axiomatic attribution for deep networks. *arXiv* 70, 3319–3328. doi: 10.48550/arXiv.1703.01365

Tan, R. H., Wong, S., Kril, J. J., Piguet, O., Hornberger, M., Hodges, J. R., et al. (2014). Beyond the temporal pole: limbic memory circuit in the semantic variant of primary progressive aphasia. *Brain* 137, 2065–2076. doi: 10.1093/brain/awu118

Tang, Y., Yang, D., Li, W., Roth, H., Landman, B., and Xu, D., (2022). Self-supervised pre-training of Swin transformers for 3D medical image analysis. In Conference on computer vision and pattern recognition. IEEE: New Orleans, LA.

Themistocleous, C., Ficek, B., Webster, K., den Ouden, D.-B., Hillis, A. E., and Tsapkini, K. (2021). Automatic subtyping of individuals with primary progressive aphasia. *J. Alzheimers Dis.* 79, 1185–1194. doi: 10.3233/JAD-201101

Torso, M., Bozzali, M., Cercignani, M., Jenkinson, M., and Chance, S. A. (2020). Using diffusion tensor imaging to detect cortical changes in fronto-temporal dementia subtypes. *Sci. Rep.* 10:11237. doi: 10.1038/s41598-020-68118-8

Tsai, R. M., and Boxer, A. L. (2016). Therapy and clinical trials in frontotemporal dementia: past, present, and future. *J. Neurochem.* 138, 211–221. doi: 10.1111/jnc.13640

Vijverberg, E. G., Wattjes, M. P., Dols, A., Krudop, W. A., Möller, C., Peters, A., et al. (2016). Diagnostic accuracy of MRI and additional [18F] FDG-PET for behavioral variant frontotemporal dementia in patients with late onset behavioral changes. *J. Alzheimers Dis.* 53, 1287–1297. doi: 10.3233/JAD-160285

Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J. C., et al. (2007). Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE Trans. Med. Imaging* 26, 462–470. doi: 10.1109/TMI.2005.853923

Wang, J., Redmond, S. J., Bertoux, M., Hodges, J. R., and Hornberger, M. (2016). A comparison of magnetic resonance imaging and neuropsychological examination in the diagnostic distinction of Alzheimer's disease and behavioral variant frontotemporal dementia. *Front. Aging Neurosci.* 8:119. doi: 10.3389/fnagi.2016.00119

Wilson, S. M., Ogar, J. M., Laluz, V., Growdon, M., Jang, J., Glenn, S., et al. (2009). Automated MRI-based classification of primary progressive aphasia variants. *Neuroimage* 47, 1558–1567. doi: 10.1016/j.neuroimage.2009.05.085

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 19 February 2024
DOI 10.3389/fnins.2024.1341734

OPEN ACCESS

EDITED BY
Da Ma,
Wake Forest University, United States

REVIEWED BY
Rajat Dhar,
Washington University in St. Louis, United States
Moisey Aronov,
Federal Medical and Biological Agency, Russia

*CORRESPONDENCE
Ernest J. Bobeff
✉ ernest.bobeff@umed.lodz.pl

RECEIVED 20 November 2023
ACCEPTED 29 January 2024
PUBLISHED 19 February 2024

CITATION
Puzio T, Matera K, Wiśniewski K, Grobelna M, Wanibuchi S, Jaskólski DJ and Bobeff EJ (2024) Automated volumetric evaluation of intracranial compartments and cerebrospinal fluid distribution on emergency trauma head CT scans to quantify mass effect.
Front. Neurosci. 18:1341734.
doi: 10.3389/fnins.2024.1341734

Automated volumetric evaluation of intracranial compartments and cerebrospinal fluid distribution on emergency trauma head CT scans to quantify mass effect

Tomasz Puzio[1], Katarzyna Matera[1], Karol Wiśniewski[2], Milena Grobelna[3], Sora Wanibuchi[2,4], Dariusz J. Jaskólski[2], and Ernest J. Bobeff[2,5]*

[1]Department of Diagnostic Imaging, Polish Mothers' Memorial Hospital Research Institute, Łódź, Poland, [2]Department of Neurosurgery and Neuro-Oncology, Barlicki University Hospital, Medical University of Lodz, Łódź, Poland, [3]Pixel Technology, Łódź, Poland, [4]Department of Anatomy, Aichi Medical University, Nagakute, Aichi, Japan, [5]Department of Sleep Medicine and Metabolic Disorders, Medical University of Lodz, Łódź, Poland

Background: Intracranial space is divided into three compartments by the falx cerebri and tentorium cerebelli. We assessed whether cerebrospinal fluid (CSF) distribution evaluated by a specifically developed deep-learning neural network (DLNN) could assist in quantifying mass effect.

Methods: Head trauma CT scans from a high-volume emergency department between 2018 and 2020 were retrospectively analyzed. Manual segmentations of intracranial compartments and CSF served as the ground truth to develop a DLNN model to automate the segmentation process. Dice Similarity Coefficient (DSC) was used to evaluate the segmentation performance. Supratentorial CSF Ratio was calculated by dividing the volume of CSF on the side with reduced CSF reserve by the volume of CSF on the opposite side.

Results: Two hundred and seventy-four patients (mean age, 61 years ± 18.6) after traumatic brain injury (TBI) who had an emergency head CT scan were included. The average DSC for training and validation datasets were respectively: 0.782 and 0.765. Lower DSC were observed in the segmentation of CSF, respectively 0.589, 0.615, and 0.572 for the right supratentorial, left supratentorial, and infratentorial CSF regions in the training dataset, and slightly lower values in the validation dataset, respectively 0.567, 0.574, and 0.556. Twenty-two patients (8%) had midline shift exceeding 5 mm, and 24 (8.8%) presented with high/mixed density lesion exceeding >25 ml. Fifty-five patients (20.1%) exhibited mass effect requiring neurosurgical treatment. They had lower supratentorial CSF volume and lower Supratentorial CSF Ratio (both $p < 0.001$). A Supratentorial CSF Ratio below 60% had a sensitivity of 74.5% and specificity of 87.7% (AUC 0.88, 95%CI 0.82–0.94) in identifying patients that require neurosurgical treatment for mass effect. On the other hand, patients with CSF constituting 10–20% of the intracranial space, with 80–90% of CSF specifically in the supratentorial compartment, and whose Supratentorial CSF Ratio exceeded 80% had minimal risk.

Conclusion: CSF distribution may be presented as quantifiable ratios that help to predict surgery in patients after TBI. Automated segmentation of intracranial compartments using the DLNN model demonstrates a potential of artificial

intelligence in quantifying mass effect. Further validation of the described method is necessary to confirm its efficacy in triaging patients and identifying those who require neurosurgical treatment.

KEYWORDS

mass effect, automated segmentation, deep-learning neural network, intracranial compartments, cerebrospinal fluid reserve, traumatic brain injury

1 Introduction

The intracranial (IC) compartments, formed by the falx cerebri and tentorium cerebelli, have limited capacity to accommodate volume changes of the brain, blood, and cerebrospinal fluid (CSF) (Wilson, 2016). Brain injury results in reduction of CSF reserve that may lead to mass effect. This phenomenon can contribute to secondary injury, including cerebral edema, ischemia, and herniation. Further investigation is needed to understand the anatomical and pathological aspects of compartmental distribution of IC contents and its consequences.

Cerebrospinal fluid reserve is researched in terms of IC pressure (ICP) which is measured using intraventricular sensors, and volume which can be assessed on imaging studies (Dhar et al., 2021). However, there is no widely accepted method to quantify mass effect. The Marshall scale integrates qualitative aspects such as basal cistern effacement, midline shift exceeding 5 mm and high density lesion larger than 25 mm^3, for prognostic assessment (Marshall et al., 1992). However, interrater variability may affect the results (Maas et al., 2005), and simplified formulas to ascertain the volumetric criterion may be imprecise (Vos et al., 2001). Automated volumetric evaluation may enhance its accuracy and assist in clinical decision-making at emergency departments without delays in diagnosis (Jain et al., 2019).

The growing demand for CT to detect IC hemorrhages and assess mass effect can be addressed through the use of artificial intelligence (AI) and machine learning (Chang et al., 2016; Heit et al., 2017; Raju et al., 2020; Brossard et al., 2021; Colasurdo et al., 2022). They are utilized in emergency care for various purposes, including triage, injury prediction, and outcome evaluation (Hunter et al., 2023). The ongoing efforts aim to automate lesion identification and segmentation, and assess CSF reserve (Monteiro et al., 2020; Colasurdo et al., 2022; Schmitt et al., 2022; Hunter et al., 2023; Yamada et al., 2023).

We conducted manual segmentation of IC compartments and threshold segmentation of CSF on emergency CT scans, which served as the ground truth. This data was then utilized as input for a deep-learning neural network (DLNN), which was trained to automate the segmentation task.

The main objective of this study was to develop an algorithm to quantify the mass effect requiring neurosurgical treatment on emergency head CT scans.

2 Methods

The study is in accordance with human rights declarations and regulations, and was approved by Institutional Review Board. Patient consent to the study was not required as it involved retrospective analysis of anonymized medical records. We screened head CT scans obtained from patients after traumatic brain injury (TBI) at a high-volume emergency department between 2018 and 2020. CT scans were performed on three scanners (Optima CT540, Revolution CT, Lightspeed VCT; GE Healthcare, USA). The manuscript was prepared following the CLAIM (Mongan et al., 2020) and the STROBE Guidelines.

2.1 CT screening and neurosurgical assessment

Studies with technical flaws, significant motion artifacts, or incomplete skull coverage were excluded. The presence of ischemia or hemorrhage, including subdural (SDH), epidural (EDH), intracerebral (ICH), cerebellar (CBH), subarachnoid (SAH), intraventricular (IVH), and contusions, was recorded. We undertook a thorough investigation to identify radiological criteria for mass effect necessitating neurosurgical treatment, drawing from the literature of the past two decades (Bullock et al., 2006a,b,c,d; Carney et al., 2017; Greenberg, 2019; Hawryluk et al., 2020; Greenberg et al., 2022). A summary of the criteria is shown in Supplementary Table S1. Neurosurgical assessment was independently carried out by three investigators, following the radiological criteria and clinical experience.

2.2 Manual segmentation

Two investigators segmented brain series of ≤1.25 mm slice thickness. The sagittal plane was manually adjusted to closely align with the falx cerebri, serving as the delineation between the left and right supratentorial compartments. During IC space segmentation we utilized the two-dimensional smart brush tool in Exhibeon3 DICOM viewer (Pixel Technology, Lodz, Poland) in the bone window (W: 2500 L: 800). The boundary with the spinal canal was drawn along the transverse plane, perpendicular to the

Abbreviations: AI, artificial intelligence; CBH, cerebellar hemorrhage; CSF, cerebrospinal fluid; CT, computed tomography; DLNN, deep-learning neural network; EDH, epidural hemorrhage; HCA, hierarchical clustering analysis; ICH, intracerebral hemorrhage; IC, intracranialICPintracranial pressure; IVH, intraventricular hemorrhage; SDH, subdural hematoma; SAH, subarachnoid hemorrhage; TBI, traumatic brain injury.

established sagittal plane of the falx cerebri, intersecting the McRae line connecting *basion* and *opisthion* craniometric points. The boundaries with cranial openings were drawn in line with the inner surface of the cranium. The boundary between supra- and infratentorial compartments was delineated in the brain window (W: 80 L: 40) using multiplanar reconstructions, taking into account the course of the tentorium cerebelli. The tentorial notch was identified on coronal reconstructions as a line connecting the free edges of the tentorium cerebelli, and refined on transverse reconstructions. The three resulting compartments – right and left supratentorial and one infratentorial – covered everything inside the cranium, including the brain, CSF, and any potential pathologies. The sum of the volumes of the three compartments constitutes the IC space. Voxels exhibiting Hounsfield Unit (HU) values ranging from −5 to 15 were labeled as CSF. Presence of artifacts like beam hardening and pervasive noise, often led to misidentifying voxels as CSF, which was manually excluded.

2.3 Network architecture

We used a convolutional neural network with basic UNet architecture in a 3D version (Falk et al., 2019). The model takes in a single-channel image as input and produces seven channels of output with segmentation. The model's encoder comprised five levels, with feature sizes of 32, 32, 64, 128, and 256, respectively. Leaky ReLU was employed as the activation layer (Xu et al., 2015). The total number of parameters in the model was 5.7 million. During the training process, the sum of Dice Loss and Cross Entropy was minimized using the AdamW optimizer. To schedule the learning rate, the One Cycle Scheduler technique was utilized with a maximum learning rate value of 0.001. PyTorch was used as a training framework. Augmentation and image processing was done in MonAI. Model weights were initialized randomly at the start of the training process.

2.4 Image preprocessing

The training and validation datasets were randomly selected to ensure representative coverage of the entire available data (Table 1). We conducted several preprocessing steps before utilizing medical images as inputs for our model. Firstly, the images were resampled to a spacing of 1 millimeter to ensure consistency in resolution. Secondly, based on the Hounsfield Scale a threshold value of 100 was applied to retain only the most relevant information. Specifically, any pixel values above 100 were set to this value. Following this, the intensities of the remaining pixels were normalized to range between −1 and 1. To ensure the model was exposed to a diverse range of inputs during training, randomly selected preprocessed images were used with augmentations such as Gaussian Noise, random contrast adjustments, and rotations. This helped to train a robust model capable of handling varied inputs. Single voxels marked as CSF by the initial threshold, which might have corresponded to artifacts or small post-ischemic lesions, were excluded from CSF. This augmentation resulted in a more faithful representation of the ventricular system and subarachnoid reserve on the CT scans, aligning with human perception (Figure 1). Upon visual inspection, the final model, which exhibited the smallest variations in studies with the greatest ground truth discrepancies, was selected. The performance metrics of the optimal model across all data partitions are provided in Table 1 and remained similar and consistent across both the training (dependent) and validation (independent) datasets. Consequently, we used the DLNN predictions from both the training and validation datasets to evaluate the clinical efficacy of CSF Distribution Ratios.

TABLE 1 Patients characteristics and comparison of the DCS between training and validation datasets.

	Training dataset *n* = 189	Validation dataset *n* = 85
Patients characteristics		
Mean age	62 years ±18.7	59 years ±18.1
Acute SDH	64 (33.9%)	25 (29.4%)
Chronic SDH	27 (14.3%)	13 (15.3%)
EDH	9 (4.8%)	6 (7.1%)
ICH	53 (28%)	15 (17.6%)
CBH	5 (2.6%)	1 (1.2%)
Traumatic SAH	28 (14.8%)	4 (4.7%)
Spontaneous SAH	12 (6.3%)	3 (3.5%)
IVH	27 (14.3%)	7 (8.2%)
Contusions	63 (33.3%)	20 (23.5%)
Ischemia	60 (31.7%)	21 (24.7%)
Marshall classification		
- Diffuse injury I (no pathology)	21 (11.1%)	22 (25.9%)
- Diffuse injury II	113 (59.8%)	45 (52.9%)
- Diffuse injury III (swelling)	9 (4.8%)	4 (4.7%)
- Diffuse injury IV (shift)	1 (0.5%)	0
- Evacuated mass lesion	42 (22.2%)	13 (15.3%)
- Nonevacuated mass lesion	3 (1.6%)	1 (1.2%)
DSC		
Average	0.782	0.765
Right supratentorial compartment	0.935	0.927
Left supratentorial compartment	0.932	0.927
Infratentorial compartment	0.905	0.903
Right supratentorial CSF	0.589	0.567
Left supratentorial CSF	0.615	0.574
Infratentorial CSF	0.572	0.556

The datasets were randomly selected to ensure representative coverage of the entire available data. CBH, cerebellar hemorrhage, CSF, cerebrospinal fluid, EDH, epidural hematoma, CT, computed tomography, DSC, Dice Similarity Coefficient, ICH, intracerebral hemorrhage, IVH, intraventricular hemorrhage, SAH, subarachnoid hemorrhage, SDH, subdural hematoma.

2.5 CSF distribution ratios

Volumetric data obtained from automated segmentation performed by the DLNN model was used to compute a series of quantitative indicators in each patient (Figure 2). The ratio "CSF/IC" refers to the proportion of CSF volume in relation to the IC space

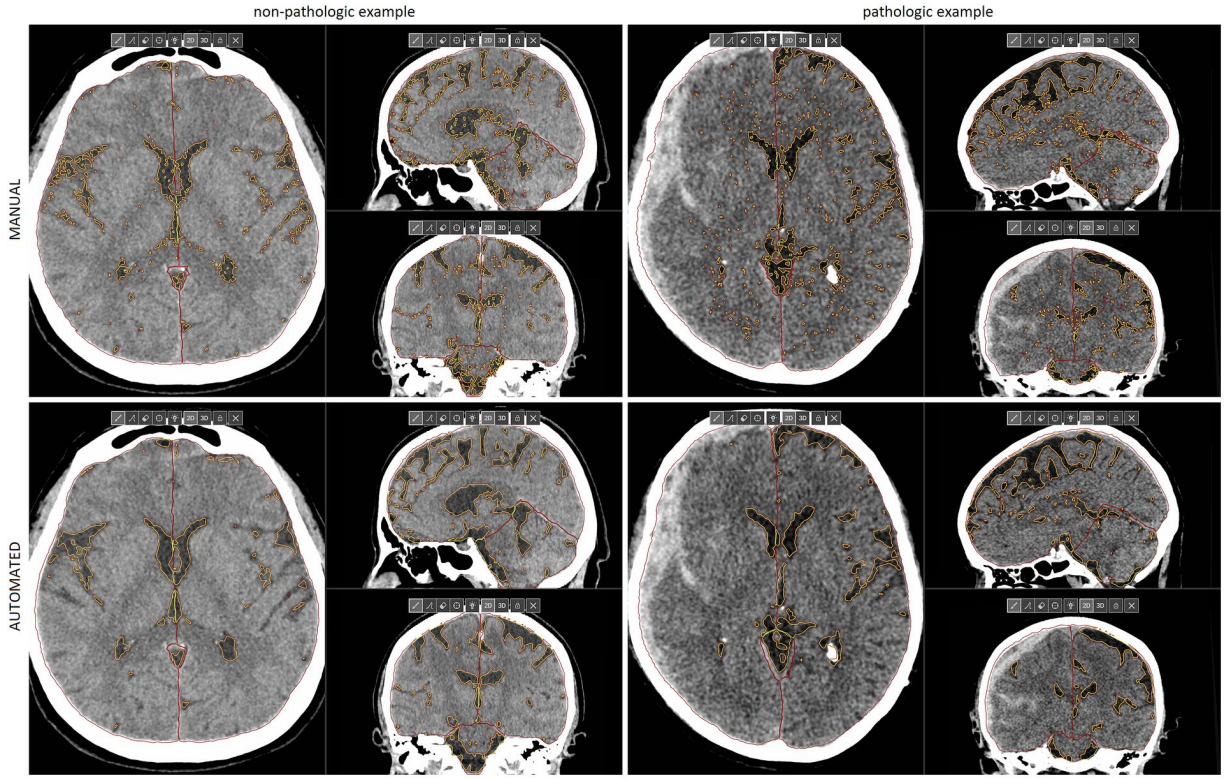


FIGURE 1
Multiplanar reconstructions of manual **(upper)** and automated **(lower)** segmentations of IC compartments and CSF in the non-pathologic **(left)** and pathologic example **(right)** of emergency CT scans. The latter example shows a right-sided acute SDH with significant mass effect that requires neurosurgical treatment, despite a relatively low midline shift. Reduced IC reserve in the right supratentorial compartment is well visualized. SDH, subdural hematoma, CSF, cerebrospinal fluid, CT, computed tomography, IC, intracranial. This figure is original to this submission so no credit or license is needed.

volume. The "Supratentorial CSF/IC CSF" represent the proportion of CSF in the supratentorial compartments relative to the CSF volume. The "Supratentorial CSF Ratio" quantifies the asymmetry in CSF distribution within the supratentorial compartments by dividing the volume of CSF on the side with reduced CSF reserve by the volume of CSF on the opposite side.

2.6 Statistical analysis

We used StatSoft Statistica (Tulsa, OK) and R Programming. Continuous variables were compared using Mann–Whitney *U* test. Categorical variables were compared using either Pearson's chi-squared test or two-sided Fisher's exact test. Predictive model was developed using logistic regression modelling with backward stepwise feature selection with likelihood ratio-test and with *p*-value of greater than 0.01 needed for stepwise feature removal. The heatmap was generated using unsupervised hierarchical clustering analysis with the *pheatmap* package in R Studio. Bland–Altman plots were generated using the ggplot2 package in R Studio. Power analysis for the test group was done using the pROC package. It yielded a required sample size of approximately 31 cases and 154 controls, with control-to-case ratio of 5, an anticipated area under the ROC curve of 0.7 and a desired power of 0.95 at a significance level of 0.05.

3 Results

The study included 274 patients, mean age 61 years ±18.6. Example segmentations of the IC compartments and CSF are presented in Figure 1. The mean volumes are provided in Table 2. The intraclass correlation coefficient (ICC) values between manual and automated segmentations were all above 0.92 (Figure 3).

Mass effect that required neurosurgical treatment was present in 55 patients (20.1%). Supratentorial CSF Ratio below 60% demonstrated a sensitivity of 74.5% and specificity of 87.7% in accurately identifying these patients. The ROC curve illustrated an AUC of 0.88 (Supplementary Figure S1). Noteworthy, neurosurgery for mass effect was never indicated in patients whose CSF constituted 10–20% of the IC space, with 80–90% being supratentorial, and whose Supratentorial CSF Ratio was larger than 80%. Uni- and multivariate analyses of radiological predictors of mass effect requiring neurosurgical treatment is provided in Table 3. Based on the selected CSF Distribution Ratios, we created a triage protocol for patients at the emergency department (Table 4).

By utilizing unsupervised hierarchical clustering analysis (HCA), patients (columns) were grouped according to A the triage protocol based on the selected CSF Distribution Ratios (Figure 4A) the presence and type of IC bleeding, any high or mixed density lesion larger than 25 mL, midline shift greater than 5 mm, and appearance of basal cisterns (Figure 4B).

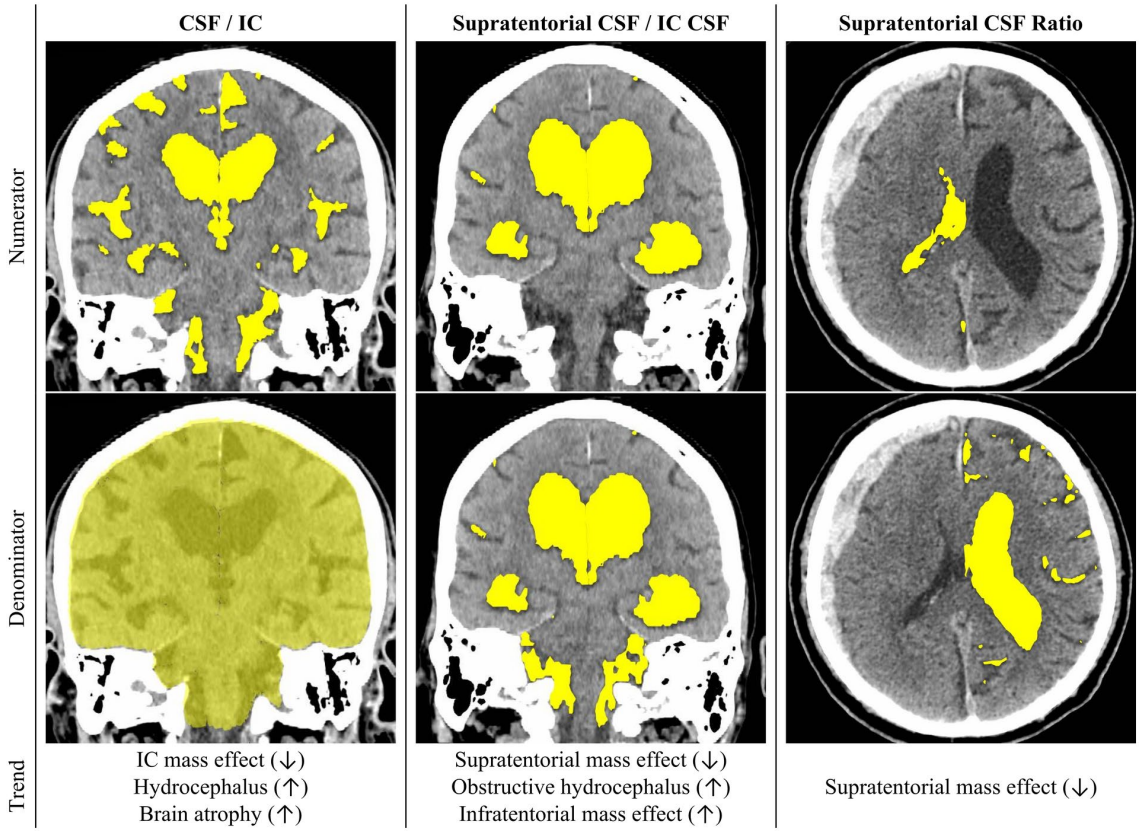


FIGURE 2
The illustration showcases the proposed CSF Distribution Ratios, with their definitions outlined in the manuscript. Rows one and two visualize the numerator and denominator, respectively, for each CSF Distribution Ratio. In row three, the figure delineates potential applications of each ratio and elucidates the directional changes associated with specific pathologies. CSF, cerebrospinal fluid, IC, intracranial, This figure is original to this submission so no credit or license is needed.

TABLE 2 Comparison of the volumes of IC compartments and CSF obtained from manual and automated segmentations of 274 head trauma CT scans performed at the emergency department.

Segmentation	Manual (ml)	Automated (ml)	ICC
Mean IC vol.	1415.9 ± 151	1416.3 ± 149.7	0.9948
Mean right supratentorial vol.	615 ± 68.3	613.9 ± 67.3	0.9882
Mean left supratentorial vol.	616.8 ± 67.9	615.8 ± 67.1	0.9905
Mean infratentorial vol.	184.1 ± 20.9	186.6 ± 21	0.9381
Mean CSF vol.	127.1 ± 65.2	120 ± 63	0.9561
Mean right supratentorial CSF vol.	52.9 ± 31.5	48.6 ± 31	0.9549
Mean left supratentorial CSF vol.	56.7 ± 32.1	54.6 ± 31.5	0.9678
Mean infratentorial CSF vol.	17.5 ± 6.7	16.8 ± 6.6	0.9213

CSF, cerebrospinal fluid, IC, intracranial, ICC, intraclass correlation coefficient, vol., volume.

Quantitative assessment (Figure 4A) associated with the triage protocol revealed three clusters of patients. The first cluster contained patients marked in red according to triage protocol, among whom 41 (60.3%) required neurosurgical treatment. In this group, all patients had a Supratentorial CSF Ratio below 60%. The second cluster contained patients marked in green who did not require neurosurgical treatment. All showed a balanced CSF distribution between IC compartments, and a Supratentorial CSF Ratio close to 1. The third cluster contains patients marked in yellow, among whom 14 (8%) required neurosurgical treatment. This is the largest and most heterogeneous group.

HCA based on the qualitative assessment is provided in Figure 4B. Clusters one and two were composed of patients with compressed basal cisterns, most of whom required surgery, whereas, patients in cluster three usually did not require surgery and were characterized by bilateral lesions, contusions, ischemia, traumatic SAH, and acute SDH. Cluster four included more than three-quarters of patients with either unremarkable head CT or surgical indications due to various lesions. HCA analysis highlights that incorporating the triage protocol based on the selected CSF Distribution Ratios could improve the accuracy of determining the need for neurosurgical treatment.

4 Discussion

Automated segmentation of IC compartments and CSF might contribute to fast, accurate, and consistent diagnosis of neurological emergencies. The underlying hypothesis is that various pathologies that require neurosurgical treatment, such as hemorrhage, brain edema, hydrocephalus or infarction, present as a mass effect associated with CSF displacement (Chen et al., 2016; Bobeff et al., 2018; Mönch

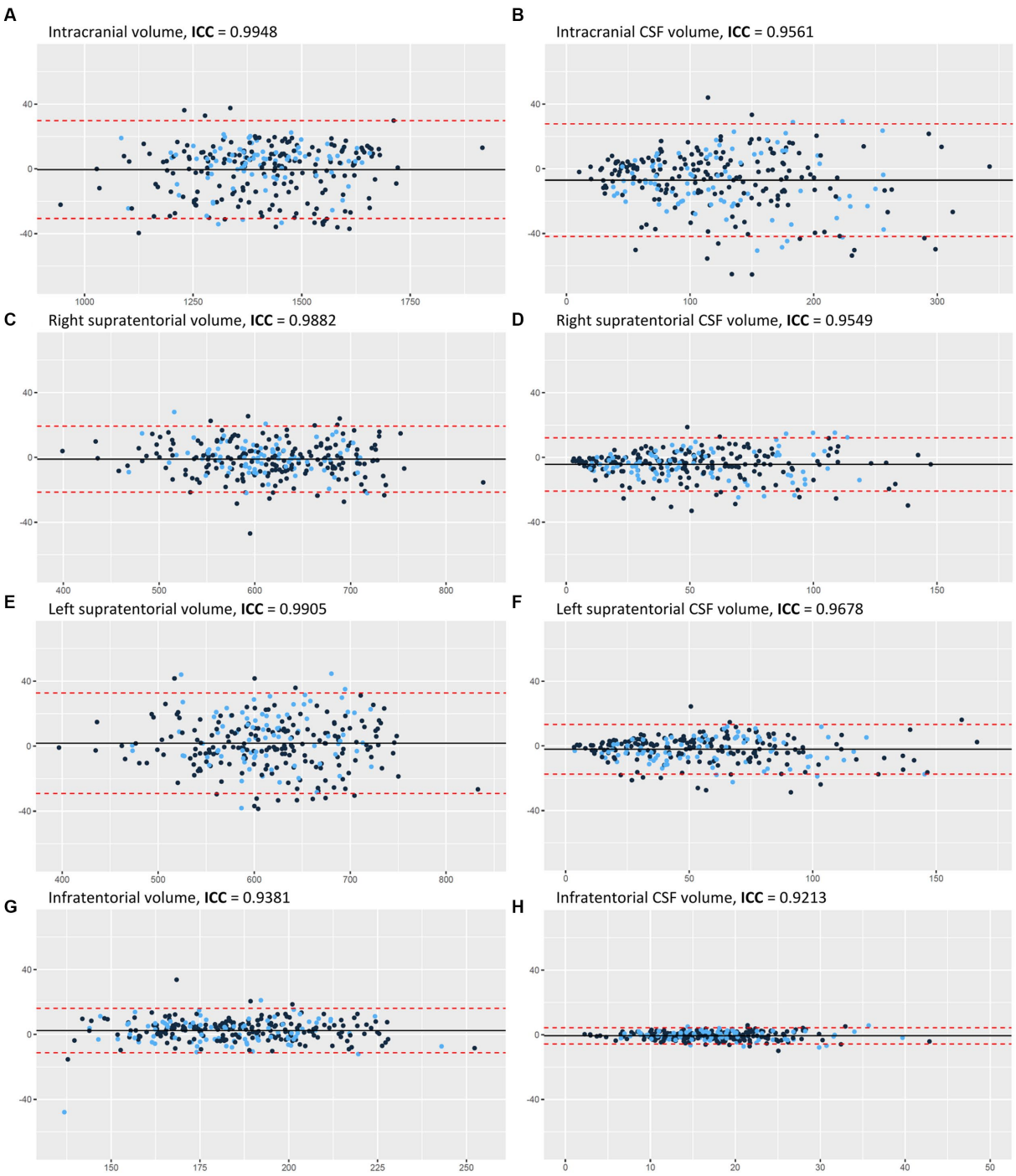


FIGURE 3
Bland–Altman plots for the manual and automated measurements of: **(A)** IC volume, **(B)** CSF volume, **(C)** right supratentorial volume, **(D)** right supratentorial CSF volume, **(E)** left supratentorial volume, **(F)** left supratentorial CSF volume, **(G)** infratentorial volume, and **(H)** infratentorial CSF volume. *Y* axes represent the difference between manual and automated measurements. *X* axes represent the average of manual and automated measurements. The color of each dot signifies the training (black) and the validation (blue) datasets. The black horizontal line indicates the mean measurement difference (bias), and if it is below zero it means that the average automated measurement was lower than the average manual measurement. The two red dashed horizontal lines represent the limit of agreement (1.96 × SD). AI, artificial intelligence, CSF, cerebrospinal fluid, IC, intracranial, ICC, intraclass correlation coefficient, SD, standard deviation. This figure is original to this submission so no credit or license is needed.

et al., 2020; Dhar et al., 2021) (Figure 5). Our key findings are: (1) there was strong agreement between manual and automated segmentations of IC compartments and CSF that support further validation of the latter and its use in clinical scenario, (2) CSF Distribution Ratios may help quantify mass effect and improve radiological reports without increasing time burdens.

Evaluation of automated segmentations was done in the context of Dice Similarity Coefficient (DSC), volumetric measurements, and

TABLE 3 Radiological predictors of mass effect requiring neurosurgical treatment in 274 patients who were diagnosed at the emergency department.

	All	Neurosurgical treatment	Univariate OR (95% CI)	Multivariate OR (95% CI)	*value of p*
Total	274	55 (20.1%)			
Bilateral Lesions	81	25 (30.9%)	2.4 (1.3–4.5)	–	
Acute SDH	89	21 (23.6%)	1.4 (0.7–2.5)	–	
Chronic SDH	40	19 (47.5%)	4.9 (2.4–10.2)	15.1 (3.9–58.9)	<0.001
EDH	15	7 (46.7%)	3.8 (1.3–11.1)	–	
ICH	68	27 (39.7%)	4.2 (2.2–7.9)	8.0 (2.3–28.1)	0.001
CBH	6	3 (50.0%)	4.2 (0.8–21.2)	–	
Contusions	83	15 (18.1%)	0.8 (0.4–1.6)	–	
Traumatic SAH	32	6 (18.8%)	0.9 (0.3–2.3)	–	
Spontaneous SAH	15	4 (26.7%)	1.5 (0.5–4.9)	–	
IVH	34	13 (38.2%)	2.9 (1.4–6.3)	–	
Ischemia	82	19 (23.2%)	1.3 (0.7–2.5)	–	
Basal Cisterns Compressed	41	25 (61.0%)	10.6 (5.1–22.1)	–	
Basal Cisterns Absent	11	10 (90.9%)	48.4 (6.0–388)	388 (24.7–6,111)	<0.001
MLS > 5 mm	22	21 (95.5%)	134 (17.5–1,033)	–	
High/Mixed Density Lesion>25 mL	24	20 (83.3%)	30.7 (9.9–95.2)	14 (2.9–67)	<0.001
Supratentorial CSF Ratio (continuous variable)	81% (62–92%)	40% (28–65%)	>999	1,072 (87.1–13,221)	<0.001
Supratentorial CSF Ratio < 60%	68	41 (60.3%)	20.8 (10.1–43.1)	–	
Supratentorial CSF Ratio > 80%	147	6 (4.1%)	0.07 (0.03–0.17)	–	
Supratentorial CSF/IC CSF (continuous variable)	86% (80–89%)	81% (70–87%)	758 (33.1–17,363)	–	
Supratentorial CSF/IC CSF 80–90%	153	25 (16.3%)	0.59 (0.33–1.07)	–	
IC CSF/IC volume (continuous variable)	8% (5–11%)	6% (3–9%)	>999	-	
IC CSF/IC volume 10–20%	94	9 (9.6%)	0.31 (0.14–0.66)	-	

CSF Distribution Ratios were calculated using automated segmentation by the DLNN model developed specifically for this study. Continuous variables are presented as medians and IQR. Supratentorial CSF Ratio was defined as a ratio of ipsilateral and contralateral supratentorial CSF volumes. "Ipsilateral" and "contralateral" refer to the supratentorial compartment with reduced CSF reserve. CBH, cerebellar hemorrhage; CSF, cerebrospinal fluid; EDH, epidural hemorrhage; IC, intracranial; ICH, intracerebral hemorrhage; IQR, interquartile range; IVH, intraventricular hemorrhage; MLS, midline shift; NS, not significant; SAH, subarachnoid hemorrhage; SDH, subdural hemorrhage.

through an unmediated evaluation of images with a focus on the most outliers. DSC for training and validation datasets were broadly equivalent (Table 1) and ICC very high (Table 2); furthermore, upon visual assessment, automated segmentation excelled in accurately identifying CSF and effectively partitioning IC compartments (Figure 1).

Emergency CT imaging aims to identify primary injuries, such as extraaxial hematomas, cerebral hemorrhage, contusion, and skull fractures. It also assesses their impact on IC contents, resulting in cerebral edema and increased ICP (Rincon et al., 2016). Both primary and secondary injuries reduce CSF reserve in the affected IC compartment or reduce the overall IC reserve in case of diffuse injury. In fact, radiological manifestations such as sulcal marking obliteration and brain displacement into sulci, cisterns, and ventricles, can be more challenging to observe than primary injuries itself.

Quantifying mass effect can improve the interpretation of radiological findings and reduce reliance on subjective descriptions with variable agreement among raters. Currently, there is no standardized method for quantitatively assessing mass effect, apart

TABLE 4 Triage protocol for mass effect that requires neurosurgical treatment based on the three selected CSF Distribution Ratios obtained from automated segmentation using the DLNN model developed specifically for this study.

Triage	Criteria	Mass effect requiring neurosurgical treatment
Red *Immediate*	Supratentorial CSF Ratio < 60%	41/68 (60%)
Yellow *Urgent*	*other patients*	14/168 (8%)
Green *Low risk*	Supratentorial CSF Ratio > 80% *and* CSF/ IC vol. 10–20% *and* Supratentorial CSF/ IC CSF 80–90%	0/38

CSF, cerebrospinal fluid, DLNN, deep-learning neural network, IC, intracranial, vol., volume.

from midline shift. The evaluation of radiological findings indicating increased ICP relies on the expertise of neurosurgeons and radiologists. Common terms used in radiological reports include "CSF reserve reduction/loss," "sulci effacement/loss," accompanied by specifying the location, such as "right-sided supratentorial" or "infratentorial." They are primarily qualitative and may not convey precise information. Our results show that CSF Distribution Ratios offer a valuable and potentially reproducible method to quantify mass effect.

Triaging imaging studies becomes increasingly important with the spread of teleradiology that potentially leads to delays in diagnosis. The use of CSF Distribution Ratios can prioritize cases with the utmost urgency, expedite radiology reports, and facilitate consultation between clinicians and radiologists, especially in centers with large numbers of CT scans (O'Neill et al., 2020). Possible triage criteria for categorizing patients into 3 risk groups of mass effect are outlined in Table 4. To comprehensively represent IC conditions, the protocol includes prognostic factors validated in univariate analysis and describes CSF reserve, supratentorial CSF asymmetry, and infra- and supratentorial CSF distribution.

Remote neurosurgical consultations frequently take place in distant hospitals and, if patient transportation is required, entail substantial costs and time. Frequently conservative therapy is preferred, still the stigma associated with IC hemorrhage, even without the need for neurosurgical treatment, can result in unnecessary patient transport. Automated segmentation and quantitative evaluation offer a precise and timely approach. This approach can be critical in situations where patient transport is risky and immediate surgery is being considered. It could facilitate remote neurosurgical consultations and aid in early-stage diagnosis at the emergency department.

The ratios "CSF/IC" and "supratentorial CSF/IC CSF" may capture nuances in mass effect resulting from infratentorial lesions and hydrocephalus due to *aqueductal stenosis*. The diagnosis of hydrocephalus requires clinical expertise and careful evaluation of signs and symptoms. CSF Distribution Ratios could enable more precise assessment of subsequent examinations within the same patient to achieve a more accurate and comprehensive disease monitoring.

Other pathologies that should be considered during the assessment of CSF reserve, where no localized primary injury is evident, include inflammatory or infectious processes, demyelinating diseases, vascular malformations, and metabolic disorders. Unless there is previous CT, it is often difficult to judge whether CSF reserve is within normal limits, diminished or severely diminished as a result of edema and mild brain swelling. Percentile grids for IC contents normalized by IC volume, gender, and age could guide radiologists by highlighting values outside established thresholds. For example, if normalized CSF reserve is below 3rd percentile, general brain swelling could be considered in impressions of radiologic report. Percentile grids could also help in cases of brain atrophy, a natural phenomenon associated with aging but not directly measured in clinical practice. In cases of cerebral atrophy, there is a notable reduction in the volume of both white and grey matter, which is subsequently supplanted by CSF. This phenomenon manifests radiologically as an enlargement of the lateral ventricles and widening of the arachnoid space fissures. Consequently, volumetric assessments reveal an increased CSF volume, leading to an increased CSF / IC ratio. Hence, the extent of cerebral atrophy can be quantitatively evaluated through our method, which leverages these radiological and volumetric changes.

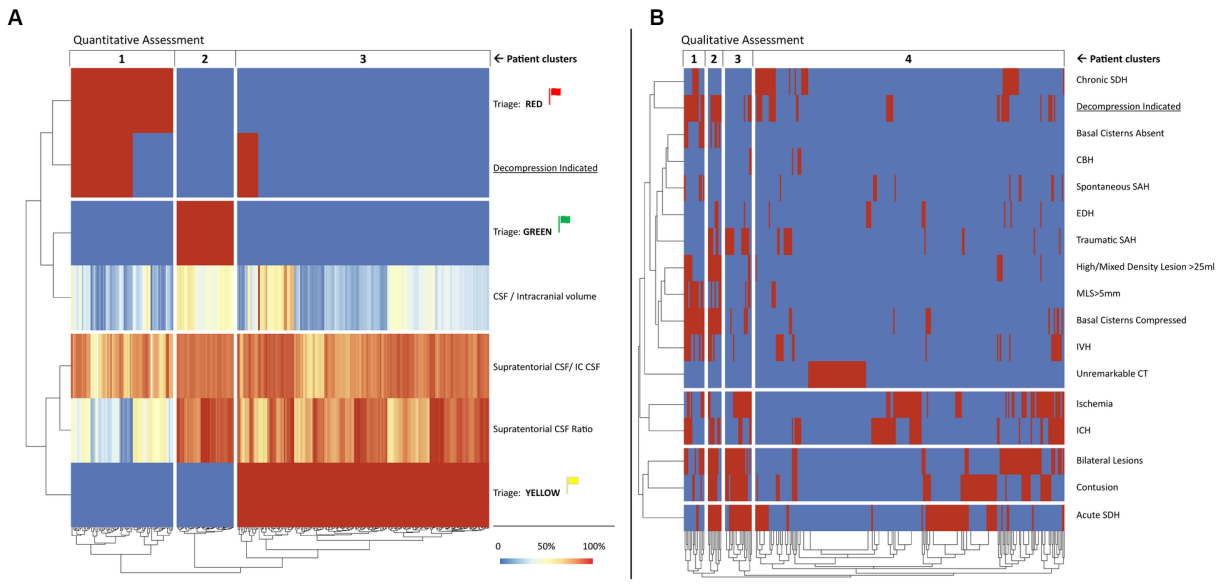


FIGURE 4
Heatmap representation of unsupervised HCA of selected quantitative **(A)** and qualitative **(B)** predictors of mass effect that requires neurosurgical treatment in patients after emergency head CT scans. Each column represents one patient, and they are grouped into clusters according to unsupervised HCA. The quantitative assessment **(A)** shows the selected CSF Distribution Ratios calculated from automated segmentation of IC compartments and CSF volumes and the proposed triage system presented in the Table 4, whereas the qualitative assessment **(B)** was based on the radiological reports and simplified formulas used to ascertain the volumetric criterion. Red indicates "yes" and blue indicates "no." Color legend for the continuous variables is provided in the diagram. Explanation and interpretation of the findings depicted in the figure can be found in the Results section of the article. CSF, cerebrospinal fluid leak, CT, computed tomography, HCA, hierarchical clustering analysis, IC, intracranial. This figure is original to this submission so no credit or license is needed.

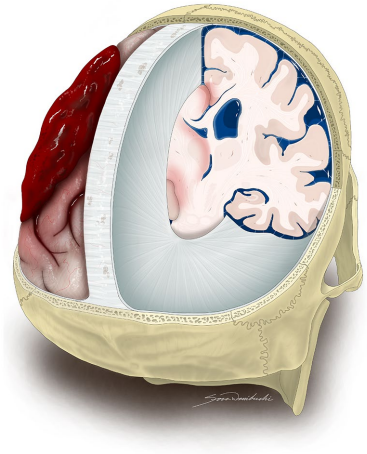


FIGURE 5
An artistic representation of oblique posterior view of IC contents revealing mass effect that requires neurosurgical treatment due to a left-sided acute SDH. The illustration exhibits the removal of both sides of the skull, as well as the dura mater covering the right and left hemispheres, except for the regions of the falx cerebri and tentorium cerebelli in order to show the limited capacity of the IC compartments to accommodate volume changes of the brain, blood, and CSF. The acute SDH over the left cerebral hemisphere causes midline shift, while the CSF reserve remains unaffected on the right side. SDH, subdural hematoma, CSF, cerebrospinal fluid leak, IC, intracranial. This figure is original to this submission so no credit or license is needed.

4.1 Limitations

It was a single center study. We acknowledge the heterogeneity of our patient cohort, consisting of individuals who experienced TBI and were diagnosed at the emergency department. On one hand this contributed to the diversity of mass effect presentations, including CBH, SDH, global edema, and hydrocephalus, on the other highlighted the method's versatility as the precise cutoff points could be tailored in specific pathologies. Another limitation of our study is the absence of detailed information on which patients with hydrocephalus required drainage procedures, limiting our ability to robustly assess the effectiveness of the presented ratios in predicting the need for such interventions. Our sample had a small number of infratentorial lesions. Reproducibility of our DLNN model was not subject to test–retest assessment. HCA of quantitative variables and one qualitative variable is very likely to split the group based on the latter; however, our goal was to show correlations between CSF Distribution Ratios and mass effect requiring neurosurgical treatment. We did not consider clinical factors related to patients condition that may influence the decision to perform neurosurgery, such as patient age, functional status, Glasgow Coma Scale (GCS) score, and comorbidities; however, this was our assumption that the model should identify radiological predictors, and the final treatment decision is made by clinicians, who take into consideration all available information. The role of the DLNN is to provide accurate and timely information, but not to replace a trained neuroradiologist. Going forward, we plan to integrate lesion volume calculations into our algorithm to enhance its capabilities and provide precise cutoff points for particular lesions.

5 Conclusion

Automated segmentation of IC compartments and calculation of CSF Distribution Ratios may enhance clinical decision-making and improve emergency management. The DLNN model effectively partitions the IC space into supra- and infratentorial compartments. CSF Distribution Ratios offer timely estimation of CSF reserve thus may enhance the predictive value of radiological reports. The integration of AI into the medical field can enhance the accuracy and speed of clinical diagnosis. Further research and implementation of AI into the healthcare system present an area of great interest bearing in mind their promising potential.

Data availability statement

The datasets presented in this article are not readily available because the dataset consisted only of computed tomography images. Requests to access the datasets should be directed to m.grobelna@pixel.com.pl.

Ethics statement

The studies involving humans were approved by Institutional Review Board approval RNN/211/16/KE. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because there was a retrospective study design.

Author contributions

TP: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Writing – original draft. KM: Data curation, Formal analysis, Writing – original draft. KW: Formal analysis, Investigation, Writing – original draft. MG: Investigation, Software, Writing – original draft. SW: Visualization, Writing – original draft. DJ: Methodology, Supervision, Writing – review & editing. EB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The

research presented in this article was funded from the project "RADi – asystent radiologa," co-financed by the European Union from the European Regional Development Fund under the Smart Growth Operational Programme 2014–2020, Priority Axis "Support for R&D activities of enterprises," Action 1.1 "R&D projects of enterprises," Sub-measure 1.1.1 "Industrial research and development work carried out by enterprises." The project provided financial support for the preparation of the deep-learning neural network and technical support necessary for the study.

Acknowledgments

We would like to acknowledge Pixel Technology for their support in making this research possible.

Conflict of interest

MG was employed by Pixel Technology. TP, KM, and EB participated in the project "RADi – asystent radiologa" described below.

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. The funding source did not have any influence on the study findings, and the authors declare no conflicts of interest related to the research. The results presented in this article are based solely on the data collected and analyzed in accordance with the study objectives and methods.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024.1341734/full#supplementary-material

References

Bobeff, E. J., Fortuniak, J., Bobeff, K. Ł., Wiśniewski, K., Wójcik, R., Stefańczyk, L., et al. (2018). Diagnostic value of lateral ventricle ratio: a retrospective case-control study of 112 acute subdural hematomas after non-severe traumatic brain injury. *Brain Inj.* 33, 226–232. doi: 10.1080/02699052.2018.1539871

Brossard, C., Lemasson, B., Attyé, A., de Busschère, J. A., Payen, J. F., Barbier, E. L., et al. (2021). Contribution of CT-scan analysis by artificial intelligence to the clinical care of TBI patients. *Front. Neurol.* 12:666875. doi: 10.3389/fneur.2021.666875

Bullock, M. R., Chesnut, R., Ghajar, J., Gordon, D., Hartl, R., Newell, D. W., et al. (2006a). Surgical management of acute epidural hematomas. *Neurosurgery* 58, S2-7–S2-15. doi: 10.1227/01.NEU.0000210363.91172.A8

Bullock, M. R., Chesnut, R., Ghajar, J., Gordon, D., Hartl, R., Newell, D. W., et al. (2006c). Surgical management of posterior fossa mass lesions. *Neurosurgery* 58, S2-47–S2-55. doi: 10.1227/01.NEU.0000210366.36914.38

Bullock, M. R., Chesnut, R., Ghajar, J., Gordon, D., Hartl, R., Newell, D. W., et al. (2006d). Surgical management of traumatic parenchymal lesions. *Neurosurgery* 58, S2-25–S2-46. doi: 10.1227/01.NEU.0000210365.36914.E3

Bullock, M. R., Chesnut, R., Ghajar, J., Gordon, D., Hartl, R., Newell, D. W., et al. (2006b). Surgical management of acute subdural hematomas. *Neurosurgery* 58, S16–S24.

Carney, N., Totten, A. M., O'Reilly, C., Ullman, J. S., Hawryluk, G. W. J., Bell, M. J., et al. (2017). Guidelines for the management of severe traumatic brain injury. *Neurosurgery* 80, 6–15. doi: 10.1227/NEU.0000000000001432

Chang, J. C., Lin, Y. Y., Hsu, T. F., Chen, Y. C., How, C. K., and Huang, M. S. (2016). Trends in computed tomography utilisation in the emergency department: a 5 year experience in an urban medical Centre in northern Taiwan. *Emerg. Med. Australas.* 28, 153–158. doi: 10.1111/1742-6723.12557

Chen, Y., Dhar, R., Heitsch, L., Ford, A., Fernandez-Cadenas, I., Carrera, C., et al. (2016). Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *Neuroimage Clin.* 12, 673–680. doi: 10.1016/j.nicl.2016.09.018

Colasurdo, M., Leibushor, N., Robledo, A., Vasandani, V., Luna, Z. A., Rao, A. S., et al. (2022). Automated detection and analysis of subdural hematomas using a machine learning algorithm. *J. Neurosurg.* 138, 1–8. doi: 10.3171/2022.8.JNS22888

Dhar, R., Hamzehloo, A., Kumar, A., Chen, Y., He, J., Heitsch, L., et al. (2021). Hemispheric CSF volume ratio quantifies progression and severity of cerebral edema after acute hemispheric stroke. *J. Cereb. Blood Flow Metab.* 41, 2907–2915. doi: 10.1177/0271678X211018210

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., 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

Greenberg, M. S. (2019). *Handbook of neurosurgery*. *9th Edn*, New York: Thieme Medical Publishers.

Greenberg, S. M., Ziai, W. C., Cordonnier, C., Dowlatshahi, D., Francis, B., Goldstein, J. N., et al. (2022). 2022 guideline for the Management of Patients with Spontaneous Intracerebral Hemorrhage: a guideline from the American Heart Association/American Stroke Association. *Stroke* 53, e282–e361. doi: 10.1161/STR.0000000000000407

Hawryluk, G. W. J., Rubiano, A. M., Totten, A. M., O'Reilly, C., Ullman, J. S., Bratton, S. L., et al. (2020). Guidelines for the Management of Severe Traumatic Brain Injury: 2020 update of the decompressive Craniectomy recommendations. *Neurosurgery* 87, 427–434. doi: 10.1093/neuros/nyaa278

Heit, J. J., Iv, M., and Wintermark, M. (2017). Imaging of intracranial hemorrhage. *J Stroke* 19, 11–27. doi: 10.5853/jos.2016.00563

Hunter, O. F., Perry, F., Salehi, M., Bandurski, H., Hubbard, A., Ball, C. G., et al. (2023). Science fiction or clinical reality: a review of the applications of artificial intelligence along the continuum of trauma care. *World J. Emerg. Surg.* 18:16. doi: 10.1186/s13017-022-00469-1

Jain, S., Vyvere, T. V., Terzopoulos, V., Sima, D. M., Roura, E., Maas, A., et al. (2019). Automatic quantification of computed tomography features in acute traumatic brain injury. *J. Neurotrauma* 36, 1794–1803. doi: 10.1089/neu.2018.6183

Maas, A. I., Hukkelhoven, C. W., Marshall, L. F., and Steyerberg, E. W. (2005). Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. *Neurosurgery* 57, 1173–1182. doi: 10.1227/01.neu.0000186013.63046.6b

Marshall, L. F., Marshall, S. B., Klauber, M. R., Van Berkum Clark, M., Eisenberg, H., Jane, J. A., et al. (1992). The diagnosis of head injury requires a classification based on computed axial tomography. *J. Neurotrauma* 9, S287–S292.

Mönch, S., Sepp, D., Hedderich, D., Boeckh-Behrens, T., Berndt, M., Maegerlein, C., et al. (2020). Impact of brain volume and intracranial cerebrospinal fluid volume on the clinical outcome in endovascularly treated stroke patients. *J. Stroke Cerebrovasc. Dis.* 29:104831. doi: 10.1016/j.jstrokecerebrovasdis.2020.104831

Mongan, J., Moy, L., and Kahn, C. E. Jr. (2020). Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol. Artif. Intell.* 2:e200029. doi: 10.1148/ryai.2020200029

Monteiro, M., Newcombe, V. F. J., Mathieu, F., Adatia, K., Kamnitsas, K., Ferrante, E., et al. (2020). Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. *Lancet Digit Health.* 2, e314–e322. doi: 10.1016/S2589-7500(20)30085-6

O'Neill, T. J., Xi, Y., Stehel, E., Browning, T., Ng, Y. S., Baker, C., et al. (2020). Active reprioritization of the Reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. *Radiol. Artif. Intell.* 3:e200024. doi: 10.1148/ryai.2020200024

Raju, B., Jumah, F., Ashraf, O., Narayan, V., Gupta, G., Sun, H., et al. (2020). Big data, machine learning, and artificial intelligence: a field guide for neurosurgeons [published online ahead of print, 2020 Oct 2]. *J. Neurosurg.* 135, 1–11. doi: 10.3171/2020.5.JNS201288

Rincon, S., Gupta, R., and Ptak, T. (2016). "Imaging of head trauma" in *Neuroimaging part I. Handbook of clinical neurology*, *vol.* 135 Eds. J. C. Masdeu and R. Gilberto González (Elsevier) 447–477.

Schmitt, N., Mokli, Y., Weyland, C. S., Gerry, S., Herweh, C., Ringleb, P. A., et al. (2022). Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients. *Eur. Radiol.* 32, 2246–2254. doi: 10.1007/s00330-021-08352-4

Vos, P. E., van Voskuilen, A. C., Beems, T., Krabbe, P. F., and Vogels, O. J. (2001). Evaluation of the traumatic coma data bank computed tomography classification for severe head injury. *J. Neurotrauma* 18, 649–655. doi: 10.1089/089771501750357591

Wilson, M. H. (2016). Monro-Kellie 2.0: the dynamic vascular and venous pathophysiological components of intracranial pressure. *J. Cereb. Blood Flow Metab.* 36, 1338–1350. doi: 10.1177/0271678X16648711

Xu, B., Naiyan, W., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *ArXiv.org*. 1505.00853. doi: 10.48550/arXiv.1505.00853,

Yamada, S., Otani, T., Ii, S., Kawano, H., Nozaki, K., Wada, S., et al. (2023). Aging-related volume changes in the brain and cerebrospinal fluid using artificial intelligence-automated segmentation [published online ahead of print, 2023 Apr 15]. *Eur. Radiol.* 33, 7099–7112. doi: 10.1007/s00330-023-09632-x

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 19 February 2024
DOI 10.3389/fnins.2024.1288274

OPEN ACCESS

EDITED BY
Da Ma,
Wake Forest University, United States

REVIEWED BY
Zejun Zhang,
Zhejiang Normal University, China
Yanchao Gong,
Xi'an University of Posts and Telecommunications, China

*CORRESPONDENCE
Wenna Chen
✉ chenwenna0408@163.com
Hongwei Jiang
✉ jianghw@haust.edu.cn

RECEIVED 04 September 2023
ACCEPTED 05 February 2024
PUBLISHED 19 February 2024

CITATION
Chen W, Tan X, Zhang J, Du G, Fu Q and Jiang H (2024) A robust approach for multi-type classification of brain tumor using deep feature fusion.
Front. Neurosci. 18:1288274.
doi: 10.3389/fnins.2024.1288274

A robust approach for multi-type classification of brain tumor using deep feature fusion

Wenna Chen[1]*, Xinghua Tan[2], Jincan Zhang[2], Ganqin Du[1], Qizhi Fu[1] and Hongwei Jiang[1]*

[1]The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China, [2]College of Information Engineering, Henan University of Science and Technology, Luoyang, China

Brain tumors can be classified into many different types based on their shape, texture, and location. Accurate diagnosis of brain tumor types can help doctors to develop appropriate treatment plans to save patients' lives. Therefore, it is very crucial to improve the accuracy of this classification system for brain tumors to assist doctors in their treatment. We propose a deep feature fusion method based on convolutional neural networks to enhance the accuracy and robustness of brain tumor classification while mitigating the risk of over-fitting. Firstly, the extracted features of three pre-trained models including ResNet101, DenseNet121, and EfficientNetB0 are adjusted to ensure that the shape of extracted features for the three models is the same. Secondly, the three models are fine-tuned to extract features from brain tumor images. Thirdly, pairwise summation of the extracted features is carried out to achieve feature fusion. Finally, classification of brain tumors based on fused features is performed. The public datasets including Figshare (Dataset 1) and Kaggle (Dataset 2) are used to verify the reliability of the proposed method. Experimental results demonstrate that the fusion method of ResNet101 and DenseNet121 features achieves the best performance, which achieves classification accuracy of 99.18 and 97.24% in Figshare dataset and Kaggle dataset, respectively.

KEYWORDS

brain tumor classification, deep learning, transfer learning, ResNet101, DenseNet121, EfficientNetB0, feature fusion

1 Introduction

In recent years, the rising incidence and mortality rates of brain tumor diseases have posed significant threats to human well-being and life (Satyanarayana, 2023). Because of the different causes and locations of brain tumors, the treatment methods for brain tumors are very different. Additionally, the severity of lesions significantly impacts the efficacy of treatment methods. Therefore, it is very important to determine the type and severity of brain tumor lesions prior to treatment development. With the development of modern technology, Computer-Aided Diagnosis (CAD) technology plays an increasingly important role in the medical diagnosis process (Fujita, 2020; Gudigar et al., 2020; Sekhar et al., 2022). The diagnosis and analysis of brain tumor magnetic resonance imaging (MRI) images by physicians based solely on personal experience is not only inefficient but also subjective and prone to errors, leading to misleading results (Chan et al., 2020; Arora et al., 2023). Consequently, enhancing the efficiency and accuracy of computer-aided diagnosis for brain tumors has emerged as a

prominent research hotspot in the field of brain tumor-assisted diagnosis.

Traditionally, the classification method of medical images consists of several stages, including image pre-processing, image segmentation, feature extraction, feature selection, training of classifiers and image classification (Muhammad et al., 2021; Yu et al., 2022). Nevertheless, in recent years, with the emergence of deep learning theory, more and more researchers applied the deep learning theory into medical image processing (Maurya et al., 2023). Deep learning has been employed widely in the analysis and diagnosis of diverse diseases (Cao et al., 2021; Gu et al., 2021; Lin et al., 2022; Yang, 2022; Yao et al., 2022; Zolfaghari et al., 2023). Convolutional Neural Networks (CNNs) are widely recognized as one of the most prominent deep learning techniques. By utilizing the images as input, CNNs mitigate the issue of low classification accuracy resulting from the selection of unrepresentative features by humans.

Medical images are usually difficult to obtain, and the amount of image data is relatively small (Shah et al., 2022). Although training an effective deep learning model typically necessitates a substantial amount of data, transfer learning can address the issue of limited dataset size and expedite the training process. Therefore, transfer learning has been widely used in the medical field (Yu et al., 2022). Yang et al. (2018) utilized AlexNet and GoogLeNet for glioma grade classification. Experimental results demonstrated that CNNs trained using transfer learning and fine-tuning were employed for glioma grading, achieving improved performance compared to traditional machine learning methods reliant on manual features, as well as compared to CNNs trained from scratch. Swati et al. (2019) and Zulfiqar et al. (2023) employed VGG19 and EfficientNetB2, respectively for the classification of brain tumors. Arora et al. (2023) examined the classification performance of 14 pre-trained models for the identification of skin diseases. DenseNet201 obtained superior classification performance, achieving an accuracy of 82.5%. Meanwhile, ResNet50 exhibits the second-highest classification accuracy at 81.6%. Aljuaid et al. (2022), ResNet 18, ShuffleNet, and Inception-V3Net models were used to classify breast cancer, with ResNet 18 showing excellent performance with an accuracy of 97.81%.

However, only relying on a single model often results in over-fitting on the training set and poor generalization on the test set, in turn to diminish the model's robustness. Therefore, in this paper, to addresses the limitations associated with only relying on a single model, model integration techniques are proposed. In this paper, three pre-trained models namely ResNet101, DenseNet121, and EfficientNetB0 are used to extract the features of brain tumor images. Subsequently, the extracted features are fused using a summation method, followed by classification of the fused features. The main contributions of this paper are as follows:

1 An image classification method for brain tumors based on feature fusion is proposed.
2 The feature outputs of the three pre-trained models were adjusted to have consistent dimensions.
3 Feature fusion was accomplished through summation.
4 The validity of the method was verified on two publicly available datasets including Figshare dataset (Cheng et al., 2015) referred to as dataset 1, and Kaggle dataset (Bhuvaji et al., 2020) referred to as dataset 2, and the model outperformed other state-of-the-art models.

2 Related work

There have been many studies on the classification of brain tumors. Alanazi et al. (2022) constructed a 22-layer CNN architecture. Initially, the model underwent training with a large dataset utilizing binary classification. Subsequently, the model's weights were adjusted, and it was evaluated on dataset 1 and dataset 2 using migration learning. The model achieved accuracy of 96.89 and 95.75% on dataset 1 and dataset 2, respectively. Hammad et al. (2023) constructed a CNN model with 8 layers. The model achieved an accuracy of 99.48% for binary classification of brain tumors and 96.86% for three-class classification. Liu et al. (2023) introduced the self-attention similarity-guided graph convolutional network (SASG-GCN) model to classify multi-type low-grade gliomas. The model incorporates a convolutional depth setting signal network and a self-attention-based method for chart construction on a 3D MRI water surface, which achieved an accuracy of 93.62% on the TCGA-LGG dataset. Kumar et al. (2021) employed the pre-trained ResNet50 model for brain tumor classification, achieving a final accuracy of 97.48% on dataset 1. Swati et al. (2019) presented an exposition on the merits and demerits of conventional machine learning and deep learning techniques. They introduced a segmented fine-tuning approach leveraging a pre-trained deep convolutional neural network model. Through fine-tuning, they achieved an accuracy of 94.82% on dataset 1 using the VGG19 architecture. Ghassemi et al. (2020) employed a pre-trained generative adversarial network (GAN) for feature extraction in the classification of brain tumors. The experiment was conducted on dataset 1, yielding an accuracy of 95.6%. Saurav et al. (2023) introduced a novel lightweight attention-guided convolutional neural network (AG-CNN). This network incorporates a channel attention mechanism. The model achieves accuracies of 97.23 and 95.71% on dataset 1 and dataset 2, respectively.

Integration through models is a feasible solution. In Hossain et al. (2023), an ensemble model IVX16 was proposed based on the average of the classification results of three pre-trained models (VGG16, InceptionV3, Xception).The model achieved a classification accuracy of 96.94% on dataset 2. A comparison between IVX16 and Vison Transformer (ViT) models reveals that IVX16 outperforms the ViT models. Tandel et al. (2021) presented a method of majority voting. Firstly, five pre-trained convolutional neural networks and five machine learning models are used to classify brain tumor MRI images into different grades and types. Next, a majority voting-based ensemble algorithm is utilized to combine the predictions of the ten models and optimize the overall classification performance. In Kang et al. (2021), nine pre-trained models including ResNet, DenseNet, VGG, AlexNet, InceptionV3, ResNeXt, ShuffleNetV2, MobileNetV2, and MnasNet were employed. The pre-trained models were utilized to extract features, which were then forwarded to a machine learning classifier. From the extracted features, three deep features with excellent performance were selected and concatenated along the channel dimension. The resulting feature representation was subsequently sent to both the machine learning classifier and fully connected (FC) layer. On dataset 2, the model achieved an accuracy of 91.58%. Alturki et al. (2023) employed a voting-based approach to classify brain tumors as either healthy or tumorous. They utilized a CNN to extract tumor features, and employed logistic regression and stochastic gradient descent as the classifiers. To achieve high accuracy of tumor classification, a soft voting method was employed.

Furthermore, the combination of CNNs and machine learning classifiers offers the potential ways to enhance the model's performance. Sekhar et al. (2022), image features were extracted using GoogLeNet, and feature classification was performed using both support vector machines (SVM) and K-Nearest Neighbor (KNN). Ultimately, KNN outperformed SVM, achieving a model accuracy of 98.3% on dataset 1. Deepak and Ameer (2021) employed a hybrid approach combining CNN and SVM to effectively classify three distinct types of brain tumors. The researchers introduced a CNN architecture comprising five convolutional layers and two fully-connected layers. Subsequently, they extracted features from the initial fully connected layer of the designed CNN model, and ultimately performed classification using SVM. Remarkably, this approach achieved an impressive classification accuracy of 95.82% on dataset 1. Özyurt et al. (2019), the researchers utilized a hybrid approach called Neutrosophy and Convolutional Neural Network (NS-CNN) to classify tumor regions that were segmented from brain images into benign and malignant categories. Initially, the MRI images undergo segmentation employing the Neutral Set Expert Maximum Fuzzy Determination Entropy (NS-EMFSE) method. Subsequently, the features of the segmented brain images are extracted through a CNN and then classified using SVM and K-Nearest Neighbors (KNN) classifiers. The experimental results demonstrated that the utilization of CNN features in conjunction with SVM yielded superior classification performance, achieving an average accuracy of 95.62%. Gumaei et al. (2019) introduced the classification method of brain tumors based on the hybrid feature extraction method of regularized extreme learning machine (RELM). In this paper, the mixed feature extraction method is used to extract the features of brain tumors, and RELM is used to classify the types of brain tumors. This method achieves 94.233% classification accuracy on dataset 1. Öksüz et al. (2022) introduced a method that combines deep and shallow features. Deep features of brain tumors were extracted using pre-trained models: AlexNet, ResNet-18, GoogLeNet, and ShuffleNet. Subsequently, a shallow network is developed to extract shallow features from brain tumors, followed by fusion with the deep features. The fused features are utilized to train SVM and KNN classifiers. This method achieves a classification accuracy of 97.25% on dataset 1. In their work, Demir and Akbulut (2022) developed a Residual Convolutional Neural Network (R-CNN) to extract profound features. Subsequently, they applied the L1-Norm SVM ReliefF (L1NSR) algorithm to identify the 100 most discriminative features and utilized SVM for classification. The achieved classification accuracies for 2-categorized and 4-categorized data were 98.8 and 96.6%, respectively.

Moreover, the hyperparameters of the model can be optimized through the utilization of an optimization algorithm. Ren et al. (2023), the study employed preprocessing, feature selection, and artificial neural networks for the classification of brain tumors. Furthermore, the authors utilized a specific optimization algorithm known as water strider courtship learning to optimize both the feature selection and neural network parameters. The effectiveness of the proposed method was evaluated on the "Brain-Tumor-Progression" database, obtaining a final classification accuracy of 98.99%. SbDL was utilized by Sharif et al. (2020) for saliency map construction, while deep feature extraction was performed using the pre-trained Inception V3 CNN model. The connection vector was optimized using Particle Swarm Optimization (PSO) and employed for classification with the softmax classifier. The proposed method was validated on Brats2017 and Brats2018 datasets with an average accuracy of more than 92%. In Nirmalapriya et al. (2023), employed a combination of U-Net and CFPNet-M for segmenting brain tumors into four distinct classes. The segmentation process was conducted using the Aquila Spider Monkey Optimization (ASMO) to optimize segmentation model and the Spider Monkey Optimization (SMO), Aquila Optimizer (AO), and Fractional Calculus (FC) optimized SqueezeNet models. The model achieved a tested accuracy of 92.2%. The authors introduced a model, referred to in Nanda et al. (2023) as the Saliency-K-mean-SSO-RBNN model. This model comprises the K-means segmentation technique, radial basis neural network, and social spider optimization algorithm. The tumor region is segmented using the k-means clustering method. The segmented image then undergoes feature extraction through multiresolution wavelet transform, principal component analysis, kurtosis, skewness, inverse difference moment (IDM), and cosine transforms. The clustering centers are subsequently refined using the social spider optimization (SSO) algorithm, followed by processing the feature vectors for efficient classification using the radial basis neural network (RBNN). The final model achieves classification accuracies of 96, 92, and 94% on the three respective datasets.

3 Materials and methods

This paper utilizes three pre-trained models, namely ResNet101, DenseNet121, and EfficientNetB0. The outputs of these models are adjusted to ensure consistent data size, and then the extracted features from these models are fused. Subsequently, feature classification is performed. To achieve consistent output from the feature extraction modules across all models, we harmonized the feature extraction modules of EfficientNetB0 and ResNet101 with DenseNet121 by utilizing a 1×1 convolutional layer.

3.1 Datasets and Preprocessing

The study employed two datasets. Dataset 1, introduced by Cheng et al. (2015), is a publicly available dataset comprising 3,064 T1 MRI images. It includes three different types of brain tumors: glioma (1,426 images), meningioma (708 images), and pituitary tumor (930 images). Dataset 2, a widely used open-source dataset (Bhuvaji et al., 2020), encompasses 3,264 MRI images which consist of four categories: glioma (926 images), meningioma (937 images), pituitary tumor (901 images), and normal (500 images).

The MRI data consists of two-dimensional images with a size of 512×512. However, the input of the pre-training model is necessary to be RGB image. Therefore, the images were resized to dimensions of $224 \times 224 \times 3$. Furthermore, the min-max normalization method was adopted to scale the intensity values of the image to the range of $[0, 1]$. The dataset 2 was processed in the same way. We divided the dataset into a training set and a test set with a ratio of 8:2.

3.2 Architecture of the proposed method

Transfer learning is a kind of machine learning technique, which leverages the knowledge acquired during training on one problem to

train on another task or domain. The transfer learning approach, which utilizes pre-trained network knowledge obtained from extensive visual data, is very advantageous in terms of time-saving and achieving superior accuracy compared with training a model from scratch (Yu et al., 2022; Arora et al., 2023).

ResNet, DenseNet and EfficietNet have been proved to be very effective brain tumor classification models (Zhang et al., 2023; Zulfiqar et al., 2023). The accuracy of brain tumor classification of VGG19 and ResNet50 is 87.09 and 91.18%, respectively (Zhang et al., 2023). The accuracy of GoogLeNet is 94.9% (Sekhar et al., 2022). We also have tested the ability of ResNet101 and EfficientNetB0 for brain tumor classification, whose accuracy is 96.57, 96.41%, respectively. The comparison shows that ResNet101, DenseNet121 and EfficientNetB0 are more accurate, so they are chosen as the basic models.

Figure 1 depicts the framework of the proposed method in this paper. Firstly, the brain tumor data was processed and the images were adjusted. Secondly, features are extracted from brain tumor images using pre-trained models. Finally, the extracted features are then aggregated for feature fusion, followed by classification. Specifically, ResNet101, DenseNet121, and EfficientNetB0 serve as pre-trained models. The outputs of the ResNet101 and EfficientNetB0 feature extraction layers are adjusted to dimensions of (1,024, 7, 7). Brain tumor feature fusion is accomplished by pairwise summation of the extracted features. Finally, the fused features are classified using a linear classifier.

3.3 Pre-trained models

As a fundamental component of neural network architecture, the convolutional layer extracted features by sliding a fixed-size convolutional kernel over the original image and performing multiplication operations between the kernel parameters and the image. To achieve different effects, the convolution operation relies on additional parameters, primarily the step size, padding, and size of the convolution kernel. The size of the output features from the convolutional layer can be calculated using Equation (1).

$$H_{out} = \frac{H_{\text{in}} + 2 \times padding[0] - \text{ker}nel\_size[0]}{stride[0]} + 1$$
$$W_{out} = \frac{W_{\text{in}} + 2 \times padding[1] - \text{ker}nel\_size[1]}{stride[1]} + 1 \quad (1)$$

where H_{in} and W_{in} represent the dimensions of the input data, padding refers to the number of zero-padding layers, *Kernel_size* represents the dimensions of the convolution kernel. And stride represents the step size of the convolution operation. The formula indicates that when the *kernel_size* is set to (1,1), the stride is set to 1 and padding is set to 0, the output dimension of the convolutional layer remains unchanged.

3.3.1 ResNet101

Residual network (ResNet) is a widely recognized and straightforward model used for deep learning tasks, particularly in image recognition (He et al., 2016). Previously, as the number of network layers increases, a common issue of vanishing gradients may arise, resulting in performance saturation and degradation of the model. Deep residual networks address this issue by incorporating jump connections between layers to mitigate information loss. The core idea of the deep residuals network is to add a path parallel to the main convolution path, which combines the features from the subsequent convolution layer with those from the previous layer within the same residuals block, in turn to can achieve a deeper network model. Within the residual network, each building block performs an identity mapping, and the resulting features are element-wise summed across the convolutional layers preceding and following

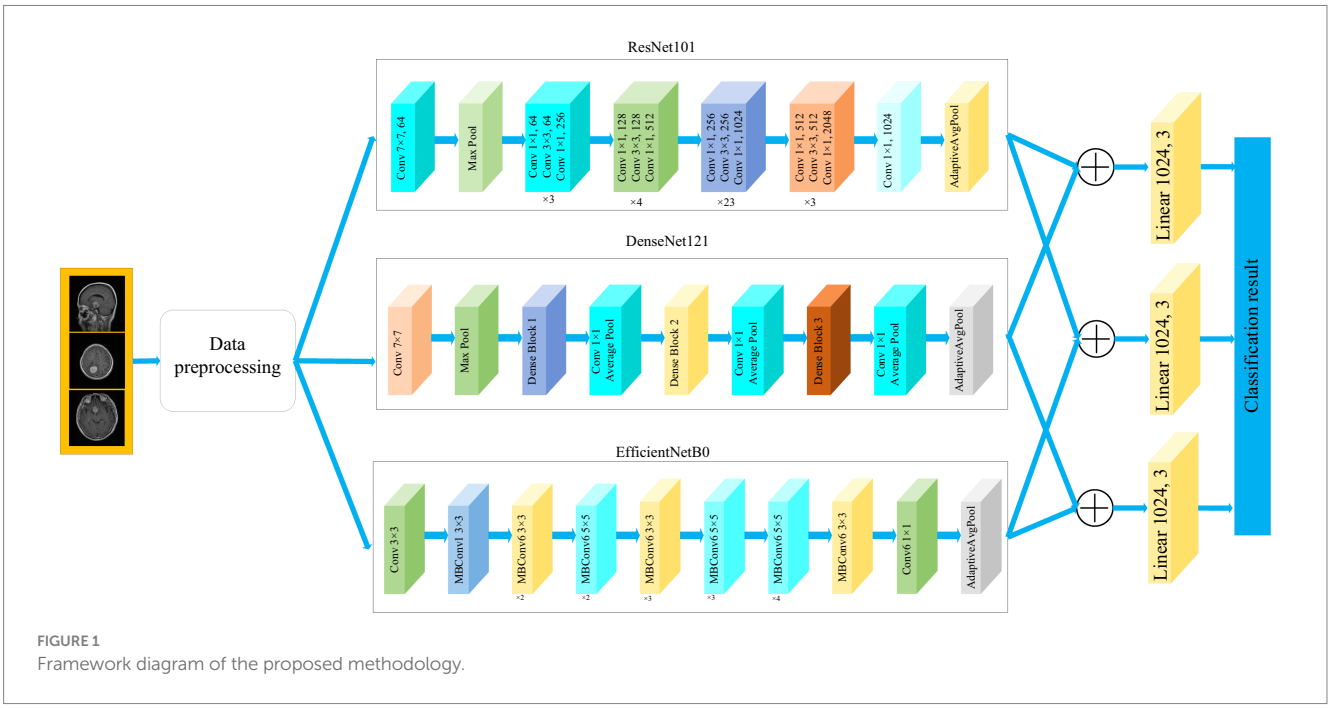


FIGURE 1
Framework diagram of the proposed methodology.

the identity connection. Figure 2 illustrates the foundational architecture of ResNet101. The feature extraction layer of the ResNet101 model produces an output with dimensions of (2048, 7, 7). Subsequently, a 1 × 1 convolutional layer with 1,024 convolutional kernels is added to the base model, which modifies the output dimension to (1,024, 7, 7).

3.3.2 DenseNet121

The DenseNet convolutional neural network model was proposed by Huang et al. (2017). The network is based on the ResNet structure, but it incorporates dense connections (i.e., summed variable joins) between all preceding and subsequent layers. Another significant aspect of DenseNet is the reuse of features through channel connections. In DenseNet, every layer receives feature maps as input from all preceding layers, and its output feature maps are subsequently utilized as input for each subsequent layer. In ResNet, the features of each block are combined by summation, whereas in DenseNet, feature aggregation is accomplished through concatenation. Figure 3 shows the fundamental framework of the DenseNet121 model. The core of the network is the reused combination of Dense Blocks and Transition Layers, forming the intermediate structure of DenseNet. Additionally, the topmost part of DenseNet consists of a 7 × 7 convolutional layer with a stride of 2, and a 3 × 3 MaxPool2d layer with a stride of 2. The output dimension of the feature extraction layer of the model is (1,024, 7, 7).

3.3.3 EfficientNetB0

The EfficientNet model was proposed by the Google AI research team in 2019 (Tan and Le, 2019). In contrast to traditional scaling methods used in previous studies, where the width, depth, and resolution of the deep CNN architecture are arbitrarily increased to enhance model performance, EfficientNets achieve network performance improvement through a fixed-scale approach that scales the width, depth, and resolution of the network's input images. The calculations are as follows [Equations (2–6)]:

$$Depth: d = \alpha^{\varphi} \tag{2}$$

$$Width: w = \beta^{\varphi} \tag{3}$$

$$resolution\,ratio: r = \gamma^{\varphi} \tag{4}$$

$$s.t. \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \tag{5}$$

$$\alpha \geq 1, \beta \geq, \gamma \geq 1 \tag{6}$$

where, α, β, and γ are obtained by hyperparametric mesh search techniques and can determine the allocation of additional resources to the width, depth, and resolution of the network. φ is a user-specified coefficient that controls the amount of additional resources used for model scaling. In Figure 4, the structure of the EfficientNetB0 model is shown. In order to transform the feature output of the EfficientNetB0 model from its original dimension of (1,280, 7, 7) to the desired dimension of (1,024, 7, 7), a 1×1 convolution with 1,024 convolution kernels is applied so that the output is (1,024, 7, 7).

3.4 Training of CNNs

The convolutional neural network training process is a combination of forward and backward propagation. It starts at the input layer and propagates forward from layer to layer until it reaches the classification layer. The error is then propagated back to the first layer of the network. In layer L of the network, input from layer L-1 neuron j is received in a forward propagation path. The weighted sums are calculated as follows [Equation (7)]:

$$In = \sum_{j=1}^{n} W_{ij}^{l} x_j + b_i \tag{7}$$

Here, the letters $W\,l\,ij$ stand for weights, x_j stand for training samples, and b_i stand for bias. The nonlinearity of the model can be increased by the activation function to make the network fit the data better. Equation (8) shows how the Relu function is calculated.

$$R_i^l = \max\left(0, In_i^l\right) \tag{8}$$

In the classification layer of the convolutional neural network, the probability of categorization is calculated by the following softmax function. This classification layer evaluates the probability score of each category by softmax function. Equation (9) shows the method of calculation.

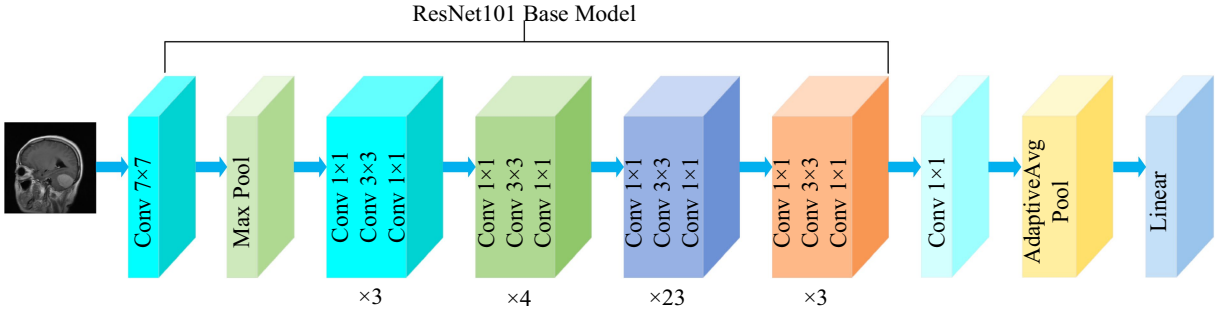


FIGURE 2
Structure of the ResNet101 model.

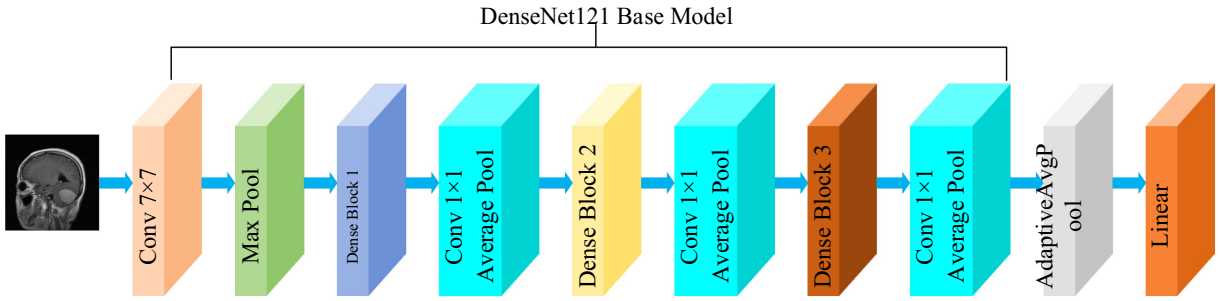


FIGURE 3
Structure of the DenseNet121 model.

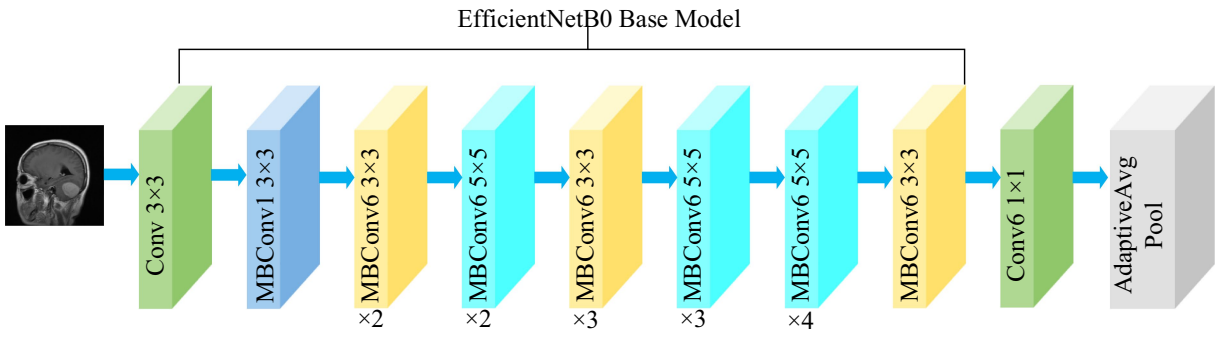


FIGURE 4
Structure of the EfficientNetB0 model.

$$out_i^l = \frac{e^{In_i^l}}{\sum_i e^{out_i^l}} \tag{9}$$

CNN weights are updated by Backpropagation. The algorithm uses unknown weight W to minimize the tracking cost function. The loss function is calculated as follows [Equation (10)]:

$$C = -\frac{1}{m}\sum_i^m \ln\left(P\left(\frac{y_i}{x_i}\right)\right) \tag{10}$$

Here, m represents the total count of training samples. x_i represents the initial training sample. y_i represents the label associated with the sample x_i. And $P\left(\frac{y_i}{x_i}\right)$ represents the probability of x_i belonging to class y_i.

Stochastic gradient descent on small batches of size N is used to minimize the cost function C and approximate the training cost by the small batch cost. W denotes the weights at iteration t of the l convolutional layer, and C denotes the small batch cost. The weights are then updated in the next iteration as follows [Equation (11)]:

$$\gamma^t = \gamma^{\left\lfloor \frac{tN}{m} \right\rfloor}$$

$$V^{t+1} = \mu V_l^t - \gamma^t \alpha_l \frac{\partial C}{\partial W} \tag{11}$$

$$W_i^{t+1} = W_l^t + V_t^{t+1}$$

In this case, α_l is the learning rate of layer l. γ is the scheduling rate that reduces the initial learning rate at the end of a specified number of periods. And μ stands for the momentum factor, which indicates the effect of the previously updated weights on the current iteration.

4 Results and discussion

The experiments were conducted on a Windows 10 system with 64 GB of Random Access Memory (RAM). The graphics card utilized was RTX 4070, and the programming language employed was Python, with PyTorch serving as the framework. The hyperparameters of the model in the experiment are shown in Table 1.

4.1 Evaluation metrics

To comprehensively assess the effectiveness of the model, the evaluation metrics including accuracy, precision, recall, and F1-score are employed in this paper. The expressions of the evaluation metrics are shown in Equations (12–15) (Yeung et al., 2022; Alyami et al., 2023).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

TABLE 1 Hyperparameters.

Parameters	Setting
Epoch	25
Learning rate	0.0001
Batch size	32
Optimizer	Adam
Loss function	Cross entropy

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{15}$$

where, true positive (*TP*) represents the count of accurately classified sick images in each respective category. True negative (*TN*) denotes the total number of correctly classified images in all categories, excluding the relevant category. False negative (*FN*) represents the count of incorrectly classified images in the relevant category. False positive (*FP*) denotes the count of misclassified images in all categories, excluding the relevant category.

4.2 Classification results

This section presents the classification results of the proposed method and includes a comparative analysis with and without the utilization of feature fusion methods.

4.2.1 The representation of a single model

The confusion matrix illustrating the classification results of models, which was pre-trained through fine-tuning on the test set of the dataset 1, is presented in Figure 5. To analyze the classification outcomes of the three pre-trained models on the test set of the dataset 2, Figure 6 shows the corresponding confusion matrix. Additionally, Table 2 lists the specific values of accuracy, precision, recall, and F1-score, calculated using Equations (12–15) respectively. According to Table 2, on dataset 1, DenseNet121 has the best classification performance for brain tumor with 98.53% accuracy, while on dataset 2, ResNet101 has excellent classification performance with 95.71% accuracy.

4.2.2 With feature fusion

Figures 7, 8 display the confusion matrices of the brain tumor classification results achieved by feature fusion on dataset 1 and dataset 2, respectively. Furthermore, Table 3 present detailed values of the classification indexes for dataset 1 and dataset 2. It can be seen that ResNet101 + DenseNet121 attains optimal classification results on both datasets, with an accuracy of 99.18% on dataset 1 and 97.24% on dataset 2.

Figures 9A, B show the average evaluation metrics for brain tumor classification of every model on dataset 1 and dataset 2, respectively. On the dataset 1, from Figure 9A, it can be observed that the combination of ResNet101 and DenseNet121 (ResNet101 + DenseNet121) achieved the best classification accuracy, precision, recall, and F1-score, with values of 99.18, 99.07, 99.11, and 99.08%, respectively. Additionally, among the individual models, EfficientNetB0 exhibits the best classification results for brain tumor classification. Notably, DenseNet121 outperforms ResNet101 + EfficientNetB0 but is outperformed by both ResNet101 + DenseNet121 and DenseNet121 + EfficientNetB0. In Figure 9B (i.e., dataset 2), the ResNet101 + DenseNet121 model also achieves the best performance. However, among the individual models, DenseNet121 exhibits the best classification results, with accuracy, precision, recall, and F1-score of 97.24, 97.06, 97.58, and 97.28%, respectively. Unlike dataset 1, where DenseNet121 showed strong performance, it appears to have the weakest classification ability on the dataset 2. Conversely, ResNet101 + DenseNet121, ResNet101 + EfficientNetB0, and DenseNet121 + EfficientNetB0 all outperform the individual models. The experimental results validate the effectiveness of combining features from different models through feature fusion, thus providing a more reliable approach for brain tumor classification than relying on a single model. In addition, the average improvement of ResNet101 + DenseNet121 is 2.085% (dataset 1 is 2.61%, dataset 2 is 1.56%) and 1.32% (dataset 1 is 0.65%, dataset 2 is 1.99%) compared with ResNet101 and DenseNet121, respectively. Similarly, the accuracy improvement for ResNet101 + EfficientNetB0 is 1.035% (1.31% for dataset 1 and 0.76% for dataset 2) and 1.345% (1.47% for dataset 1 and 1.22% for dataset 2) compared with ResNet101and EfficientNetB0 alone. In comparison with Densenet121 and EfficientNetB0, the average accuracy improvement for DenseNet121 + EfficientNetB0 is 1.225% (0.61% for dataset 1 and 1.84% for data set 2) and 1.985% (2.28% for dataset 1 and 1.69% for dataset 2), respectively. The modeled results strongly support the efficacy of employing feature fusion in brain tumor classification. In addition, it is evident that ResNet101 achieves the most favorable classification results, while DenseNet121 yields the terrible results on dataset 2. But the classification effectiveness of ResNet101 + DenseNet121 surpasses that of ResNet101 + EfficientNetB0 and DenseNet121 + EfficientNetB0. This suggests that the combination of ResNet101 and DenseNet121 outperforms configurations involving EfficientNetB0. The possible reason for this phenomenon is the inferior feature matching effect of ResNet101 + EfficientNetB0 and DenseNet121 + EfficientNetB0 compared to ResNet101 + DenseNet121.

A subject Receiver Operating Curve (ROC) is also utilized in the analysis process. It is a curve that illustrates the relationship between the true positive rate and the false positive rate. The size of the Area Under Curve (AUC) of the ROC curve indicates the strength of the model's ability to differentiate between different types of tumors, with a larger AUC value indicating better classification performance. As shown in Figure 10, the ROC curves of ResNet101 + DenseNet121 for the model are demonstrated and the values of AUC for the three types of brain tumors in dataset 1 are 0.9987, 0.9952, and 0.9999, respectively. In dataset 2, the values of AUC are 0.9991, 0.9971, 0.9999, and 0.9998, respectively.

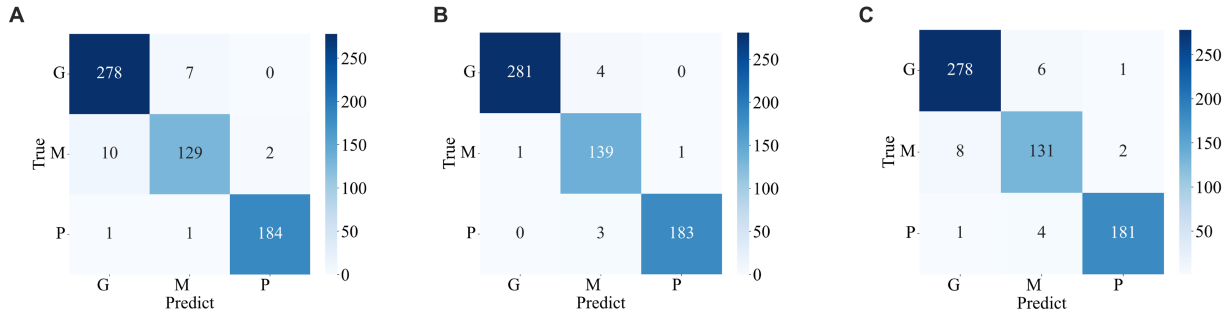


FIGURE 5
Confusion matrix of predicted results for a single model on the test set of the dataset 1. **(A)** ResNet101 **(B)** DenseNet121 **(C)** EfficientNetB0.

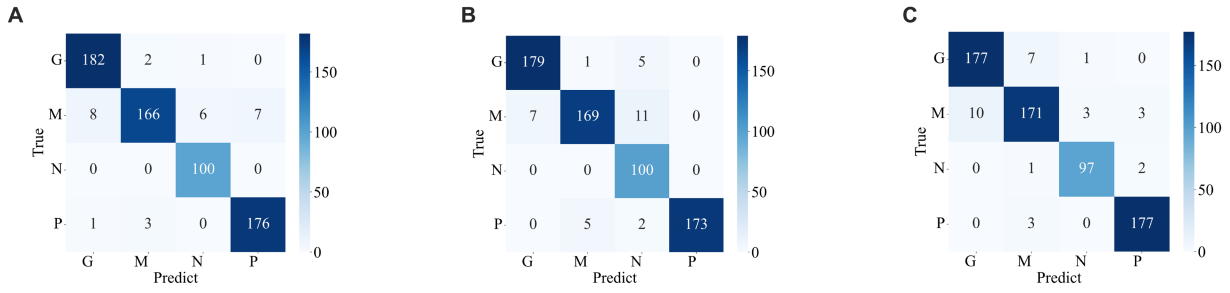


FIGURE 6
Confusion matrix of the predicted results of a single model on the test set of the dataset 2 **(A)** ResNet101 **(B)** DenseNet121 **(C)** EfficientNetB0.

4.2.3 Cross-dataset validation and robustness validation

Based on the foregoing, it is evident that the ResNet101 + DenseNet121 yields superior classification results across the two public datasets. This section aims to assess the robustness of ResNet101 + DenseNet121. To further assess the model's robustness, a cross-data verification method was employed. The normal class in Dataset 2 was excluded, and data from the remaining three brain tumor classes were utilized to evaluate the dataset 1 trained model, ResNet101 + DenseNet121. The precision, recall, F1-score and accuracy of ResNet101 + DenseNet121 are verified to be 94.71, 94.44, 94.41, and 94.38%, respectively, which indicates its good robustness.

4.3 Discussion

There have been many studies on brain tumor classification. Among these methods, the key is the extracted features. Generally, there is a relationship between the effectiveness of the model and the amount of data. Whereas the acquisition of medical images is usually difficult and expensive. Transfer learning can take full advantage of its advantages on tasks with small datasets to improve model performance, accelerate the training process, and reduce the risk of overfitting. In addition, model integration is a technique that combines the prediction results of multiple independently trained models to obtain more powerful and robust global predictions, which can improve the upper limit of performance. In our work, the pre-trained model is used to extract the features of the image, and then the extracted features are fused using the model integration method of feature fusion to enhance the ability of the model.

From the previous analysis, it can be found that among the three fused models, ResNet101 + DenseNet121 achieves the best classification results. ResNet101 adopts the method of residual learning to construct residual blocks, which makes the network easier to train and reduces the problem of gradient vanishing. Densenet121, on the other hand, uses the idea of dense connectivity, where each layer's input contains the output of all previous layers. This kind of connection is helpful to the transmission of information and the flow of gradients, and slows down the problem of information bottleneck. Dense connectivity also facilitates feature reuse. The features extracted by ResNet101 and those extracted by Densenet121 are fused to realize the complementary feature, which makes the feature more abundant and diversified, and thus achieves better classification effect. To demonstrate the effectiveness of the proposed method, we use the method of t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the features extracted by the model ResNet101 + DenseNet121 trained on dataset 1, and the visualization results are shown in Figure 11. The feature set of ResNet101 is shown in Figure 11A. It can be seen that some gliomas and meningiomas are nested with each other. The mean and standard deviation of the feature set are−0.0057 and 0.6141, respectively. The feature set of DenseNet121 is shown in Figure 11B, which shows that only a few gliomas and meningiomas are nested with each other. The mean and standard deviation of the feature set are 0.2323 and 0.652795, respectively. Figure 11C displays the feature set of ResNet101 + DenseNet121, indicating minimal nested classes. The mean and standard deviation of the feature set are 0.2267 and 0.9604,

TABLE 2 Indicators for the classification of a single model.

Dataset	Model	Tumor type	Precision	Recall	F1-score	Accuracy
Dataset 1	ResNet101	Glioma	96.19%	97.54%	96.86%	
		Meningioma	94.16%	91.49%	92.81%	
		Pituitary	98.92%	98.92%	98.92%	
		average	96.43%	95.99%	96.20%	96.57%
	DenseNet121	Glioma	99.65%	98.60%	99.12%	
		Meningioma	95.21%	98.58%	96.86%	
		Pituitary	99.46%	98.39%	98.92%	
		average	98.10%	98.52%	98.30%	98.53%
	EfficientNetB0	Glioma	96.86%	97.54%	97.20%	
		Meningioma	92.91%	92.91%	92.91%	
		Pituitary	98.37%	97.31%	97.84%	
		average	96.05%	95.92%	95.98%	96.41%
Dataset 2	ResNet101	Glioma	95.29%	98.38%	96.81%	
		Meningioma	97.08%	88.77%	92.74%	
		NoTumor	93.46%	100.0%	96.62%	
		Pituitary	96.17%	97.78%	96.97%	
		Average	95.50%	96.23%	95.78%	95.71%
	DenseNet121	Glioma	96.24%	96.76%	96.50%	
		Meningioma	96.57%	90.37%	93.37%	
		NoTumor	84.75%	100.0%	91.74%	
		Pituitary	100.0%	96.11%	98.02%	
		Average	94.39%	95.81%	94.91%	95.25%
	EfficientNetB0	Glioma	94.65%	95.68%	95.16%	
		Meningioma	93.96%	91.44%	92.68%	
		NoTumor	96.04%	97.00%	96.52%	
		Pituitary	97.25%	98.33%	97.79%	
		Average	95.48%	95.61%	95.54%	95.40%

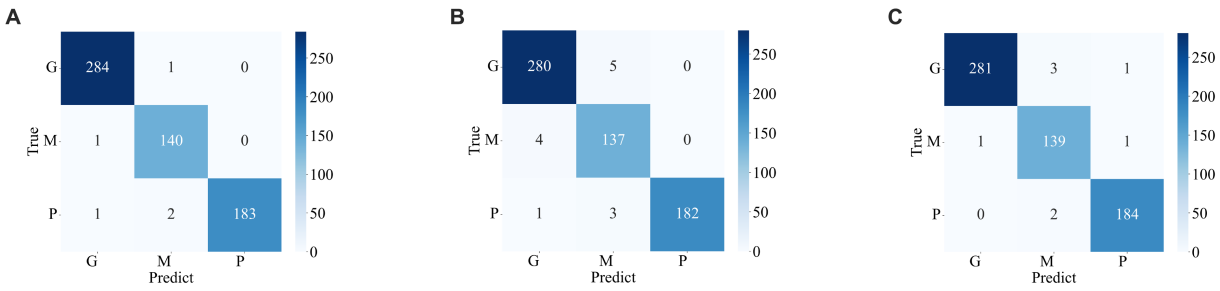


FIGURE 7
Classification results of brain tumors on the test set of the dataset 1 **(A)** ResNet101 + DenseNet121 **(B)** ResNet101 + efficientNetB0 **(C)** DenseNet121 + EfficientNetB0.

respectively. Additionally, the analysis shows that the standard deviation of the feature set of ResNet101 + Densenet121 is the highest, which also shows that ResNet101 + Densenet121 increases the uniqueness of extracting the image features of brain tumors and enhances the ability to distinguish brain tumors.

4.4 Comparison with other state of the art methods

We compared the classification results obtained in this study with those reported in the literature using the same

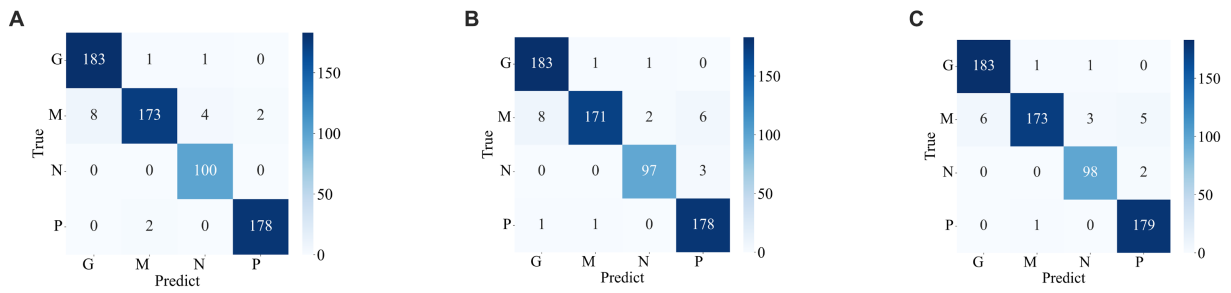


FIGURE 8
Classification results of brain tumors on the test set of the dataset 2 **(A)** ResNet101 + DenseNet121 **(B)** ResNet101 + efficientNetB0 **(C)** DenseNet121 + EfficientNetB0.

TABLE 3 The classification results of feature fusion methods.

Dataset	Model	Tumor type	Precision	Recall	F1-score	Accuracy
Dataset 1	ResNet101 + DenseNet121	Glioma	99.30%	99.65%	99.47%	
		Meningioma	97.90%	99.29%	98.58%	
		Pituitary	1.00%	98.39%	99.19%	
		average	99.07%	99.11%	99.08%	99.18%
	ResNet101 + EfficientNetB0	Glioma	98.25%	98.25%	98.25%	
		Meningioma	94.48%	97.16%	95.80%	
		Pituitary	1.00%	97.85%	98.91%	
		average	97.58%	97.75%	97.65%	97.88%
	DenseNet121 + EfficientNetB0	Glioma	99.65%	98.60%	99.12%	
		Meningioma	96.53%	98.58%	97.54%	
		Pituitary	98.92%	98.92%	98.92%	
		average	98.37%	98.70%	98.53%	98.69%
Dataset 2	ResNet101 + DenseNet121	Glioma	95.81%	98.92%	97.34%	
		Meningioma	98.30%	92.51%	95.32%	
		NoTumor	95.24%	1.00%	97.56%	
		Pituitary	98.89%	98.89%	98.89%	
		Average	97.06%	97.58%	97.28%	97.24%
	ResNet101 + EfficientNetB0	Glioma	95.31%	98.92%	97.08%	
		Meningioma	98.84%	91.44%	95.00%	
		NoTumor	97.00%	97.00%	97.00%	
		Pituitary	95.19%	98.89%	97.00%	
		Average	96.59%	96.56%	96.52%	96.47%
	DenseNet121 + EfficientNetB0	Glioma	96.83%	98.92%	97.86%	
		Meningioma	98.86%	92.51%	95.58%	
		NoTumor	96.08%	98.00%	97.03%	
		Pituitary	96.24%	99.44%	97.81%	
		Average	97.00%	97.22%	97.07%	97.09%

dataset. The compared results shown in Table 4 demonstrate that our study achieved competitive classification performance when compared to the state-of-the-art approaches in the current literature.

5 Conclusion

This paper proposes a novel method for brain tumor classification, utilizing feature fusion to improve performance. Three advanced

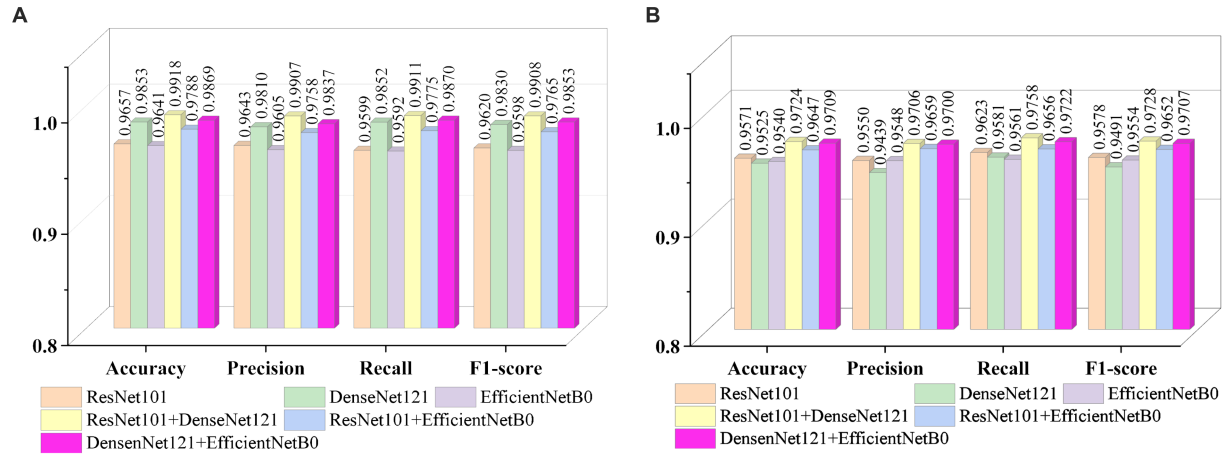


FIGURE 9
Visualization of brain tumor classification metrics **(A)** dataset 1 **(B)** dataset 2.

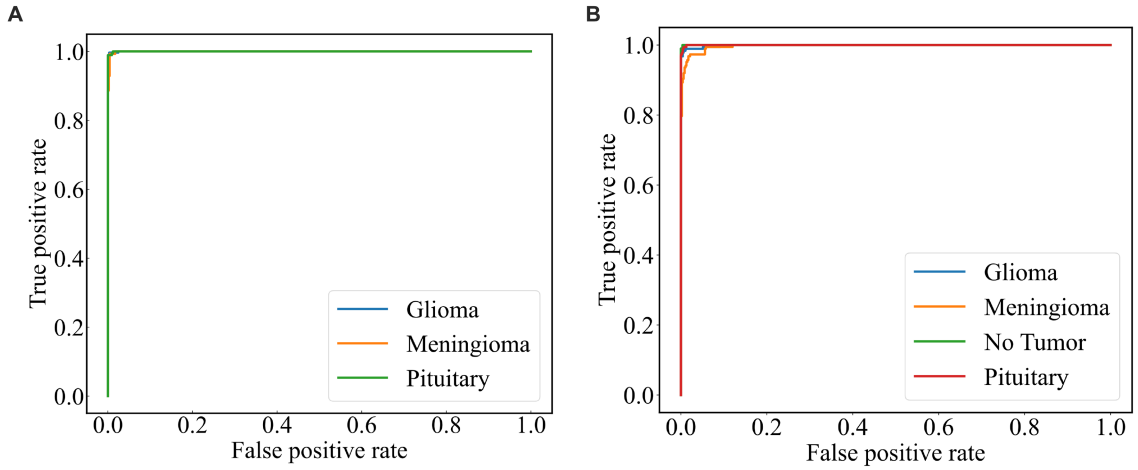


FIGURE 10
ROC curve of the model **(A)** dataset 1 **(B)** dataset 2.

pre-trained models including ResNet101, DenseNet121, and EfficientNetB0, were selected as base models and adjusted to have the same output size (1,024, 7, 7). Brain tumor images were fed into these models to extract their respective features, and then feature fusion was achieved by pairwise combination of the models through feature summation. The fused features were subsequently used for the final classification. The method was validated on two publicly available datasets, and evaluation metrics such as accuracy, precision, recall, and F1-score were employed. Experimental Results indicated that the combination of ResNet101 and DenseNet121 (ResNet101 + DenseNet121) achieved the best classification results for both dataset 1 and dataset 2. On dataset 1, accuracy of 99.18%, precision of 99.07%, recall of 99.11%, and F1-score of 99.08% were achieved. For dataset 2, the corresponding metrics values including accuracy of 97.24%, precision of 97.06%, recall of 97.58%, and F1-score of 97.28% were obtained. Comparing our method with other state-of-the-art techniques, our approach exhibits superior classification performance. In the future, we plan to study two important works. On one hand, we will expand the experimentation by incorporating additional models to validate the effectiveness of feature fusion through summation for brain tumor classification. On the other hand, we aim to extend this method to encompass other brain diseases, thus enhancing the model's capacity to recognize multiple classes of brain diseases.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427 and https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

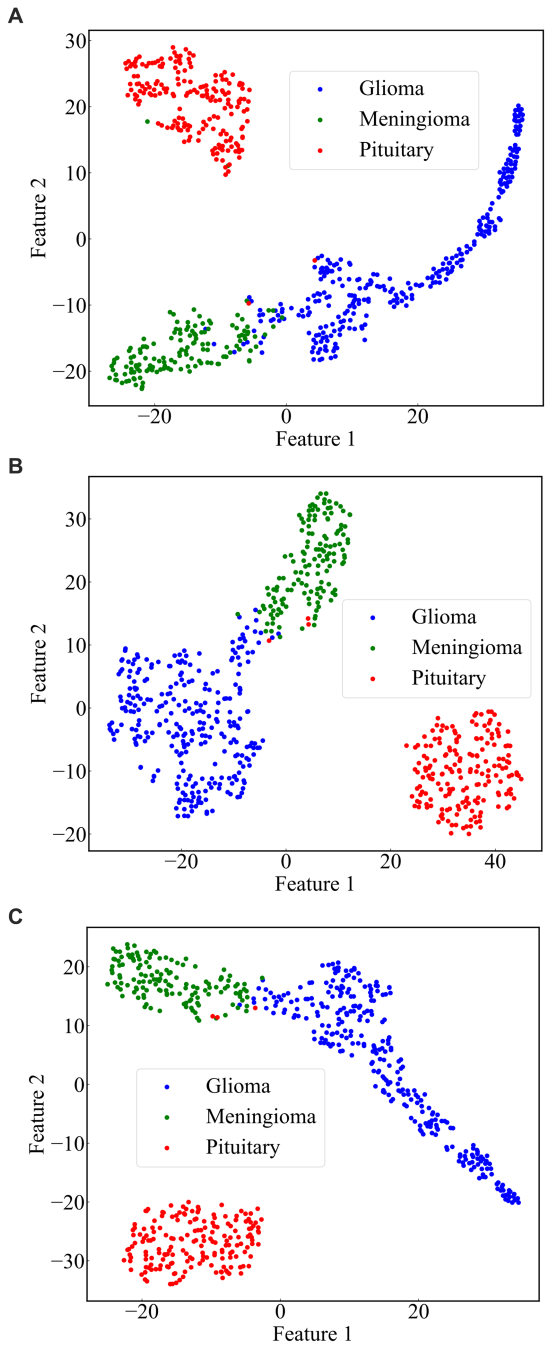


FIGURE 11
Scatterplot of the feature set. **(A)** ResNet101 **(B)** DenseNet121 **(C)** ResNet101 + DenseNet121.

TABLE 4 Comparison with other state-of-the-art models.

Reference	Dataset	Method	Accuracy
Gumaei et al. (2019)	Dataset 1	RELM	94.233%
Swati et al. (2019)	Dataset 1	Fine-tuning the VGG19 model.	94.82%
Ghassemi et al. (2020)	Dataset 1	Pre-trained GAN	95.6%
Deepak and Ameer (2021)	Dataset 1	CNN + SVM	98%
Sekhar et al. (2022)	Dataset 1	GoogLeNet+KNN	98.3%
Öksüz et al. (2022)	Dataset 1	Deep and shallow feature fusion	97.25%
Hammad et al. (2023)	Dataset 1	CNN model with 8 layers	96.86%
Saurav et al. (2023)	Dataset 1	Pre-trained ResNet50	97.48%
Kang et al. (2021)	Dataset 2	Feature connection	91.8%
Demir and Akbulut (2022)	Dataset 2	R-CNN	96.6%
Hossain et al. (2023)	Dataset 2	Ensemble Model	96.94%
Alanazi et al. (2022)	Dataset 1	CNN	96.89%
	Dataset 2		95.75%
Saurav et al. (2023)	Dataset 1	AG-CNN	97.23%
	Dataset 2		95.71%
Proposed model	Dataset 1	ResNet101 + DenseNet121	99.18%
	Dataset 2		97.24%

Author contributions

WC: Formal analysis, Software, Validation, Visualization, Writing – review & editing. XT: Software, Writing – original draft. JZ: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft. GD: Investigation, Project administration, Visualization, Writing – review & editing. QF: Validation, Writing – review & editing. HJ: Investigation, Methodology, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Major Science and Technology Projects of Henan Province (Grant No. 221100210500), the Foundation of Henan Educational Committee (No. 24A320004), the Medical and Health Research Project in Luoyang (Grant No. 2001027A), and the Construction Project of Improving Medical Service Capacity of Provincial Medical Institutions in Henan Province (Grant No. 2017-51).

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

References

Alanazi, M. F., Ali, M. U., Hussain, S. J., Zafar, A., Mohatram, M., Irfan, M., et al. (2022). Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* 22:372. doi: 10.3390/s22010372

Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., and Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Comput. Methods Prog. Biomed.* 223:106951. doi: 10.1016/j.cmpb.2022.106951

Alturki, N., Umer, M., Ishaq, A., Abuzinadah, N., Alnowaiser, K., Mohamed, A., et al. (2023). Combining CNN features with voting classifiers for optimizing performance of brain tumor classification. *Cancers* 15:1767. doi: 10.3390/cancers15061767

Alyami, J., Rehman, A., Almutairi, F., Fayyaz, A. M., Roy, S., Saba, T., et al. (2023). Tumor localization and classification from MRI of brain using deep convolution neural network and Salp swarm algorithm. *Cogn. Comput.* doi: 10.1007/s12559-022-10096-2

Arora, G., Dubey, A. K., Jaffery, Z. A., and Rocha, A. (2023). A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data. *Neural Comput. & Applic.* 35, 7989–8015. doi: 10.1007/s00521-022-06922-1

Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S, and Kanchan, S. (2020). Brain Tumor Classification (MRI)Available at: https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-m. Accessed August, 1 2020

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

Chan, H.-P., Hadjiiski, L. M., and Samala, R. K. (2020). Computer-aided diagnosis in the era of deep learning. *Med. Phys.* 47, e218–e227. doi: 10.1002/mp.13764

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

Deepak, S., and Ameer, P. M. (2021). Automated Categorization of Brain Tumor from MRI Using CNN features and SVM. *J. Ambient. Intell. Human Comput.* 12, 8357–8369. doi: 10.1007/s12652-020-02568-w

Demir, F., and Akbulut, Y. (2022). A new deep technique using R-CNN model and L1NSR feature selection for brain MRI classification. *Biomed. Signal Process. Control* 75:103625. doi: 10.1016/j.bspc.2022.103625

Fujita, H. (2020). AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiol. Phys. Technol.* 13, 6–19. doi: 10.1007/s12194-019-00552-4

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., 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

Gudigar, A., Raghavendra, U., Hegde, A., Kalyani, M., Kalyani, M., Ciaccio, E. J., et al. (2020). Brain pathology identification using computer aided diagnostic tool: a systematic review. *Comput. Methods Prog. Biomed.* 187:105205. doi: 10.1016/j.cmpb.2019.105205

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

Hammad, M., ElAffendi, M., Ateya, A. A., and El-Latif, A. A. A. (2023). Efficient brain tumor detection with lightweight end-to-end deep learning model. *Can. Underwrit.* 15:2837. doi: 10.3390/cancers15102837

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. Las Vegas, NV, USA

Hossain, S., Chakrabarty, A., Gadekallu, T. R., Alazab, M., and Piran, M. J. (2023). Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. *IEEE J. Biomed. Health Inform.* 99, 1–14. doi: 10.1109/JBHI.2023.3266614

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *IEEE Conf. Comp. Vision Pattern Recog.* 2017, 2261–2269. doi: 10.1109/CVPR.2017.243

Kang, J., Ullah, Z., and Gwak, J. (2021). MRI-based brain tumor classification using Ensemble of Deep Features and Machine Learning Classifiers. *Sensors* 21:2222. doi: 10.3390/s21062222

Kumar, R. L., Kakarla, J., Isunuri, B. V., and Singh, M. (2021). Multi-class brain tumor classification using residual network and global average pooling. *Multimed. Tools Appl.* 80, 13429–13438. doi: 10.1007/s11042-020-10335-4

Lin, Y., Xiao, Y., Wang, L., Guo, Y., Zhu, W., Dalip, B., et al. (2022). Experimental exploration of objective human pain assessment using multimodal sensing signals. *Front. Neurosci.* 16:831627. doi: 10.3389/fnins.2022.831627

Liu, L., Chang, J., Zhang, P., Qiao, H., and Xiong, S. (2023). SASG-GCN: self-attention similarity guided graph convolutional network for multi-type lower-grade glioma classification. *IEEE J. Biomed. Health Inform.* 27, 3384–3395. doi: 10.1109/JBHI.2023.3264564

Maurya, S., Tiwari, S., Mothukuri, M. C., Tangeda, C. M., Nandigam, R. N. S., and Addagiri, D. C. (2023). A review on recent developments in cancer detection using machine learning and deep learning models. *Biomed. Signal Process. Control* 80:104398. doi: 10.1016/j.bspc.2022.104398

Muhammad, K., Khan, S., Ser, J. D., and Albuquerque, V. H. C. D. (2021). Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 507–522. doi: 10.1109/TNNLS.2020.2995800

Nanda, A., Barik, R. C., and Bakshi, S. (2023). SSO-RBNN driven brain tumor classification with saliency-K-means segmentation technique. *Biomed. Signal Process. Control* 81:104356. doi: 10.1016/j.bspc.2022.104356

Nirmalapriya, G., Agalya, V., Regunathan, R., and Belsam Jeba Ananth, M. (2023). Fractional Aquila spider monkey optimization based deep learning network for classification of brain tumor. *Biomed. Signal Process. Control* 79:104017. doi: 10.1016/j.bspc.2022.104017

Öksüz, C., Urhan, O., and Güllü, M. K. (2022). Brain tumor classification using the fused features extracted from expanded tumor region. *Biomed. Signal Process. Control* 72:103356. doi: 10.1016/j.bspc.2021.103356

Özyurt, F., Sert, E., Avci, E., and Dogantekin, E. (2019). Brain tumor detection based on convolutional neural network with neutrosophic expert maximum fuzzy sure entropy. *Measurement* 147:106830. doi: 10.1016/j.measurement.2019.07.058

Ren, W., Hasanzade Bashkandi, A., Afshar Jahanshahi, J., AlHamad A, Q. M., Javaheri, D., and Mohammadi, M. (2023). Brain tumor diagnosis using a step-by-step methodology based on courtship learning-based water strider algorithm. *Biomed. Signal Process. Control* 83:104614. doi: 10.1016/j.bspc.2023.104614

Satyanarayana, G. (2023). A mass correlation based deep learning approach using deep convolutional neural network to classify the brain tumor. *Biomed. Signal Process. Control* 81:104395. doi: 10.1016/j.bspc.2022.104395

Saurav, S., Sharma, A., Saini, R., and Singh, S. (2023). An attention-guided convolutional neural network for automated classification of brain tumor from MRI. *Neural Comput. Applic.* 35, 2541–2560. doi: 10.1007/s00521-022-07742-z

Sekhar, A., Biswas, S., Hazra, R., Sunaniya, A. K., Mukherjee, A., and Yang, L. (2022). Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system. *IEEE J. Biomed. Health Inform.* 26, 983–991. doi: 10.1109/JBHI.2021.3100758

Shah, H. A., Saeed, F., Yun, S., Park, J.-H., Paul, A., and Kang, J.-M. (2022). A robust approach for brain tumor detection in magnetic resonance images using Finetuned EfficientNet. *IEEE Access* 10, 65426–65438. doi: 10.1109/ACCESS.2022.3184113

Sharif, M. I., Li, J. P., Khan, M. A., and Saleem, M. A. (2020). Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images. *Pattern Recognit. Lett.* 129, 181–189. doi: 10.1016/j.patrec.2019.11.019

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., et al. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput. Med. Imaging Graph.* 75, 34–46. doi: 10.1016/j.compmedimag.2019.05.001

Tan, M., and Le, Q. V. (2019). EfficientNet: rethinking model scaling for convolutional Neural Networks. *International Conference on Machine Learning.* 97. Long Beach, CA.

Tandel, G. S., Tiwari, A., and Kakde, O. G. (2021). Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. *Comput. Biol. Med.* 135:104564. doi: 10.1016/j.compbiomed.2021.104564

Yang, J. (2022). Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput. Struct. Biotechnol. J.* 20, 333–342. doi: 10.1016/j.csbj.2021.12.028

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

Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., et al. (2022). Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE Trans. Med. Imaging* 41, 1242–1254. doi: 10.1109/TMI.2021.3136682

Yeung, M., Sala, E., Schönlieb, C.-B., and Rundo, L. (2022). Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* 95:102026. doi: 10.1016/j.compmedimag.2021.102026

Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., and Zhang, Y.-D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing* 489, 230–254. doi: 10.1016/j.neucom.2021.08.159

Zhang, J., Tan, X., Chen, W., Du, G., Fu, Q., Zhang, H., et al. (2023). EFF_D_SVM: a robust multi-type brain tumor classification system. *Front. Neurosci.* 17:1269100. doi: 10.3389/fnins.2023.1269100

Zolfaghari, B., Mirsadeghi, L., Bibak, K., and Kavousi, K. (2023). Supplementary materials for: cancer prognosis and diagnosis methods based on ensemble learning. *ACM Comput. Surv.* 55, 1–34. doi: 10.1145/3580218

Zulfiqar, F., Ijaz Bajwa, U., and Mehmood, Y. (2023). Multi-class classification of brain tumor types from MR images using EfficientNets. *Biomed. Signal Process. Control* 84:104777. doi: 10.1016/j.bspc.2023.104777

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 08 March 2024
DOI 10.3389/fnins.2024.1303741

OPEN ACCESS

EDITED BY
Hao Zhang,
Central South University, China

REVIEWED BY
Xiaoke Hao,
Hebei University of Technology, China
Yangding Li,
Hunan Normal University, China

*CORRESPONDENCE
Bin Wang
✉ wangbin01@tyut.edu.cn
Tianyi Yan
✉ yantianyi@bit.edu.cn

RECEIVED 28 September 2023
ACCEPTED 07 February 2024
PUBLISHED 08 March 2024

CITATION
Huang Y, Li Y, Yuan Y, Zhang X, Yan W, Li T, Niu Y, Xu M, Yan T, Li X, Li D, Xiang J, Wang B and Yan T (2024) Beta-informativeness-diffusion multilayer graph embedding for brain network analysis. *Front. Neurosci.* 18:1303741. doi: 10.3389/fnins.2024.1303741

Beta-informativeness-diffusion multilayer graph embedding for brain network analysis

Yin Huang[1], Ying Li[1], Yuting Yuan[1], Xingyu Zhang[1], Wenjie Yan[1], Ting Li[2], Yan Niu[1], Mengzhou Xu[3], Ting Yan[4], Xiaowen Li[5], Dandan Li[1], Jie Xiang[1], Bin Wang[1]*and Tianyi Yan[3]*

[1]College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Taiyuan, China, [2]School of Life Science, Beijing Institute of Technology, Beijing, China, [3]School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China, [4]Translational Medicine Research Center, Shanxi Medical University, Taiyuan, China, [5]Computer Information Engineering Institute, Shanxi Technology and Business College, Taiyuan, China

Brain network analysis provides essential insights into the diagnosis of brain disease. Integrating multiple neuroimaging modalities has been demonstrated to be more effective than using a single modality for brain network analysis. However, a majority of existing brain network analysis methods based on multiple modalities often overlook both complementary information and unique characteristics from various modalities. To tackle this issue, we propose the Beta-Informativeness-Diffusion Multilayer Graph Embedding (BID-MGE) method. The proposed method seamlessly integrates structural connectivity (SC) and functional connectivity (FC) to learn more comprehensive information for diagnosing neuropsychiatric disorders. Specifically, a novel beta distribution mapping function (beta mapping) is utilized to increase vital information and weaken insignificant connections. The refined information helps the diffusion process concentrate on crucial brain regions to capture more discriminative features. To maximize the preservation of the unique characteristics of each modality, we design an optimal scale multilayer brain network, the inter-layer connections of which depend on node informativeness. Then, a multilayer informativeness diffusion is proposed to capture complementary information and unique characteristics from various modalities and generate node representations by incorporating the features of each node with those of their connected nodes. Finally, the node representations are reconfigured using principal component analysis (PCA), and cosine distances are calculated with reference to multiple templates for statistical analysis and classification. We implement the proposed method for brain network analysis of neuropsychiatric disorders. The results indicate that our method effectively identifies crucial brain regions associated with diseases, providing valuable insights into the pathology of the disease, and surpasses other advanced methods in classification performance.

KEYWORDS

brain network, beta-informativeness-diffusion, graph embedding, schizophrenia, bipolar disorder

1 Introduction

The human brain represents an intricate network comprising interconnected regions in both structure and function (Cao et al., 2020). Anomalous wiring within the brain network may result in brain dysfunction (Van Den Heuvel et al., 2013). Neuropsychiatric disorders encompass a range of neurological diseases affecting the brain, characterized by cognitive dysfunction as a central symptom. Previous research has suggested that many neuropsychiatric disorders (such as schizophrenia, bipolar disorder, and Alzheimer's disease) are caused by damage to the brain's internal nervous system (Liu et al., 2018; Lian et al., 2020), leading to dysconnectivity between distinct brain regions (Yan et al., 2018; Wang et al., 2022). In medical physiology, neuroimaging techniques have rapidly evolved to provide critical insights into the diagnosis of neuropsychiatric disorders (Dubois and Adolphs, 2016; Cui et al., 2022; Liu et al., 2022a).

Brain networks derived from various neuroimaging modalities have been extensively used to analyze neuropsychiatric disorders. According to graph theory, a brain network comprises nodes and edges, with nodes denoting distinct brain regions, and edges signifying either physical connections or pairwise similarity. Diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) are two frequently employed neuroimaging techniques. DTI reveals the physical connections between distinct brain regions, serving as a structural connectivity (SC) to build the structural brain network. fMRI captures the temporal correlation between blood-oxygen-level-dependent (BOLD) signals across various brain regions, which is normally treated as functional connectivity (FC) to establish the functional brain network (Osipowicz et al., 2016). Some methods relying on structural or functional brain networks have been effectively employed to identify potential biomarkers in the diagnosis of neuropsychiatric disorders. For example, Zhang et al. (2018) proposed ordinal patterns (e.g., subgraphs and motifs) containing weighted edge sequences for the connectivity analysis of brain networks. Huang et al. (2020a) employed SGNS to extract embedding features of structured brain networks and aligned these node representations through orthogonal transformations, then computed feature distances for brain disease diagnosis. Graph embedding methods, such as node2vec, are also widely used to extract node-level feature vectors of brain networks for brain disease analysis, which capture subtle structural changes in the brain network and contain richer information (Rahimiasl et al., 2021; Ramesh Kumar Lama and Kwon, 2021). These approaches are typically focused on either SC or FC, thereby only considering node interactions within a single modality. In practice, different modalities provide possibilities to analyze brain diseases from multiple perspectives (Dai et al., 2019; Zhang et al., 2021); integrating multiple modalities has been shown to be more effective than using a single modality in brain network analysis (Yan et al., 2020).

In recent years, A variety of approaches have emerged to combine SC and FC to perform brain network analysis (Huang et al., 2020b; Song et al., 2023). These methods typically can be divided into two categories. The first category involves a data fusion strategy, considering SC and FC as multi-modal data and combining their features by employing established machine learning techniques. For example, Gao et al. (2020) proposed a multi-kernel SVM to integrate multi-modal MRI by exploiting the subspace similarity of the decomposition components in each modality. Lei et al. (2020) combined low-order self-calibrated functional and structural brain networks to perform joint multitask learning for the early diagnosis of Alzheimer's disease. Mill et al. (2021) used univariate and multivariate methods to fuse structural MRI and functional connectivity features for diagnosing patients with prescription opioid use disorder. These methods view SC and FC as separate modalities to extract latent node representations, neglecting the potential complementary information that exists between the modalities. The other category refers to a guiding strategy, which involves utilizing one modality to aid another in extracting features or leveraging multi-modal data to construct a unified brain network. For instance, Huang et al. (2020b) proposed an attention-diffusion-bilinear neural network for brain network analysis, in which node interactions in structural brain networks are used to further guide diffusion processes in functional brain networks to generate new node representations. Zhu et al. (2021) proposed a unified brain network construction framework, using a low-rank representation to build correlation models of all brain regions in functional data, simultaneously embedding local manifolds with structural data into the model to fuse multi-modal features. Liu et al. (2022b) utilized machine learning to extract important features from a structural graph network and exploited these features to adjust the corresponding edge weights in a functional graph network, which serves as an input to a multilayer GCN to achieve disease classification. However, these methods lead to each subject ultimately having only one brain network, thereby losing the unique characteristics of each modality's brain network (Zhu et al., 2022). It has been proved that some internal properties within the brain network play a pivotal role in the analysis of brain networks (Wang et al., 2017; Yan et al., 2019). However, these multi-modal brain network analysis methods cannot adequately balance both the utilization of complementary information and the preservation of unique characteristics from various modalities.

To tackle this challenge, we propose a Beta-Informativeness-Diffusion Multilayer Graph Embedding (BID-MGE) method to learn holistic information for brain network analysis. Specifically, to maximize the preservation of each modality's unique characteristics, we design a multilayer brain network, the functional layer of which is built through the guidance of its structural layer, and inter-layer connections are defined by node informativeness. Then, the multilayer informativeness diffusion first selects a more informative layer depending on node informativeness to exploit complementary information between modalities through wider node interactions. Within each layer, traversing nodes based on SC or FC capture the unique characteristics of each modality. Through propagating node features from a selected node to all its linked nodes in a diffusion manner, more comprehensive information is therefore considered in feature learning. In addition, beta mapping further assists the diffusion process to extract more discriminative features by refining crucial connectivity. Finally, to compare and analyze differences between different groups, we reconfigure node representations by PCA and then compute cosine distances with reference to multiple templates for statistical analysis and classifications. The statistical analysis is conducted on the node distances. For the classifications, the network distance serves as input into the Support Vector Machine (SVM) for identifying the label of each network.

The principal contributions of this study are as follows:

1. Beta mapping to refine the connectivity information of each modality. The refined information helps direct the diffusion process towards important brain region to capture discriminative features.

2. We proposed a novel framework for constructing a multilayer brain network, in which the inter-layer connections are based on node informativeness, and the network scale is optimized by the structural layer.
3. The multilayer informativeness diffusion learns complementary information and unique characteristics from various modalities. It is also an unsupervised embedding technique that only needs low time and space complexity and has no sample size limitations.
4. We validated the efficacy of our method on actual neuropsychiatric disorder datasets through two group-level analyses.

2 Proposed method

The entire processes of our method are depicted in Figure 1, comprising three primary components: data preprocessing, node representation learning, statistical analysis, and disease classifications. We describe each component of the BID-MGE method in detail below.

2.1 Data preprocessing

Throughout the experiments, we utilized two types of data: MRI images and clinical scores. The MRI images encompass both DTI and resting-state fMRI (rs-fMRI), which require different preprocessing. The specific steps are described below.

DTI is preprocessed using PANDA toolboxes (Cui et al., 2013). First, the initial images go through head motion correction and eddy current distortion. Second, the fractional anisotropy (FA) is computed for every voxel, followed by registering the FA images in the original space to the T1-weighted images using an affine transformation. Third, we employ the Anatomical Automatic Labeling (AAL) atlas to delineate and mark the regions of interest (ROI) within the DTI data, and then reconstruct WM pathways (fibers or tracts) via a deterministic white matter tractography method (Mori and van Zijl, 2002). Finally, we acquire the count of fibers that connected any two brain regions from DTI data.

The rs-fMRI data is preprocessed using DPABI (Yan et al., 2016). Before starting the preprocessing, we discarded the initial 10 time points due to the incipient signal fluctuation. Subsequently, head motion and slice timing corrections are applied to each subject. Then, the T1 image is aligned with the central rs-fMRI image with corrected head movement. The functional images are resampled to 3-mm isotropic voxels and then subjected to spatial smoothing using a 4-mm full-width half-maximum (FWHM) Gaussian kernel. Several interfering signals, such as head motion signals, and cerebrospinal fluid are regressed from the image. Low-frequency drift and high-frequency noise are removed by linear detrending and bandpass filtering (0.01–0.25 Hz). Ultimately, the average time series are extracted from brain regions parcellated according to the AAL atlas.

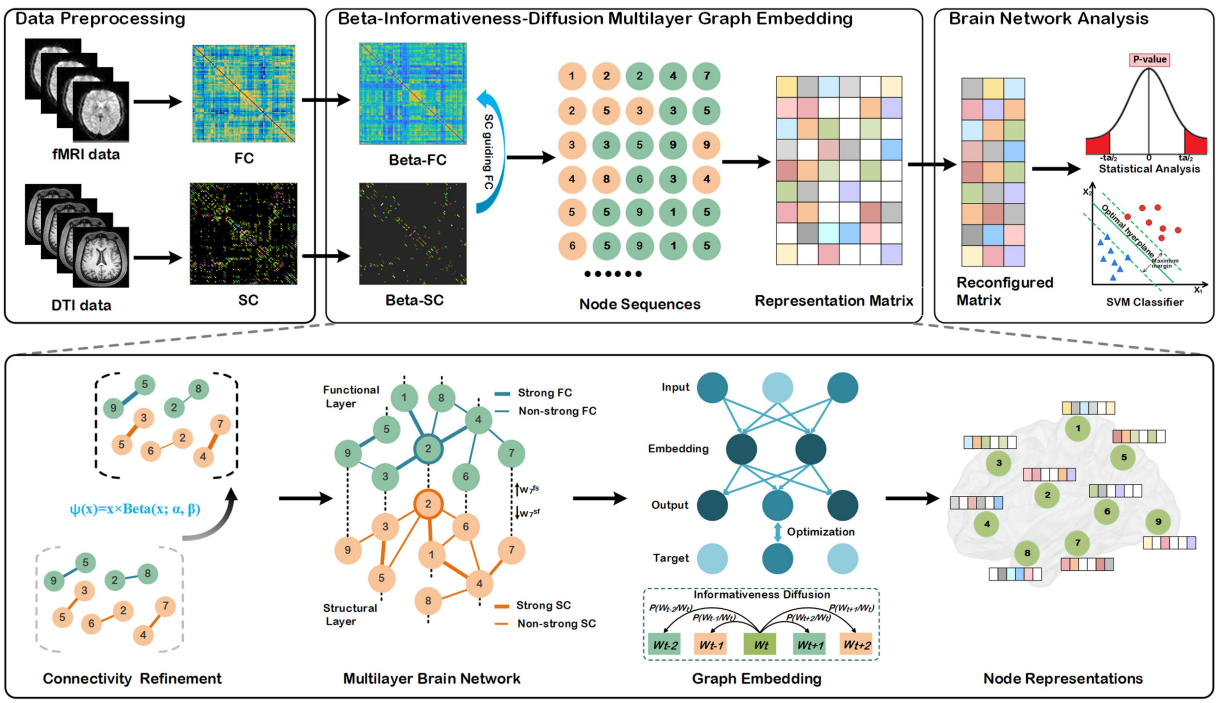


FIGURE 1
Architecture of the proposed BID-MGE method for brain network analysis. There are three modules in our method: a data preprocessing module, beta-informativeness-diffusion multilayer graph embedding module, and brain network analysis module. The data preprocessing module transforms the DTI and fMRI data into a structural and functional connectivity matrix. The Beta-Informativeness-Diffusion multilayer graph embedding module integrates SC and FC for generating node representations with comprehensive information of the brain network. The brain network analysis module consists of a statistical analysis and classifications.

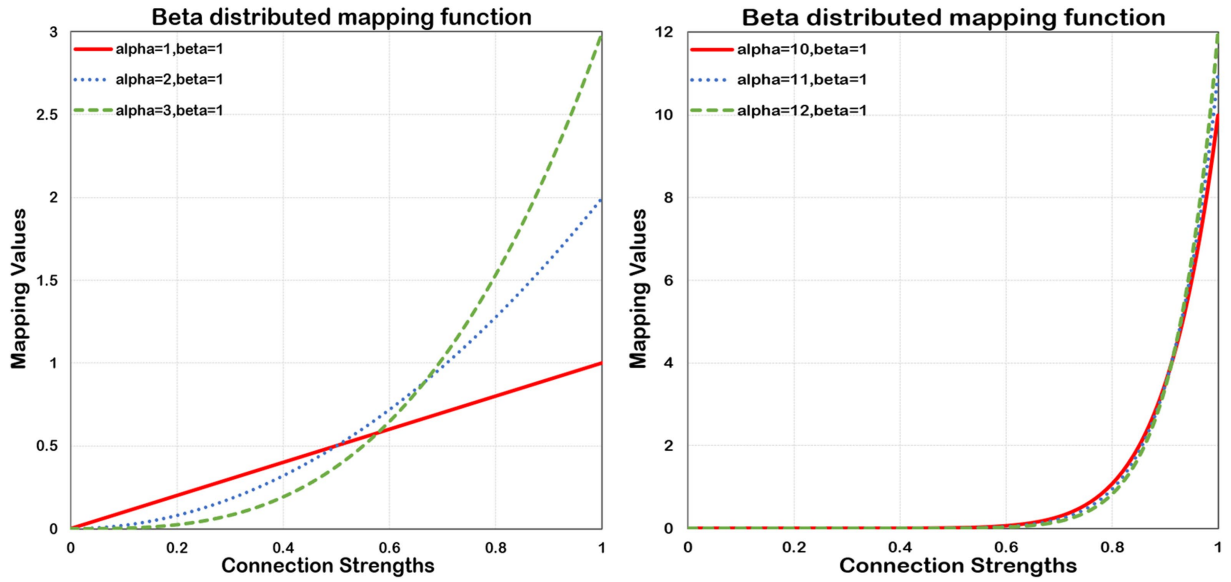


FIGURE 2
Beta distributed mapping function. Beta mapping with different values of α and a fixed $\beta = 1$. As α increases, the squeezing and expanding properties become stronger.

2.2 Structural and functional brain network construction

Graphs provide a useful abstraction for representing many complex relationships in reality. In general, a weighted graph is denoted as $G = (V, E, W)$, where $V = \{v_1, v_2, \ldots, v_n\}$ defines the set of the nodes, $E = \{e_{ij}\}_{(i,j=1,2,\ldots,n)}$ denotes the set of the edges, and W represents a connectivity matrix reflecting the strength of connectivity between any two nodes within the graph. Likewise, the human brain network can be abstractly denoted as such a graph. The graph's nodes symbolize brain regions, while the edges represent the connections linking these regions. In our experiment, we adopt triples, $G_s = (V_s, E_s, W_s)$ and $G_f = (V_f, E_f, W_f)$, to represent the structural and functional brain networks, respectively. Here, $V = V_s = V_f$. Among them, $v^s \in V_s$ denotes a brain region in the structural brain network. W_s signifies a structural connectivity matrix, with the weight $w_{ij}^s \in W_s$ for each $e_{ij}^s \in E_s$ calculated by the count of fibers divided by the sum of two interconnected surface areas of ROIs. $v^f \in V_f$ represents a brain region within the functional brain network, W_f refers to a functional connectivity matrix, with the weight $w_{ij}^f \in W_f$ for each $e_{ij}^f \in E_f$ determined by computing the Pearson correlation among the average time series of the brain regions. Notably, as the negative correlation coefficients have no clear biological explanations, it is common practice to set these negative values to zero (Murphy et al., 2009; Cao et al., 2020). Additionally, the self-correlations coefficients are also set to zero (Rubinov and Sporns, 2010).

2.3 Connectivity information refinement

To extract more discriminative features, the following mapping function (beta mapping) as shown in Eq. 1, has been proposed to refine the connectivity information of the brain.

$$\psi(x) = x \times Beta(x; \alpha, \beta). \quad (1)$$

Where Beta is a continuous probability distribution function on the range [0,1]. The parameters α and β, both more than zero, determine the shape of its distribution. The shape can be concave, convex, monotonically increasing, monotonically decreasing, and curved or straight. However, the probability density function (PDF) of Beta is monotonically ascending only in the case of $\alpha \geq 1$ and $\beta \leq 1$, which maps smaller values to nearly zero numbers, and larger values to more significant numbers, thereby allowing for its compression and expansion properties. The Beta's compression and expansion properties enable $\psi(x)$ to scale the input values. Considering two typical values of connection strength, 0.5 and 0.9, the value 0.5 normally happens between nodes. In contrast, the value 0.9 rarely occurs, and it also implies a strong connection between connected nodes. Without using beta mapping, the latter value is merely 80% stronger than the former. However, by employing beta mapping with α =2 with β=2, the latter transforms to 1.62, signifying a 224% increase in strength. In Figure 2, we present the beta mapping ψ(x) for different values of α with β constant 1. The larger α means more significant compression and expansion properties. The maximum value of $\psi(x)$ is equal to α when $x = 1$. $\psi(x)$ makes it possible to refine the essential connections and eliminate negligible information. Eventually, the connectivity matrices W_s and W_f are converted to BW_s and BW_f, respectively.

2.4 Structure-guided multilayer brain network construction

A multilayer brain network comprises two layers: a structural layer and a functional layer that correspond to the structural and functional brain networks, respectively. For the structural layer, its

edges are determined from the structural connectivity matrix BW_s. Therefore, this layer is inherently a sparse network, and the number of edges is also fixed. For the functional layer, the edges are derived from the functional connectivity matrix BW_f, which is almost fully connected, and some of the connections are negligible, which also increases the computation time of the multilayer brain network, so only some of the important connections in BW_f will be used to build the functional layer instead of all of them. In this study, we adopt the structural layer to guide the selection of edges for building the functional layer, determining its network scale so that it is comparable in scale to the structural layer. Specifically, we first calculate the average edge number of all nodes within the structural layer, denoted by avg_s. If the given network is undirected, $avg_s = 2\times|E_s|/n$, otherwise, $avg_s = |E_s|/n$. Then, for each node v^f, the top $\theta \times avg_s$ edges are selected to construct the functional layer in terms of the connection values in BW_f, where θ is the network scale parameter, $\theta \in \mathbb{R}_+$. Finally, inter-layer edges (directed and weighted) are used to connect the corresponding nodes in the structural and functional layers to constitute a multilayer brain network. The weights of these edges depend on node informativeness. The notion of node informativeness will be explained later.

2.5 Multilayer informativeness diffusion

We propose a graph embedding technique based on multilayer informativeness diffusion, which learns node representations by intelligently traversing the nodes between structural and functional layers in a diffusion manner. Whenever the diffusion process reaches a node, our goal is to select a more informative layer by assessing the informativeness of the current node in its corresponding layer.

A node that has strong connections to many nodes is less similar to its neighbors, while a node strongly connected to only a few nodes is more similar to its neighbors. The latter node also means more informativeness (Ribeiro et al., 2017). For the diffusion process, it is crucial to traverse nodes that have more informativeness. In this study, we suppose that a strong connection refers to an edge with a weight exceeding the average weight of its network layer. Consequently, we define T_i^s as the collection of neighbors' non-strong connection with node v_i^s in the structural layer, denoted as Eq. 2.

$$\mathrm{T}_i^s = \{\upsilon_j^s \in V_s \mid w_{ij}^s \le \frac{1}{|E_s|}\sum_{e'\in E_s} w_{e'}^s\}. \tag{2}$$

Each node in T_i^s has an edge connected to v_i^s with a weight not exceeding the mean weight of the structural layer. $|\mathrm{T}_i^s|$ denotes the count of nodes that belong to T_i^s. Similarly, T_i^f for the functional layer is defined as Eq. 3:

$$\mathrm{T}_i^f = \{\upsilon_j^f \in V_f \mid w_{ij}^f \le \frac{1}{|E_f|}\sum_{e'\in E_f} w_{e'}^f\}. \tag{3}$$

Given the sets T_i^s and T_i^f, the informativeness of nodes v_i^s and v_i^f is defined as Eq. 4.

$$\mathrm{I}_i^s = \ln\left(e + |\mathrm{T}_i^s|\right), \mathrm{I}_i^f = \ln\left(e + |\mathrm{T}_i^f|\right). \tag{4}$$

Now, let us consider the inter-layer directed weighted edges. The weight is set as I_i^s from the functional layer to the structural layer, and vice versa as I_i^f. The diffusion process starts with selecting the structural or functional layer according to the weights of inter-layer directed edges. If the value of I_i^s is high, the diffusion process will step into the structural layer. Otherwise, the functional layer will be chosen. We aim to step into a layer where the node possesses greater informativeness.

Subsequently, we formulate the probabilities of inter-layer and intra-layer diffusion for multilayer informativeness diffusion. Given a node v_i, the probability of inter-layer diffusion is defined as Eq. 5:

$$P\left(v_i^s \mid v_i^f\right) = \frac{\mathrm{I}_i^s}{\mathrm{I}_i^s + \mathrm{I}_i^f}, P\left(v_i^f \mid v_i^s\right) = \frac{\mathrm{I}_i^f}{\mathrm{I}_i^s + \mathrm{I}_i^f}. \tag{5}$$

Where the likelihood of moving to a structural layer is represented as $P\left(v_i^s|v_i^f\right)$, and vice versa for $P\left(v_i^f|v_i^s\right)$. The probability of intra-layer diffusion delineates the likelihood of transitioning from the present vertex to the subsequent vertex within the layer. Suppose the diffusion process visited node $v_{k-1}^{l_i}$ at time t−1 and propagated to node $v_k^{l_j}$ at current time t, where l_i and l_j denote the corresponding layers $l_i, l_j \in \{s, f\}$. If the diffusion process steps into another layer at time t (i.e., $l_i \neq l_j$), $e_{(k-1,k)}^{l_j} \notin E_{l_j}$, otherwise (i.e., $l_i = l_j$), $e_{(k-1,k)}^{l_j} \in E_{l_j}$. For $e_{(k-1,k)}^{l_j} \notin E_{l_j}$, The selection probability of the next node depends entirely on the weight of the edges connecting to $v_k^{l_j}$ in layer l_j. In other cases, the intra-layer diffusion probabilities follow the unnormalized transition probabilities in node2vec (Grover and Leskovec, 2016). Hence, we define the probability of intra-layer diffusion (i.e., the probability of selecting the next node $v_{k+1}^{l_j}$ in layer l_j at time $t+1$) as Eq. 6:

$$P\left(v_{k+1}^{l_j} \mid v_k^{l_j}, v_{k-1}^{l_j}\right) = \begin{cases} w_{(k,k+1)}^{l_j}, if\ e_{(k-1,k)}^{l_j} \notin E_{l_j} \\ \frac{1}{p} w_{(k,k+1)}^{l_j}, if\ e_{(k-1,k)}^{l_j} \in E_{l_j} \wedge d_{(k-1,k+1)}^{l_j} = 0 \\ w_{(k,k+1)}^{l_j}, if\ e_{(k-1,k)}^{l_j} \in E_{l_j} \wedge d_{(k-1,k+1)}^{l_j} = 1 \\ \frac{1}{q} w_{(k,k+1)}^{l_j}, if\ e_{(k-1,k)}^{l_j} \in E_{l_j} \wedge d_{(k-1,k+1)}^{l_j} = 2 \end{cases}. \tag{6}$$

Here, $d_{(k-1,k+1)}^{l_j}$ represents the unweighted path length between two nodes, $v_{k-1}^{l_j}$ and $v_{k+1}^{l_j}$. For parameters p and q, both are greater than 0. Parameter p determines the probability of traversing the recently visited node $v_{k-1}^{l_j}$, and parameter q controls the search to proceed in either a BFS or DFS manner. If $q > 1$, the diffusion process prefers nodes closer to node $v_{k-1}^{l_j}$. If $q < 1$, the diffusion process tends to visit nodes farther away from it.

The multilayer informativeness diffusion is performed as follows: at a given time point of the diffusion process, a node is on either the structural or functional layer. The diffusion process first evaluates the informativeness of the node in each layer to determine which layer to enter next, then traverses the node

according to the transition probabilities. The selected node is added to node sequences after discarding its layer information, which ensures each node corresponds to only one node representation. We repeatedly perform the above steps λ times, where λ signifies the truncated walk length starting from a node.

After generating the necessary number of node sequences for every node, learning node representation is achieved using the following objective function (Eq. 7), optimizing the log-probability of a node observing its context within the node sequence, given by F:

$$\max_{F} \sum_{v \in V} \log P(N(v) \mid F(v)) = \max_{F} \sum_{v \in V} \sum_{u \in N(v)} \log P(u \mid F(v)). \quad (7)$$

Let $F : \mathrm{V} \to \mathbb{R}^{d}$ be a learnable projection function mapping nodes to vector representations. Here, parameter d fixes the dimensions of the node representation. Correspondingly, F specifies a parameter matrix of size $\mathrm{n} \times \mathrm{d}$, representing the node representation. $N(v)$ is the neighborhood of node v in a diffusion process. To render the optimization problem tractable, we also apply two criterion assumptions: conditional independence and feature space symmetry (Grover and Leskovec, 2016). The above optimization function is simplified (Eq. 8):

$$\max_{F} \sum_{v \in V} \sum_{u \in N(v)} \left(-\log Z_v + F(u).F(v)\right). \quad (8)$$

The partition function $Z_v = \sum_{v' \in V} \exp\left(F(v').F(v)\right)$ can be estimated using negative sampling. The model parameters denoting the feature F in Eq. 8 can be optimized through stochastic gradient ascent.

2.6 Node representation reconfiguration

A particular dimension within a node representation may encompass varying latent concepts across different networks. Hence, these representations have to be reconfigured sequentially to ascertain the importance of individual features (Salsabilian and Najafizadeh, 2020). To accomplish this objective, we adopt PCA, which also serves as information compression. We retain top k principal components ($k < d$) and transform the representation matrix $F^{n \times d}$ into a reconfigured representation matrix $A^{n \times k}$ in an important sequential manner $(p_1, p_2, \ldots p_k)$, where p_i represents the ith principal component as a column vector and the row j of A, A_j, denotes the j th reconfigured node representation.

2.7 Cosine distance computation

Given two vector representations, $A = (x_1, x_1, \ldots, x_t)$ and $B = (y_1, y_1, \ldots, y_t)$, the cosine distance between A and B can be calculated as Eq. 9:

$$\mathrm{Cos}\, Dist(A, B) = 1 - \cos(A, B) = \frac{A_2 B_2 - A \cdot B}{\|\mathrm{A}\|_2 \|\mathrm{B}\|_2}. \quad (9)$$

which reflects the differences between vector representations. The smaller the distance is, the more similar the vector representations are. Nevertheless, because of lacking shared reference coordinates, such pairwise distances are not directly employed in the group-level analysis (Huang et al., 2020a). To compare differences between different groups, we propose node distance and network distance, with reference to common coordinates at the node-level and network-level, respectively.

2.7.1 Node distance

After reconfiguring node representations, we calculate the node distance. This node distance becomes smaller if nodes i and j are more similar in structure or function. First, we construct the reference template $\{\tau_1, \tau_1, \cdots \tau_n\}^T$, where $\tau_i \in \mathbb{R}^k$ is the centroid node representation ($\tau_i = \frac{1}{m^c} \sum_{r=1}^{m^c} A_i^r$, where m^c is the count of subjects with the same labeling). Second, we calculate the distances between nodes in the target network and those in the template. Given a target network G_t and the reference template $\{\tau_1, \tau_1, \cdots \tau_n\}^T$, a node distance vector $\ell = \{\iota_1, \iota_2, \cdots \iota_n\}$ can be obtained, here ι_i is the node distance between nodes v_i in both networks (i.e., $\iota_i = CosDist\left(A_i^t, \tau_i\right)$). Notably, the template can be designated as the HC template $\{\tau_1^h, \tau_2^h, \cdots \tau_n^h\}^{\mathrm{T}}$, the SZ template $\{\tau_1^s, \tau_2^s, \cdots \tau_n^s\}^{\mathrm{T}}$, and the BD template $\{\tau_1^b, \tau_2^b, \cdots \tau_n^b\}^{\mathrm{T}}$. Third, we utilize the node distance vector, generated for each subject, to compose a node distance matrix, $\mathcal{L}^{m \times n} = \{\ell_1, \ell_2, \cdots \ell_m\}^T$, where $m = m^h + m^s + m^b$ and m^h, m^s, and m^b are the number of HC, SZ and BD subjects, respectively. Each column, $\mathcal{L}[i]$, can be subdivided into three parts based on the label of each network: $\mathcal{L}^{\mathrm{h}}[i], \mathcal{L}^{\mathrm{s}}[i]$, and $\mathcal{L}^{\mathrm{b}}[i]$. Using these node distances, the two-tailed t-test will be employed to recognize brain regions exhibiting structural or functional differences.

2.7.2 Network distance

Moreover, the network distance can also be computed using reconfigured representations. First, node representations, $\mathrm{A}^{n \times k}$, are concatenated to generate a network representation $\mathrm{A}'^{1 \times (n \times k)}$ for each network. To find the all-round network-level differences between groups, we construct the positive template $C^+ = \{c_1^+, c_2^+, \cdots c_{\mathrm{n} \times \mathrm{k}}^+\}$ and the negative template $C^- = \{c_1^-, c_2^-, \cdots c_{\mathrm{n} \times \mathrm{k}}^-\}$ (i.e., $C^+ = \frac{1}{m^+} \sum_{i=1}^{m^+} A_i'$, $C^- = \frac{1}{m^-} \sum_{i=1}^{m^-} A_i'$, where m^+, m^- is the respective count of positive and negative samples). According to these templates, a network distance matrix $\mathcal{H} \in \mathbb{R}^{(m^+ + m^-) \times 2}$ is proposed to depict the network distance between each network and reference templates. For instance, the network distance between the target network G_a and two templates can be computed as $\mathcal{H}_{(a,1)} = CosDist\left(A_a', C^+\right), \mathcal{H}_{(a,2)} = CosDist\left(A_a', C^-\right)$. $\mathcal{H}$ reflects the network distance between each network and the corresponding positive and negative templates, with the first and second columns of $\mathcal{H}$ representing the two kinds of distances.

2.8 Statistical analysis and classification

This study performs t-tests on each column $\mathcal{L}[i]$ to identify significantly different brain regions, considering different templates as

the references. The Bonferroni correction (Bonferroni $p<0.05$) is employed to address the issue of node-level multiple comparisons. For disease classification, the network distance matrix, $\mathcal{H}$, serves as the input for the SVM classifier to determine the corresponding labels.

3 Experiments

3.1 Dataset

The proposed method is evaluated using the Consortium for Neuropsychiatric Phenomics (CNP) database (Poldrack et al., 2016), which is hosted on OpenfMRI (www.openfmri.org). In addition, the CNP dataset also contained substantial demographic information, neuropsychological assessments, and neurocognitive task results. The study collected 147 subjects with DTI and rs-fMRI brain imaging data, including 50 healthy controls (HC), 48 SZ patients, and 49 BD patients. All participants were between 21 and 50 years of age. A two-tailed t-test was performed for age and sex, both of which were not significantly different. Table 1 presents detailed demographic information about the subjects. All brain imaging data were acquired using a Siemens Trio scanner. The parameters for obtaining DTI data were as follows: slices = 176, slice thickness = 1 mm, TR = 1,900 ms, echo TE = 2.26 ms, FOV = 250 mm, flip angle = 90°, and the acquisition matrix = 256 × 256. The parameters of collecting rs-fMRI data were as follows: slices = 34, slice thickness = 4 mm, TR = 2,000 ms, TE = 30 ms, FOV = 192 mm; flip angle = 90°, and the acquisition matrix = 64 × 64.

3.2 Node distance analysis

We first calculated node distances between each network and the reference templates (i.e., the HC template, SZ template, and BD template). These average node distances for each group (i.e., the HC group, SZ group, and BD group) are presented in Figure 3. A larger node distance means greater individual differences in that brain region. Node distances between each group and their homologous templates are consistently small, as shown in the main diagonal line of Figure 3. Some regions of the brain exhibit larger node distances between each group and their heterogeneous templates. In addition, along the main diagonal line, node distances show a similar distribution in symmetrical positions. For example, HC subjects refer to the SZ template and SZ patients refer to the HC template, as the node distance reflects the same node differences from opposite perspectives. These detailed node differences are revealed through the following statistical analysis.

After obtaining the node distance matrix $\mathcal{L}$, we performed the statistical test on each column of $\mathcal{L}$ (i.e., $\mathcal{L}^{\mathrm{h}}[i]$, $\mathcal{L}^{s}[i]$, and $\mathcal{L}^{b}[i]$). The nodes with significant differences between any two of the HC, SZ, and BD groups are presented in Figure 4. We discovered that only a few nodes are significantly different on their common heterogenous templates for two groups, as shown in the sub-diagonal line in Figure 4. Most of the nodes with significant differences are concentrated on any homologous template for two groups. As shown in Figure 4A, nodes with differences between SZ and HC groups are concentrated in the thalamus, gyrus rectus, precuneus, posterior cingulate gyrus, middle frontal gyrus orbital and motor area. From Figure 4B, these nodes exhibiting differences between BD and HC groups primarily localize in the frontal lobe, cuneus, lingual gyrus, rolandic operculum, and hippocampus. Figure 4C shows nodes with differences between the SZ and BD groups are mainly the posterior cingulate gyrus, parahippocampal gyrus, precuneus, and hippocampus. Additionally, we observed that brain regions with significant differences in the homologous templates related to both groups are not completely consistent. For example, the superior parietal gyrus and postcentral gyrus only show differences on the HC template, whereas the amygdala and parahippocampal gyrus orbital only present differences on the SZ template. This might be attributed to the following factors: (1) The diverse causes of different neuropsychiatric disorders and (2) the inherent large distances between templates.

TABLE 1 The detailed demographic information of participants used in this study.

Name	Number	Age (mean ± std)	Gender (female / male)
Healthy controls (HC)	50	32.9 ± 8.2	20 / 30
Schizophrenia (SZ)	48	35.8 ± 8.7	13 / 35
Bipolar disorder (BD)	49	35.3 ± 8.9	21 / 28

3.3 Network distance visualization

To visualize the network distance, we mapped the distance matrix $\mathcal{H}$ onto a two-dimensional plane, where the first and second columns of $\mathcal{H}$ are assigned to the horizontal and vertical axes, respectively. To facilitate comparison, we also visualized the network distance for structural and functional brain networks, the node representations of which are extracted by node2vec, and the parameter settings are the same as our method. The merit of network distance is estimated by observing how clustered the points belonging to the same class are. Figure 5 visualizes the 2D scatter plots of these distance matrices in three classification combinations. The distance matrix generated by building a multilayer brain network with our approach outperforms using single-modal brain networks. Consequently, based on this distance matrix $\mathcal{H}$, distinct groups can be easily distinguished by employing some machine learning methods (e.g., SVM).

3.4 Performance evaluation

For the evaluation of classification performance, we employed classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic (ROC) curve (AUC). These metrics are defined as Eqs. 10-12:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{10}$$

$$SEN = \frac{TP}{TP + FN} \tag{11}$$

$$SPE = \frac{TN}{TN + FP}. \tag{12}$$

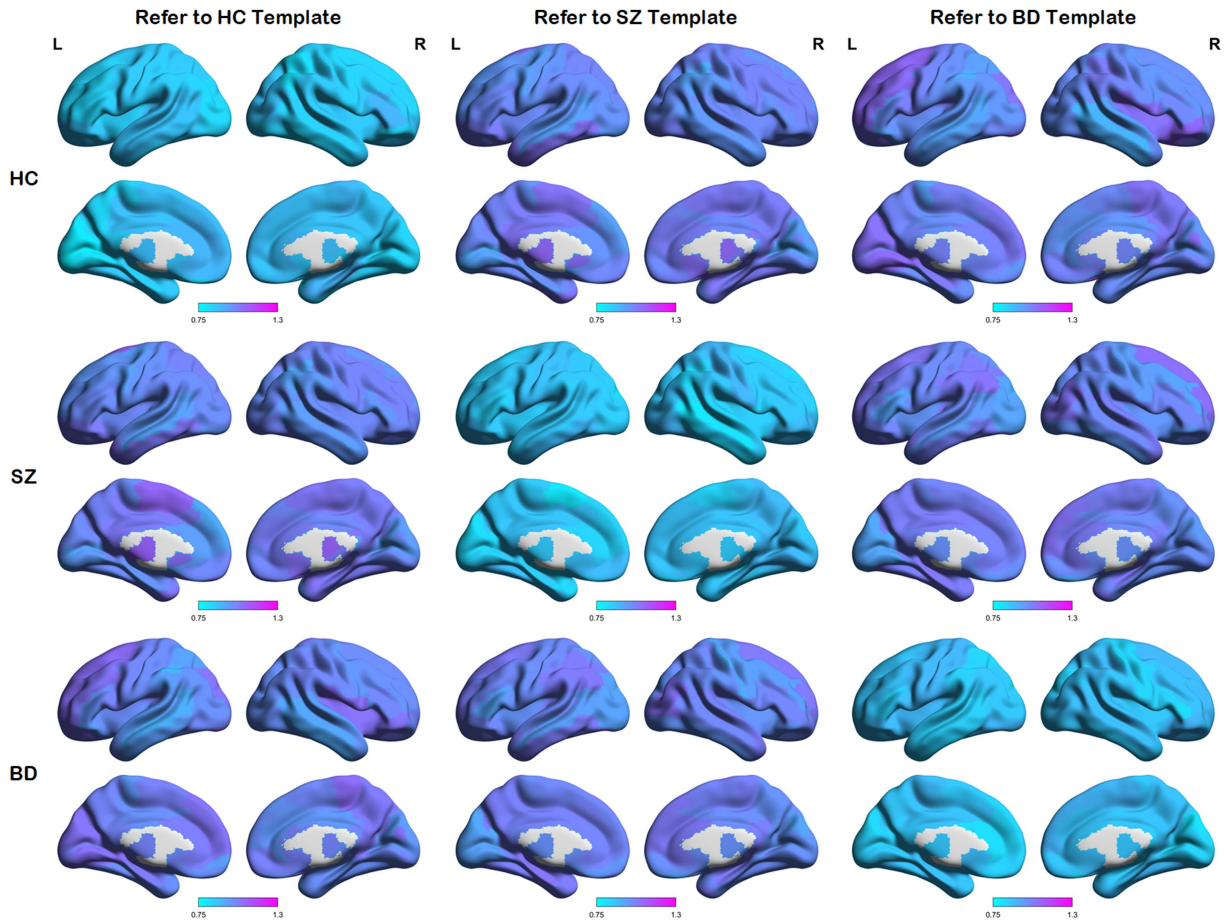


FIGURE 3
Maps of average node distances. Average node distances between each group and three templates (i.e., the HC template, SZ template, and BD template).

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

3.5 Classification performance

To evaluate the efficacy of our method in distinguishing patients from healthy controls (i.e., SZ vs. HC and BD vs. HC), we conducted a comparison with several baseline methods. The baseline models include state-of-the-art brain network analysis methods.

SVM (Atlas-based) (Tripathi et al., 2017): uses an atlas-based segmentation method to extract multiple known disease-related regions of interest and then employs gray-matter voxel-based intensity variations and structural changes extracted with a spherical harmonic framework to learn the discriminative features.

H-FCN (Lian et al., 2020): proposes a hierarchical full convolutional network to automatically identify discriminative local plaques and regions, then jointly learns and fuses multi-scale feature representations to construct hierarchical classification models for AD diagnosis.

nSEAL (Huang et al., 2020a): defines a node-level structural embedding and alignment representation to accurately characterize the node-level structural information, and calculates distances at different scales based on the embedding representation for brain disease analysis.

DCNs (Jie et al., 2018): uses manifold regularized multi-task feature learning and multi-kernel learning to integrate both temporal and spatial variabilities of DCNs for brain disease diagnosis.

N2EN (Zhu et al., 2018): proposes a non-negative elastic-net based method to extract changes in brain functional connectivity. Then, a kernel discriminant analysis (KDA) is utilized to classify subjects with the selected discriminative brain connectivity features.

SVM (Multi-kernel) (Shao et al., 2020): uses a group-sparsity regularizer with a hypergraph-based regularization term to jointly select the common features of multiple modalities. Then, a multi-kernel SVM is utilized to integrate the features selected from different modalities for final classification.

3D-CNN (Masoudi et al., 2021): proposes a multimodal hierarchical fusion method based on attention mechanisms, selectively extracting features from MRI and PET while suppressing irrelevant information.

HebrainGNN (Shi et al., 2022): models the brain network as a heterogeneous graph with multiple types of nodes and edges. Then, a self-supervised pre-training strategy based on the heterogeneous brain network is proposed to solve the potential overfitting problem.

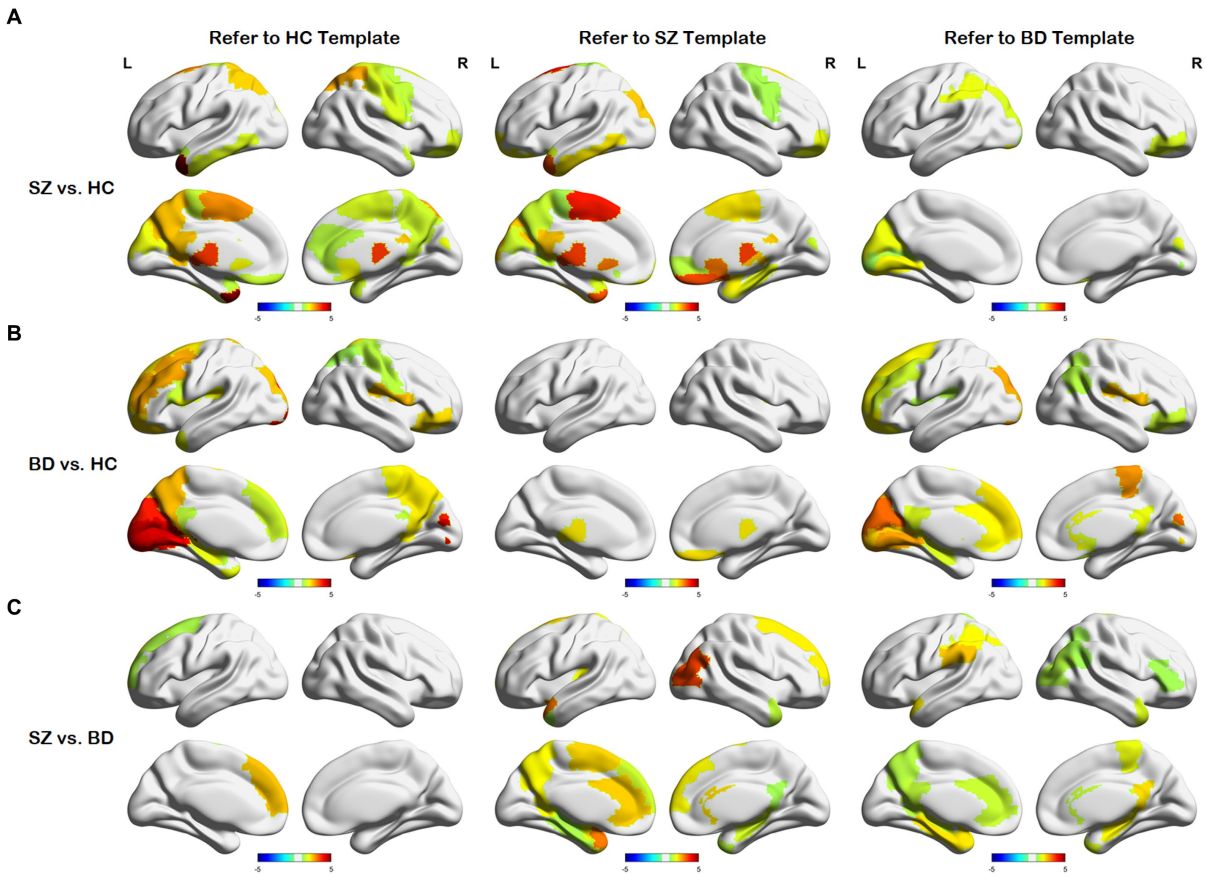


FIGURE 4
Differences in node distances between different groups with reference to the three templates. **(A)** Node differences between the SZ and HC groups. **(B)** Node differences between the BD and HC groups. **(C)** Node differences between the SZ and BD groups.

MME-GCN (Liu et al., 2022b): adopts XGBoost to extract important features from the structural brain network. These features are used to adjust the corresponding edge weights in the functional brain network. Finally, a multi-layer GCN is trained and applied to binary classification tasks.

OLFG (Chen et al., 2023): projects multiple modalities into a common latent space by orthogonal constrained projection with learning graph regularization terms to capture discriminative information, and adaptively ranks feature importance using a feature weighting matrix. Finally, the representations in the latent space are mapped to the target space for AD diagnosis.

Based on the inputs, we categorized these methods into two classes. One category only employs single-modal data as input, while the other incorporates multi-modal data. For a fair comparison, we either precisely reproduced these methods as mentioned in the article or utilized the code provided by the authors. In addition, all methods used identical training and test sets. The 10-fold cross-validation is employed to assess classification performance, repeating 10 times to derive the average performance.

The results of all methods are presented in Table 2. The accuracy values obtained from the proposed method in SZ vs. HC and BD vs. HC classification tasks achieve 99.07 and 98.80% respectively, which consistently outperforms all methods compared. Most multi-modal methods incorporating DTI and fMRI exhibit superior performance to single-modal methods using the DTI or fMRI. The accuracy of the majority of single-modal methods is below 95%, whereas multi-modal methods achieve an accuracy exceeding 95%. This verifies that combining SC and FC can offer complementary information, thereby enhancing the classification performance. Moreover, among all multi-modal methods, SVM (Multi-kernel) yields the lowest accuracy at 95.60 and 95.82%. The proposed BID-MGE method attains optimal performance on most evaluation metrics, surpassing the highest comparison method (OLFG) by approximately 2.00%. In addition, we observed that employing the embedding features directly as inputs to SVM for classification has a lower performance than some multi-modal brain network analysis methods (e.g., MME-GCN, 3D-CNN, and OLFG). This discrepancy arises from the substantial feature dimensionality resulting from concatenating all nodes, which is prone to causing a "dimensional disaster" and negatively impacting classification performance. Neural network methods, however, are better equipped to handle high-dimensional features. To further examine the sensitivity of the BID-MGE method for diverse neuropsychiatric disorders, we conducted a binary classification between SZ and BD. As shown in Figures 6A,B, our method also achieves a promising result with an ACC of 96.88, SEN of 95.94%, SPE of 97.11%, and AUC of 0.9682, which exceeds the latest neuroimaging and brain network research (Chen et al., 2017; Du et al., 2020).

The superior performance of our method compared with those multi-modal approaches may stem from the following facts. First,

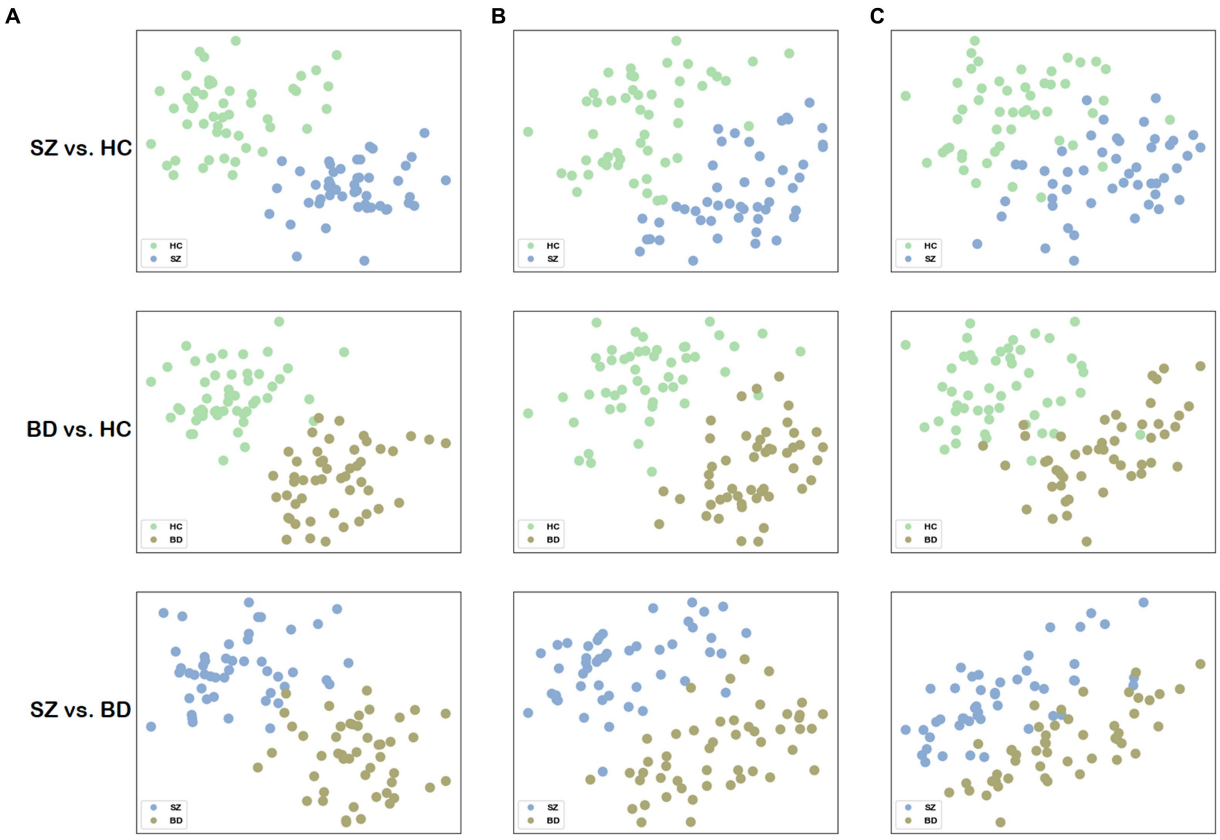


FIGURE 5
Visualization of the network distance matrix. **(A)** Scatter plots of the network distance matrix for our method. **(B)** Scatter plots of the network distance matrix for the structural brain network. **(C)** Scatter plots of the network distance matrix for the functional brain network.

TABLE 2 **Performance of all comparative methods in SZ vs. HC and BD vs. HC classification.**

Method	Modality	SZ vs. HC				BD vs. HC			
		ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
SVM (Atlas-based)	DTI	85.87	86.88	84.82	0.8571	86.18	86.94	85.40	0.8585
H-FCN	DTI	86.75	86.70	86.80	0.8675	85.98	84.54	87.44	0.8606
nSEAL	DTI	87.46	84.17	88.46	0.8632	88.86	92.14	85.42	0.8878
DCNs	fMRI	90.54	89.65	91.46	0.9083	91.64	91.24	92.04	0.9192
N2EN	fMRI	93.45	92.27	94.67	0.9392	93.76	92.60	94.94	0.9396
SVM (Multi-kernel)	DTI & fMRI	95.60	94.20	97.05	0.9594	95.82	96.52	95.11	0.9596
HebrainGNN	DTI & fMRI	95.64	93.06	97.50	0.9528	95.97	95.83	96.28	0.9605
MME-GCN	DTI & fMRI	95.93	97.98	94.25	0.9612	95.88	95.83	96.33	0.9608
3D-CNN	DTI & fMRI	96.06	96.70	95.39	0.9592	96.03	95.84	96.22	0.9631
OLFG	DTI & fMRI	96.78	96.25	98.00	0.9712	96.73	96.08	97.77	0.9693
BID-MGE (without distances)	DTI & fMRI	95.91	97.43	92.84	0.9514	94.55	90.90	98.00	0.9445
BID-MGE	DTI & fMRI	99.07	98.47	99.97	0.9923	98.80	99.92	97.65	0.9897

these multi-modal methods typically emphasize the internal relationships within brain networks, often overlooking the potential interactions between nodes across modalities. By contrast, our method can capture wider node interactions and preserve the characteristics unique to each modality through multilayer informativeness diffusion. Second, our method employs beta mapping to refine the vital connectivity of brain networks, which facilitates the extraction of more discriminative features during the diffusion process and plays a crucial role in improving classification performance. In summary, our results suggest that alterations in structural and functional connections are crucial for diagnosing neuropsychiatric disorders. Moreover, incorporating multi-modal brain networks significantly improves classification performance. It also implies that exploring wider node interactions between brain structures and

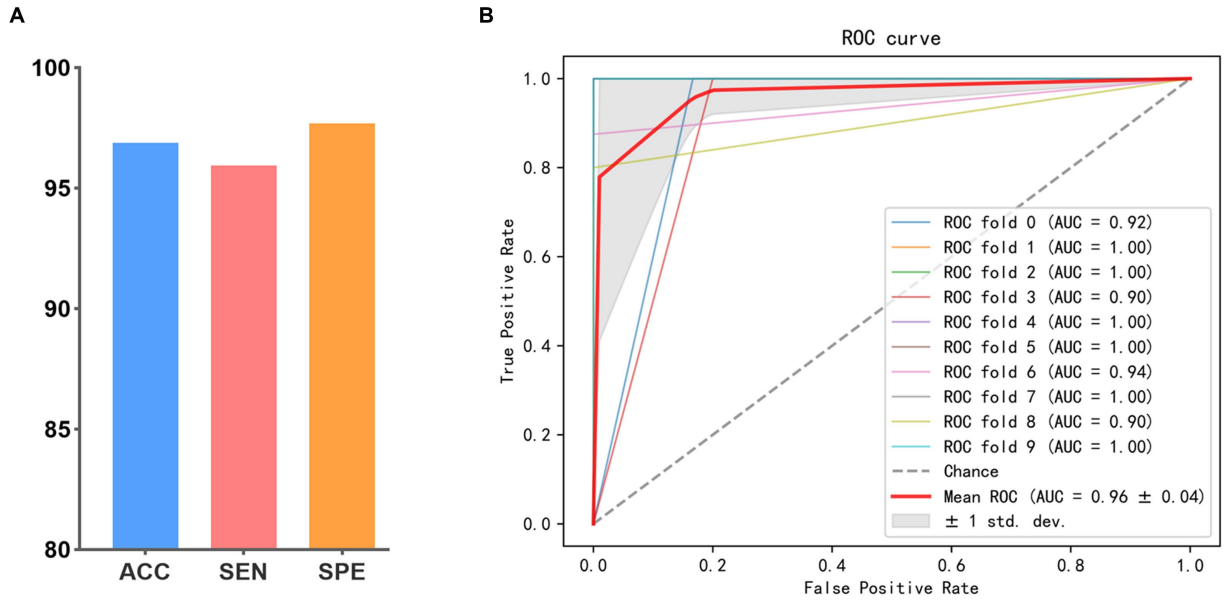


FIGURE 6
Classification performance in SZ vs. BD. **(A)** ACC, SEN, and SPE. **(B)** ROC curve.

TABLE 3 Performance of our method and previous studies on the COBRE dataset (SZ vs. HC).

Study	Modality	Subject	ACC (%)	SEN (%)	SPE (%)
Huang et al. (2020a)	fMRI	67 HC, 53 SZ	82.4	91.30	72.50
Aggarwal et al. (2017)	fMRI	50 HC, 50 SZ	89	–	–
Chyzhyk et al. (2015)	fMRI	72 HC, 74 SZ	91.2	–	–
Silva et al. (2014)	sMRI and fMRI	75 HC, 69 SZ	94	–	–
Qureshi et al. (2017)	sMRI and fMRI	72 HC, 72 SZ	99.29	100.00	98.57
Masoudi and Danishvar (2022)	DTI and sMRI	81 HC, 64 SZ	99.50	99.75	97.13
BID-MGE (without beta mapping)	DTI and fMRI	37 HC, 36 SZ	97.60	95.36	99.60
BID-MGE (without distances)	DTI and fMRI	37 HC, 36 SZ	98.57	98.33	99.65
BID-MGE	DTI and fMRI	37 HC, 36 SZ	99.71	99.67	99.75

functions and mining intrinsic characteristics of brain networks could further enhance the diagnosis of neuropsychiatric disorders.

3.6 Comparison with previous studies

In this section, we conducted a comparison with several available methods using neuroimaging data from the COBRE dataset (Mayer et al., 2013). The dataset includes structural magnetic resonance imaging (sMRI), fMRI, and DTI modalities. We collected 73 subjects for whom both DTI data and resting-state fMRI data are available, participants consist of 37 HC and 36 SZ. The ages of all subjects ranged from 20 to 65 years, and their age and gender distributions were not significantly different. Data acquisition parameters of DTI and fMRI can be found in Masoudi and Danishvar (2022). Data preprocessing is described above. The methods compared include single-modal methods and multi-modal methods. Table 3 reported the results of previous studies. Notably, the results of different methods are not directly comparable due to variations in the sample sizes, preprocessing methods, and data division. From Table 3, we observed the following points. First, multi-modal methods outperform single-modal methods due to the utilization of complementary information between modalities. Second, the performance of the BID-MGE method surpasses that of the existing method for most evaluation metrics. The enhancements attained by BID-MGE can be due to the incorporation of both complementary information and unique characteristics from various modalities. Third, beta mapping enhances the performance of our method, which further proves that beta mapping is effective in refining structural and functional connectivity information.

4 Discussion

4.1 Significance of results

The node representation proves to be a useful form for brain network analysis. Previous studies showed that neuropsychiatric disorders may result from abnormalities in some specific brain regions, thereby leading to alterations in structural and functional connectivity among brain regions (Klauser et al., 2017; Kim et al.,

TABLE 4 The ROIs with significant differences (corrected value of p <0.05).

Group	Type	ROI	Full name	Related studies
SZ vs. HC	Structure	Precentral_R	Precentral gyrus	Zhou et al. (2005)
		Rectus_R	Gyrus rectus	Masaoka et al. (2020)
		Thalamus_L	Thalamus	Shimizu et al. (2008)
	Function	Precuneus_L	Precuneus	Hoptman et al. (2010)
		Supp_Motor_Area_L area	Supplementary motor area	Mashal et al. (2014)
BD vs. HC	Structure	Cuneus_L	Cuneus	Qiu et al. (2014)
		Frontal_Sup_Medial_L	Superior frontal gyrus, medial	Repple et al. (2017)
	Function	Paracentral_Lobule_R	Paracentral lobule	Zhang et al. (2020)
		Rolandic_Oper_R	Rolandic operculum	Lin et al. (2018)
		Lingual_L	Lingual gyrus	Zhong et al. (2016)
SZ vs. BD	Structure	Cingulum_Post_R	Posterior cingulate gyrus	Koo et al. (2008)
	Function	ParaHippocampal_L	Parahippocampal gyrus	Lui et al. (2015)
		Amygdala_L	Amygdala	Mahon et al. (2012)
		Bilateral Hippocampus	Hippocampus	Hall et al. (2010)

2019). To capture these changes, the BID-MGE method generates node representations with comprehensive information to characterize brain connectivity. BID-MGE exhibits three key differences compared with existing methods: (1) our method considers both complementary information and unique features from various modalities. (2) The traditional graph embedding methods are generally used for node classification and link prediction, rather than specifically for brain network analysis. Thus, these methods fail to take into account the integration of diverse neuroimaging modalities (e.g., SC and FC). (3) Our method incorporates beta mapping to refine SC and FC, effectively steering the diffusion process toward key brain regions that cause disease. The results in Tables 2, 3 illustrate that the proposed method enhances the classification performance. Additionally, our method also discovers several crucial brain regions associated with the disease, as depicted in Figure 4. For further details, Table 4 lists several brain regions exhibiting a value of p less than 0.05 after Bonferroni correction, consistent with previous research findings. The value of p is derived from a two-tailed t-test. Specifically, several brain regions have abnormalities in SZ and BD as displayed in Figures 4A,B, such as the middle frontal gyrus, orbital, cuneus, and paracentral lobule. This may be due to shared structural and functional dysfunctions in SZ and BD (Dong et al., 2017; Xia et al., 2019).

4.2 Prediction of clinical scores

In this part, we examine the predictive ability of node distance for scale scores using connectome-based predictive modeling (CPM) (Shen et al., 2017). We concatenate the portions of node distance matrices with the same labels (e.g., $\mathcal{L}^s$, $\mathcal{L}^b$) for three node-level templates to generate a new matrix as input to CPM. The correlation coefficient for retaining the number of nodes is $p = 0.05$. The predictive power of the node distance is estimated by the Spearman correlation between the predicted and true scale scores. All statistical tests are two-tailed. We found that node distances can effectively predict scale scores in unobserved subjects with SZ (BPRS, $r = 0.5976$, $p < 0.0001$; SANS, $r = 0.6130$, $p < 0.0001$; SAPS, $r = 0.7173$, $p < 0.0001$) and BD (HAMD, $r = 0.6352$, $p < 0.0001$; YMRS, $r = 0.5618$, $p < 0.0001$); the predicted and the true scale scores present a significant correlation as illustrated in Figures 7A–E. These results further indicate that our method effectively captures structural or functional brain alterations, and the node distance can act as an essential indicator to estimate the severity of the disease.

4.3 Time and space complexity of multilayer informativeness diffusion

For the time complexity of multilayer informativeness diffusion, the sampling process of the proposed method is the same as the standard random walk. During each iteration, sampling according to the transition probability, only one node sequence is generated per node. The sampling strategy uses alias sampling, which can complete one-step diffusion in $O(1)$ time complexity (Grover and Leskovec, 2016), assuming that the count of iterations starting with every node and each truncated walk length is constant. Hence, the time complexity of completing the entire graph sampling is $O(|V|)$. For the space complexity of multilayer informativeness diffusion, the first is the space needed to store the multilayer brain network. As mentioned above, the edge number of the functional layer is θ times that of the structural layer (θ is a constant). Hence, our method needs $O((\theta+1)(|V|+|E|)) = O(|V|+|E|)$ space to store the graph in the adjacency list format. In addition, alias sampling requires an additional $O(|E|)$ space complexity. Thus, the total space complexity is $O(|V|+2|E|) = O(|V|+|E|)$. The approximate time and space complexity of our method has no increase compared with classic random walk algorithms typically used for networks with single structural data.

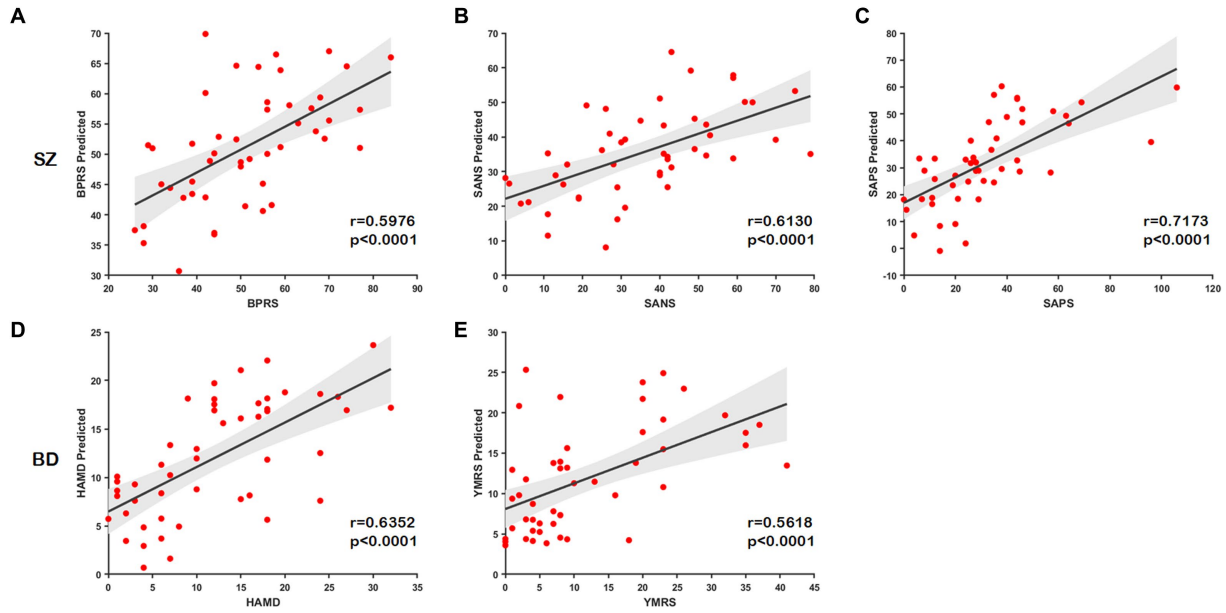


FIGURE 7
Scatter plots show correlations between the true scale scores and predictions. **(A–C)** The predicted scores of the scale of SZ. **(D,E)** The predicted scores of the scale of BD.

4.4 Parameter sensitivity

The localized diffusion tends to capture higher-order proximity more effectively. Therefore, smaller values for p and larger values for q are typically favored for graph embedding within brain networks to learn superior node representation. In our experiments, we first fixed p and q at 0.1 and 1.6, respectively. Additionally, the two other parameters, λ and k, were set to 10 and $d/2$, respectively. Then, we tested three main parameters of BID-MGE, including the functional layer network scale, distribution of beta mapping, and embedding dimension of BID-MGE. The network scale of the functional layer influences the computational time to process the multilayer brain network and the specificity of the learned node representations. The distribution of beta mapping determines its squeezing and expanding properties. The embedding dimension controls the integrity of reserving information.

4.4.1 The functional layer network scale

To minimize the computation time in processing the multilayer brain network without compromising essential connectivity information, we use the structural layer as a benchmark to select the edges that form the functional layer. Figure 8 presents classification accuracies with different network scales of the functional layer. The best performance is obtained at $\theta = 0.5$ for the three binary classifications (i.e., the functional layer is half the network scale of the structural layer). However, if the network scale of the functional layer is as small as $\theta = 0.25$, it may lead to an incomplete aggregation of the semantic neighborhood information of the nodes. Consequently, we set $\theta = 0.5$ as the optimal parameter of the network scale.

4.4.2 The distribution of beta mapping

In beta mapping, the parameters α and β are used to control the shape of the distribution, thereby altering its compression and expansion properties. We want to strengthen the connections that matter and weaken the ones that do not. In addition, for $\alpha \geq 1$ and $\beta < 1$, the value of Beta tends to move toward infinity as x is close to 1 and so does ψ(x), thereby causing irrational connections existing in the brain network. Therefore, we only consider the case in which the beta mapping monotonically grows with an upper bound (i.e., $\alpha > 1$ and $\beta = 1$). Figure 9A presents the results for α values ranging from 1 to 12 and β values of 1 in all cases. The best performance for the three binary classifications is achieved at $\alpha = 10$. When $\alpha > 10$, the classification accuracies are gradually decreased. In our study, 10 is finally chosen as the value of parameter α.

4.4.3 The embedding dimension of node representation

To explore the impact of the embedding dimension on the proposed method, we tested the BID-MGE method with different embedding dimensions and the results are depicted in Figure 9B. We noticed that optimal performance occurs at $d = 80$ for all classifications. Beyond this dimension, the accuracies decline due to the involvement of redundant or interfering features.

4.5 The effectiveness of beta mapping

The beta mapping's squeezing and expanding properties make it possible to increase critical connectivity and weaken negligible information. In Figures 10A,B, the SC and FC of a healthy subject are illustrated. These images display the changes with and without beta mapping. We observed that the number of strength connections decreased, which promotes the diffusion process to focus more on key brain regions. From Figure 10C, we can find that the classification accuracies are remarkably improved after employing beta mapping; the results indicate that beta mapping contributes to

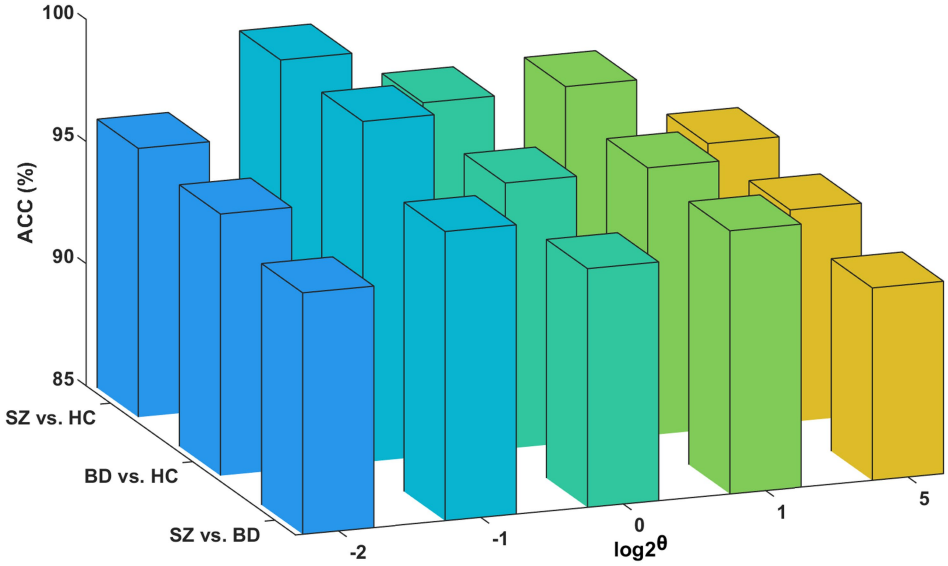


FIGURE 8
Influence of the functional layer network scale. Classification accuracies for the functional layer with different network scales.

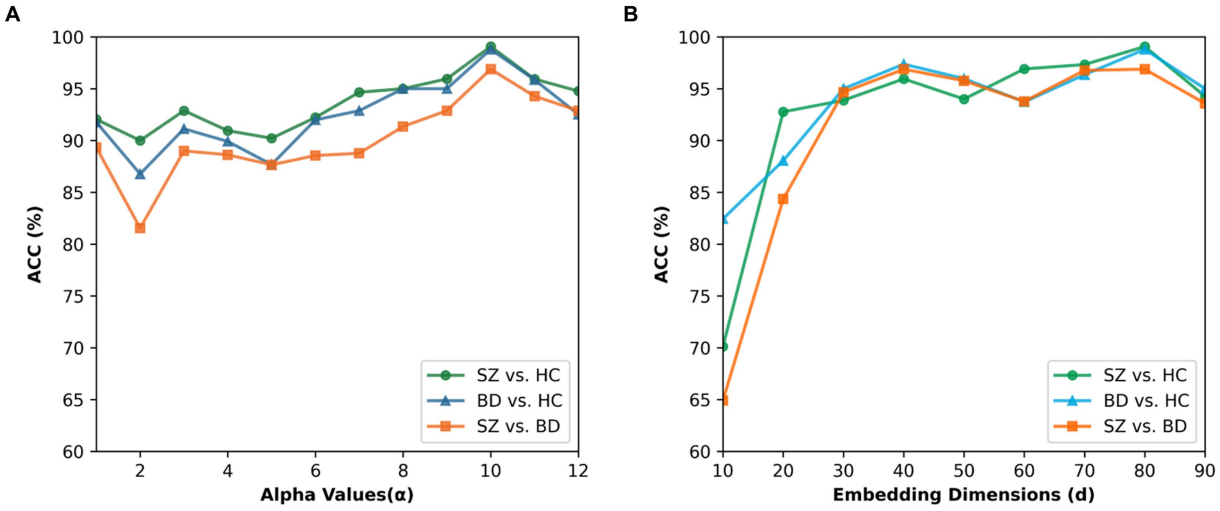


FIGURE 9
Effect of the parameter alpha and embedding dimension. **(A)** Classification accuracies for different alpha values of beta mapping. **(B)** Classification accuracies for different embedding dimensions.

the identification of diseases. Specifically, beta mapping significantly improves the accuracy of classification by structural brain networks. The reason is the small differences in the connection strengths of the original structural connectivity. After applying beta mapping, these differences are amplified and some interfering information is removed, allowing more discriminative features to be extracted in the diffusion process.

4.6 Limitations and future work

There are three primary limitations in the current study. First, brain regions are defined using only the AAL template. In future studies, we will validate the efficacy of the proposed method using other brain region templates, such as the Human Brainnetome Atlas (Fan et al., 2016). Second, our method only considers connectivity information among brain regions even though brain regions still have some attributes, such as cortical thickness, anisotropy index, ReHo, and ALFF, which are also crucial for diagnosing neuropsychiatric disorders. Therefore, we will combine brain attributes and brain connectivity to further improve neuropsychiatric disorder diagnosis. Third, BD episodes include different phases (e.g., manic, depressive, or mixed). In our study, we do not consider the different phases of BD. Different phases may have different brain activities, necessitating further studies in the future.

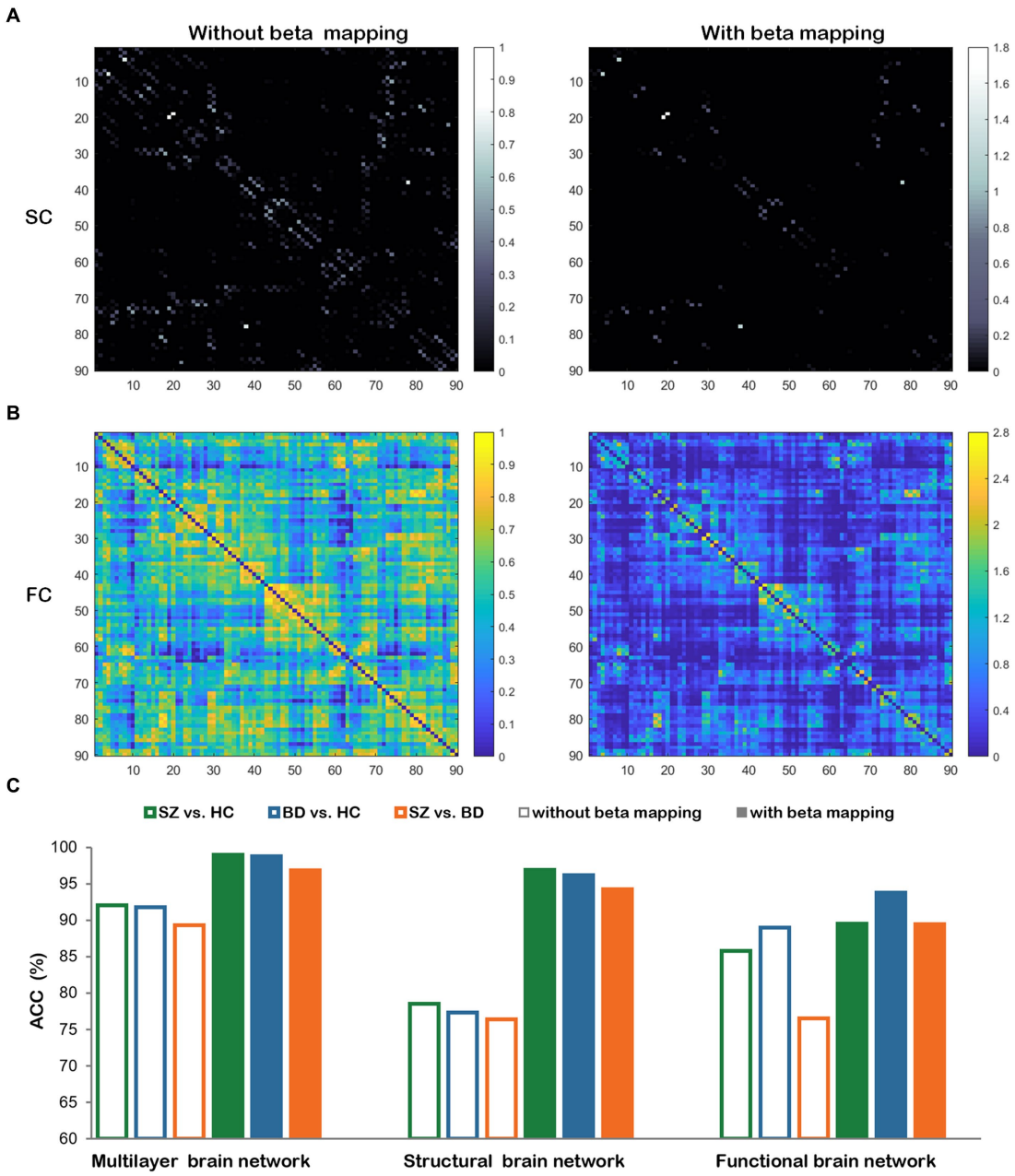


FIGURE 10
The comparisons with and without beta mapping. **(A)** Structural connectivity matrix. **(B)** Functional connectivity matrix. **(C)** Classification accuracies for three classification tasks with and without beta mapping in three brain networks.

5 Conclusion

In this study, we propose a novel brain network analysis method based on multiple modalities, which integrates SC and FC by intelligently traversing the nodes between structural and functional layers in a diffusion manner. Our approach takes full advantage of the complementary information and unique characteristics provided by various modalities and generates node representations with holistic information. Moreover, beta mapping allows the refined connectivity to encompass more valuable information, which further guides the diffusion process to concentrate on crucial brain regions to learn discriminative features. Experimental results on neuropsychiatric disorders validate the efficacy of our method.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving humans were approved by the Poldrack Lab and Center for Reproducible Neuroscience at Stanford University. The studies were conducted in accordance with the local legislation and

institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YH: Methodology, Writing – original draft. YL: Formal analysis, Writing – review & editing. YY: Data curation, Writing – original draft. XZ: Data curation, Writing – original draft. WY: Data curation, Writing – original draft. TL: Validation, Writing – review & editing. YN: Formal analysis, Writing – review & editing. TinY: Validation, Writing – review & editing. XL: Formal analysis, Writing – review & editing. DL: Supervision, Writing – review & editing. JX: Supervision, Writing – review & editing. BW: Conceptualization, Writing – review & editing. TiaY: Project administration, Writing – review & editing. MX: Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (62176177 and 20A20191); the 2018AAA0102601 (2018AAA0102604); the STI 2030—Major Projects (2022ZD0208500); the Natural Science Foundation of Shanxi (20210302123112); and the Research Project Supported by the Shanxi Scholarship Council of China (2021–039).

Conflict of interest

The authors declare that this study was implemented without financial involvement that could be perceived as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024.1303741/full#supplementary-material

References

Aggarwal, P., Gupta, A., and Garg, A. (2017). Multivariate brain network graph identification in functional MRI. *Med. Image Anal.* 42, 228–240. doi: 10.1016/j.media.2017.08.007

Cao, R., Wang, X., Gao, Y., Li, T., Zhang, H., Hussain, W., et al. (2020). Abnormal anatomical Rich-Club organization and structural-functional coupling in mild cognitive impairment and Alzheimer's disease. *Front. Neurol.* 11:53. doi: 10.3389/fneur.2020.00053

Chen, Y.-L., Kao, Z.-K., Wang, P.-S., Huang, C.-W., Chen, Y.-C., and Wu, Y.-T. (2017). *Resilience of functional networks: A potential Indicator for classifying bipolar disorder and schizophrenia. In 2017 international automatic control conference (cacs)* (New York: Ieee).

Chen, Z., Liu, Y., Zhang, Y., and Li, Q. (2023). Orthogonal latent space learning with feature weighting and graph learning for multimodal Alzheimer's disease diagnosis. *Med. Image Anal.* 84:102698. doi: 10.1016/j.media.2022.102698

Chyzhyk, D., Savio, A., and Graña, M. (2015). Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM. *Neural Netw.* 68, 23–33. doi: 10.1016/j.neunet.2015.04.002

Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A. A. C., Lukemire, J., et al. (2022). BrainGB: a benchmark for brain network analysis with graph neural networks. *IEEE Trans. Med. Imaging* 42, 493–506. doi: 10.1109/TMI.2022.3218745

Cui, Z., Zhong, S., Xu, P., Gong, G., and He, Y. (2013). PANDA: a pipeline toolbox for analyzing brain diffusion images. *Front. Hum. Neurosci.* 7, 42. doi: 10.3389/fnhum.2013.00042

Dai, Z., Lin, Q., Li, T., Wang, X., Yuan, H., Yu, X., et al. (2019). Disrupted structural and functional brain networks in Alzheimer's disease. *Neurobiol. Aging* 75, 71–82. doi: 10.1016/j.neurobiolaging.2018.11.005

Dong, D., Wang, Y., Chang, X., Jiang, Y., Klugah-Brown, B., Luo, C., et al. (2017). Shared abnormality of white matter integrity in schizophrenia and bipolar disorder: a comparative voxel-based meta-analysis. *Schizophr. Res.* 185, 41–50. doi: 10.1016/j.schres.2017.01.005

Du, Y., Hao, H., Wang, S., Pearlson, G. D., and Calhoun, V. D. (2020). Identifying commonality and specificity across psychosis sub-groups via classification based on features from dynamic connectivity analysis. *NeuroImage-Clin.* 27:102284. doi: 10.1016/j.nicl.2020.102284

Dubois, J., and Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends Cogn. Sci.* 20, 425–443. doi: 10.1016/j.tics.2016.03.014

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., et al. (2016). The human Brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* 26, 3508–3526. doi: 10.1093/cercor/bhw157

Gao, S., Calhoun, V., and Sui, J. (2020). Multi-modal component subspace-similarity-based multi-kernel SVM for schizophrenia classification. *Med. Imag. 2020*:139. doi: 10.1117/12.2550339

Grover, A., and Leskovec, J. (2016). "Node2vec: scalable feature learning for networks" in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (San Francisco, California, USA: ACM), 855–864.

Hall, J., Whalley, H. C., Marwick, K., McKirdy, J., Sussmann, J., Romaniuk, L., et al. (2010). Hippocampal function in schizophrenia and bipolar disorder. *Psychol. Med.* 40, 761–770. doi: 10.1017/S0033291709991000

Hoptman, M. J., Zuo, X.-N., Butler, P. D., Javitt, D. C., D'Angelo, D., Mauro, C. J., et al. (2010). Amplitude of low-frequency oscillations in schizophrenia: a resting state fMRI study. *Schizophr. Res.* 117, 13–20. doi: 10.1016/j.schres.2009.09.030

Huang, J., Wang, M., Xu, X., Jie, B., and Zhang, D. (2020a). A novel node-level structure embedding and alignment representation of structural networks for brain disease analysis. *Med. Image Anal.* 65:101755. doi: 10.1016/j.media.2020.101755

Huang, J., Zhou, L., Wang, L., and Zhang, D. (2020b). Attention-diffusion-bilinear neural network for brain network analysis. *IEEE Trans. Med. Imaging* 39, 2541–2552. doi: 10.1109/TMI.2020.2973650

Jie, B., Liu, M., and Shen, D. (2018). Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. *Med. Image Anal.* 47, 81–94. doi: 10.1016/j.media.2018.03.013

Kim, D.-J., Schnakenberg Martin, A. M., Shin, Y.-W., Jo, H. J., Cheng, H., Newman, S. D., et al. (2019). Aberrant structural-functional coupling in adult cannabis users. *Hum. Brain Mapp.* 40, 252–261. doi: 10.1002/hbm.24369

Klauser, P., Baker, S. T., Cropley, V. L., Bousman, C., Fornito, A., Cocchi, L., et al. (2017). White matter disruptions in schizophrenia are spatially widespread and topologically converge on brain network hubs. *Schizophr. Bull.* 43, sbw100–sbw435. doi: 10.1093/schbul/sbw100

Koo, M.-S., Levitt, J. J., Salisbury, D. F., Nakamura, M., Shenton, M. E., and McCarley, R. W. (2008). A cross-sectional and longitudinal magnetic resonance imaging study of cingulate gyrus gray matter volume abnormalities in first-episode schizophrenia and first-episode affective psychosis. *Arch. Gen. Psychiatry* 65, 746–760. doi: 10.1001/archpsyc.65.7.746

Lei, B., Cheng, N., Frangi, A. F., Tan, E.-L., Cao, J., Yang, P., et al. (2020). Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. *Med. Image Anal.* 61:101652. doi: 10.1016/j.media.2020.101652

Lian, C., Liu, M., Zhang, J., and Shen, D. (2020). Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 880–893. doi: 10.1109/TPAMI.2018.2889096

Lin, K., Shao, R., Lu, R., Chen, K., Lu, W., Li, T., et al. (2018). Resting-state fMRI signals in offspring of parents with bipolar disorder at the high-risk and ultra-high-risk stages and their relations with cognitive function. *J. Psychiatr. Res.* 98, 99–106. doi: 10.1016/j.jpsychires.2018.01.001

Liu, L., Chang, J., Wang, Y., Liang, G., Wang, Y.-P., and Zhang, H. (2022a). Decomposition-based correlation learning for multi-modal MRI-based classification of neuropsychiatric disorders. *Front. Neurosci.* 16:832276. doi: 10.3389/fnins.2022.832276

Liu, L., Wang, Y.-P., Wang, Y., Zhang, P., and Xiong, S. (2022b). An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Med. Image Anal.* 81:102550. doi: 10.1016/j.media.2022.102550

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

Lui, S., Yao, L., Xiao, Y., Keedy, S. K., Reilly, J. L., Keefe, R. S., et al. (2015). Resting-state brain function in schizophrenia and psychotic bipolar probands and their first-degree relatives. *Psychol. Med.* 45, 97–108. doi: 10.1017/S003329171400110X

Lama, R. K., and Kwon, G.-R. (2021). Diagnosis of Alzheimer's disease using brain network. *Front. Neurosci.* 15:605115. doi: 10.3389/fnins.2021.605115

Mahon, P. B., Eldridge, H., Crocker, B., Notes, L., Gindes, H., Postell, E., et al. (2012). An MRI study of amygdala in schizophrenia and psychotic bipolar disorder. *Schizophr. Res.* 138, 188–191. doi: 10.1016/j.schres.2012.04.005

Masaoka, Y., Velakoulis, D., Brewer, W. J., Cropley, V. L., Bartholomeusz, C. F., Yung, A. R., et al. (2020). Impaired olfactory ability associated with larger left hippocampus and rectus volumes at earliest stages of schizophrenia: a sign of neuroinflammation? *Psychiatry Res.* 289:112909. doi: 10.1016/j.psychres.2020.112909

Mashal, N., Vishne, T., and Laor, N. (2014). The role of the precuneus in metaphor comprehension: evidence from an fMRI study in people with schizophrenia and healthy participants. *Front. Hum. Neurosci.* 8:818. doi: 10.3389/fnhum.2014.00818

Masoudi, B., Daneshvar, S., and Razavi, S. N. (2021). Multi-modal neuroimaging feature fusion via 3D convolutional neural network architecture for schizophrenia diagnosis. *Intell. Data Anal.* 25, 527–540. doi: 10.3233/IDA-205113

Masoudi, B., and Danishvar, S. (2022). Deep multi-modal schizophrenia disorder diagnosis via a GRU-CNN architecture. *NNW* 32, 147–161. doi: 10.14311/NNW.2022.32.009

Mayer, A. R., Ruhl, D., Merideth, F., Ling, J., Hanlon, F. M., Bustillo, J., et al. (2013). Functional imaging of the hemodynamic sensory gating response in schizophrenia: fMRI and gating in schizophrenia. *Hum. Brain Mapp.* 34, 2302–2312. doi: 10.1002/hbm.22065

Mill, R. D., Winfield, E. C., Cole, M. W., and Ray, S. (2021). Structural MRI and functional connectivity features predict current clinical status and persistence behavior in prescription opioid users. *Neuroimage Clin.* 30:102663. doi: 10.1016/j.nicl.2021.102663

Mori, S., and van Zijl, P. C. M. (2002). Fiber tracking: principles and strategies - a technical review. *NMR Biomed.* 15, 468–480. doi: 10.1002/nbm.781

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* 44, 893–905. doi: 10.1016/j.neuroimage.2008.09.036

Osipowicz, K., Sperling, M. R., Sharan, A. D., and Tracy, J. I. (2016). Functional MRI, resting state fMRI, and DTI for predicting verbal fluency outcome following resective surgery for temporal lobe epilepsy. *J. Neurosurg.* 124, 929–937. doi: 10.3171/2014.9.JNS131422

Poldrack, R. A., Congdon, E., Triplett, W., Gorgolewski, K. J., Karlsgodt, K. H., Mumford, J. A., et al. (2016). A phenome-wide examination of neural and cognitive function. *Sci. Data* 3:160110. doi: 10.1038/sdata.2016.110

Qiu, L., Lui, S., Kuang, W., Huang, X., Li, J., Li, J., et al. (2014). Regional increases of cortical thickness in untreated, first-episode major depressive disorder. *Transl. Psychiatry* 4:e378. doi: 10.1038/tp.2014.18

Qureshi, M. N. I., Oh, J., Cho, D., Jo, H. J., and Lee, B. (2017). Multimodal discrimination of schizophrenia using hybrid weighted feature concatenation of brain functional connectivity and anatomical features with an extreme learning machine. *Front. Neuroinform.* 11:59. doi: 10.3389/fninf.2017.00059

Rahimiasl, M., Charkari, N. M., and Ghaderi, F. (2021). Random walks on B distributed resting-state functional connectivity to identify Alzheimer's disease and mild cognitive impairment. *Clin. Neurophysiol.* 132, 2540–2550. doi: 10.1016/j.clinph.2021.06.036

Repple, J., Meinert, S., Grotegerd, D., Kugel, H., Redlich, R., Dohm, K., et al. (2017). A voxel-based diffusion tensor imaging study in unipolar and bipolar depression. *Bipolar Disord.* 19, 23–31. doi: 10.1111/bdi.12465

Ribeiro, L. F. R., Savarese, P. H. P., and Figueiredo, D. R. (2017). "Struc2vec: learning node representations from structural identity" in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 385–394.

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

Salsabilian, S., and Najafizadeh, L. (2020). "Detection of mild traumatic brain injury via topological graph embedding and 2D convolutional neural networks" in *2020 42nd annual international conference of the IEEE engineering in Medicine & Biology Society (EMBC)* (Montreal, QC, Canada: IEEE), 3715–3718.

Shao, W., Peng, Y., Zu, C., Wang, M., and Zhang, D. (2020). Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Comput. Med. Imaging Graph.* 80:101663. doi: 10.1016/j.compmedimag.2019.101663

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., et al. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12, 506–518. doi: 10.1038/nprot.2016.178

Shi, G., Zhu, Y., Liu, W., Yao, Q., and Li, X. (2022). Heterogeneous graph-based multimodal brain network learning. Available at: http://arxiv.org/abs/2110.08465 (Accessed March 11, 2023).

Shimizu, M., Fujiwara, H., Hirao, K., Namiki, C., Fukuyama, H., Hayashi, T., et al. (2008). Structural abnormalities of the adhesio interthalamica and mediodorsal nuclei of the thalamus in schizophrenia. *Schizophr. Res.* 101, 331–338. doi: 10.1016/j.schres.2007.12.486

Silva, R. F., Castro, E., Gupta, C. N., Cetin, M., Arbabshirani, M., Potluru, V. K., et al. (2014). "The tenth annual MLSP competition: schizophrenia classification challenge" in *2014 IEEE international workshop on machine learning for signal processing (MLSP)* (Reims, France: IEEE), 1–6.

Song, X., Zhou, F., Frangi, A. F., Cao, J., Xiao, X., Lei, Y., et al. (2023). Multicenter and multichannel pooling GCN for early AD diagnosis based on dual-modality fused brain network. *IEEE Trans. Med. Imaging* 42, 354–367. doi: 10.1109/TMI.2022.3187141

Tripathi, S., Nozadi, S. H., Shakeri, M., and Kadoury, S. (2017). Sub-cortical shape morphology and voxel-based features for Alzheimer's disease classification. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 991–994.

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

Wang, B., Zhang, S., Yu, X., Niu, Y., Niu, J., Li, D., et al. (2022). Alterations in white matter network dynamics in patients with schizophrenia and bipolar disorder. *Hum. Brain Mapp.* 43, 3909–3922. doi: 10.1002/hbm.25892

Wang, S., He, L., Cao, B., Lu, C.-T., Yu, P. S., and Ragin, A. B. (2017). Structural deep brain network mining. Kdd'17: Proceedings of the 23rd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, (New York: Assoc Computing Machinery). 475–484.

Xia, M., Womer, F. Y., Chang, M., Zhu, Y., Zhou, Q., Edmiston, E. K., et al. (2019). Shared and distinct functional architectures of brain networks across psychiatric disorders. *Schizophr. Bull.* 45, 450–463. doi: 10.1093/schbul/sby046

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

Yan, T., Wang, W., Yang, L., Chen, K., Chen, R., and Han, Y. (2018). Rich club disturbances of the human connectome from subjective cognitive decline to Alzheimer's disease. *Theranostics* 8, 3237–3255. doi: 10.7150/thno.23772

Yan, Y., Zhu, J., Duda, M., Solarz, E., Sripada, C., and Koutra, D. (2019). GroupINN: grouping-based interpretable neural network for classification of limited, Noisy brain data. Kdd'19: Proceedings of the 25th Acm Sigkdd International Conferencce on Knowledge Discovery and Data Mining, (New York: Assoc Computing Machinery). 772–782.

Yan, N., Mengzhou, X., Yan, T., Xiaowen, L., Dandan, L., Jie, X., et al. (2020). "Unified brain network with functional and structural data" in *Medical image computing and computer assisted intervention – MICCAI 2020 lecture notes in computer science*. eds. A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga and S. K. Zhou et al. (Cham: Springer International Publishing), 114–123.

Zhang, D., Huang, J., Jie, B., Du, J., Tu, L., and Liu, M. (2018). Ordinal pattern: a new descriptor for brain connectivity networks. *IEEE Trans. Med. Imaging* 37, 1711–1722. doi: 10.1109/TMI.2018.2798500

Zhang, L., Li, W., Wang, L., Bai, T., Ji, G.-J., Wang, K., et al. (2020). Altered functional connectivity of right inferior frontal gyrus subregions in bipolar disorder: a resting state fMRI study. *J. Affect. Disord.* 272, 58–65. doi: 10.1016/j.jad.2020.03.122

Zhang, L., Wang, L., Gao, J., Risacher, S. L., Yan, J., Li, G., et al. (2021). Deep fusion of brain structure-function in mild cognitive impairment. *Med. Image Anal.* 72:102082. doi: 10.1016/j.media.2021.102082

Zhong, X., Pu, W., and Yao, S. (2016). Functional alterations of fronto-limbic circuit and default mode network systems in first-episode, drug-naive patients with major depressive disorder: a meta-analysis of resting-state fMRI data. *J. Affect. Disord.* 206, 280–286. doi: 10.1016/j.jad.2016.09.005

Zhou, S.-Y., Suzuki, M., Hagino, H., Takahashi, T., Kawasaki, Y., Matsui, M., et al. (2005). Volumetric analysis of sulci/gyri-defined in vivo frontal lobe regions in schizophrenia: precentral gyrus, cingulate gyrus, and prefrontal region. *Psychiatry Res. Neuroimaging* 139, 127–139. doi: 10.1016/j.pscychresns.2005.05.005

Zhu, Q., Huang, J., and Xu, X. (2018). Non-negative discriminative brain functional connectivity for identifying schizophrenia on resting-state fMRI. *Biomed. Eng. Online* 17:32. doi: 10.1186/s12938-018-0464-x

Zhu, Q., Yang, J., Wang, S., Zhang, D., and Zhang, Z. (2022). Multi-modal non-Euclidean brain network analysis with community detection and convolutional autoencoder. *IEEE Trans. Emerg. Top. Comput. Intell.* 7, 436–446. doi: 10.1109/TETCI.2022.3171855

Zhu, Q., Yang, J., Xu, B., Hou, Z., Sun, L., and Zhang, D. (2021). Multimodal brain network jointly construction and fusion for diagnosis of epilepsy. *Front. Neurosci.* 15:734711. doi: 10.3389/fnins.2021.734711

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 23 April 2024
DOI 10.3389/fnins.2024.1380886

OPEN ACCESS

EDITED BY
Hao Zhang,
Central South University, China

REVIEWED BY
Dahua Yu,
Inner Mongolia University of Science and Technology, China
Yinglong Dai,
Hunan Normal University, China
Jun Hu,
Hunan University, China

*CORRESPONDENCE
Dan Yu
✉ yudan202310@163.com

RECEIVED 02 February 2024
ACCEPTED 01 April 2024
PUBLISHED 23 April 2024

CITATION
Yu D and Fang Jh (2024) Using artificial intelligence methods to study the effectiveness of exercise in patients with ADHD. *Front. Neurosci.* 18:1380886. doi: 10.3389/fnins.2024.1380886

Using artificial intelligence methods to study the effectiveness of exercise in patients with ADHD

Dan Yu[1]* and Jia hui Fang[2]

[1]The College of Education and Liberal Arts, Adamson University, Manila, Philippines, [2]Medical Faculty, Ludwig Maximilian University of Munich, Munich, Germany

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental disorder that significantly affects children and adults worldwide, characterized by persistent inattention, hyperactivity, and impulsivity. Current research in this field faces challenges, particularly in accurate diagnosis and effective treatment strategies. The analysis of motor information, enriched by artificial intelligence methodologies, plays a vital role in deepening our understanding and improving the management of ADHD. The integration of AI techniques, such as machine learning and data analysis, into the study of ADHD-related motor behaviors, allows for a more nuanced understanding of the disorder. This approach facilitates the identification of patterns and anomalies in motor activity that are often characteristic of ADHD, thereby contributing to more precise diagnostics and tailored treatment strategies. Our approach focuses on utilizing AI techniques to deeply analyze patients' motor information and cognitive processes, aiming to improve ADHD diagnosis and treatment strategies. On the ADHD dataset, the model significantly improved accuracy to 98.21% and recall to 93.86%, especially excelling in EEG data processing with accuracy and recall rates of 96.62 and 95.21%, respectively, demonstrating precise capturing of ADHD characteristic behaviors and physiological responses. These results not only reveal the great potential of our model in improving ADHD diagnostic accuracy and developing personalized treatment plans, but also open up new research perspectives for understanding the complex neurological logic of ADHD. In addition, our study not only suggests innovative perspectives and approaches for ADHD treatment, but also provides a solid foundation for future research exploring similar complex neurological disorders, providing valuable data and insights. This is scientifically important for improving treatment outcomes and patients' quality of life, and points the way for future-oriented medical research and clinical practice.

KEYWORDS

ADHD, artificial intelligence, motor information, Random Forest, TCN, ACT-R

1 Introduction

ADHD is a common neurodevelopmental disorder that widely affects children and adults worldwide. Its main characteristics include persistent inattention, hyperactivity and impulsive behaviors, which often have a significant impact on an individual's ability to learn, socialize, and work (Tang et al., 2020). The diagnosis of ADHD is complex and varied and often requires a combination of medical, psychological and behavioral evaluations (Loh et al., 2023). Currently, the exact cause of ADHD is not fully understood, and it is

widely believed that a combination of genetics, environmental factors, and variations in brain development play a role (Tan et al., 2023). This complexity makes accurate diagnosis and effective treatment of ADHD a challenge (Amado-Caballero et al., 2020). Traditional diagnosis of ADHD relies on behavioral observations and psychological assessments, but these methods carry the potential for subjective judgments that can lead to diagnostic inconsistencies and accuracy issues (Shoeibi et al., 2023). In addition, due to the diversity of ADHD symptoms and their similarity to other disorders, it is often difficult for a single diagnostic approach to fully capture the full picture of the disease (Berrezueta-Guzman et al., 2021a). Therefore, researchers have been seeking more objective and accurate diagnostic tools.

The analysis of motor information plays a pivotal role in ADHD research and treatment, as hyperactive behavior significantly influences a patient's daily functioning and learning capabilities (Enriquez-Geppert et al., 2019). Motor control issues and hyperactivity, essential for diagnosis and treatment planning, offer insights into behavioral and neurophysiological changes in individuals with ADHD (Chen et al., 2020; Slobodin et al., 2020; Berrezueta-Guzman et al., 2021b). Movement tracking technologies and comprehensive analysis of motor behaviors can elucidate ADHD's neurobiological foundations (Amado-Caballero et al., 2023), enhancing diagnostic accuracy and aiding the development of more effective treatments (Berrezueta-Guzman et al., 2022). Additionally, advancements in Artificial Intelligence (AI) have transformed ADHD diagnosis and treatment strategies, with machine learning techniques uncovering complex patterns in data, facilitating preliminary feature selection and analysis (Moghaddari et al., 2020; Zhang et al., 2021a; Tang et al., 2022). This evolving AI landscape necessitates sophisticated, integrative models for a more nuanced understanding of ADHD (Leontyev et al., 2019; Yeh et al., 2020).

However, challenges remain in harnessing AI for ADHD research, notably in data acquisition, processing, and model comprehensiveness and interpretability. High-quality data collection and processing are critical for reliable research outcomes, but standardized, comprehensive datasets are difficult to obtain due to data diversity, complexity, and privacy concerns (Öztekin et al., 2021). Furthermore, the significant individual variability in ADHD symptoms and behaviors requires models that can integrate various data sources and analytical methods to accurately reflect these differences (Chen et al., 2021), highlighting the need for continued innovation in AI methodologies to address these challenges effectively.

In response to the identified gaps in existing research, we have developed an innovative network model that seamlessly integrates Random Forest, Temporal Convolutional Network (TCN), and Adaptive Control of Thought-Rational (ACT-R) to examine the effects of physical exercise on ADHD patients. This integrated framework is designed to transcend the limitations of traditional methodologies by leveraging the distinct strengths of each component (Speiser et al., 2019), thereby enhancing diagnostic accuracy and efficiency in handling complex ADHD-related data. The Random Forest algorithm, recognized for its prowess in managing high-dimensional data, plays a pivotal role in our model by identifying and isolating key features associated with ADHD symptoms. This process not only aids in refining input data for deeper analysis but also capitalizes on its capability to navigate non-linear and intricate data relationships (Dimov et al., 2020). Concurrently, the TCN model, with its specialization in processing time-series data, adeptly captures the dynamic changes in behavior and physiology characteristic of ADHD, thus offering a nuanced reflection of the patients' behavioral patterns and physiological states over time.

The model performs feature extraction and selection of multi-source data through the Random Forest algorithm to effectively identify key features associated with ADHD symptoms. Next, TCN is used to analyze time-series data from these features to capture behavioral and physiological signals over time. The ACT-R model is used to simulate the cognitive processes of ADHD patients to help predict their behavioral responses and symptom performance. Finally, the results of these analyses are synthesized and optimized for diagnosis and treatment prediction of ADHD using deep learning algorithms. By integrating these three models, our network model is able to provide an in-depth understanding of the behavioral and cognitive characteristics of ADHD patients from multiple dimensions and optimize the diagnosis and treatment prediction of ADHD using deep learning algorithms. This multidimensional and multimodal integrated approach is not only more accurate and effective in dealing with complex ADHD data, but also improves the accuracy of diagnosis and personalization of treatment. In addition, this approach helps to reveal the complex pathological mechanisms of ADHD, providing new perspectives and methods for future research and treatment strategies. This fusion model not only deepens the understanding of ADHD, but also provides a new, more precise and comprehensive analytical tool for clinical practice, which has important application value. In the subsequent sections of this thesis, we will detail our model architecture and experimental results to validate its effectiveness in studying the effects of exercise in patients with ADHD.

The contribution points of this paper are as follows:

- We have successfully developed a novel fusion model that integrates Random Forest, TCN, and ACT-R algorithms. This innovative integration approach has demonstrated outstanding performance in processing ADHD data, particularly in enhancing diagnostic accuracy and understanding the pathophysiology.
- Our research is the first to combine deep learning techniques with cognitive psychology models in the analysis of ADHD, providing a new perspective for the diagnosis and treatment of ADHD. This interdisciplinary approach allows us to gain a deeper understanding of the behavioral and cognitive characteristics of ADHD patients, laying the groundwork for developing more effective personalized treatment strategies.
- Our model has been validated on actual clinical data and has shown efficient computational performance and good scalability. This achievement not only proves the practicality of our model but also provides a reliable reference for applying deep learning and cognitive models in future research on similar complex neurological disorders.

2 Related work

2.1 DNN for analyzing ADHD patients' response to exercise

Recent endeavors in the realm of ADHD research have seen the application of Deep Neural Networks (DNN) to parse through complex, multidimensional datasets (Baxi et al., 2022), ranging from biometric readings to comprehensive behavioral assessments (Gupta et al., 2022). By harnessing the power of DNN, researchers aim to uncover the nuanced effects that physical activities exert on the ADHD phenotype, hoping to identify patterns that correlate with symptom alleviation or exacerbation (Wang et al., 2024). The capability of DNN to process vast arrays of input data and to learn from these inputs in an unsupervised or semi-supervised manner has opened up new avenues for predicting the outcomes of various therapeutic interventions, including exercise and movement-based therapies.

However, deploying DNN in ADHD research is fraught with challenges. The primary issue revolves around the interpretability of the models. The intrinsic complexity of DNN architectures, while a boon for navigating large data sets, renders the extraction of clear, actionable insights difficult (Ahmadi et al., 2021). Clinicians and therapists seeking to apply these findings are often met with a gap between statistical significance and practical applicability. Moreover, the reliance on extensive computational resources for data processing and model training limits the accessibility of DNN methodologies, particularly in resource-constrained research environments. The demand for vast, meticulously annotated data sets further complicates research efforts (Hernández-Capistran et al., 2023), given the inherent variability in ADHD manifestations across individuals and the ethical considerations tied to patient data privacy.

2.2 SVM in identifying ADHD biomarkers from physical activity data

The use of Support Vector Machines (SVM) in analyzing behavioral data presents a focused approach to understanding ADHD, especially in the context of physical activity interventions (Mohd et al., 2022). SVM's robust classification capabilities allow for the distinction between ADHD-affected individuals and their neurotypical peers based solely on quantified behavioral metrics derived from physical activity patterns (Wang et al., 2018). Such analyses are instrumental in pinpointing potential behavioral biomarkers for ADHD, facilitating a deeper comprehension of the disorder's external manifestations and the ways in which targeted physical interventions might ameliorate or modify these behaviors.

Despite the strengths of SVM in classification tasks, the model's application in ADHD research is not devoid of limitations. The necessity for labeled data poses a significant bottleneck (Chen et al., 2023), especially in early-stage research where diagnostic ambiguity prevails. Additionally, SVM models, traditionally linear, may struggle with the complex, non-linear behavioral patterns characteristic of ADHD, even though kernel methods can offer some mitigation (Eslami et al., 2021). The focus on behavioral data, to the exclusion of neurophysiological or cognitive data, might also narrow the scope of findings, potentially overlooking multifaceted aspects of ADHD symptomatology.

2.3 CNN for processing EEG data in ADHD exercise studies

Convolutional Neural Networks (CNN) have revolutionized the analysis of neurophysiological data, such as EEG, offering fresh perspectives on the neurological aspects of ADHD (TaghiBeyglou et al., 2022). The application of CNN to EEG data pre- and post-physical activity interventions has shed light on the neurophysiological shifts that might underlie observed behavioral changes in ADHD patients. CNN's adeptness at detecting spatial hierarchies in data makes it uniquely suited to identifying patterns within the complex signals characteristic of EEG recordings (Ribas et al., 2023), providing a conduit for exploring the neurobiological impact of exercise on individuals with ADHD.

The implementation of CNN in the study of ADHD through neurophysiological data is not without challenges. The model's sensitivity to the specificities of the training data raises concerns about overfitting (Delvigne et al., 2021), particularly acute in neurophysiological studies where sample sizes are often limited. The preprocessing required to adapt EEG data for CNN analysis is both intricate and labor-intensive, risking the introduction of bias or the loss of critical information (Sawangjai et al., 2019). Moreover, the complexity of CNN outputs complicates their translation into clinically relevant insights, presenting an ongoing challenge for bridging the divide between advanced AI-driven analyses and actionable treatment strategies for ADHD.

By elaborating on these studies, we gain a nuanced understanding of the current landscape of AI in ADHD research concerning physical activity, acknowledging the progress made and the hurdles that lie ahead. This comprehensive view serves as a critical stepping stone for future investigations aimed at harnessing AI's full potential in this domain.

3 Materials and methods

3.1 Overview of our network

In this study, we have developed an integrated model combining Random Forest, Temporal Convolutional Network (TCN), and Adaptive Control of Thought-Rational (ACT-R) to investigate the effects of physical activity in patients with ADHD.

We developed a comprehensive model that integrates the strengths of Random Forest, TCN and ACT-R to cope with the complexity of ADHD. Random Forest is crucial for feature selection and extraction. It processes the initial input data, identifying and isolating key features that are most relevant to ADHD symptoms and motor activities. The strength of Random Forest lies in its ability to handle high-dimensional data and uncover complex, non-linear relationships, making it ideal for the initial analysis stage. TCN serves as the core component for analyzing time-series data, particularly motor monitoring and neurophysiological data. Its architecture, designed to handle

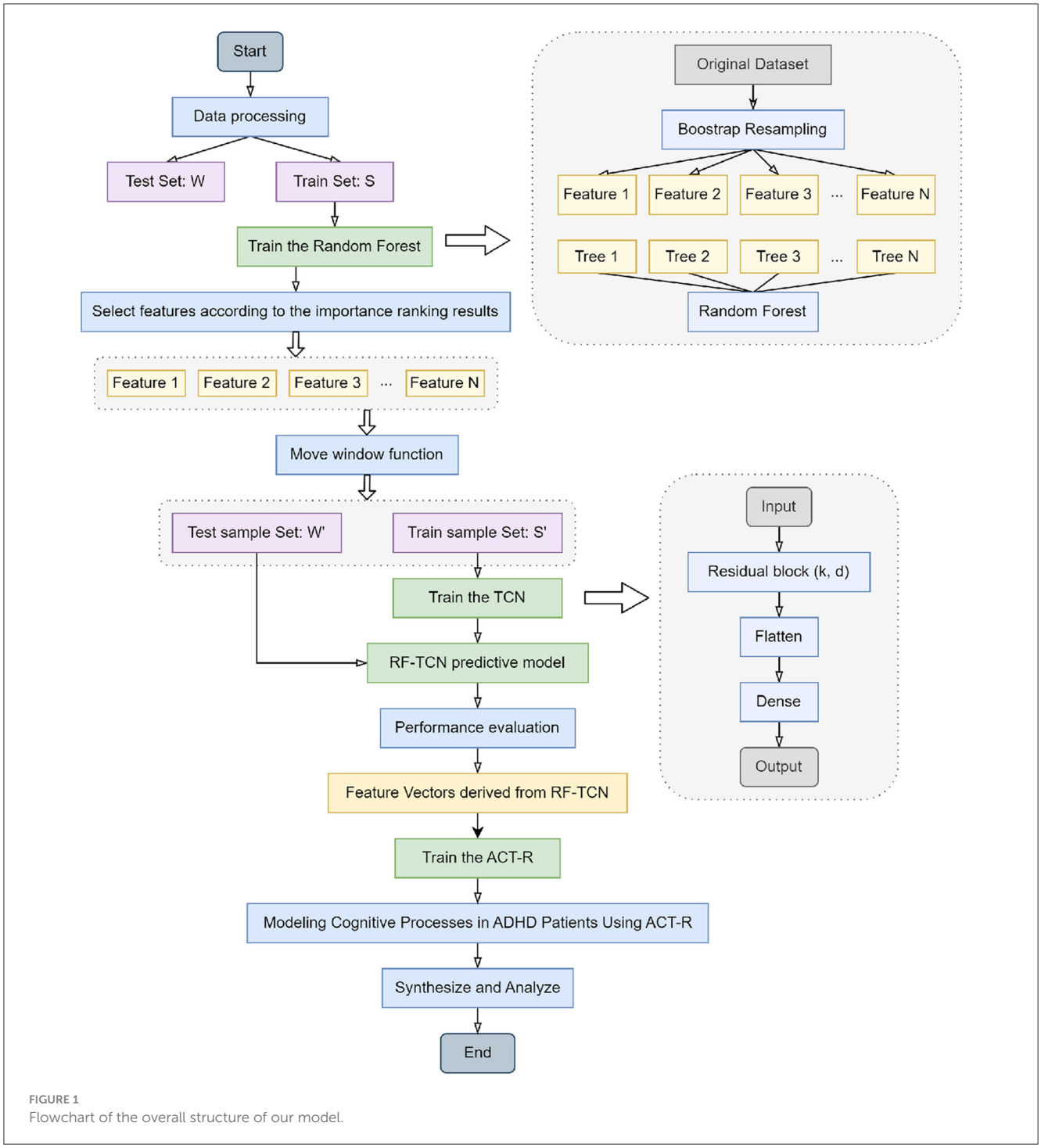


FIGURE 1
Flowchart of the overall structure of our model.

sequential data, captures temporal dependencies and dynamic changes in ADHD patients' behavior and physiological responses. TCN's effectiveness in our model stems from its deep, dilated convolutional structure, enabling detailed analysis of intricate time-related patterns. ACT-R is utilized to simulate and interpret the cognitive processes of ADHD patients. This model integrates the outputs from the Random Forest and TCN, providing a cognitive perspective to the analysis. It helps in understanding how ADHD affects cognitive functions and how physical activities might influence these cognitive patterns.

The workflow of our integrated model, detailing the collaborative functions of Random Forest, TCN, and ACT-R in the context of ADHD physical activity research, is systematically illustrated in the flowchart presented in Figure 1. In constructing our integrated network model, we began with the data processing of the original dataset, which included Bootstrap resampling to ensure consistency of data across the training and test sets. The Random Forest algorithm was trained on dataset S, selecting key features based on importance rankings. These features were then transformed into time series datasets S' and W' using a moving

window function, preparing them for TCN training. This process readied TCNs for training by sliding a fixed-size window along the time axis of the dataset and capturing local data features within each window. The generated time series datasets were then fed into the TCN, which was trained using its residual blocks defined by kernel size k and dilation coefficient d to build the RF-TCN predictive model. Data processed through these residual blocks passed through Flatten and Dense layers to generate the final output.

Upon the training and performance evaluation of the RF-TCN model, we proceeded to integrate the ACT-R model. The input to the ACT-R model consists of feature vectors derived from the Random Forest and Temporal Convolutional Network (RF-TCN) model. These vectors encapsulate the significant features related to ADHD symptoms' behaviors and physiological responses, pinpointed through our initial analyses. Utilizing this input, the ACT-R model was further trained to simulate the cognitive processes of ADHD patients. This process aimed at blending the time series analytical capabilities of the RF-TCN model with the cognitive simulations facilitated by the ACT-R framework. The output of the ACT-R model encompasses predictions on cognitive states and potential behavioral responses of ADHD patients to various exercise regimens. By providing a comprehensive analysis of the behavioral and cognitive patterns of ADHD patients under physical intervention, this output is invaluable for understanding how specific exercises can influence cognitive functions and behavioral patterns in patients with ADHD. This integrated approach not only enhances our ability to understand and evaluate the impact of physical activity on ADHD patients but also lays the groundwork for further personalized treatment approaches, aiming to tailor individualized exercise-based treatment plans based on the predictive insights generated by the ACT-R model.

The significance of our model lies in its multifaceted approach to understanding ADHD. By combining the strengths of Random Forest, TCN, and ACT-R, our model offers a comprehensive analysis of ADHD patients' motor activities and their cognitive implications. This integrated approach allows for a deeper understanding of how physical activity affects ADHD patients, not just in terms of immediate motor responses but also in long-term cognitive and behavioral changes. The model's ability to process complex data and provide insights into the temporal dynamics of ADHD presents a significant advancement in researching effective treatment and management strategies for ADHD, particularly in the realm of physical interventions.

To ensure the trustworthiness and transparency of our AI model, we incorporated an interpretability and reliability analysis into our methodology. For interpretability, we utilized SHapley Additive exPlanations (SHAP) values to quantify the impact of each feature on the model's predictions. This approach helps in identifying the most influential factors contributing to the model's decision-making process. Additionally, to assess the reliability of our model, we employed a rigorous cross-validation technique, along with an external validation on a separate dataset, ensuring the model's robustness and its capability to generalize across different populations.

The interpretability analysis revealed that certain features, such as the duration and intensity of exercise, played a significant role in the model's predictions regarding the effectiveness of exercise in ADHD patients. SHAP value plots highlighted these features' positive influence on the model's confidence in predicting improvement in ADHD symptoms, offering insights into how exercise routines can be optimized for therapeutic purposes.

The reliability analysis, conducted through 10-fold cross-validation and further validated on an external dataset, demonstrated consistent accuracy levels, underscoring the model's robustness. The slight variations observed across different folds were within acceptable limits, indicating the model's capability to generalize and perform reliably in diverse settings.

3.2 Random Forest

Random Forest is a machine learning classifier composed of multiple decision trees. It is capable of handling classification, regression, and dimensionality reduction problems (Borup et al., 2023). In a Random Forest, each decision tree operates independently and without correlation to others (Sheykhmousa et al., 2020). For classification tasks, each tree classifies the test sample, and the final category is determined by the mode of the outputs from the forest, essentially using a voting mechanism to decide the category of the test sample. For regression tasks, the final result is the average of the outputs from all trees. Compared to a single decision tree, Random Forest exhibits a stronger tolerance to outliers and noise and shows better performance in both prediction and classification (Cheng et al., 2019).

A decision tree is a commonly used algorithm for classification and regression (Maji and Arora, 2019). It constructs a tree-like structure by dividing the dataset into different subsets, where each node represents a feature, each branch represents a value of that feature, and each leaf node represents a category or a value. In building a decision tree, the optimal feature for splitting must be selected, which necessitates the concept of entropy. Entropy is a measure of the uncertainty of a dataset (Li et al., 2021); the greater the entropy, the higher the uncertainty. In decision trees, we aim to select the optimal feature that minimizes the entropy of the subsets post-split, thereby enhancing the accuracy of the decision tree. Therefore, entropy can be used to measure the information gain of each feature, aiding in the selection of the optimal feature.

The entropy in a decision tree can be calculated using Equation 1:

$$H(D) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{1}$$

Here, $H(D)$ denotes the entropy of dataset D, n is the number of categories in D, and π represents the proportion of samples of the ith category in D. To calculate entropy, we compute the proportion of each category in the dataset and substitute these into the formula.

In decision trees, it is also necessary to calculate the information gain of each feature for optimal splitting. Information gain can be calculated using Equation 2:

$$Gain(D, a) = H(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} H(D^v) \tag{2}$$

where $Gain\ (D, a)$ denotes the information gain of dataset D on feature a, $H(D)$ is the entropy of dataset D, V represents the number of values for feature a, D^v is the subset of samples where feature a has the value v, $|D^v|$ is the number of samples in D^v, and $|D|$ is the number of samples in dataset D. When calculating information gain, we compute it for each feature in the dataset and select the feature with the highest information gain for splitting.

The process of Random Forest involves several key steps. Initially, it includes a random sampling process where the model samples both rows and columns from the input data. For row sampling, it employs a bootstrap method, meaning that the sampled dataset may contain duplicate samples. If the input sample size is N, then the sampled dataset will also have N samples. This approach ensures that each tree in the training phase does not use all the samples, reducing the likelihood of over-fitting. For column sampling, out of M features, a subset of m features (where m $<<$ M) is selected. Following this, decision trees are constructed using a complete splitting method, where each leaf node either cannot be further split or contains samples belonging to the same category. Unlike many decision tree algorithms that involve a crucial step of pruning, Random Forest does not require this due to the randomness introduced in the two sampling processes, thereby preventing over-fitting even without pruning.

The procedure then involves drawing a specific number of samples from the training set randomly to form the root node samples for each tree. During the construction of the decision trees, a set number of candidate attributes are randomly selected, and the most suitable one is chosen as the splitting node. Once the Random Forest is built, for a test sample, each decision tree produces either a class output or a regression output. In classification problems, the final category is determined through a voting mechanism among the decision trees, while for regression problems, the final result is the average output of all the trees. As depicted in the Figure 2, suppose a Random Forest consists of three decision trees, with two trees classifying a sample as Category B and one as Category A, the Random Forest would classify the sample as Category B.

The randomness in Random Forest is reflected in two aspects. Firstly, it's exhibited in the randomness of sample selection, where a certain number of samples are randomly drawn from the training set with replacement to construct sub-datasets. These sub-datasets are of the same size as the original dataset, and elements within them can be repeated. Secondly, the randomness is in the selection of attributes. During the construction of each decision tree, a certain number of candidate attributes are randomly selected, from which the most suitable attribute is chosen for the splitting node. This process ensures diversity among the trees in the Random Forest, thereby enhancing the classification performance.

In our research, Random Forest, as a core tool, works in conjunction with the ACT-R model and the TCN model, playing a vital role. It employs an ensemble learning approach to comprehensively process and analyze data and insights obtained from both the ACT-R and TCN models. The ACT-R model is used to simulate the cognitive processes of ADHD patients, particularly during physical activities, while the TCN model primarily handles time-series data related to movement, such as motion monitoring or neurophysiological data.

The primary task of the Random Forest is to integrate and analyze these diverse data sources. Through its multitude of decision trees, Random Forest is capable of effectively handling high-dimensional and complex datasets, which is crucial for our research. It aids in identifying key factors affecting ADHD patients from multiple dimensions and enables precise predictions. Through the analysis conducted by Random Forest, we gain a deeper understanding of the potential cognitive and behavioral impacts of physical interventions on ADHD patients and predict the potential effects of different types of physical interventions on various patient groups. These insights are invaluable for designing more effective treatment plans and intervention measures, providing us with data-driven decision support.

Especially in the context of studying the impact of physical activities on ADHD patients, the application of Random Forest is particularly significant. It integrates various data sources, such as neuroimaging data and behavioral observation data, offering a comprehensive analytical perspective for our research. By analyzing different feature combinations, Random Forest helps to reveal the effects of physical interventions on the cognitive and behavioral patterns of ADHD patients. Its high-accuracy predictive and classification capabilities can also be used to assess the effectiveness of physical interventions for different types of ADHD patients and identify which patients may benefit most from specific types of physical activities. Thus, Random Forest becomes a powerful tool in addressing this complex issue.

3.3 Temporal convolutional networks

The Temporal Convolutional Network (TCN) is a neural network architecture specifically designed for processing time series data, with its core feature being the utilization of one-dimensional convolutional layers for handling such sequential data. A key characteristic of the TCN is causal convolution, ensuring that the model uses only the current and previous data points for predictions, effectively preventing the leakage of future information. This attribute is crucial for ensuring the accuracy and reliability of the model's predictions. Additionally, TCN incorporates a design with residual connections, similar to those used in ResNet. These residual connections help address the issue of vanishing gradients common in training deep networks (Gao et al., 2023), thereby enhancing the efficiency and stability of model training. This is particularly significant when dealing with complex time series data. Through its unique structure and functions, the TCN provides an effective method for understanding and analyzing time series data, making it particularly suitable for applications involving long-term data dependencies and complex dynamic patterns.

TCN, built upon the principles of CNN, features a dilated causal convolution architecture that maintains equal lengths for both input and output. The specific structure of this dilated causal convolution is depicted in Figure 3. This design choice in TCN, emphasizing dilated convolutions, enables the model to efficiently handle sequential data while preserving the temporal sequence length from input to output.

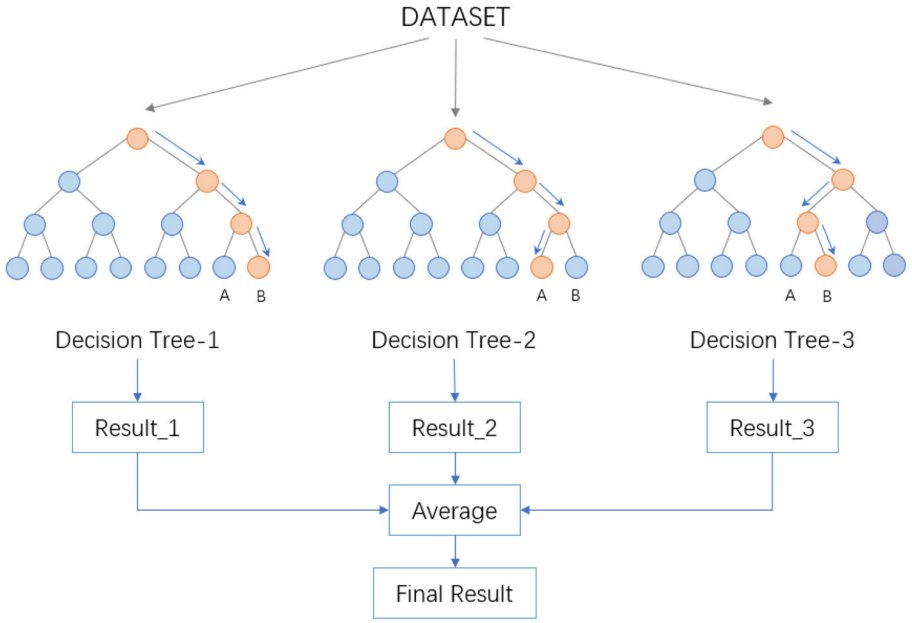


FIGURE 2
Data processing flow in random forest.

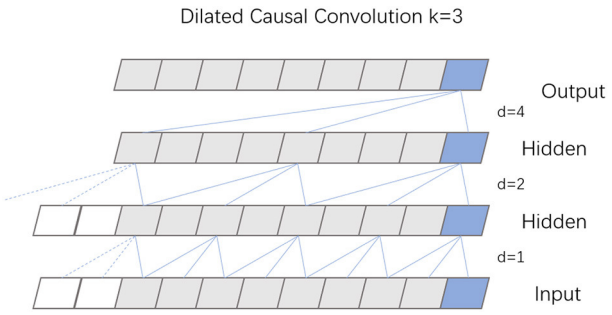


FIGURE 3
Structure diagram of dilation causal convolution.

Causal convolution is designed to exclusively consider present and past data points, disregarding any future information. This approach means that for any given time t in the output sequence, the result is influenced solely by the input sequence's elements at time t and earlier, thus preserving the integrity of historical data. Expanding on this concept, dilated convolution incorporates a dilation coefficient d, which dictates the interval at which the input is sampled, thereby enlarging the receptive field of each convolutional layer. The extent of this dilation, and consequently the sampling rate, hinges on the value of d. Typically, as a network deepens, the dilation coefficient d increases exponentially, often doubling with each added layer. To maintain uniformity in the size of the data through the network layers, and to ensure the output layer matches the width of the input layer, zero padding is employed within each layer of the dilated causal convolution.

To mitigate the problems of vanishing and exploding gradients that often arise in overly deep network structures, TCN incorporates a specifically designed residual block. This block consists of two layers of dilated causal convolution, complemented by a non-linear mapping arrangement that incorporates both a WeightNorm and a Dropout layer. A detailed illustration of this residual block structure is presented in Figure 4. The WeightNorm layer functions to standardize the weights within the network layer, thereby streamlining the training process, while the Dropout layer plays a crucial role in preventing overfitting. This configuration equips the TCN with the combined attributes of CNNs and RNNs. Its uncomplicated yet adaptable structure allows for parallel processing of input sequences, which significantly cuts down on both the memory usage and time required for network training. In medical and health research applications, TCN's capability to forecast long-step outputs from complex feature sets demonstrates a notable advantage over traditional models like LSTM and GRU.

In our experiments, the TCN works in conjunction with the Random Forest and ACT-R models, providing support for research into the impact of physical activity on patients with ADHD (Tian et al., 2023). The primary role of the TCN in this model combination is to process and analyze time series data, such as movement monitoring and neurophysiological data. These data are key to understanding the dynamic changes in the behavior and physiological responses of ADHD patients, and the TCN, with its deep and dilated convolutional structure, effectively captures these complex temporal dependencies. The results of the TCN analysis provide a rich feature input for the Random Forest and, when combined with the output from the ACT-R model, offer us a comprehensive perspective for understanding the cognitive and behavioral patterns of ADHD patients under physical activity interventions.

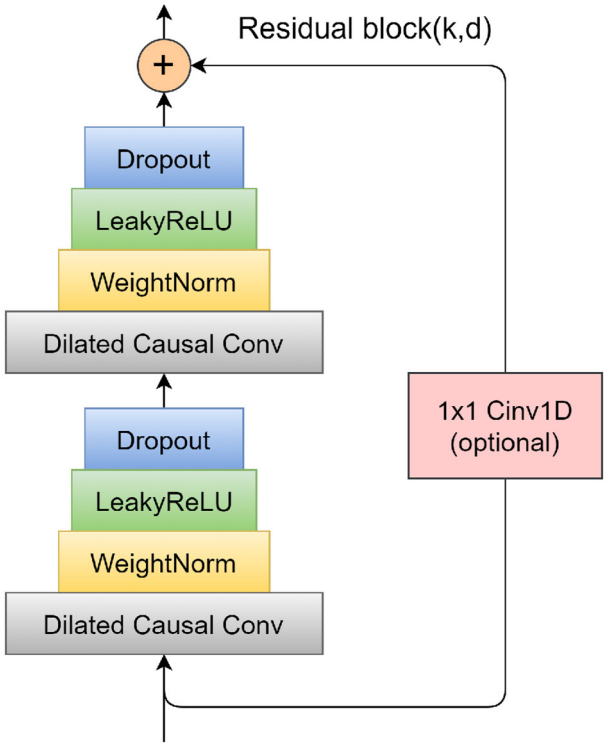


FIGURE 4
Schematic diagram of residual block.

The application of TCN has demonstrated its significance in analyzing the impact of physical activity on patients with ADHD. It not only provides a deep understanding of the immediate effects of physical interventions on the behavior of ADHD patients but also plays a crucial role in capturing the long-term effects of such interventions. Through in-depth analysis by TCN, we can uncover how physical interventions affect the daily behavior and cognitive patterns of ADHD patients, which is essential for accurately assessing the effectiveness of physical activity as a therapeutic approach. The analysis by TCN reveals both the immediate and long-term effects of physical activity on patients' cognition and behavior, and provides crucial data support for designing more personalized and effective treatment plans. This profound analysis and understanding of time series data offer a new perspective and approach for exploring the role of physical activity in treating ADHD, providing significant scientific evidence for enhancing treatment effectiveness and improving patients' quality of life. It also offers valuable data and insights for future research.

3.4 ACT-R

The ACT-R (Adaptive Control of Thought—Rational) model is a cognitive architecture specifically designed to simulate human cognitive processes (Fisher et al., 2020). This model is predicated on the assumption that human cognition is comprised of multiple interacting subsystems, each responsible for processing different types of information, such as visual and motor information. The core of the ACT-R model lies in decomposing the cognitive process into a series of modular components, each dedicated to processing specific types of information. This includes modules for storing long-term memory (Zhang et al., 2021b), buffers for processing short-term memory, and a decision center that guides behavior based on information from various modules. Additionally, ACT-R incorporates several distinct modules, each simulating specific human cognitive functions, such as thinking and decision-making processes, thereby facilitating the study and understanding of cognitive psychology phenomena.

The ACT-R system is a hybrid cognitive architecture consisting of both symbolic and sub-symbolic systems. As can be seen in Figure 5, the symbolic system is composed of several modules, with a procedural module at its core. This procedural module connects the various modules into a cohesive whole, functioning similarly to a model driven by a production system, where procedural rules in the module manipulate the buffers of different modules. The sub-symbolic system, although not explicitly represented in visualizations, controls the internal operations of modules in the symbolic system through mathematical methods. This structure allows the ACT-R model to simulate human cognitive processes with greater precision and comprehensiveness.

In the ACT-R model, different modules assume various functions and tasks. The Intentional Module (also known as the Goal Module) serves as the executive control center, responsible for planning and controlling behavior. It determines the goals of the current task and coordinates the activities of other modules to achieve these goals. The Declarative Module acts as a repository for storing facts, rules, and conceptual knowledge, supporting the model in accessing and retrieving information from long-term memory during decision-making and problem-solving processes. The Visual Module processes sensory input, simulating the human visual processing system. This module is responsible for perceiving and understanding visual information, such as objects, scenes, and symbols. The Manual Module enables the model to perform manual actions, such as moving and grasping objects. It controls the model's movements and interactions, simulating the execution of physical actions. The Production Module (also referred to as the Procedural System) is one of the core components of ACT-R, representing knowledge and decision-making. It includes production rules that describe condition-action pairs. When specific conditions are met, these production rules are triggered, executing corresponding actions and simulating the decision-making process in cognitive tasks.

In our proposed integrated model, the ACT-R model plays a crucial role in providing a deep understanding of the cognitive processes in ADHD patients. This encompasses how they process information, make decisions, and respond behaviorally to various stimuli, such as physical interventions. By leveraging the feature vectors derived from the RF-TCN model, the ACT-R model utilizes its comprehensive cognitive architecture to simulate these cognitive processes accurately. This architecture is adept at mirroring various cognitive functions, including memory processes, attention, and decision-making, thereby laying the groundwork for understanding the complex behavioral patterns that ADHD patients may exhibit in response to physical intervention.

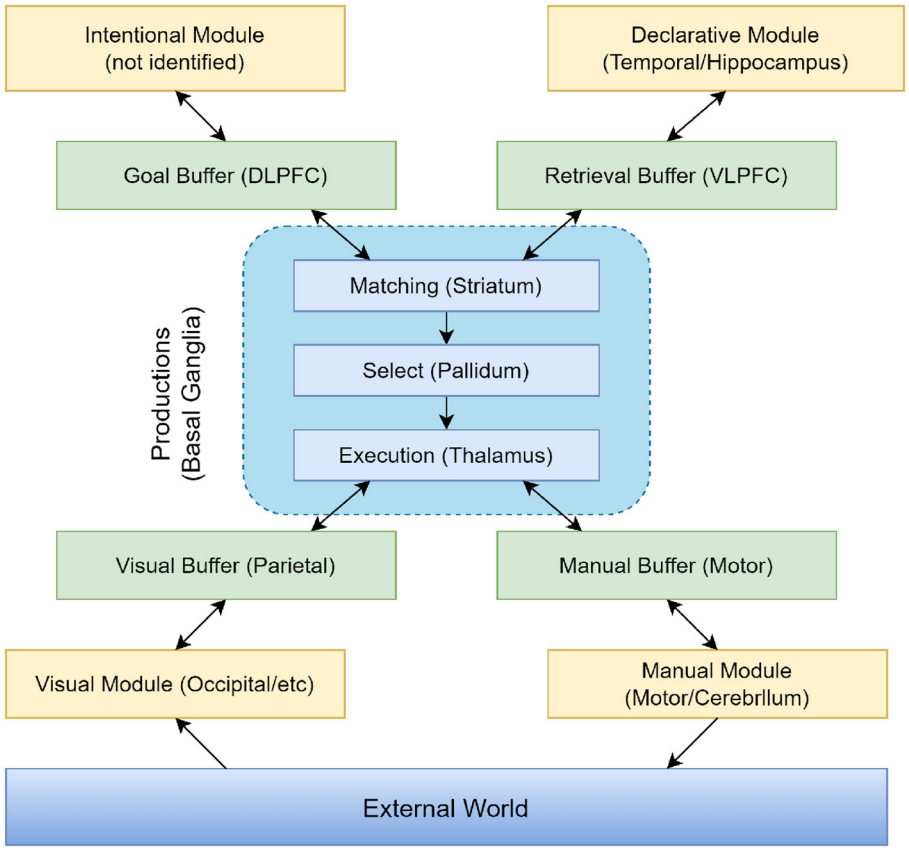


FIGURE 5
Information organization in ACT-R 5.0.

The integration of the ACT-R model with the TCN's capability in processing time-series data, like motion monitoring data, alongside the advantages of Random Forest in analyzing and integrating multi-dimensional data, culminates in a multi-faceted and multi-layered analytical framework. Through this simulation, the ACT-R model not only learns to associate specific patterns of physical activity with cognitive outcomes in ADHD patients but is also designed to simulate and understand the cognitive decision-making processes. It achieves this by mapping the input features—processed information from the RF-TCN model—to cognitive states that represent ADHD characteristics. This dynamic process enables the ACT-R model to learn the underlying cognitive patterns associated with ADHD, offering valuable insights into how various factors might influence cognitive processes in patients.

Combined, these elements highlight the integrative approach of our research, demonstrating how the ACT-R model's simulation capabilities, when enriched with data from the RF-TCN model, form a comprehensive analytical tool. This tool not only deciphers the intricate cognitive underpinnings of ADHD but also facilitates a nuanced understanding of how physical interventions can be optimized for therapeutic efficacy, based on individual cognitive responses.

The ACT-R model is vital in our experiments because it enables us to comprehend the impact of physical activities on ADHD patients from a cognitive perspective. By simulating the cognitive processes of ADHD patients, we gain a deeper understanding of their responses to physical interventions, including changes at cognitive, emotional, and behavioral levels. This in-depth understanding is crucial for assessing the effectiveness of physical interventions, especially when designing targeted treatment plans and intervention measures. Overall, the ACT-R model provides a unique perspective in our research, complementing the capabilities of TCN and Random Forest in data processing and analysis, and offers a key cognitive dimension to understand the overall impact of physical activities on ADHD patients.

ACT-R is not only a tool for simulating human cognitive processes but also a bridge linking the inner cognitive processes and external behavioral manifestations of ADHD patients. By precisely simulating the cognitive activities of ADHD patients under physical intervention, ACT-R provides insights into how they process information, make decisions, and how their attention and memory are affected by physical activities. The details of these cognitive processes are critical in evaluating the specific effects of physical interventions in areas such as improving attention, reducing impulsive behaviors, and enhancing emotional regulation. For instance, by simulating specific cognitive tasks, we can assess how physical activities influence the working memory, attention allocation, and task-switching abilities of ADHD patients. These details offer direct evidence of how physical interventions alter the brain's information processing methods in ADHD patients, aiding

in a better understanding of the potential mechanisms by which physical activities improve ADHD symptoms. Through this deep cognitive-level analysis, the ACT-R model significantly enhances our ability to design more effective treatment and intervention strategies, providing robust scientific support for improving the quality of life of ADHD patients.

4 Experiment

4.1 Datasets

To comprehensively explore the complexities of how physical activity impacts patients with ADHD, this study employs multiple datasets, aiming to provide an integrated analysis of the effects of exercise on individuals with ADHD from various perspectives. We have selected four key datasets, each with its unique value and applicability, aiding us in an in-depth understanding of the influence of physical activity on cognitive, physiological, and social behaviors in ADHD patients. These datasets include: the ADHD Dataset, the ADHD TIDAL Dataset, the ADHD-200 Dataset, and the EEG Dataset. The combined use of these datasets not only strengthens the foundation of our research but also offers robust support for subsequent data analysis and model development.

Attention-Deficit Hyperactivity Disorder Distribution (ADHD) Dataset (Cao et al., 2023): The ADHD Dataset offers a comprehensive exploration into ADHD, encompassing an extensive cohort of over 7,400 subjects. This dataset extends beyond simple ADHD symptomatology to include biosamples critical for genetic and biological research, shedding light on ADHD's hereditary aspects through family studies and enhancing our understanding of its genetic underpinnings. It incorporates detailed clinical tools, such as the Diagnostic Interview Schedule for Children, facilitating a thorough assessment of participants' conditions. With its access to genetic repositories, the dataset provides invaluable demographic, diagnostic, and genealogical data, serving as a pivotal resource for studies targeting the clinical and genetic aspects of ADHD. This aids in analyzing genetics and biological markers associated with the disorder. The dataset's vast size not only enables an in-depth analysis of ADHD's hereditary factors but also aids in identifying potential biomarkers, thereby enriching our comprehension of this complex condition.

To further prepare this rich dataset for our study, we undertook standardization processes to normalize scores across various scales and employed median imputation for missing values, drawing from similar patient profiles. This preprocessing step was crucial for ensuring data consistency and reliability. We extracted key features, such as symptom severity scores, diagnostic criteria, and patient demographics, enabling us to effectively correlate behavioral patterns with the impacts of physical activity. These steps ensured that the ADHD Dataset was meticulously prepared for our analysis, allowing for a nuanced examination of the interactions between genetic predispositions, clinical symptoms, and the benefits of physical interventions in ADHD patients.

The ADHD Teen Integrative Data Analysis Longitudinal (ADHD TIDAL) Dataset (Sibley and Coxe, 2020): Integrating data from four pivotal longitudinal studies conducted between 2010 and 2019, this dataset offers an expansive insight into the long-term effects of psychosocial treatments on 1,500 adolescent subjects diagnosed with ADHD. It provides a multifaceted view of treatment outcomes, encapsulating detailed information on academic performance, diagnostic criteria, and symptom ratings as reported by both parents and teachers. This dataset is instrumental in shedding light on various treatment modalities, including medication and special education interventions, thus delivering invaluable insights into ADHD's impact on educational outcomes and adolescents' daily lives. Such comprehensive information makes this dataset an essential tool for researchers aiming to assess the effectiveness and sustainability of ADHD treatments, offering a deep understanding of how different interventions influence the long-term wellbeing and academic success of affected adolescents.

To enhance the dataset's utility for our analysis, we undertook a meticulous preprocessing regimen. This involved aligning time-series data from multiple assessment points to ensure consistency across the longitudinal study and encoding categorical variables into a format conducive to machine learning analysis. Our preprocessing efforts concentrated on extracting pivotal features such as variations in symptom severity over time, adherence levels to prescribed treatments, and key indicators of academic performance.

ADHD-200 Dataset (Bellec et al., 2017): The ADHD-200 Dataset as a fundamental resource in neuroimaging research, shedding light on the profound impact of ADHD on brain function. Comprising 776 resting-state fMRI and anatomical datasets from eight independent imaging sites, it includes data from 285 children and adolescents with ADHD (ages 7–21) and 491 typically developing individuals. This amalgamation not only supports a wide-ranging comparative analysis but also deepens our investigation into the neurological underpinnings of ADHD and its developmental trajectory. The dataset is rich with detailed diagnostic statuses, ADHD symptom measures, and extensive demographic information, including age, sex, IQ, and medication history, making it an invaluable tool for probing into the neural basis and developmental aspects of ADHD. Furthermore, its unrestricted public access greatly enhances its utility, fostering diverse research endeavors aimed at decoding ADHD's neural correlates.

To cater to the unique requirements of MRI image analysis within this dataset, we undertook specific preprocessing steps, including skull stripping, spatial normalization, and smoothing, to refine the images for subsequent investigation. Our focal points during the analysis were on brain volume measurements in ADHD-impacted regions, connectivity patterns among these areas, and textural analysis of neural tissue for pinpointing structural differences. This meticulous approach allows for an in-depth exploration of how ADHD affects brain structure and function, laying a solid foundation for advancements in understanding, diagnosing, and treating ADHD.

EEG Data for ADHD/Control Children Dataset (Motie Nasrabadi et al., 2020): The EEG Dataset is distinguished by its focus on neurophysiological data through EEG recordings from a cohort of 61 children diagnosed with ADHD and 60 healthy controls, aged between 7 and 12 years. This dataset is enriched by comprehensive psychiatric evaluations and detailed medication histories, presenting an extensive neurological profile

of ADHD in children. Such a compilation of data is invaluable for pinpointing potential EEG biomarkers and dissecting the complex neural mechanisms underlying ADHD. These insights are crucial for developing more refined diagnostic and therapeutic strategies, particularly through AI-based research methodologies. By integrating clinical assessments with medication data, the EEG Dataset lays a robust groundwork for exploring the neurological aspects of ADHD in young patients, establishing it as a key resource in the field.

In preparing this dataset for analysis, we undertook meticulous preprocessing steps that included filtering to eliminate electrical noise and artifacts, segmenting the recordings into epochs guided by event markers, and applying baseline correction. We focused on extracting neurophysiological features such as spectral power in critical frequency bands, coherence between electrode pairs, and characteristics of event-related potentials. These selected features are designed to elucidate the neurophysiological foundations of ADHD and assess the impact of physical activities on brain function, thereby offering profound insights into the disorder and potential avenues for intervention.

These datasets offer a comprehensive perspective on ADHD, covering genetic, clinical, educational, neurological, and treatment-related aspects. For our research on the effects of exercise on ADHD patients using AI methods, this multifaceted data is crucial. It allows for a holistic analysis, integrating physical activity's impact on various dimensions of ADHD. By leveraging these diverse datasets, we could more accurately assess how exercise influences genetic predispositions, clinical symptoms, educational performance, and neurological functioning in ADHD patients. This approach is invaluable for developing a nuanced understanding of exercise's role in managing ADHD and its broader implications for patient care and family dynamics.

4.2 Experimental details

4.2.1 Experimental environment

Hardware Environment: The hardware environment used in the experiments consists of a high-performance computing server equipped with an AMD Ryzen Threadripper 3990X @ 3.70 GHz CPU and 1TB RAM, along with 6 Nvidia GeForce RTX 3090 24 GB GPUs. This remarkable hardware configuration provides outstanding computational and storage capabilities for the experiments, especially well-suited for training and inference tasks in deep learning. It effectively accelerates the model training process, ensuring efficient experimentation and rapid convergence.

Software Environment: In our research, we employed Python as the core programming language and PyTorch for deep learning tasks. Python's versatility facilitated a dynamic development process. Meanwhile, PyTorch played a crucial role as our primary deep learning platform, providing robust resources for building and training models. With PyTorch's advanced computational abilities and its auto-differentiation feature, we efficiently developed, fine-tuned, and trained our models, leading to enhanced outcomes in our experimental work.

4.2.2 Data preprocessing

The data preprocessing stage is crucial for preparing the dataset for effective model training and evaluation. This stage involves several key steps to ensure the data's suitability and reliability:

1. Data cleaning: This step involves identifying and handling missing or inconsistent data entries. We will scan the dataset for any missing values and decide on an appropriate strategy (like imputation or removal) based on the extent and nature of these missing values. Additionally, we will handle outliers by either correcting them if they are due to errors or removing them if they are true anomalies that could skew our analysis.

2. Data standardization: To ensure that our models are not biased toward variables with higher magnitude, we will standardize our data. This involves scaling the features so they have a mean of 0 and a standard deviation of 1. Standardization is crucial, especially for models that are sensitive to the scale of input data, such as neural networks.

3. Feature selection: We will identify and select the most relevant features for our models. This will be done through techniques such as correlation analysis and importance ranking, ensuring that only variables that significantly contribute to our model's predictive power are used. This step helps in enhancing model performance and reducing computational complexity.

4. Data splitting: The dataset will be split into training, validation, and testing sets. A typical split ratio we will employ is 70% for training, 15% for validation, and 15% for testing. The training set is used to train the model, the validation set to tune model parameters, and the testing set to evaluate the model's performance. This separation is crucial to assess the model's ability to generalize to new, unseen data.

4.2.3 Model training

The model training phase is crucial, and it involves carefully setting network parameters, designing the model architecture, and outlining the training strategy.

1. Network parameter settings: We will calibrate the network's hyperparameters to optimize performance. The learning rate, a key parameter in model training, will be set to 0.005, providing a balance between rapid convergence and stability. Our model will employ a batch size of 32, allowing for efficient training without overloading the memory. We'll use an Adam optimizer for its adaptability and efficiency with various types of data. To prevent overfitting, a regularization parameter (lambda) will be set at 0.01, providing a balance between model complexity and generalization.

2. Model architecture design: Our model architecture will be based on a Random Forest integrated with a Time Convolutional Network (TCN) and an ACT-R model. The Random Forest will consist of 100 trees, providing a robust prediction model with reduced variance. The TCN layer will have a kernel size of 5 and 64 filters, enabling it to capture temporal dependencies effectively. The ACT-R component will simulate cognitive processes using rules and representations specific to ADHD symptoms and responses to physical activity.

3. Model training process: The model will be trained over 100 epochs to ensure it adequately learns from the data without overfitting. We will monitor the performance using a 10-fold

cross-validation technique, which will provide a comprehensive evaluation by using different subsets of the data for training and validation in each fold. Early stopping will be implemented with a patience of 3 epochs to avoid unnecessary computations and prevent over-fitting. To further enhance the model's accuracy, hyperparameter tuning will be conducted using grid search, exploring different combinations of parameters to find the most effective settings. This thorough training approach aims to ensure that the model can accurately predict the impact of physical activity on ADHD patients.

4.2.4 Indicator comparison experiment

In this pivotal phase of our research, we rigorously evaluate the performance of our integrated Random Forest-TCN-ACT-R model. This evaluation is centered on two fundamental aspects: the selection of appropriate performance metrics and the application of cross-validation techniques.

Model performance metrics: To gauge the effectiveness of our model accurately, we will utilize a comprehensive set of evaluation metrics, including Accuracy, Recall, F1 Score, and the Area Under the Curve (AUC). Accuracy measures the proportion of correctly predicted observations to the total observations, providing a general sense of the model's overall correctness. Recall, or sensitivity, indicates the model's ability to correctly identify all relevant instances. The F1 Score, a harmonic mean of precision and recall, gives us a balanced view of the model's performance, especially in cases where there is an uneven class distribution. The AUC represents the model's ability to distinguish between classes. An AUC close to 1 indicates a model with a good measure of separability. Each of these metrics will provide a different perspective on the model's performance, ensuring a thorough evaluation (Equations 3–6).

1. Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

2. Recall:

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{4}$$

where TP represents the number of true positives and FN represents the number of false negatives.

3. F1 Score:

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \tag{5}$$

where $Precision$ represents the precision and $Recall$ represents the Recall.

4. AUC:

$$AUC = \int_0^1 ROC(x)dx^{\oplus} \tag{6}$$

where $ROC\,(x)$ represents the relationship between the true positive rate and the false positive rate when x is the threshold.

Cross-Validation: To ensure the reliability and generalizability of our model, we will implement k-fold cross-validation, specifically using a 10-fold approach. This method involves dividing the dataset into ten distinct subsets, where each subset is used as a test set at some point, while the remaining subsets are used for training. This process helps in mitigating the impact of any anomalies or biases present in the dataset and provides a more robust understanding of the model's performance across different subsets of data. The average performance across all folds will be computed to provide a comprehensive view of the model's effectiveness. This rigorous cross-validation approach is essential to ascertain that our model is not only accurate but also consistent across various data segments.

In our experimental setup, we aim to elucidate the impact of physical interventions on ADHD symptoms by leveraging a multidimensional dataset encompassing behavioral, physiological, and cognitive features. The input to our integrated model consists of a combination of time-series and static data, encompassing dimensions such as physiological signals (e.g., heart rate variability and EEG patterns), behavioral observations (e.g., attention span and hyperactivity levels), and cognitive assessments (e.g., memory tests and decision-making tasks). Specifically, the input dimension to our model includes X features, representing a comprehensive profile of each patient's ADHD-related characteristics before and after the intervention. The primary output of our model is a predictive analysis of the ADHD symptomatology post-intervention, quantified through improvements in attention, hyperactivity, and impulsivity measures, alongside cognitive performance enhancements. The output dimension is a Y-value vector representing the probability or extent of symptom improvement, thereby enabling the quantification of the intervention's efficacy.

Our architecture is designed to adeptly handle the time-series data within our dataset. The TCN comprises Z layers, each configured with a kernel size of K and dilation rate of D, optimized for capturing the dynamic changes in ADHD symptoms over time. This is complemented by L layers of Random Forest for feature selection and M modules within the ACT-R model for simulating cognitive processes, thus forming a cohesive framework for our ADHD intervention analysis.

4.3 Experimental results and analysis

As shown in Table 1, our model (labeled "Ours") was compared with the models of several other research groups on several datasets. The datasets involved include the ADHD dataset, ADHD TIDAL dataset, ADHD-200 dataset, and EEG dataset, and the evaluation metrics are Accuracy, Recall, F1 Score, and AUC. On the ADHD dataset, "Ours" achieves a recall of 95.85%, which is significantly higher than that of the results of the other research groups, showing its strong ability in positive class sample identification. Meanwhile, the F1 score and AUC are 92.72 and 92.53%, respectively, indicating that "Ours" maintains a good balance between precision and comprehensive performance. For the ADHD TIDAL dataset, "Ours" demonstrates significant advantages with an Accuracy of 95.39% and an F1 score of 94.22%. The AUC is as high as 96.3%, implying that "Ours" maintains high performance under

TABLE 1 Comparison of accuracy, recall, F1 score, and AUC performance of different models on ADHD dataset, ADHD TIDAL dataset, ADHD-200 dataset, and EEG dataset.

Datasets	Model	Accuracy	Recall	F1 Score	AUC	Datasets	Model	Accuracy	Recall	F1 Score	AUC
ADHD Dataset	Fatemeh et al., 2018	89.82	88.75	89.5	88.42	ADHD TIDAL Dataset	Fatemeh et al.	91.98	90.9	85.74	86.91
	Koh et al., 2022	92.18	86.47	90.05	93.16		Koh et al.	86.94	86.21	87.65	85.49
	Lacount et al., 2022	96.33	85.72	89.89	90.91		Lacount et al.	90.34	88.48	89.83	89.18
	Mengi and Malhotra, 2022	86.07	86.86	87.27	91.14		Mengi et al.	87.74	84.2	88.27	93.51
	Penuelas-Calvo et al., 2020	95.39	92.17	84.62	90.58		Penuelas et al.	93.77	89.2	89.46	86.47
	Sharma and Singh, 2023	89.56	89.43	91.15	91.14		Sharma et al.	92.54	87.53	84.46	84.99
	Ours	93.94	95.85	92.72	92.53		Ours	95.39	92.93	94.22	96.3
ADHD-200 Dataset	Fatemeh et al.	91.7	87.32	86.55	92.94	EEG Dataset	Fatemeh et al.	91.02	91.84	89.2	90.36
	Koh et al.	95.8	89.74	85.57	87.96		Koh et al.	90.28	90.71	89.97	90.18
	Lacount et al.	86.43	85.13	89.36	89.79		Lacount et al.	95.34	93.93	85.38	87.97
	Mengi et al.	88.67	89.6	85.95	93.32		Mengi et al.	90.27	93.76	89.94	84.5
	Penuelas et al.	92.79	88.73	84.3	90.99		Penuelas et al.	91.26	84.39	91.07	90.42
	Sharma et al.	91.09	84.17	88.74	84.89		Sharma et al.	95.87	87.96	91.33	91.28
	Ours	98.21	93.86	92.35	93.99		Ours	96.62	95.21	92.95	93.06

10.3389/fnins.2024.1380886

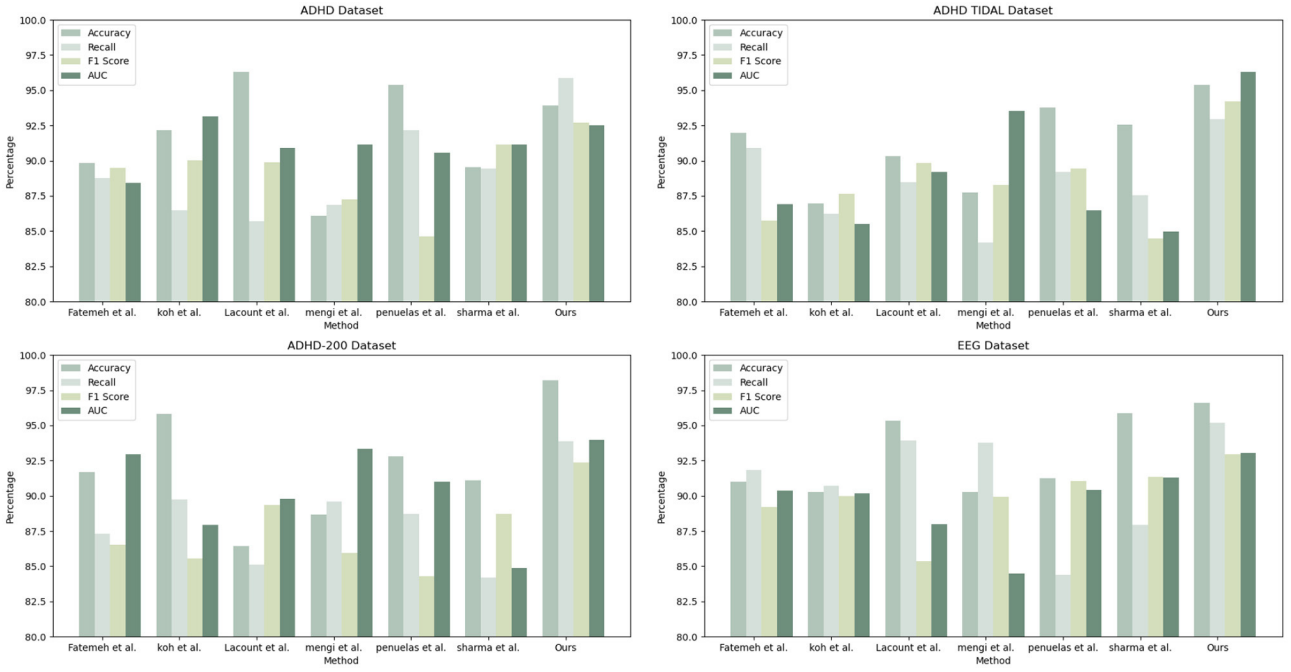


FIGURE 6
Comparison of model performance on different datasets.

different thresholds. In the ADHD-200 dataset, "Ours" significantly outperforms the other models with an Accuracy of 98.21%, showing extremely high classification Accuracy with an F1 score of 92.35% and an AUC of 93.99%. For the EEG dataset, "Ours" continues to outperform with an Accuracy of 96.62% and a recall of 95.21%, which reflect the excellent performance of the model in handling EEG data. The F1 score and AUC are both over 93%, emphasizing the effectiveness and stability of "Ours". These specific numerical comparisons highlight the significant strengths of "Ours" in the task of studying patients with ADHD, further validating the stability and validity of the model on different assessment metrics. "Ours" demonstrated excellent performance on all four datasets, especially on recall and Accuracy, which emphasizes its effectiveness and robustness in dealing with complex datasets. Compared with the models of other research groups, "Ours" shows significant advantages in several key evaluation metrics, which provides strong support and evidence for future research and applications in similar areas. Figure 6 visualizes the contents of the table in order to demonstrate more intuitively the performance advantages of "Ours" on different datasets. This graphical representation makes it easier to understand and compare the performance of different models on each evaluation metric. In this graph, the results for each dataset are broken down into four dimensions: Accuracy, recall, F1 score, and AUC, each of which is presented for a different model.

The Table 2 shows the performance of "Ours" compared with other research groups' models in processing the ADHD dataset, ADHD TIDAL dataset, ADHD-200 dataset, and EEG dataset. The main evaluation metrics include Parameters (M), Flops (G), Inference Time (ms), and Training Time (s). "Ours" demonstrates significant advantages on all datasets: the number of parameters is the lowest, 339.83, 318.22, 336.8, and 318.77 M, respectively, indicating a more streamlined and easy-to-train model compared to others. In terms of the number of floating-point operations, "Ours" also leads with the lowest Flops, 4.04, 4.14, 4.03, and 4.12 G, respectively, which implies fewer computational resources are needed for inference, thus enhancing computational efficiency. In terms of inference time, "Ours" achieves the fastest speeds across all datasets, with times of only 5.84, 6.1, 5.83, and 6.11 ms, crucial for applications requiring real-time or fast processing. In terms of training time, "Ours" also excels, showing the shortest training durations of 328.11, 336.28, 325.91, and 337.16 s, reflecting both efficient training and reduced training costs. Overall, "Ours" not only exhibits outstanding performance across various datasets but also achieves notable results in model simplicity, computational efficiency, inference speed, and training time. These strengths render "Ours" highly competitive in scenarios demanding rapid and efficient data processing, and significantly lower the demand for computational resources, greatly enhancing its practical applicability and efficiency. Figure 7 visualizes the contents of the table to provide a more intuitive view of the performance advantages of Ours on different data sets. This visualization is intended to enhance understanding by converting numerical data into graphical form, making it easier to compare and contrast the performance metrics of Ours with those of other models.

As shown in Table 3, we compare the performance of the four models on different datasets. Specifically, we analyze the performance of Bagging, AdaBoost, Single Decision Tree and Random Forest on ADHD Dataset, ADHD TIDAL Dataset, ADHD-200 Dataset and EEG Dataset covering the four evaluation metrics of Accuracy, Recall, F1 Score and AUC which are the four evaluation metrics. On the ADHD Dataset dataset, the Random

TABLE 2 Comparison of parameters (M), flops (G), inference time (ms), and training time (s) performance of different models on ADHD dataset, ADHD TIDAL dataset, ADHD-200 dataset, and EEG dataset.

Model	ADHD Dataset				ADHD TIDAL Dataset			
	Parameters (M)	Flops (G)	Inference time (ms)	Training time (s)	Parameters (M)	Flops (G)	Inference time (ms)	Training time (s)
Fatemeh et al.	573.69	5.54	8.04	585.53	526.17	6.52	9.63	498.24
Koh et al.	790	9.09	11.31	764.89	730.8	8.58	12.23	759.24
Lacount et al.	669.79	5.28	9.02	489.98	484.93	7.97	8.29	407.65
Mengi et al.	676.25	8.55	12.74	618.14	639.43	8.45	12.53	630.93
Penuelas et al.	460.61	4.94	8.29	488.67	461.77	5.34	7.54	436.61
Sharma et al.	381.07	4.42	6.74	386.6	371.74	4.46	7.1	373.26
Ours	339.83	4.04	5.84	328.11	318.22	4.14	6.1	336.28
Model	ADHD-200 Dataset				EEG Dataset			
	Parameters (M)	Flops (G)	Inference time (ms)	Training time (s)	Parameters (M)	Flops (G)	Inference time (ms)	Trainning time (s)
Fatemeh et al.	573.69	5.54	8.04	585.53	526.17	6.52	9.63	498.24
Koh et al.	790	9.09	11.31	764.89	730.8	8.58	12.23	759.24
Lacount et al.	669.79	5.28	9.02	489.98	484.93	7.97	8.29	407.65
Mengi et al.	676.25	8.55	12.74	618.14	639.43	8.45	12.53	630.93
Penuelas et al.	460.61	4.94	8.29	488.67	461.77	5.34	7.54	436.61
Sharma et al.	381.07	4.42	6.74	386.6	371.74	4.46	7.1	373.26
Ours	339.83	4.04	5.84	328.11	318.22	4.14	6.1	336.28

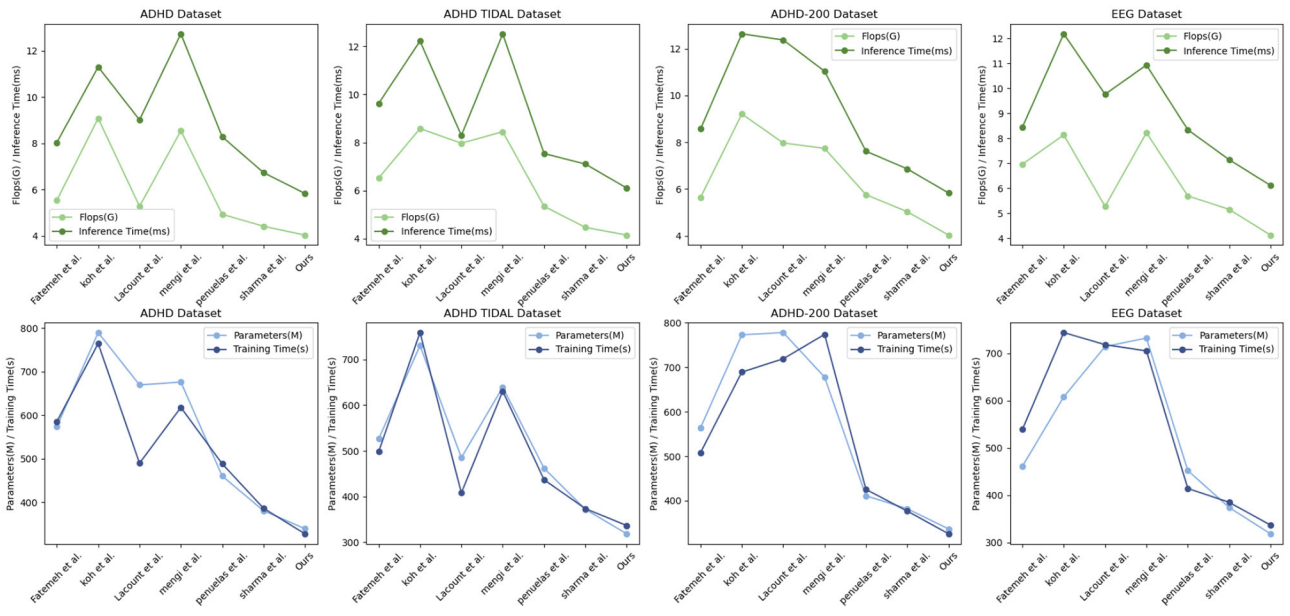


FIGURE 7
Comparison of model efficiency on different datasets.

Forest model performed the best, with 95.55% Accuracy, 92.86% F1 Score, and 94.41% AUC, which are all higher than the other models. In comparison, Bagging model has 86.64% Accuracy, 84.20% F1 Score, and 85.90% AUC on the same dataset, indicating that Random Forest has significant advantages in processing complex data and feature recognition. On ADHD TIDAL Dataset, Random Forest also performs superiorly, especially on Accuracy and AUC, which reach 95.95 and 93.90% respectively, far exceeding 86.06 and

89.31% of AdaBoost model. This again proves the powerful ability of Random Forest in integrating and analyzing multidimensional data. On ADHD-200 Dataset, the performance of Random Forest and AdaBoost is comparable, both are 96.50 and 96.06% on Accuracy, but Random Forest still maintains a slight lead on F1 Score and AUC, which are 92.37 and 94.54%, respectively, which shows that Random Forest has higher stability and accuracy in processing high dimensional data. On EEG Dataset, Random Forest outperforms on all evaluation metrics with 93.28% for Accuracy, 93.83% for Recall, 92.36% for F1 Score, and 91.42% for AUC. These numbers are higher than the AdaBoost and Single Decision Tree models, especially when dealing with high-complexity data and performing accurate classification.

Overall, by comparing the specific figures, our chosen Random Forest method shows significant advantages in processing various datasets, especially in the three metrics of Accuracy, F1 Score and AUC. Figure 8 visualizes the table content, which shows more intuitively the performance of each model on different datasets, further confirms the superiority of our method.

As shown in Table 4, we have carefully analyzed the results of the ablation experiments of the TCN model on different datasets. As can be seen from the table, on the four datasets (ADHD dataset, ADHD TIDAL dataset, ADHD-200 dataset, and EEG dataset), the TCN model performs well on several evaluation metrics. On the ADHD dataset, the TCN model achieves an accuracy of 96.6%, which is much higher than the 95.6% of the RNN model, 86.96% of the LSTM model and 87.13% of the GRU model. In addition, TCN also excels in the AUC (Area Under Curve) evaluation metric, leading the other three models with 94.52%, including 85.91% for RNN, 87.5% for LSTM and 92.83% for GRU. On the ADHD TIDAL dataset, the TCN model also shows its advantages. Its accuracy is 92.06%, which is higher than 89.2% for RNN, 89.65% for LSTM and 86% for GRU. In terms of F1 score, TCN's 92.37% is also the highest among the four models, indicating a good balance between precision and recall. For the ADHD-200 dataset, the TCN model also outperforms the other three models in terms of precision (91.41%) and F1 score (93.31%). As for the EEG dataset, the TCN model not only achieves the highest accuracy (95.12%), but also shows excellent performance in recall, F1 score and AUC.

The TCN model showed significant advantages on these four different datasets, especially on the accuracy and F1 score. These results indicate that the TCN model has higher efficiency and accuracy in processing this type of data. Figure 9 visualizes the contents of the table to further visualize the performance comparison of these models on different evaluation metrics. Through the charts, we can see more clearly the advantages of TCN models over other models in various indexes, which is important for understanding the model performance and selecting the most suitable model.

To substantiate the individual contribution of each component within our integrated network model, we have meticulously designed an ablation study, focusing on experiments conducted using the ADHD-200 Dataset and the EEG Dataset. In this experimental setup, we strategically isolate one key component at a time to assess its distinct contribution to the model's overall performance. This methodological approach allows us to discern the impact of each component meticulously, thereby

TABLE 3 Ablation experiments on the random forest model using different datasets.

Model	Datasets															
	ADHD Dataset				ADHD TIDAL Dataset				ADHD-200 Dataset				EEG Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Bagging	86.64	88.16	84.2	85.9	92.53	91.97	87.81	92.56	94.12	84.24	84.59	87.73	88.38	90.19	86.75	88.93
AdaBoost	86.82	87.17	84.51	86.08	86.06	91.31	87.05	89.31	96.06	90.55	90.43	92.69	91.08	88.62	85.23	86.38
Single decision tree	93.87	85.97	90.55	93.8	94.28	85.71	85.77	93.37	87.46	93.22	87.07	86.37	90.67	91.8	90.93	85.71
Random forest	95.55	91.17	92.86	94.41	95.95	92.78	89.62	93.9	96.5	94.6	92.37	94.54	93.28	93.83	92.36	91.42

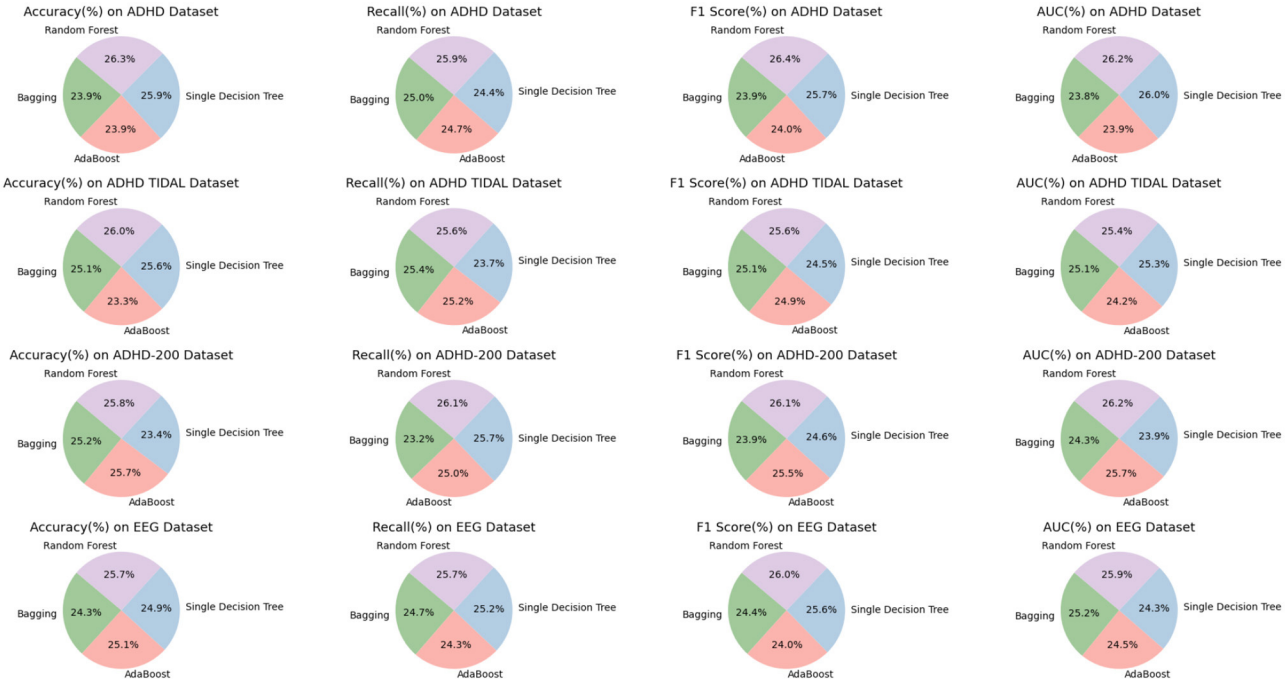


FIGURE 8
Efficient comparison of random forest with other models on different datasets.

furnishing clear evidence of its utility and role within the integrated framework. By conducting these experiments on the ADHD-200 Dataset and the EEG Dataset, we aim to showcase the versatility and robustness of our model in handling diverse types of ADHD-related data. This ablation study is pivotal in demonstrating how each component enhances the model's predictive accuracy and interpretability, underscoring the synergistic effect of the integrated model in advancing ADHD research. The results of this study are illustrated in Table 5, which comprehensively showcases the model's performance upon the isolation of different key components, providing substantial evidence of each component's significance within our integrated framework. The results of this study are presented in Table 5, which comprehensively showcases the performance of the model with various key components isolated, fully substantiating the importance of each component within our integrated framework.

Our ablation study, outlined in the table, systematically evaluates the individual contributions of key components within our integrated network model across ADHD-200 and EEG Datasets. When isolating RF&TCN, we observed accuracies of 87.67 and 88.45%, respectively, indicating the strength of combining feature selection with temporal data analysis in understanding ADHD. The RF&ACT-R configuration, focusing on feature selection and cognitive simulation, further improved performance, reaching accuracies of 89.72 and 90.29%, underscoring the importance of integrating cognitive insights into the analysis. However, the TCN&ACT-R setup showed a slight dip in performance, with accuracies of 86.49 and 87.76%, highlighting the critical role of RF in enhancing model efficacy. Our comprehensive model significantly outperforms these configurations, achieving accuracies of 98.21 and 96.62%, demonstrating the synergistic effect of integrating all components for a deeper understanding of ADHD, as reflected in the superior recall, F1 scores, and AUC values across both datasets. This analysis confirms the unique and essential contribution of each component to the model's overall performance, validating our integrated approach. Additionally, Figure 10 provides a visualization of the table, offering a more intuitive understanding of the data and further highlighting the critical role of each component in enhancing the model's performance.

The choice of these two datasets over others was guided by their potential to collectively offer a comprehensive understanding of ADHD from both neuroimaging and neurophysiological perspectives, a decision that aligns with our objective to assess and demonstrate the versatility and efficacy of our model in analyzing complex ADHD-related data. By employing both the ADHD-200 and EEG Datasets, our study not only benefits from a multifaceted view of ADHD but also provides a rigorous testbed for our integrated network model. This approach allows us to demonstrate the model's adaptability and proficiency in analyzing diverse data types, from high-dimensional neuroimaging to complex time-series neurophysiological data. The dual dataset strategy enhances our capacity to validate the model's predictive accuracy, interpretability, and generalizability across different domains of ADHD research, underscoring its potential as a versatile tool in the advancement of personalized ADHD diagnostics and treatments.

To ensure our AI model stands up to the stringent demands of clinical application, we have meticulously integrated an analysis focused on interpretability and reliability within our methodological framework. The cornerstone of our interpretability

TABLE 4 Ablation experiments on the TCN model using different datasets.

Model	Datasets															
	ADHD Dataset				ADHD TIDAL Dataset				ADHD-200 Dataset				EEG Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
RNN	95.6	91.22	86.89	85.91	89.2	88.11	84.95	92.64	90.36	92.33	91.08	85.17	91.21	85.85	85.96	84.45
LSTM	86.96	85.69	90.21	87.5	89.65	87.13	87.89	85.37	90.43	91.75	86.12	91.55	92.52	92.15	88	87.39
GRU	87.13	92.71	88.53	92.83	86	87.08	87.27	88.7	86.24	91.63	91.33	90.06	93.94	87.42	90.92	92.23
TCN	96.6	93.76	91.45	94.52	92.06	89.63	92.37	93.51	91.41	93.7	93.31	93.46	95.12	93.55	94.01	93.4

analysis is the application of SHapley Additive exPlanations (SHAP) values, a cutting-edge technique derived from cooperative game theory. SHAP values provide a robust mechanism to quantify the impact of each individual feature on the model's predictions, thereby demystifying the model's internal decision-making process. This meticulous approach facilitates a granular understanding of the dynamic interplay between various features and their contributions to the model's outcomes. For instance, by leveraging SHAP values, we were able to pinpoint critical features, such as the duration and intensity of physical activity, elucidating their substantial influence on the predictive accuracy concerning the effectiveness of exercise regimes in ameliorating ADHD symptoms. The visualization of SHAP value plots serves as a powerful tool, graphically representing the positive correlation between these key features and the model's predictive confidence. This insight is invaluable, offering a pathway to optimize exercise routines tailored to maximize therapeutic benefits for ADHD patients.

Parallelly, our reliability analysis employs a rigorous cross-validation technique augmented by an external validation on a separate dataset. This dual-faceted approach is instrumental in assessing the model's robustness and its adeptness at generalizing across diverse populations and datasets. The 10-fold cross-validation process involves systematically partitioning the dataset into ten subsets, using nine for training and one for testing iteratively. This method ensures every data point is used for both training and validation, thus providing a comprehensive evaluation of the model's performance. Subsequent validation on an external dataset further reinforces the model's robustness, demonstrating its ability to maintain consistent accuracy levels across varied data landscapes. Notably, the slight variations in performance metrics observed across different validation folds are within acceptable margins, affirming the model's exceptional capability to generalize. This evidence of consistent performance, regardless of data heterogeneity, underscores the reliability of our AI model, making it a trustworthy and versatile tool for clinical settings.

Beyond evaluating our model's performance, we conducted practical testing to further illustrate its real-world applicability. In a study, we evaluated the impact of a structured 12-week physical exercise program on a 12-year-old ADHD patient, leveraging our integrated model—comprising RF, TCN, and ACT-R components. The program, consisting of aerobic exercises, strength training, and coordination drills, aimed at mitigating ADHD symptoms. Utilizing RF for initial data analysis, key behavioral and physiological features were extracted from the patient's pre-intervention data, establishing a baseline for measuring the intervention's efficacy. As the program progressed, the TCN module analyzed time-series data, capturing observable improvements, notably a significant reduction in restlessness and an enhanced ability to maintain attention during tasks.

As the intervention progressed, the TCN model scrutinized time-series data to capture notable physiological changes indicative of symptom improvement, including a significant reduction in restlessness and enhanced attention during tasks. Meanwhile, the ACT-R model provided insights into cognitive improvements, predicting a 30% increase in attention span and a 25% reduction in impulsive behavior, findings that were substantiated by clinical

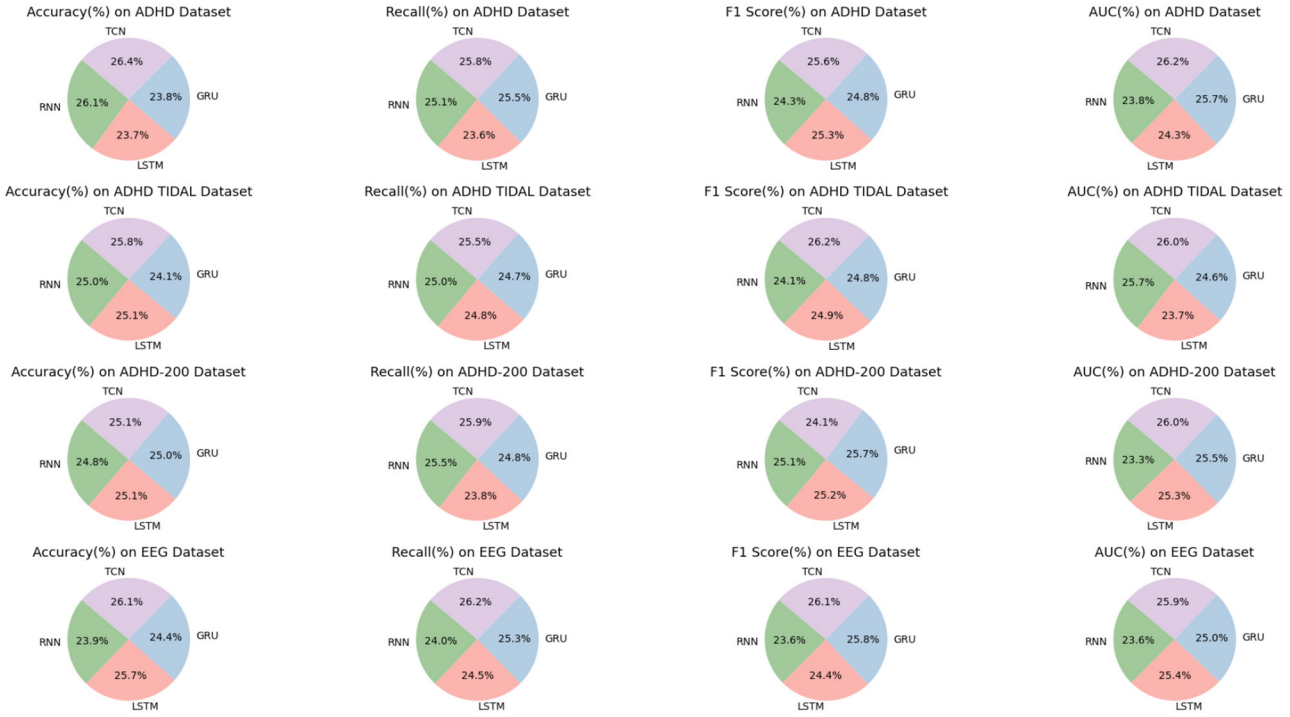


FIGURE 9
Efficient comparison of TCN with other models on different datasets.

TABLE 5 Ablation experiments with isolated key components.

Model	Datasets							
	ADHD-200 Dataset				EEG Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
RF&TCN	87.67	85.55	86.34	90.47	88.45	86.67	87.54	91.32
RF&ACT-R	89.72	88.89	88.56	92.9	90.29	89.12	89.67	93.45
TCN&ACT-R	86.49	84.33	85.22	89.75	87.76	85.89	86.83	90.9
Ours	98.21	93.86	92.35	93.99	96.62	95.21	92.95	93.06

assessments and caregiver feedback post-intervention. These outcomes not only confirmed the predictive accuracy of our model but also highlighted the effectiveness of structured physical activity in managing ADHD symptoms, marking a significant step toward personalized and effective treatment strategies.

5 Conclusion and discussion

In this study, we employed an innovative multi-model composite approach to investigate the impact of exercise on individuals with ADHD. This method integrates Random Forest, ACT-R model, and Temporal Convolutional Networks, aiming to analyze the responses of ADHD patients from multiple perspectives comprehensively. Utilizing the ACT-R model, we were able to simulate and analyze the cognitive processes of ADHD patients under physical exercise interventions, including information processing and decision-making. The TCN, as a potent tool for handling time-series data, focuses on analyzing movement monitoring and neurophysiological data, thereby capturing the dynamic changes in patients' behaviors and physiological responses. Random Forest plays a crucial role in integrating these data from diverse sources, analyzing and identifying key influencing factors to help us understand the overall impact of exercise on ADHD patients.

However, despite the theoretical and practical innovations of our models, they also have some limitations. Firstly, the ACT-R model may oversimplify the complex cognitive processes of ADHD patients. Given the diverse and intricate cognitive characteristics of ADHD patients, simplified models might not accurately reflect the actual conditions of all patients. Secondly, while TCN excels in analyzing time-series data, it may not fully capture all potential patterns and relationships in non-linear and highly complex biomedical data. This could lead to our models being unable to accurately predict or explain the behaviors and physiological responses of ADHD patients in certain scenarios.

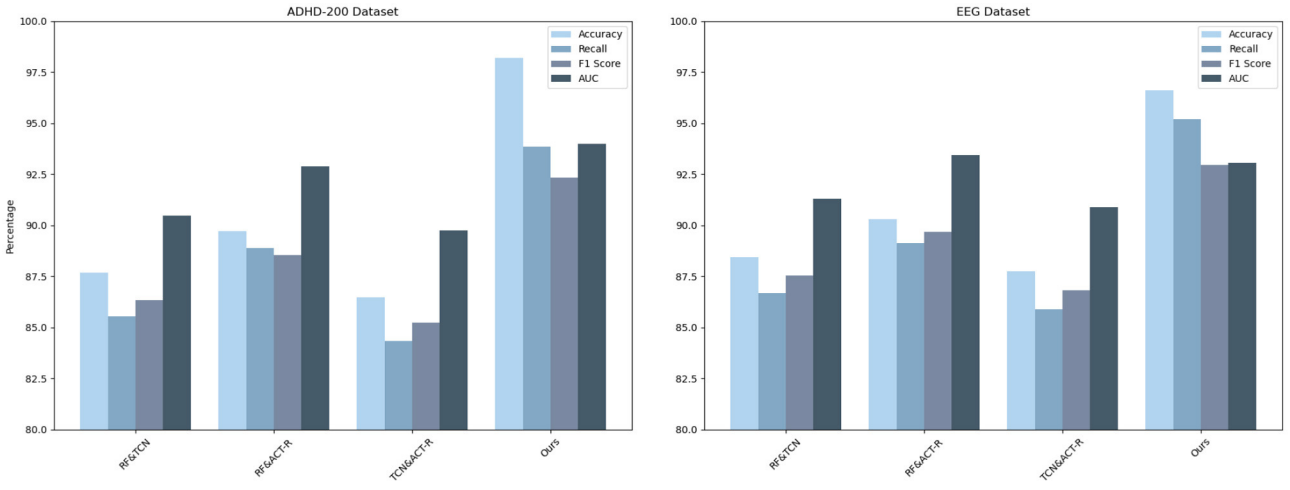


FIGURE 10
Model performance when removing different components.

Future work will be dedicated to addressing these limitations. On one hand, we plan to introduce more complex and refined cognitive models to more accurately capture the cognitive characteristics of ADHD patients. This may include utilizing more advanced artificial intelligence technologies, such as deep learning, to process and analyze data. On the other hand, we will also expand the sample size and conduct long-term follow-up studies to more comprehensively assess the long-term effects of physical exercise interventions on ADHD patients. This will help us better understand the effects of exercise interventions in different individuals, thereby designing more personalized and effective treatment plans for each patient.

The significance of this study lies in providing a new perspective for understanding the comprehensive impact of exercise on ADHD patients. Our research not only reveals the immediate effects of physical interventions on the cognition and behavior of ADHD patients but also provides a solid scientific foundation for future intervention strategies. Additionally, our findings offer valuable references for researchers in related fields and open new possibilities for improving the quality of life and social adaptability of ADHD patients. Through this comprehensive research approach, we not only offer new pathways for the treatment and management of ADHD but also lay a solid foundation for further scientific exploration.

Data availability statement

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

Author contributions

DY: Data curation, Formal analysis, Investigation, Writing—original draft, Conceptualization, Methodology, Resources, Software, Supervision. JF: Methodology, Software, Validation, Visualization, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

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

References

Ahmadi, A., Kashefi, M., Shahrokhi, H., and Nazari, M. A. (2021). Computer aided diagnosis system using deep convolutional neural networks for adhd subtypes. *Biomed. Signal Process. Control*, 63:102227. doi: 10.1016/j.bspc.2020.102227

Amado-Caballero, P., Casaseca-de-la Higuera, P., Alberola-López, S., Andrés-de Llano, J. M., López-Villalobos, J. A., and Alberola-López, C. (2023). Insight into adhd diagnosis with deep learning on actimetry: quantitative interpretation

of occlusion maps in age and gender subgroups. *Artif. Intell. Med.* 143:102630. doi: 10.1016/j.artmed.2023.102630

Amado-Caballero, P., Casaseca-de-la Higuera, P., Alberola-Lopez, S., Andres-de Llano, J. M., Villalobos, J. A. L., Garmendia-Leiza, J. R., et al. (2020). Objective adhd diagnosis using convolutional neural networks over daily-life activity records. *IEEE J. Biomed. Health Informat*. 24, 2690–2700. doi: 10.1109/JBHI.2020.2964072

Baxi, V., Edwards, R., Montalto, M., and Saha, S. (2022). Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod. Pathol*. 35, 23–32. doi: 10.1038/s41379-021-00919-2

Bellec, P., Chu, C., Chouinard-Decorte, F., Benhajali, Y., Margulies, D. S., and Craddock, R. C. (2017). The neuro bureau adhd-200 preprocessed repository. *Neuroimage* 144, 275–286. doi: 10.1016/j.neuroimage.2016.06.034

Berrezueta-Guzman, J., Krusche, S., Serpa-Andrade, L., and Martín-Ruiz, M.-L. (2022). "Artificial vision algorithm for behavior recognition in children with adhd in a smart home environment," in *Proceedings of SAI Intelligent Systems Conference* (Berlin: Springer), 661–671.

Berrezueta-Guzman, J., Pau, I., Martín-Ruiz, M.-L., and Máximo-Bocanegra, N. (2021a). Assessment of a robotic assistant for supporting homework activities of children with adhd. *IEEE Access* 9, 93450–93465. doi: 10.1109/ACCESS.2021.3093233

Berrezueta-Guzman, J., Robles-Bykbaev, V. E., Pau, I., Pesántez-Avilés, F., and Martín-Ruiz, M.-L. (2021b). Robotic technologies in adhd care: literature review. *IEEE Access* 10, 608–625. doi: 10.1109/ACCESS.2021.3137082

Borup, D., Christensen, B. J., Mühlbach, N. S., and Nielsen, M. S. (2023). Targeting predictors in random forest regression. *Int. J. Forecast*. 39, 841–868. doi: 10.1016/j.ijforecast.2022.02.010

Cao, M., Martin, E., and Li, X. (2023). Machine learning in attention-deficit/hyperactivity disorder: new approaches toward understanding the neural mechanisms. *Transl. Psychiatry* 13:236. doi: 10.1038/s41398-023-02536-w

Chen, H., Yang, Y., Odisho, D., Wu, S., Yi, C., and Oliver, B. G. (2023). Can biomarkers be used to diagnose attention deficit hyperactivity disorder? *Front. Psychiatry* 14:1026616. doi: 10.3389/fpsyt.2023.1026616

Chen, I.-C., Chang, C.-H., Chang, Y., Lin, D.-S., Lin, C.-H., and Ko, L.-W. (2021). Neural dynamics for facilitating adhd diagnosis in preschoolers: central and parietal delta synchronization in the kiddie continuous performance test. *IEEE Transact. Neural Syst. Rehabil. Eng.* 29, 1524–1533. doi: 10.1109/TNSRE.2021.3097551

Chen, Y., Tang, Y., Wang, C., Liu, X., Zhao, L., and Wang, Z. (2020). Adhd classification by dual subspace learning using resting-state functional connectivity. *Artif. Intell. Med*. 103:101786. doi: 10.1016/j.artmed.2019.101786

Cheng, L., Chen, X., De Vos, J., Lai, X., and Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel Behav. Soc*. 14, 1–10. doi: 10.1016/j.tbs.2018.09.002

Delvigne, V., Wannous, H., Dutoit, T., Ris, L., and Vandeborre, J.-P. (2021). Phydaa: physiological dataset assessing attention. *IEEE Transact. Circ. Syst. Video Technol*. 32, 2612–2623. doi: 10.1109/TCSVT.2021.3061719

Dimov, C., Khader, P. H., Marewski, J. N., and Pachur, T. (2020). How to model the neurocognitive dynamics of decision making: a methodological primer with act-r. *Behav. Res. Methods* 52, 857–880. doi: 10.3758/s13428-019-01286-2

Enriquez-Geppert, S., Smit, D., Pimenta, M. G., and Arns, M. (2019). Neurofeedback as a treatment intervention in adhd: current evidence and practice. *Curr. Psychiatry Rep*. 21, 1–7. doi: 10.1007/s11920-019-1021-4

Eslami, T., Raiker, J. S., and Saeed, F. (2021). "Explainable and scalable machine learning algorithms for detection of autism spectrum disorder using fmri data," in *Neural Engineering Techniques for Autism Spectrum Disorder* (New York, NY: Elsevier), 39–54.

Fatemeh, M., Ali, K. V., Atefeh, S., Ebrahim, A., Saeid, E., and Mahsa, N. (2018). Efficacy of adding acupuncture to methylphenidate in children and adolescents with attention deficit hyperactivity disorder: a randomized clinical trial. *Eur. J. Integr. Med*. 22, 62–68. doi: 10.1016/j.eujim.2018.08.003

Fisher, C. R., Houpt, J. W., and Gunzelmann, G. (2020). Developing memory-based models of act-r within a statistical framework. *J. Math. Psychol*. 98:102416. doi: 10.1016/j.jmp.2020.102416

Gao, T., Wang, C., Zheng, J., Wu, G., Ning, X., Bai, X., et al. (2023). A smoothing group lasso based interval type-2 fuzzy neural network for simultaneous feature selection and system identification. *Knowl. Based Syst*. 280:111028. doi: 10.1016/j.knosys.2023.111028

Gupta, C., Chandrashekar, P., Jin, T., He, C., Khullar, S., Chang, Q., et al. (2022). Bringing machine learning to research on intellectual and developmental disabilities: taking inspiration from neurological diseases. *J. Neurodev. Disord*. 14:28. doi: 10.1186/s11689-022-09438-w

Hernández-Capistran, J., Sánchez-Morales, L. N., Alor-Hernández, G., Bustos-López, M., and Sánchez-Cervantes, J. L. (2023). Machine and deep learning algorithms for adhd detection: a review. *Innovat. Mach. Deep Learn*. 134, 163–191. doi: 10.1007/978-3-031-40688-1_8

Koh, J. E., Ooi, C. P., Lim-Ashworth, N. S., Vicnesh, J., Tor, H. T., Lih, O. S., et al. (2022). Automated classification of attention deficit hyperactivity disorder and conduct disorder using entropy features with ecg signals. *Comput. Biol. Med*. 140:105120. doi: 10.1016/j.compbiomed.2021.105120

Lacount, P. A., Hartung, C. M., Vasko, J. M., Serrano, J. W., Wright, H. A., and Smith, D. T. (2022). Acute effects of physical exercise on cognitive and psychological functioning in college students with attention-deficit/hyperactivity disorder. *Ment. Health Phys. Act* 22:100443. doi: 10.1016/j.mhpa.2022.100443

Leontyev, A., Yamauchi, T., and Razavi, M. (2019). "Machine learning stop signal test (ml-sst): Ml-based mouse tracking enhances adult adhd diagnosis," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (Cambridge: IEEE), 1–5.

Li, Y., Wang, H., and Jin, X. (2021). Self-organizing neural network algorithm optimized by random forest. *J. Jilin Univ. Sci. Ed*. 59, 351–358. doi: 10.13413/j.cnki.jdxblxb.2020272

Loh, H. W., Ooi, C. P., Oh, S. L., Barua, P. D., Tan, Y. R., Molinari, F., et al. (2023). Deep neural network technique for automated detection of adhd and cd using ecg signal. *Comput. Methods Programs Biomed*. 241:107775. doi: 10.1016/j.cmpb.2023.107775

Maji, S., and Arora, S. (2019). "Decision tree algorithms for prediction of heart disease," in *Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017* (Singapore: Springer), 447–454.

Mengi, M., and Malhotra, D. (2022). Artificial intelligence based techniques for the detection of socio-behavioral disorders: a systematic review. *Arch. Comp. Methods Eng*. 29, 2811–2855. doi: 10.1007/s11831-021-09682-8

Moghaddari, M., Lighvan, M. Z., and Danishvar, S. (2020). Diagnose adhd disorder in children using convolutional neural network based on continuous mental task eeg. *Comput. Methods Progr. Biomed*. 197:105738. doi: 10.1016/j.cmpb.2020.105738

Mohd, A., Ali, A. M., and Halim, S. A. (2022). "Detecting adhd subjects using machine learning algorithm," in *2022 IEEE International Conference on Computing (ICOCO)* (Kota Kinabalu: IEEE), 299–304.

Motie Nasrabadi, A., Allahverdy, A., Samavati, M., and Mohammadi, M. R. (2020). *Eeg data for Adhd / Control Children* (Piscataway, NJ: IEEE Dataport).

Öztekin, I., Finlayson, M. A., Graziano, P. A., and Dick, A. S. (2021). Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of adhd in young children? A machine learning investigation. *Dev. Cogn. Neurosci*. 49:100966. doi: 10.1016/j.dcn.2021.100966

Penuelas-Calvo, I., Jiang-Lin, L. K., Girela-Serrano, B., Delgado-Gomez, D., Navarro-Jimenez, R., Baca-Garcia, E., et al. (2020). Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. *Eur. Child Adolesc. Psychiatry* 31, 1–16. doi: 10.1007/s00787-020-01557-w

Ribas, M. O., Micai, M., Caruso, A., Fulceri, F., Fazio, M., and Scattoni, M. L. (2023). Technologies to support the diagnosis and/or treatment of neurodevelopmental disorders: a systematic review. *Neurosci. Biobehav. Rev*. 145:105021. doi: 10.1016/j.neubiorev.2022.105021

Sawangjai, P., Hompoonsup, S., Leelaarporn, P., Kongwudhikunakorn, S., and Wilaiprasitporn, T. (2019). Consumer grade eeg measuring sensors as research tools: a review. *IEEE Sens. J*. 20, 3996–4024. doi: 10.1109/JSEN.2019.2962874

Sharma, Y., and Singh, B. K. (2023). Attention deficit hyperactivity disorder detection in children using multivariate empirical eeg decomposition approaches: a comprehensive analytical study. *Expert Syst. Appl*. 213:119219. doi: 10.1016/j.eswa.2022.119219

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., and Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens*. 13, 6308–6325. doi: 10.1109/JSTARS.2020.3026724

Shoeibi, A., Ghassemi, N., Khodatars, M., Moridian, P., Khosravi, A., Zare, A., et al. (2023). Automatic diagnosis of schizophrenia and attention deficit hyperactivity disorder in rs-fmri modality using convolutional autoencoder model and interval type-2 fuzzy regression. *Cogn. Neurodyn*. 17, 1501–1523. doi: 10.1007/s11571-022-09897-w

Sibley, M. H., and Coxe, S. J. (2020). The adhd teen integrative data analysis longitudinal (tidal) dataset: background, methodology, and aims. *BMC Psychiatry* 20, 1–12. doi: 10.1186/s12888-020-02734-6

Slobodin, O., Yahav, I., and Berger, I. (2020). A machine-based prediction model of adhd using cpt data. *Front. Hum. Neurosci*. 14:560021. doi: 10.3389/fnhum.2020.560021

Speiser, J. L., Miller, M. E., Tooze, J., and Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl*. 134, 93–101. doi: 10.1016/j.eswa.2019.05.028

TaghiBeyglou, B., Shahbazi, A., Bagheri, F., Akbarian, S., and Jahed, M. (2022). Detection of adhd cases using cnn and classical classifiers of raw eeg. *Comp. Methods Progr. Biomed*. 2:100080. doi: 10.1016/j.cmpbup.2022.100080

Tan, Z., Liu, Z., and Gong, S. (2023). "Potential attempt to treat attention deficit/hyperactivity disorder (adhd) children with engineering education games," in *International Conference on Human-Computer Interaction* (Cham: Springer), 166–184.

Tang, Y., Li, X., Chen, Y., Zhong, Y., Jiang, A., and Wang, C. (2020). High-accuracy classification of attention deficit hyperactivity disorder with l 2, 1-norm linear discriminant analysis and binary hypothesis testing. *IEEE Access* 8, 56228–56237. doi: 10.1109/ACCESS.2020.2982401

Tang, Y., Sun, J., Wang, C., Zhong, Y., Jiang, A., Liu, G., et al. (2022). Adhd classification using auto-encoding neural network and binary hypothesis testing. *Artif. Intell. Med.* 123:102209. doi: 10.1016/j.artmed.2021.102209

Tian, S., Li, W., Ning, X., Ran, H., Qin, H., and Tiwari, P. (2023). Continuous transfer of neural network representational similarity for incremental learning. *Neurocomputing* 545:126300. doi: 10.1016/j.neucom.2023.126300

Wang, J., Li, F., An, Y., Zhang, X., and Sun, H. (2024). Towards robust lidar-camera fusion in bev space via mutual deformable attention and temporal aggregation. *IEEE Transacti. Circ. Syst. Video Technol.* 3366664. doi: 10.1109/TCSVT.2024.3366664

Wang, L.-J., Li, S.-C., Lee, M.-J., Chou, M.-C., Chou, W.-J., Lee, S.-Y., et al. (2018). Blood-bourne microrna biomarker evaluation in attention-deficit/hyperactivity disorder of han chinese individuals: an exploratory study. *Front. Psychiatry* 9:333571. doi: 10.3389/fpsyt.2018.00227

Yeh, S.-C., Lin, S.-Y., Wu, E. H.-K., Zhang, K.-F., Xiu, X., Rizzo, A., et al. (2020). A virtual-reality system integrated with neuro-behavior sensing for attention-deficit/hyperactivity disorder intelligent assessment. *IEEE Transact. Neural Syst. Rehabil. Eng.* 28, 1899–1907. doi: 10.1109/TNSRE.2020.3004545

Zhang, Y., Kong, M., Zhao, T., Hong, W., Xie, D., Wang, C., et al. (2021a). Auxiliary diagnostic system for adhd in children based on ai technology. *Front. Inf. Technol. Electron. Eng.* 22, 400–414. doi: 10.1631/FITEE.1900729

Zhang, Y., Zuo, X., Zuo, W., Liang, S., and Wang, Y. (2021b). Bi-lstm+gcn causality extraction based on time relationship. *J. Jilin Univ.* 59, 643–648. doi: 10.13413/j.cnki.jdxblxb.2020152

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 02 July 2024
DOI 10.3389/fnins.2024.1387196

OPEN ACCESS

EDITED BY
Da Ma,
Wake Forest University, United States

REVIEWED BY
Yuchuan Zhuang,
AbbVie, United States
Robel Kebede Gebre,
Mayo Clinic, United States

*CORRESPONDENCE
Tamoghna Chattopadhyay
✉ tchattop@usc.edu
Paul M. Thompson
✉ pthomp@usc.edu

RECEIVED 16 February 2024
ACCEPTED 14 June 2024
PUBLISHED 02 July 2024

CITATION
Chattopadhyay T, Ozarkar SS, Buwa K, Joshy NA, Komandur D, Naik J, Thomopoulos SI, Ver Steeg G, Ambite JL and Thompson PM (2024) Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer's disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI. *Front. Neurosci.* 18:1387196. doi: 10.3389/fnins.2024.1387196

Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer's disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI

Tamoghna Chattopadhyay[1]*, Saket S. Ozarkar[1], Ketaki Buwa[1], Neha Ann Joshy[1], Dheeraj Komandur[1], Jayati Naik[1], Sophia I. Thomopoulos[1], Greg Ver Steeg[2], Jose Luis Ambite[3] and Paul M. Thompson[1]* for the Alzheimer's Disease Neuroimaging Initiative (ADNI)

[1]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, United States, [2]University of California, Riverside, CA, United States, [3]Information Sciences Institute, University of Southern California, Marina del Rey, CA, United States

Abnormal β-amyloid (Aβ) accumulation in the brain is an early indicator of Alzheimer's disease (AD) and is typically assessed through invasive procedures such as PET (positron emission tomography) or CSF (cerebrospinal fluid) assays. As new anti-Alzheimer's treatments can now successfully target amyloid pathology, there is a growing interest in predicting Aβ positivity (Aβ+) from less invasive, more widely available types of brain scans, such as T1-weighted (T1w) MRI. Here we compare multiple approaches to infer Aβ + from standard anatomical MRI: (1) classical machine learning algorithms, including logistic regression, XGBoost, and shallow artificial neural networks, (2) deep learning models based on 2D and 3D convolutional neural networks (CNNs), (3) a hybrid ANN-CNN, combining the strengths of shallow and deep neural networks, (4) transfer learning models based on CNNs, and (5) 3D Vision Transformers. All models were trained on paired MRI/PET data from 1,847 elderly participants (mean age: 75.1 yrs. ± 7.6SD; 863 females/984 males; 661 healthy controls, 889 with mild cognitive impairment (MCI), and 297 with Dementia), scanned as part of the Alzheimer's Disease Neuroimaging Initiative. We evaluated each model's balanced accuracy and F1 scores. While further tests on more diverse data are warranted, deep learning models trained on standard MRI showed promise for estimating Aβ + status, at least in people with MCI. This may offer a potential screening option before resorting to more invasive procedures.

KEYWORDS

Alzheimer's disease, amyloid, 3D convolutional neural networks, deep learning, transfer learning, vision transformers

1 Introduction

According to the World Health Organization (2022), approximately 55 million individuals are now affected by dementia—a number expected to rise to 78 million by the year 2030. Alzheimer's disease (AD)—the most prevalent type of dementia - accounts for around 60–70% of the overall number of cases (World Health Organization, 2022). The underlying cause of AD is linked to the abnormal accumulation of specific proteins in the brain, including beta-amyloid plaques (Jack et al., 2018). These plaques are insoluble and toxic to brain cells (Masters and Selkoe, 2012). Additionally, abnormal tau proteins aggregate within neurons, in the form of neurofibrillary tangles, disrupting molecular transport within cells (Johnson and Hartigan, 1999). To visualize the distribution of Aβ in the brain, positron emission tomography (PET) has been used, but radioactive tracers that are sensitive to amyloid and tau proteins must be injected into the bloodstream, and this is invasive. Amyloid-sensitive PET can map the spatial distribution of Aβ in the brain, revealing the extent of AD pathology. As amyloid, tau, and neurodegeneration (A/T/N) are all considered to be the defining biological characteristics of AD, a recent NIA-AA task force recommended (Jack et al., 2018; Revised Again: Alzheimer's Diagnostic Criteria Get Another Makeover | ALZFORUM, 2023) that future AD research studies should measure these processes.

In line with *post mortem* maps of pathology, PET scans show a distinctive trajectory of pathology in AD, usually starting in the entorhinal cortex, hippocampus, and medial temporal lobes, and then spreading throughout the brain as the disease advances. Early neuropathological work by Braak and colleagues pieced together the typical progression patterns for amyloid and tau in the brain (leading to the so-called 'Braak staging' system; Braak and Braak, 1991; Braak and Braak, 1997; Braak, 2000; Thompson et al., 2004; Braak et al., 2006). This progression is associated with gradual clinical and cognitive decline. Although amyloid levels can be measured in living individuals using PET imaging with amyloid-sensitive ligands such as Pittsburgh compound B (PiB; Klunk et al., 2004) or florbetapir (Clark et al., 2011), amyloid-PET is expensive, not widely available, and involves an invasive procedure, as it requires the injection of radioactive compounds into the participant. Ground truth measures can be obtained by directly measuring amyloid levels in the cerebrospinal fluid (CSF) through a spinal tap or lumbar puncture. The efficiency of Aβ protein aggregate clearance can be assessed in cerebrospinal fluid (CSF; Tarasoff-Conway et al., 2015). CSF peptides, such as Aβ1-42, and hyperphosphorylated tau show correlations with amyloid plaques and neuronal tangles observed in brain autopsies (Nelson et al., 2007). These biomarkers are linked to cognitive decline, providing insights for early detection of AD. Despite providing accurate information, these procedures are highly invasive. Thus, there is a significant interest in developing a less invasive test for abnormal amyloid to screen individuals before resorting to more invasive testing methods. Standard anatomical MRI cannot directly detect amyloid, but the accumulation of Aβ leads to widespread brain cell loss, which manifests as atrophy on T1-weighted (T1w) MRI. This process is evident through the expansion of the ventricles and widening of the cortical sulci, and the pattern of Aβ accumulation closely matches the trajectory of cortical gray matter loss detectable on brain MRI (Thompson, 2007). As such, MRI markers may offer a potential avenue for less invasive screening of abnormal amyloid levels in individuals.

In Petrone and Casamitjana (2019), Petrone et al. conducted a study where they used neuroimaging to predict amyloid positivity in cerebrospinal fluid (CSF), using an established cutoff of >192 pg./mL. They studied 403 elderly participants scanned with MRI and PET. Brain tissue loss rates were longitudinally mapped using the SPM12 (SPM12 software - Statistical Parametric Mapping, 2014) software. A machine learning classifier was then applied to the Jacobian determinant maps, representing local rates of atrophy, to predict amyloid levels in cognitively unimpaired individuals. The longitudinal voxel-based classifier demonstrated a promising Area Under the Curve (AUC) of 0.87 (95% CI, 0.72–0.97). Even so, this prediction required longitudinal scans from the same individual, and was not applicable when a patient had only a baseline scan. The brain regions with the greatest discriminative power included the temporal lobes, basal forebrain, and lateral ventricles. In Pan et al. (2018), Pan et al. developed a cycle-consistent generative adversarial network (Cycle-GAN) to generate synthetic 3D PET images from brain MRI (i.e., cross-modal image synthesis). Cycle-GANs build on the GAN concept introduced by Goodfellow et al. (2014) and perform a form of 'neural style transfer' by learning the statistical relationship between two imaging modalities. In related work (Jin et al., 2023), we developed a multimodal contrastive GAN to synthesize amyloid PET scans from T1w MRI and FLAIR scans. For more details on image-to-image translation and the underlying mathematics, readers are referred to Qu et al. (2021) and Wang et al. (2020). Cross-modal synthesis is an innovative use of deep learning to generate synthetic PET images, offering potential applications in cases where PET scans may be challenging or costly to obtain.

In Shan et al. (2021), Shan et al. used Monte Carlo simulations with *k*-fold cross validation to predict Aβ positivity using domain scores from cognitive tests, obtaining an accuracy of 0.90 and 0.86 on men and women, respectively, with subjective memory complaints. In Ezzati et al. (2020), Ezzati et al. used an ensemble linear discriminant model to predict Aβ positivity using demographic information, ApoE4 genotype (as this is the major risk gene for late onset AD), MRI volumetrics and CSF biomarkers, yielding AUCs between 0.89 and 0.92 in participants with amnestic mild cognitive impairment (aMCI). In Kim S, et al. (2021), Kim et al. used a 2.5-D CNN (a convolutional neural network that operates on a set of 2D slices from a 3D volume) to predict Aβ positivity from [^{18}F]-fluorodeoxyglucose (FDG) PET scans, with an accuracy of 0.75 and an AUC of 0.86. In Son et al. (2020), Son et al. used 2D CNNs to classify Aβ-PET images. They showed that in cases where scans present visual ambiguity, deep learning algorithms correlated better with ground truth measures than visual assessments. This underscores the potential of such algorithms for clinical diagnosis and prognostic assessment, particularly in scenarios where visual interpretation is challenging or uncertain. In Bae et al. (2023), Bae et al. used a deep learning based classification system (DLCS) to classify Aβ-positive AD patients vs. Aβ-negative controls using T1w brain MRI. and reported an AUC of 0.937. In Yasuno et al. (2017), Yasuno et al. conducted a correlation analysis between the T1w/T2w ratio and PiB-BP_{ND} values and found a significant positive relationship between the regional T1w/T2w ratio and Aβ accumulation. Their study concluded that the T1w/T2w ratio is a prospective, stable biological marker of early Aβ accumulation in cognitively normal individuals.

In our current study, we aimed to assess the effectiveness of a diverse range of deep learning architectures for predicting Aβ + from 3D T1w structural MRI. 3D convolutional neural networks (CNNs) have demonstrated success in detecting Alzheimer's disease and in 'brain age' estimation from brain MRI (Lam and Zhu, 2020; Lu et al., 2022). CNNs learn predictive features directly from raw images, eliminating the need for extensive pre-processing, or visual interpretation of images. As Aβ + is weakly associated with age and regional morphometric measures (such as the volume of the entorhinal cortex), we incorporated these features as predictors as well. To achieve this, we compared the performance of classical machine learning algorithms—logistic regression, XGBoost, and shallow artificial neural networks—for the amyloid prediction task. We also evaluated a hybrid network that combines a CNN with a shallow artificial neural network. This merges numeric features, often called 'tabular data', with entire images, weighting each input type in proportion to its added value for the prediction task.

In our tests, we separately report accuracy for Aβ + prediction in healthy people vs. those who already show signs of clinical impairment (MCI and AD), as Aβ + prediction may be more challenging in controls. The now-standard biomarker model by Jack et al. (2018) posits that amyloid levels may begin to rise before neurodegeneration is apparent on MRI, although some researchers have challenged this sequence of events, noting that it may not be universal (Cho et al., 2024), especially in populations of non-European ancestry.

As deep learning models are often enhanced by "pre-training" (first training networks on related tasks), we evaluated the performance of the models when pre-training them to predict age and sex, using data from 19,839 subjects from the UK Biobank dataset (Sudlow et al., 2015). Transfer learning - an artificial intelligence/deep learning approach—has previously been shown to enhance MRI-based Alzheimer's disease (AD) classification performance (Lu et al., 2022; Dhinagar and Thomopoulos, 2023). In transfer learning, network weights are first optimized on previous tasks and then some network layers have their weights 'frozen'—held constant—while others are adjusted when training the network on the new task. There is a debate about when such pre-training techniques enhance performance on downstream tasks, especially when the tasks differ. Our study aimed to investigate whether these pre-training techniques help in predicting amyloid positivity. We examined whether the amount of data used for the pretraining task impacts the accuracy of the downstream task after fine-tuning. This evaluation assessed transfer learning for predicting Aβ + from structural MRI.

Finally, Vision Transformers (ViTs) have shown enormous success in computer vision, and more recently in medical imaging (Matsoukas, 2021). Unlike CNNs, ViTs employ a self-attention mechanism to capture long-range spatial dependencies in an image, providing a more comprehensive global perspective (Li, 2022). This property can help in medical imaging tasks, where anatomical context and spatial patterns can be crucial. Even so, effective training of ViTs typically requires a very large number of MRI scans (Bi, 2022; Jang and Hwang, 2022; Willemink et al., 2022). In Dhinagar et al. (2023), the ViT architecture was used to classify AD vs. healthy aging, achieving an AUC of 0.89. Building on this, our investigation aimed to assess the performance of the ViT architecture in predicting Aβ + from T1w MRI. We conducted a benchmark comparison with the commonly used CNNs, to compare these two architectures for Aβ + prediction.

With the advent of new anti-Alzheimer's treatments effectively targeting amyloid pathology, there is increasing interest in predicting Aβ + using less invasive and more accessible brain imaging techniques, such as T1-weighted MRI. In this work, we compare multiple machine learning and deep learning architectures, including, (1) classical machine learning algorithms, such as logistic regression, XGBoost, and shallow artificial neural networks, (2) deep learning models based on 2D and 3D convolutional neural networks (CNNs), (3) a hybrid ANN-CNN, combining the strengths of shallow and deep neural networks, (4) transfer learning models based on CNNs, and (5) 3D Vision Transformers, to infer Aβ status from standard anatomical MRI. We hypothesize that methods (1), (3) and (5) will perform best.

2 Imaging data and preprocessing steps

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a comprehensive, multisite study initiated in 2004, at 58 locations across North America. It aims to collect and analyze neuroimaging, clinical, and genetic data to identify and better understand biomarkers associated with healthy aging and AD (Veitch et al., 2019). In our analysis, we examined data from 1,847 ADNI participants with a mean age of 74.04 ± 7.40 years (863 females and 984 males). We included participants from all phases of ADNI (1, 2, GO and 3) who had both MRI and PET scans. The data was acquired across 58 sites with (both 1.5 and 3 T) GE, Siemens or Philips scanners. Forty of these sites had a change in scanner manufacturer or model across the scanning time of our subset. The distribution of participants included 661 cognitively normal (CN) individuals, 889 with mild cognitive impairment (MCI), and 297 with dementia. Overall, the dataset included 954 individuals classified as Aβ + (amyloid positive) and 893 as Aβ- (amyloid negative). A detailed table with the subject demographic breakdown can be found in Table 1.

In ADNI1, participants initially underwent PiB scans instead of florbetapir scans (ADNI, n.d.). However, the protocol was amended before the study's conclusion to transition to florbetapir scans due to processing time constraints. Consequently, PiB scans were only collected from ADNI1 participants. For participants in ADNI1 who transitioned into ADNIGO and then ADNI2, initial PET scans occurred 2 years from the date of the last successful florbetapir and FDG-PET scan conducted under ADNIGO. Additionally, in ADNI1, only a subset of participants received FDG scans. In ADNI2, subjects underwent up to 3 florbetapir scans and up to 2 FDG scans, with each scan acquired at 2-year intervals. These scans were conducted within a two-week window before or after the in-clinic assessments at Baseline and at 24 months after Baseline. In ADNI3, both Tau and Amyloid imaging were conducted on all participants during their initial ADNI3 visit. Amyloid PET imaging was carried out every 2 years using florbetapir for participants continuing from ADNI2 or florbetaben for newly enrolled participants (ADNI, n.d.). ADNI does not perform partial volume correction for amyloid PET analysis. It also does not account for off-target binding.

Mild cognitive impairment (MCI) is an intermediate state between normal aging and AD (Petersen et al., 1999), and is a significant focus in clinical trials, as many trials enroll individuals with MCI as they are assumed to be more likely to respond to therapy than people already diagnosed with AD. In the construction of the final

TABLE 1 Demographic data of individual train, validation and test set.

Individual distribution	Total N	Sex		Mean age ± St. Dev.	Amyloid classification		Diagnosis		
		M	F		+ve	-ve	CN	MCI	Dem
Train	1,292	680	612	73.99 ± 7.43	662	630	465	630	197
Validation	278	154	124	74.12 ± 6.95	146	132	105	126	47
Test	277	150	127	74.20 ± 7.74	146	131	91	133	53

dataset, we excluded participants who lacked basic clinical information or had poor-quality imaging data, such as scans with severe motion, distortion, or ringing artifacts.

ADNI has more participants with MCI compared to those with AD or CN. This is partly due to the initiative's focus on the early stages of cognitive decline and the progression to Alzheimer's disease. From ADNI phase 1 onward, twice as many MCI subjects were enrolled than AD cases or controls, with a target enrolment ratio of 1:2:1 for controls:MCI:AD. This higher proportion of MCI participants aligns with ADNI's objective to study factors that influence disease progression from MCI to AD, which is critical for early diagnosis and intervention.

Having a balanced number of participants in each diagnostic class and repeating the experiments could in principle lead to more reliable and generalizable models, reducing the bias toward the more prevalent class, MCI. But balancing the datasets can come with its own set of challenges. One issue might be the reduced amount of training data if undersampling is used to balance the classes, which can lead to loss of information, especially as the dataset is not large to begin with. Alternatively, oversampling/differential sampling methods such as SMOTE, or generative models such latent diffusion models, denoising diffusion probabilistic models (DDPMs), or VAEs might be used to generate synthetic data for the underrepresented classes, to augment the training set, but this might also introduce noise and overfitting.

T1w MRI scans were further processed using the automated segmentation software package FreeSurfer (Fisch, 2012), following the ENIGMA standardized protocol for brain segmentation and quality assurance (Van Erp and Hibar, 2016; van Erp et al., 2018).[1] The segmentations of subcortical regions (including lateralized hippocampus) and cortical regions [based on the Desikan-Killiany (DK) atlas regions (Desikan et al., 2006); including entorhinal cortex] were extracted and visually inspected for accuracy. The CSF, white and gray matter segmentations were extracted and visually inspected for each subject using FSL's Fast function.[2]

For training the CNN architectures, we used part of this dataset, so that an independent subset of the data could be reserved for testing. We focused on 3D T1w brain MRI scans (see Figure 1) from 762 subjects, with a mean age of 75.1 ± 7.6 years (394 females, 368 males). This subset included 459 cognitively normal controls, 67 individuals with MCI, and 236 with AD. These participants were selected as they also had amyloid-sensitive PET scans collected close to the time of the T1w MRI acquisition, with a maximum interval between scans set to 180 days (We note that one could consider an extension of the current problem, where the interval from the MRI to the amyloid assessment is considered as a variable, *t*, and used as input in the model, where *t* may be positive or negative). No repeated scans were used for the CNNs. The restriction on the time interval between scans was intended to help in estimating the relation between MRI features and amyloid positivity. As ViTs are more data intensive architectures, the whole dataset - with repeated scans - was used to train them. The test dataset in that case was designed to not have repeated scans, or scans from subjects in training or validation sets. Thus, the training dataset had 1,290 T1w MRI scans from 845 individual subjects, the validation dataset had 276 T1w MRI scans, and the test dataset had 275 T1w MRI scans. For the transfer learning experiments, we used data from 19,839 subjects from the UK Biobank dataset (age: 64.6 ± 7.6 years) comprising 10,294 females and 9,545 males.

As is customary when benchmarking deep learning methods, the 3D T1w brain MRI scans underwent a series of pre-processing steps (Lam and Zhu, 2020). These steps included nonparametric intensity normalization using N4 bias field correction, 'skull stripping' for brain extraction, registration to a template using 6 degrees of freedom (rigid-body) registration, and isometric voxel resampling to 2 mm. The resulting pre-processed images were of size 91x109x91. Furthermore, the T1w images underwent min-max scaling so that all values ranged between 0 and 1. This normalization process is common in image processing (and is similar to batch or instance normalization in deep learning), allowing standardized and consistent representation of image intensity values, which may aid in subsequent analyses and model training. The preprocessing pipeline applied to the 3D T1w MRI images ensures that the background of the scans is 0 intensity, and due to the normalization of input before CNN model, ideally, the effect of the original background or intensity range of the scan on performance of convolution models is negligible. To ensure a direct correspondence with the patch sizes used for the ViT models, the T1w input scans were resized to dimensions of both 64x64x64 and 128x128x128 for the ViT experiments. This resizing ensures compatibility between the image dimensions and the patch sizes employed in the ViT models, and allowed us to consistently integrate the T1w images into the analysis pipeline.

As is the convention in the ADNI dataset, two cut-off values were employed, providing alternative definitions of amyloid positivity, based on PET cortical *standardized uptake value ratio* (SUVR; denoted Aβ_1 by ADNI). For the 18F-florbetapir tracer, amyloid positivity was determined using mean 18F-florbetapir, with Aβ + defined as >1.11 for cutoff_1 and > 0.79 for cutoff_2. When florbetaben was used, Aβ + was defined as >1.20 for cutoff_1 and > 1.33 for cutoff_2. The SUVR values were normalized by using a whole cerebellum reference region (Hansson et al., 2018; Blennow et al., 2019). Each of these two cutoffs has been employed in the literature to define amyloid positivity, and

1 http://enigma.ini.usc.edu/protocols/imaging-protocols/

2 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST

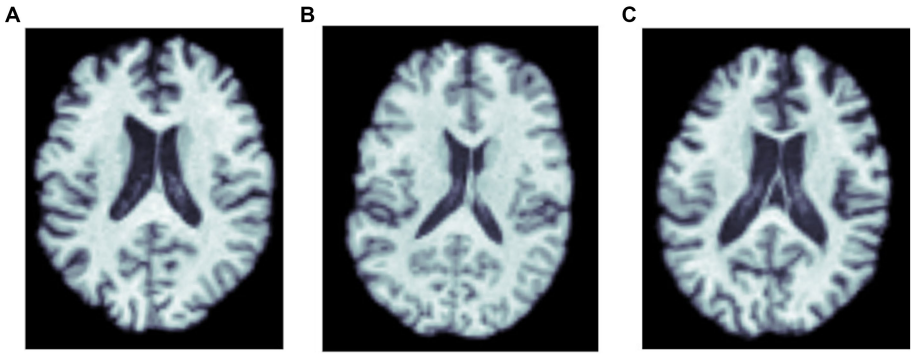


FIGURE 1
MRI scans of three amyloid positive participants: **(A)** a cognitively normal control, and participants diagnosed with **(B)** MCI, and **(C)** dementia.

to establish eligibility criteria for anti-amyloid drug treatments (van Dyck et al., 2023).

3 Models and experiments

3.1 Classical machine learning algorithms

As the first set of methods to evaluate for predicting Aβ + from anatomical MRI, we employed the following three classical machine learning algorithms: logistic regression, XGBoost, and a fully-connected artificial neural network (ANN) with 7 hidden layers. The ANN incorporated a Rectified Linear Unit (ReLU) activation function between layers. As predictors, we used measures that have previously been associated with amyloid levels in the literature: age, sex, clinical diagnosis, ApoE4 genotype values (2 for two copies of the ApoE4 allele and 1 for one E4 allele, 0 otherwise), overall volumes of cerebrospinal fluid (CSF), gray and white matter (all estimated from the brain MRI scan), as well as the left and right hippocampal and entorhinal cortex volumes. Regional volumes were extracted from the T1w MRI using FreeSurfer and were available for the entire brain. Previous studies like Kai et al. (Hu et al., 2019) and Thompson et al. (2004) show that hippocampal and entorhinal cortex volumes are among the most consistently affected in Alzheimer's disease, and as a result we focused on those two regional volumes in our study. The dataset was partitioned into independent training, validation, and testing sets, approximately in the ratio of 70:20:10. Standard performance metrics for the three algorithms (balanced accuracy and F1 Score on the test dataset), were computed to assess their effectiveness in predicting amyloid positivity.

3.2 2D CNN architecture

We implemented the 2D CNN architecture that we proposed in Gupta et al. (2021). In this model, 3D scans are used as the input, but each slice is encoded using a 2D CNN encoder (see Figure 2), which makes the training faster, requires less RAM, and allows pre-training using foundation models trained on large datasets of 2D photographic images, such as ImageNet. The encoded slices are then combined through an aggregation module that employs permutation-invariant layers, ultimately producing a single embedding for the entire scan. This embedding was then passed through feed-forward layers to predict whether the individual was amyloid positive or negative. This architecture allows for effective representation learning from 3D scans, and the aggregation module captures information from individual slices to predict amyloid status.

The 2D CNN encoder processes a single 2D slice as input and generates a d-dimensional embedding for each slice. The number of filters in the last layer of the architecture is d, determined by the dimension of the output from the aggregation module. The aggregation module incorporates permutation-invariant layers, ensuring that the output remains independent of the slice order. Specifically, the element-wise mean of all slice encodings is computed and used as the permutation-invariant layer. The value of d is fixed at 32, and a feed-forward layer with one hidden layer containing 64 activations is used. The slices in this context are sagittal. This model was trained for 100 epochs using the Adam optimizer (Kingma and Ba, 2015), a weight decay of 1×10^{-4}, a learning rate of 1×10^{-4}, and a batch size of 8. Mean squared error loss was employed as the optimization function during training. Model performance was measured using balanced accuracy.

3.3 3D CNN architecture

The 3D CNN was composed of four 3D Convolution layers with a filter size of 3×3, followed by one 3D Convolution layer with a 1×1 filter, and a final Dense layer with a sigmoid activation function (see Figure 3). A ReLU activation function and Instance normalization were applied to all layers. Dropout layers (with a dropout rate of 0.5) and a 3D Average Pooling layer with a 2×2 filter size were introduced into the 2nd, 3rd, and 4th layers. During training, models were optimized with a learning rate of 1×10^{-4}. Test performance was evaluated using balanced accuracy and F1 Score. To address overfitting, both L1 and L2 regularizers were employed, along with dropouts between layers and early stopping. Youden's J index (Youden, 1950) was used to determine the threshold for binary classification of Aβ + during testing, allowing comparison with the true cutoff values. Hyperparameter tuning was conducted through k-fold cross-validation to optimize model performance.

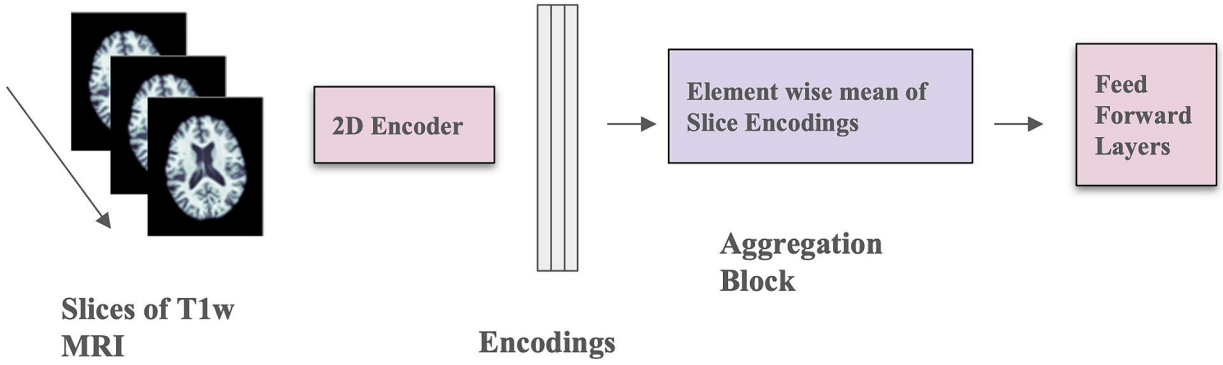


FIGURE 2
Model architecture with mean-based aggregation. The two pink blocks include trainable parameters; the purple block is a deterministic operation.

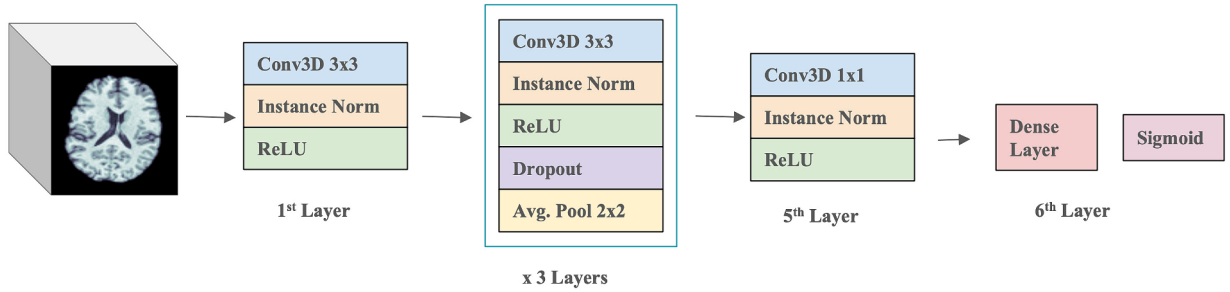


FIGURE 3
3D CNN model architecture.

3.4 Hybrid CNN architecture

The hybrid model (Figure 4) combines a 3D CNN using T1w images as input with an ANN that takes discrete, tabular data (which consists of simple values that are numeric or categorical) including age, sex, diagnosis, APOE4 values (2 for two copies of E4, 1 for one E4, and 0 for none), overall volumes of CSF, white and gray matter, and left and right hippocampal and entorhinal cortex volumes. The 3D images and the derived discrete data were fed into individual models, separately. After passing through flattening layers in the 3D CNN, the layers from the ANN are stacked with the tensors from the 3D CNN. Subsequently, the combined data passes through further Dense layers to predict Aβ+. The learning rate was set to 0.001, and the Adam Optimizer was used, with a batch size of 2. The model was trained for 200 epochs. The 3D CNN model consisted of 3 convolution blocks with increasing filter sizes (32, 64, 128, and 256) along with Batch Normalization and Max Pooling. The final convolution layer, before concatenation, had a filter size of 256 and used average pooling. The ANN had three layers with hidden layer sizes of 1,024, 512, and 64, along with instance normalization and the ReLU activation function.

This hybrid model was executed separately for both entorhinal cortex and hippocampus volumes, as well as in combination. In the combined case, we also considered the case where APOE genotype values were excluded from the discrete features input. Performance was evaluated using balanced accuracy and F1 Score, to compare the four models.

3.5 Vision transformers

We trained two variations of the ViT architecture: (i) the neuroimage transformer (NiT) and (ii) the multiple instance NiT (MINiT; Singla et al., 2022), as illustrated in Figure 5. These architectures involve several key steps. Initially, the input image is split into fixed-sized patch embeddings. These patches are then combined with learnable position embeddings and a class token. The resulting sequence of vectors is fed into a transformer encoder, consisting of alternating layers of multi-head attention and a multi-layer perceptron (MLP; *top right*, Figure 5). This architecture has been adapted to accommodate patches (cubes) from 3D scans. The NiT model was configured with a patch size of 8x8x8, without any overlap, a hidden dimension size of 256, six transformer encoder layers, and between 2 and 12 self-attention heads, with a dropout rate of 0.3.

Based on MiNiT (Singla et al., 2022), the input image, represented as $M \in R^{L \times W \times H}$, is transformed into a sequence of flattened blocks. If (B,B,B) denotes the shape of each block, the number of blocks is LWH/B^3. Non-overlapping cubiform patches are extracted from the input volume and flattened. These patches are then projected to D dimensions, the inner dimension of the transformer layers, using a learned linear projection. The generated sequence of input patches is augmented with learned positional embeddings for positional information and a learned classification token. Subsequently, this sequence is fed into a transformer encoder

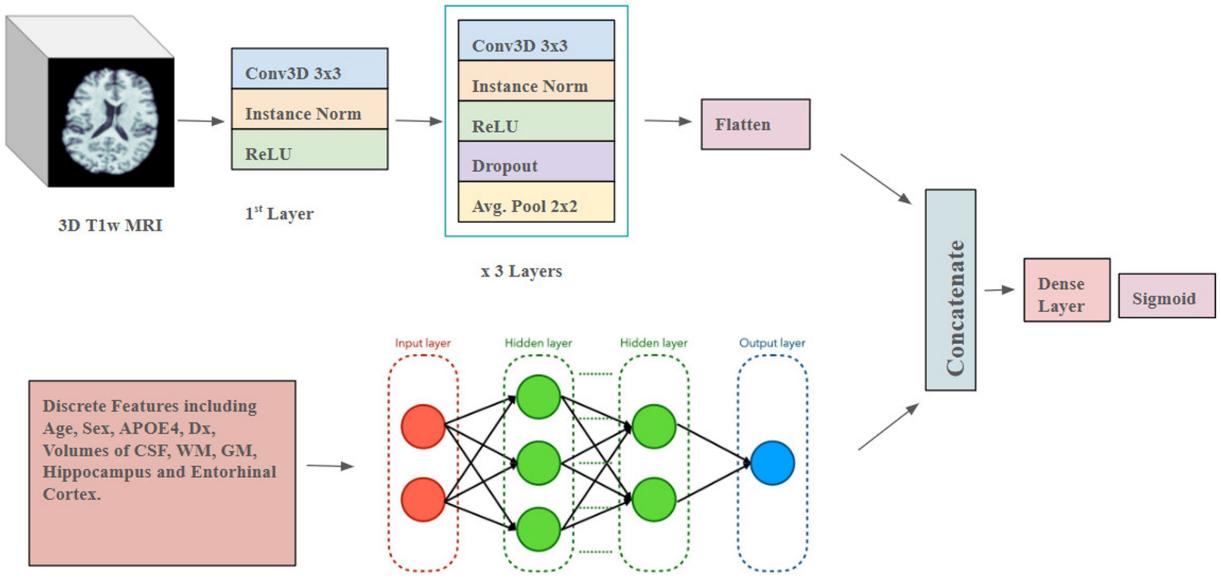


FIGURE 4
Hybrid 3D CNN model architecture.

comprising L transformer layers. Each layer consists of a multi-head self-attention block and a multi-layer perceptron (MLP) block, which incorporates two linear projections, with a Gaussian Error Gated Linear Unit (GEGLU) nonlinearity applied between them. Layer normalization is applied before - and residual connections are added after - every block in each transformer layer. Finally, a layer normalization and an MLP head consisting of a single $D \times C$ linear layer project the classification token to R^C, where C represents the number of classes (Singla et al., 2022).

The NiT architecture served as the primary model in our experiments, and we fine-tuned the default values for the number of transformer encoder layers and attention heads. In the case of MINiT, as well as incorporating a learned positional embedding on the training data to patches and adding a learned classification token to their sequence, a learned block embedding was also introduced (Singla et al., 2022). This embedding was included to retain the positional information of the block within the neuroimage of each patch. MINiT adopted similar parameters to those described for NiT.

We also performed hyperparameter selection for both models through a random search within specified upper and lower bounds. These parameters included the learning rate (chosen from a uniform distribution between 0.00001 to 0.001), weight decay (selected from a uniform distribution between 0.00001 to 0.001), the number of warm-up epochs (options included 1, 5, 16), the number of attention heads (options included 2, 4, 8, and 12), and the number of encoder layers (choices were 3, 4, and 6). These hyperparameters were defined based on the bounds typically used in ViT architectures (Bi, 2022; Jang and Hwang, 2022). We used the Adam optimizer (Kingma and Ba, 2015).

After training, we tested the model on the hold-out test dataset. We evaluated model performance with several metrics including the receiver-operator characteristic curve-area under the curve (ROC-AUC), accuracy, and F1-score. We determined the threshold for these metrics was accomplished through Youden's Index (Youden, 1950).

4 Results

In the comparison of classical machine learning models for predicting amyloid positivity, the best results were achieved with the artificial neural network (ANN), yielding a balanced accuracy of 0.771 and an F1 score of 0.771. The balanced accuracy values for the classical models ranged from 0.69 to 0.77, indicating predominantly similar classification performances across these models (Table 2).

The 2D CNN performed worse than the classical machine learning algorithms. Across an average of three runs, the model achieved a test accuracy of 0.543. In contrast, the 3D CNN architecture performed better, as indicated in Table 3. The Youden's J Index, used to determine the threshold for classifying Aβ + as 0/1 based on MRI scans, varied across different subject groups. Specifically, it was found to be 0.605 when considering only MCI and AD participants, 0.509 for cognitively unimpaired controls (CN), and 0.494 when considering all subjects. A balanced accuracy score of 0.760 was achieved for classification when all subjects were included. The accuracy increased to 0.850 when classifying individuals with only MCI or AD. In the case of CN, the balanced accuracy was 0.631. This observation aligns with expectations, as classifying Aβ + is more challenging in the earlier stages of the disease. According to the now-accepted Jack et al. model of the sequence of biomarker elevation in AD (Jack et al., 2018), abnormal amyloid accumulation typically precedes extensive brain atrophy, although individuals may vary in the order and relative intensities of these processes.

The hybrid model performed better than the 3D CNN model (Table 4). The hybrid model gave the best balanced accuracy of 0.815,

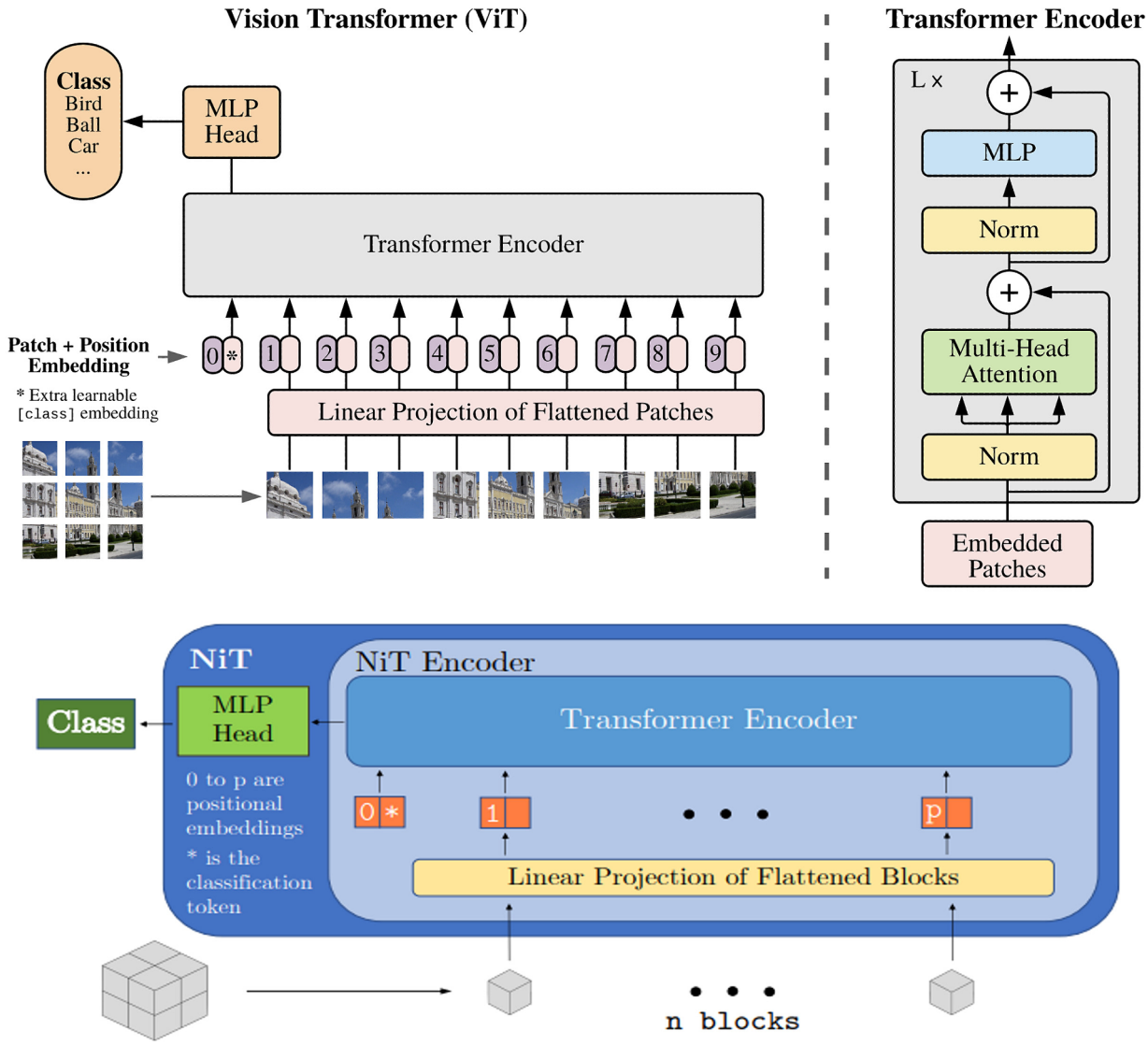


FIGURE 5
Overview of the vision transformer architecture, : reproduced from Singla et al. (2022).

when using hippocampal volume in the predictor set. Considering the CN, MCI and AD subjects in the test set separately for this model, the balanced accuracies are 0.616, 0.75 and 0.85 respectively, while the F1 Scores are 0.4, 0.969 and 0.863, respectively. This observation aligns with expectations, as classifying Aβ + is more challenging at the earlier stages of the disease.

The results comparing various hyperparameters for both NiT and MINiT model architectures are summarized in Table 5. Four different hyperparameter tunings were evaluated for both image sizes. In contrast, the NiT architecture performed more poorly, with classification accuracies close to chance (ranging between 0.5 to 0.6) across different hyperparameters and two image sizes. The MINiT architecture outperformed the NiT architectures, particularly for the image size of 64x64x64, with a test accuracy of 0.791 and a test ROC-AUC of 0.857. Therefore, the MINiT architecture improved upon the NiT architecture.

Hyperparameter tuning of attention heads, learning rate, encoder layer, and weight decay all enhanced model performance. Notably, the performance for the downscaled image of size 64x64x64 was superior to that for the upsampled image of size 128x128x128, in our experiments.

5 Discussion

This work, and several more recent amyloid-PET studies, show that the pattern of Aβ accumulation closely matches the anatomical trajectory of cortical gray matter loss detectable on brain MRI, a process that is also evident through the widening of the cortical sulci over time. Although the now widely accepted biomarker model by Jack et al. (2013) suggests that amyloid levels become statistically abnormal earlier than MRI measures of atrophy, all the processes occurring, to some extent, simultaneously. The order in which we detect them with imaging also depends, to some extent, on the sensitivity of our measurement techniques. Magnetic resonance imaging (MRI)

TABLE 2 Balanced accuracy (BA) and F1 scores for classical machine learning models.

	XGBoost	Logistic regression	ANN
	BA / F1 score	BA / F1 score	BA / F1 score
Data except for EC volume	0.742 / 0.678	0.770 / 0.734	0.711 / 0.696
Data except for HP volume	0.742 / 0.689	0.770 / 0.734	0.711 / 0.696
Data except for GM, WM and CSF volumes	0.697 / 0.656	0.770 / 0.734	0.771 / 0.771
Data with all features	0.756 / 0.701	0.770 / 0.734	0.725 / 0.716

The best performance was obtained with the ANN where all data except GM, WM and CSF volumes are considered, giving a balanced accuracy of 0.771.

TABLE 3 3D CNN results for all subjects, and with CN and MCI/AD groups considered separately.

	All subjects	CN	MCI and AD
Balanced accuracy	0.760	0.631	0.850
F1 score	0.746	0.480	0.824

measures of atrophy may not be as sensitive as amyloid positron emission tomography (PET) in detecting early changes, as amyloid levels typically become statistically abnormal earlier than structural atrophy becomes abnormal on MRI. The sensitivity of the imaging modality used plays a role in determining the order in which the pathological changes are observed, in addition to the temporal ordering of the underlying biological processes. There have been successful attempts to predict amyloid positivity in patients with MCI using radiomics and structural MRI (Petrone and Casamitjana, 2019; Kim J P, et al., 2021). To the best of our knowledge, we are the first to focus on predicting brain amyloid using deep learning architectures and T1-weighted structural MRIs. We know from work on related diseases (Kochunov et al., 2022) that even linear multivariate measures pick up disease effects with greater effect sizes than univariate measures, so a deep learning model could in theory produce a biomarker of atrophy that becomes abnormal or offers earlier anomaly detection and greater group differentiation than univariate measures such as hippocampal volume. As the amyloid accumulation and atrophy co-occur in the brain, it is plausible that our deep learning models could pick up on these signals to predict Aβ+. Thus, in early-stage patients who are Aβ+, the models attempt to detect any MRI-based anomalies that might separate them from healthy Aβ- subjects and combine them into a more accurate discriminator.

One potential issue with using amyloid and tau PET for molecular characterization of AD is off-target binding. While this may be a greater issue for tau PET than amyloid PET (Young et al., 2021), it is still an area of active research (Lemoine et al., 2018), because off-target binding may increase with age, affecting the SUVR metrics.

TABLE 4 Balanced accuracy and F1 score for the hybrid model architecture.

	Entorhinal cortex volume	Hippocampus volume	Entorhinal cortex and hippocampus volume
Balanced accuracy	0.759	0.815	0.787
F1 score	0.746	0.793	0.769

From our experiments, we can see that both deep and shallow neural networks, along with traditional classical machine learning models, showed promise in predicting amyloid positivity from standard structural brain MRI. Classical machine learning models, including XGBoost, logistic regression, and ANNs, exhibited promising balanced accuracy and F1 scores: best scores reached around 0.77. There is potential for further improvement with larger training samples and additional data modalities like Diffusion Tensor Images, which have shown significant associations with amyloid (Chattopadhyay and Singh, 2023a; Nir et al., 2023). Deep learning models, such as the 3D CNN tested, showed slightly better performance than classical machine learning models. The 2D CNN, while inferior to the 3D CNN architecture, may perform better with pre-training.

In the Alzheimer's disease (AD) progression model proposed by Jack et al. (2013), brain amyloid typically accumulates before pervasive brain atrophy is visible on MRI. As such, predicting Aβ + may be more challenging in controls than in individuals with mild cognitive impairment (MCI) and AD, where abnormalities are already evident on both PET and MRI scans. The hybrid model achieved the highest balanced accuracy of 0.815 when incorporating hippocampal volume in the predictor set. Further enhancements may be possible by increasing the size and diversity of the training data. and incorporating data from additional cohorts. The now-standard biomarker model of Alzheimer's disease, proposed by Jack et al. (2013), notes that structural MRI is typically one of the last biomarkers to show detectable changes - after CSF Abeta42, Amyloid PET, and CSF Tau. Because of this sequence, it is reasonable that an amyloid classifier based on T1w may not work as well in the very early stages of AD, and may work better when all of the biomarkers are somewhat elevated.

The MINiT architecture performed better than the other architecture considered—NiT. The results are promising. The performance we obtained may even improve with more training data, as the model has a large number of parameters; increasing the training dataset size may enhance model accuracy. In conclusion, the best performing models for the experiments are as summarized in Table 6.

A key goal of deep learning methods applied to neuroimaging data is that their performance remains robust even if the scanning protocol changes. In ADNI, the MRI scanning protocols do allow different scanner vendors (Siemens, Philips, and GE), but a long preparatory phase by the ADNI MRI Core was undertaken in 2004, to optimize the scan protocols for tracking of dementia, and to align the pulse sequences to the maximum possible extent across vendors. As such the training data from ADNI was from diverse scanners across the U.S., and included multiple vendors, and although the ADNI protocol was later adopted by many large scale imaging

TABLE 5 Experimental results for NiT and MINiT models.

Arch.	Image size	Hyperparameters of transformer architectures				Test ROC-AUC	Test balanced accuracy	Test F1 score
		Transformer layers	Attention heads	Dimension	MLP dimension			
NiT	$(64)^3$	512	3	12	175	0.494	0.541	0.614
		256	6	8	64	0.579	0.592	0.609
		256	4	8	234	0.485	0.516	0.221
	$(128)^3$	512	3	12	175	0.569	0.581	0.600
		256	6	8	64	0.692	0.590	0.584
		256	4	8	234	0.692	0.468	0.495
MINiT	$(64)^3$	6	12	256	309	**0.857**	**0.791**	**0.793**
		6	8	256	309	0.755	0.697	0.674
		6	8	128	128	0.585	0.599	0.686
		6	12	258	128	0.794	0.776	0.782
	$(128)^3$	6	12	256	309	0.503	0.534	0.557
		6	8	256	309	0.668	0.649	0.688
		6	8	128	128	0.799	0.747	0.766
		6	12	258	128	0.476	0.527	0.584

Columns 3 to 6 show the hyperparameters of the transformer architectures, namely Transformer Layers, No. of Attention Heads, Dimension and MLP Dimension. The experiments are compared using test ROC-AUC, accuracy and F1 Score. Bold numbers indicate the best results.

TABLE 6 Best performing models for amyloid classification.

Model	Balanced accuracy	F1 score
Hybrid Model using Hippocampus Volume in Predictor Set	0.815	0.793
MINiT with image size $(64)^3$, 6 Transformer Layers and 12 Attention Heads	0.791	0.793

The performance can improve by increasing the amount of training data.

initiatives, there was still somewhat less heterogeneity in the protocols than would be seen in general. Future work will examine the use of *post-hoc* methods for MRI harmonization (Liu, 2021; Zuo et al., 2021; Komandur, 2023), to test whether this improves performance on data from new scanners and other scanning protocols.

The current biological categorization of Alzheimer's disease commonly relies on other data sources such as amyloid- or tau-sensitive PET scans or cerebrospinal fluid (CSF) biomarkers, all of which are more invasive than structural brain MRI. While a T1w MRI-based model may benefit from the incorporation of other data sources, it offers a promising tool for benchmarking. T1w MRIs are more widely available and cost-effective than amyloid PET. Therefore, classifying amyloid positivity from T1w MRIs may help to identify participants, particularly those with MCI, for further, more intensive testing using other modalities. Prior works (Grill et al., 2019) show that the selection of biomarker criteria should be guided by the objective of enrolling individuals who are most likely to use and benefit from the intervention being studied in a specific context. As a result, our work shows the potential of ML/DL methods in MCI participants for detection of amyloid positivity before going for further more intensive testing using other modalities such as PET scans.

5.1 Limitations and future work

This study has limitations - notably the restricted testing on the ADNI dataset. Performance may improve with an increase in the size and diversity of the training data, by including multimodal brain MRI (Chattopadhyay and Singh, 2023a, 2023b) and by adding data from supplementary cohorts. Future work will include individuals of more diverse ancestries (John et al., 2023; Chattopadhyay and Joshy, 2024) and with various comorbidities such as vascular disease, frontotemporal dementia, and other degenerative diseases. Moreover, the sensitivity of the approach to different MRI scanning protocols and PET tracers should be examined. In the context of multisite data, harmonization methods - such as using centiloids for PET and generative adversarial networks (GANs) for MRIs - may be needed for domain adaptation. These steps may help in evaluating amyloid prediction accuracy across varied scenarios and populations. There are efforts to develop cheaper ways to measure amyloid from blood (AD Blood Tests Are Here. Now, Let's Grapple With How to Use Them | ALZFORUM, 2024), but so far tau has been easier to measure accurately (pTau217). As these methods are developed, we hope to incorporate them into multimodal setups.

Author's note

Data used in preparing this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu/). As such, many 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.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://adni.loni.usc.edu; https://www.ukbiobank.ac.uk.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

TC: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft. SO: Formal analysis, Software, Validation, Visualization, Writing – review & editing. KB: Formal analysis, Software, Validation, Visualization, Writing – review & editing. NJ: Formal analysis, Software, Validation, Visualization, Writing – review & editing. DK: Formal analysis, Software, Validation, Visualization, Writing – review & editing. JN: Formal analysis, Software, Validation, Visualization, Writing – review & editing. ST: Data curation, Writing – review & editing. GS: Writing – review & editing, Supervision. JA: Writing – review & editing, Supervision. PT: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by NIH grants R01AG058854, U01AG068057 and RF1AG057892 (PI:PT). Data collection and sharing for ADNI was funded by National Institutes of Health Grant U01 AG024904 and the DOD (Department of Defense award number W81XWH-12-2-0012). ADNI 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.

Acknowledgments

We thank the ADNI investigators, and their public and private funders. For creating and publicly disseminating the ADNI dataset. This study builds on preliminary findings in a conference paper entitled, *Can Structural MRIs be used to detect Amyloid Positivity using Deep Learning,* which may be found in the conference proceedings from the 19th International Symposium on Medical Information Processing and Analysis (SIPAIM; Chattopadhyay and Singh, 2023a).

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024.1387196/full#supplementary-material

References

AD Blood Tests Are Here. Now, Let's Grapple With How to Use Them. (2024). ALZFORUM. Available at: Www.alzforum.org.

ADNI. PET Analysis. Available at: https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/

ADNI. Data types. Available at: https://adni.loni.usc.edu/data-samples/data-types/

Alzubaidi, L., and Al-Amidie, M. (2021). Novel transfer learning approach for medical imaging with limited labeled data. *Cancer* 13:1590. doi: 10.3390/cancers13071590

Bae, J. B., Lee, S., Oh, H., Sung, J., Lee, D., Han, J. W., et al. (2023). A case-control clinical trial on a deep learning-based classification system for diagnosis of amyloid-positive Alzheimer's disease. *Psychiatry Investig.* 20, 1195–1203. doi: 10.30773/pi.2023.0052

Bi, Y., "MultiCrossViT: multimodal vision transformer for schizophrenia prediction using structural MRI and functional network connectivity data," in arXiv, (2022). Available at: http://arxiv.org/abs/2211.06726.

Blennow, K., Shaw, L. M., Stomrud, E., Mattsson, N., Toledo, J. B., Buck, K., et al. (2019). Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys Abeta(1-42), pTau and tTau CSF immunoassays. *Sci. Rep.* 9:19024. doi: 10.1038/s41598-019-54204-z

Braak, H. (2000). Vulnerability of select neuronal types to Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 924, 53–61. doi: 10.1111/j.1749-6632.2000.tb05560.x

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* 112, 389–404. doi: 10.1007/s00401-006-0127-z

Braak, H., and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259. doi: 10.1007/BF00308809

Braak, H., and Braak, E. (1997). Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiol. Aging* 18, 351–357. doi: 10.1016/S0197-4580(97)00056-0

Chattopadhyay, T., and Joshy, N. A. (2024). Brain age analysis and dementia classification using convolutional neural networks trained on diffusion MRI: tests in Indian and north American cohorts. *bioRxiv*. doi: 10.1101/2024.02.04.578829v1

Chattopadhyay, T., and Singh, A. (2023b). Comparison of anatomical and diffusion MRI for detecting Parkinson′ s disease using deep convolutional neural network: IEEE EMBC. 1–6.

Chattopadhyay, T., and Singh, A., (2023a). "Predicting dementia severity by merging anatomical and diffusion MRI with deep 3D convolutional neural networks." In the 18th *international symposium on medical information processing and analysis* (Vol. 12567, pp. 90–99). SPIE.

Cho, S. H., Kim, S., Choi, S. M., and Kim, B. C. for the Alzheimer's Disease Neuroimaging Initiative (2024). ATN classification and clinical progression of the amyloid-negative Group in Alzheimer's disease neuroimaging initiative participants. *Chonnam Med. J.* 60, 51–58. doi: 10.4068/cmj.2024.60.1.51

Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., et al. (2011). Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 305, 275–283. doi: 10.1001/jama.2010.2008

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Dhinagar, N. J., and Thomopoulos, S. I. (2023) Evaluation of transfer learning methods for detecting Alzheimer's disease with brain MRI. In the 18th *international symposium on medical information processing and analysis* (Vol. 12567, pp. 504–513). SPIE. IEEE

Dhinagar, N. J., Thomopoulos, S. I., Laltoo, E., and Thompson, P. M. (2023). Efficiently training vision transformers on structural MRI scans for Alzheimer's disease detection. *EMBC*. (pp. 1–6). IEEE.

Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., et al. (2021). Contrastive learning with continuous proxy Meta-data for 3D MRI classification. *MICCAI*. Springer International Publishing. doi: 10.1007/978-3-030-87196-3_6

Ezzati, A., Harvey, D. J., and Habeck, C. (2020). Predicting amyloid-β levels in amnestic mild cognitive impairment using machine learning techniques. *J. Alzheimer's Dis.: JAD* 73, 1211–1219. doi: 10.3233/JAD-191038

Feng, X., Provenzano, F. A., and Small, S. A. (2022). A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease. *Alzheimers Res. Ther.* 14:45. doi: 10.1186/s13195-022-00985-x

Fisch, B. (2012). FreeSurfer. *NeuroImage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Gelosa, G., and Brooks, D. J. (2012). The prognostic value of amyloid imaging. *Eur. J. Nucl. Med. Mol. Imaging* 39, 1207–1219. doi: 10.1007/s00259-012-2108-x

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Grill, J. D., Nuño, M. M., and Gillen, D. L. (2019). Alzheimer's Disease Neuroimaging Initiative. Which MCI patients should be included in prodromal Alzheimer disease clinical trials? *Alzheimer Dis. Assoc. Disord.* 33, 104–112. doi: 10.1097/WAD.0000000000000303

Gupta, U., Lam, P. K., Ver Steeg, G., and Thompson, P. M. (2021). Improved brain age estimation with slice-based set networks. In *2021 IEEE 18th international symposium on biomedical imaging (ISBI)* (pp. 840–844).

Hansson, O., Seibyl, J., Stomrud, E., Zetterberg, H., Trojanowski, J. Q., Bittner, T., et al. (2018). CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* 14, 1470–1481. doi: 10.1016/j.jalz.2018.01.010

Hu, K., Li, Y., Yu, H., and Hu, Y. (2019). CTBP1 confers protection for hippocampal and cortical neurons in rat models of Alzheimer's disease. *Neuroimmunomodulation* 26, 139–152. doi: 10.1159/000500942

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

Jack, C. R. Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., et al. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* 14, 535–562. doi: 10.1016/j.jalz.2018.02.018

Jack, C. R. Jr., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., et al. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet. Neurol.* 12, 207–216. doi: 10.1016/S1474-4422(12)70291-0

Jang, J., and Hwang, D. (2022). M3T: Three-dimensional medical image classifier using multi-plane and multi-slice transformer, 20718–20729.

Jin, Y., DuBois, J., Zhao, C., and Zhan, L., (2023). "Brain MRI to PET synthesis and amyloid estimation in Alzheimer's disease via 3D multimodal contrastive GAN." In *International workshop on machine learning in medical imaging* (pp. 94–103). Cham: Springer Nature.

John, J. P., Joshi, H., Sinha, P., Harbishettar, V., Tripathi, R., Cherian, A. V., et al. (2023). India ENIGMA initiative for Global Aging & Mental Health–a globally coordinated study of brain aging and Alzheimer's disease. *Alzheimers Dement.* 19:e076394. doi: 10.1002/alz.076394

Johnson, G. V., and Hartigan, J. A. (1999). Tau protein in normal and Alzheimer's disease brain: an update. *J. Alzheimers Dis.* 1, 329–351. doi: 10.3233/JAD-1999-14-512

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., and Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC Med. Imaging* 22:69. doi: 10.1186/s12880-022-00793-7

Kim, J. P., Kim, J., Jang, H., Kim, J., Kang, S. H., Kim, J. S., et al. (2021). Predicting amyloid positivity in patients with mild cognitive impairment using a radiomics approach. *Sci. Rep.* 11:6954. doi: 10.1038/s41598-021-86114-4

Kim, S., Lee, P., Oh, K. T., Byun, M. S., Yi, D., Lee, J. H., et al. (2021). Deep learning-based amyloid PET positivity classification model in the Alzheimer's disease continuum by using 2-[^{18}F] FDG PET. *Eur J Nucl Med Mol Imag. Res.* 11:56. doi: 10.1186/s13550-021-00798-3

Kingma, D., and Ba, J. (2015). Adam: A method for stochastic optimization: ICLR.

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., et al. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Ann. Neurol.* 55, 306–319. doi: 10.1002/ana.20009

Kochunov, P., Fan, F., Ryan, M. C., Hatch, K. S., Tan, S., Jahanshad, N., et al. (2022). Translating ENIGMA schizophrenia findings using the regional vulnerability index: association with cognition, symptoms, and disease trajectory. *Hum. Brain Mapp.* 43, 566–575. doi: 10.1002/hbm.25045

Koivunen, J., Karrasch, M., Scheinin, N. M., Aalto, S., Vahlberg, T., Någren, K., et al. (2012). Cognitive decline and amyloid accumulation in patients with mild cognitive impairment. *Dement. Geriatr. Cogn. Disord.* 34, 31–37. doi: 10.1159/000341580

Komandur, D., (2023). Unsupervised harmonization of brain MRI using 3D CycleGANs and its effect on brain age prediction. *19th International symposium on medical information processing and analysis* (SIPAIM) (pp. 1–5). IEEE.

Lam, P., and Zhu, A. H. (2020). 3-D grid-attention networks for interpretable age and Alzheimer's disease prediction from structural MRI. *arXiv* preprint arXiv:2011.09115.

Landau, S. M., Breault, C., Joshi, A. D., Pontecorvo, M., Mathis, C. A., Jagust, W. J., et al. (2013). Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. *J. Nucl. Med.* 54, 70–77. doi: 10.2967/jnumed.112.109009

Landau, S. M., and Thomas, B. A. (2014). Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. *Eur. J. Nucl. Med. Mol. Imaging* 41, 1398–1407. doi: 10.1007/s00259-014-2753-3

Lemoine, L., Leuzy, A., Chiotis, K., Rodriguez-Vieitez, E., and Nordberg, A. (2018). Tau positron emission tomography imaging in tauopathies: the added hurdle of off-target binding. *Alzheimers Dement (Amst).* 10, 232–236. doi: 10.1016/j.dadm.2018.01.007

Li, J. (2022). Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives. *arXiv* 2206:01136.

Liu, M., (2021). Style transfer using generative adversarial networks for multi-site mri harmonization. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference*, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24 (pp. 313–322). Springer International Publishing.

Lu, B., Li, H.-X., Chang, Z.-K., Li, L., Chen, N. X., Zhu, Z. C., et al. (2022). A practical Alzheimer disease classifier via brain imaging-based deep learning on 85,721 samples. *J. Big Data* 9:101. doi: 10.1186/s40537-022-00650-y

Masters, C. L., and Selkoe, D. J. (2012). Biochemistry of amyloid β-protein and amyloid deposits in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2:a006262. doi: 10.1101/cshperspect.a006262

Matsoukas, C., "Is it time to replace CNNs with transformers for medical images?" (2021). Available at: http://arxiv.org/abs/2108.09038

Morid, M. A., Borjali, A., and Fiol, G. D. (2021). A scoping review of transfer learning research on medical image analysis using ImageNet. *Comput. Biol. Med.* 128:104115. doi: 10.1016/j.compbiomed.2020.104115

Nelson, P. T., Jicha, G. A., Schmitt, F. A., Liu, H., Davis, D. G., Mendiondo, M. S., et al. (2007). Clinicopathologic correlations in a large Alzheimer disease center autopsy cohort: neuritic plaques and neurofibrillary tangles" do count" when staging disease severity. *J. Neuropathol. Exp. Neurol.* 66, 1136–1146. doi: 10.1097/nen.0b013e31815c5efb

Nir, T. M., Villalón-Reina, J. E., and Salminen, L. E. (2023). Cortical microstructural associations with CSF amyloid and pTau. *Mol. Psychiatry* 1–12.

Okello, A., Koivunen, J., Edison, P., Archer, H. A., Turkheimer, F. E., Någren, K., et al. (2009). Conversion of amyloid positive and negative MCI to AD over 3 years: an 11C-PIB PET study. *Neurology* 73, 754–760. doi: 10.1212/WNL.0b013e3181b23564

Pan, Y., Liu, M., Lian, C., Zhou, T., and Xia, Y., "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," *21st International Conference, Granada, Spain, Proceedings*, Part 11072. (2018).

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* 56, 303–308. doi: 10.1001/archneur.56.3.303

Petrone, P. M., and Casamitjana, A. (2019). Prediction of amyloid pathology in cognitively unimpaired individuals using voxel-wise analysis of longitudinal structural brain MRI. *Alzheimers Res. Ther.* 11:72. doi: 10.1186/s13195-019-0526-8

Qu, C., Zou, Y., Dai, Q., Ma, Y., He, J., Liu, Q., et al. (2021). Advancing diagnostic performance and clinical applicability of deep learning-driven generative adversarial networks for Alzheimer's disease. *Psychoradiology* 1, 225–248. doi: 10.1093/psyrad/kkab017

Revised Again: Alzheimer's Diagnostic Criteria Get Another Makeover. (2023) ALZFORUM. Available at: Www.alzforum.org.

Rowe, C. C., Ellis, K. A., Rimajova, M., Bourgeat, P., Pike, K. E., Jones, G., et al. (2010). Amyloid imaging results from the Australian imaging, biomarkers and lifestyle (AIBL) study of aging. *Neurobiol. Aging* 31, 1275–1283. doi: 10.1016/j.neurobiolaging.2010.04.007

Shan, G., Bernick, C., Caldwell, J. Z. K., and Ritter, A. (2021). Machine learning methods to predict amyloid positivity using domain scores from cognitive tests. *Sci. Rep.* 11:4822. doi: 10.1038/s41598-021-83911-9

Singla, A., Zhao, Q., do, D. K., Zhou, Y., Pohl, K. M., and Adeli, E. (2022). Multiple Instance Neuroimage Transformer. *Predictive Intelligence in Medicine: 5th International Workshop, MICCAI PRIME* 13564, 36–48. doi: 10.1007/978-3-031-16919-9_4

Son, H. J., Oh, J. S., Oh, M., Kim, S. J., Lee, J. H., Roh, J. H., et al. (2020). The clinical feasibility of deep learning-based classification of amyloid PET images in visually equivocal cases. *Eur. J. Nucl. Med. Mol. Imaging* 47, 332–341. doi: 10.1007/s00259-019-04595-y

SPM12 software - Statistical Parametric Mapping. (2014). Functional Imaging Laboratory. Available at: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12:e1001779. doi: 10.1371/journal.pmed.1001779

Tarasoff-Conway, J. M., Carare, R. O., Osorio, R. S., Glodzik, L., Butler, T., Fieremans, E., et al. (2015). Clearance systems in the brain implications for Alzheimer disease. *Nat. Rev. Neurol.* 11, 457–470. doi: 10.1038/nrneurol.2015.119

Thompson, P. M. (2007). Tracking Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 1097, 183–214. doi: 10.1196/annals.1379.017

Thompson, P. M., Hayashi, K. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Rapoport, J. L., et al. (2004). Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage* 23, S2–S18. doi: 10.1016/j.neuroimage.2004.07.071

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., et al. (2023). Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* 388, 9–21. doi: 10.1056/NEJMoa2212948

Van Erp, T. G., and Hibar, D. P. (2016). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* 21, 547–553. doi: 10.1038/mp.2015.63

van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., et al. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuro imaging genetics through meta analysis (ENIGMA) consortium. *Biol. Psychiatry* 84, 644–654. doi: 10.1016/j.biopsych.2018.04.023

Veitch, D., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement.* 15, 106–152. doi: 10.1016/j.jalz.2018.08.005

Villemagne, V. L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K. A., Salvado, O., et al. (2013). Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* 12, 357–367. doi: 10.1016/S1474-4422(13)70044-9

Villemagne, V. L., Pike, K. E., Chételat, G., Ellis, K. A., Mulligan, R. S., Bourgeat, P., et al. (2011). Longitudinal assessment of Aβ and cognition in aging and Alzheimer disease. *Ann. Neurol.* 69, 181–192. doi: 10.1002/ana.22248

Wang, T., Lei, Y., Fu, Y., Wynne, J. F., Curran, W. J., Liu, T., et al. (2020). A review on medical imaging synthesis using deep learning and its clinical applications. *J. Appl. Clin. Med. Phys.* 22, 11–36. doi: 10.1002/acm2.13121

Willemink, M. J., Roth, H. R., and Sandfort, V. (2022). Toward foundational deep learning models for medical imaging in the new era of transformer networks. *Radiol. Artif. Intell.* 4:e210284. doi: 10.1148/ryai.210284

World Health Organization. "Dementia," (2022). Available at: https://www.who.int/news-room/fact-sheets/detail/dementia.

Yasuno, F., Kazui, H., Morita, N., Kajimoto, K., Ihara, M., Taguchi, A., et al. (2017). Use of T1-weighted/T2-weighted magnetic resonance ratio to elucidate changes due to amyloid β accumulation in cognitively normal subjects. *NeuroImage: Clinical* 13, 209–214. doi: 10.1016/j.nicl.2016.11.029

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Young, C. B., Landau, S. M., Harrison, T. M., Poston, K. L., and Mormino, E. C. (2021). Influence of common reference regions on regional tau patterns in cross-sectional and longitudinal [18F]-AV-1451 PET data. *NeuroImage* 243:118553. doi: 10.1016/j.neuroimage.2021.118553

Zhou, T., Ye, X. Y., Lu, H. L., Zheng, X., Qiu, S., and Liu, Y. C. (2022). Dense convolutional network and its application in medical image analysis. *Biomed. Res. Int.* 2022, 1–22. doi: 10.1155/2022/2384830

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/JPROC.2020.3004555

Zuo, L., Dewey, B. E., Carass, A., Liu, Y., He, Y., and Calabresi, P. A., (2021). Information-based disentangled representation learning for unsupervised MR harmonization. In the *international conference on information processing in medical imaging* (pp. 346–359). Cham: Springer International Publishing.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 30 July 2024
DOI 10.3389/fnins.2024.1436619

OPEN ACCESS

EDITED BY
Da Ma,
Wake Forest University, United States

REVIEWED BY
M. Sandeep Kumar,
VIT University, India
S. Saravanan,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India

*CORRESPONDENCE
Huazhong Shu
✉ shu.list@seu.edu.cn

RECEIVED 22 May 2024
ACCEPTED 16 July 2024
PUBLISHED 30 July 2024

CITATION
Nie J, Shu H and Wu F (2024) An epilepsy classification based on FFT and fully convolutional neural network nested LSTM. *Front. Neurosci.* 18:1436619. doi: 10.3389/fnins.2024.1436619

An epilepsy classification based on FFT and fully convolutional neural network nested LSTM

Jianhao Nie, Huazhong Shu* and Fuzhi Wu

Laboratory of Image Science and Technology, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing, China

Background and objective: Epilepsy, which is associated with neuronal damage and functional decline, typically presents patients with numerous challenges in their daily lives. An early diagnosis plays a crucial role in managing the condition and alleviating the patients' suffering. Electroencephalogram (EEG)-based approaches are commonly employed for diagnosing epilepsy due to their effectiveness and non-invasiveness. In this study, a classification method is proposed that use fast Fourier Transform (FFT) extraction in conjunction with convolutional neural networks (CNN) and long short-term memory (LSTM) models.

Methods: Most methods use traditional frameworks to classify epilepsy, we propose a new approach to this problem by extracting features from the source data and then feeding them into a network for training and recognition. It preprocesses the source data into training and validation data and then uses CNN and LSTM to classify the style of the data.

Results: Upon analyzing a public test dataset, the top-performing features in the fully CNN nested LSTM model for epilepsy classification are FFT features among three types of features. Notably, all conducted experiments yielded high accuracy rates, with values exceeding 96% for accuracy, 93% for sensitivity, and 96% for specificity. These results are further benchmarked against current methodologies, showcasing consistent and robust performance across all trials. Our approach consistently achieves an accuracy rate surpassing 97.00%, with values ranging from 97.95 to 99.83% in individual experiments. Particularly noteworthy is the superior accuracy of our method in the AB versus (vs.) CDE comparison, registering at 99.06%.

Conclusion: Our method exhibits precise classification abilities distinguishing between epileptic and non-epileptic individuals, irrespective of whether the participant's eyes are closed or open. Furthermore, our technique shows remarkable performance in effectively categorizing epilepsy type, distinguishing between epileptic ictal and interictal states versus non-epileptic conditions. An inherent advantage of our automated classification approach is its capability to disregard EEG data acquired during states of eye closure or eye-opening. Such innovation holds promise for real-world applications, potentially aiding medical professionals in diagnosing epilepsy more efficiently.

KEYWORDS

electroencephalogram, fast Fourier transformation, seizure detection, convolutional neural network, long-short term memory

1 Introduction

Epilepsy is a very common neurological disorder in humankind that affects roughly 50 million people worldwide (Tuncer et al., 2021; World Health Organization, 2021). It is characterized by abnormal electrical activity in the nerve cells of the brain, resulting in recurrent seizures, unusual behavior, and possibly loss of consciousness (Fisher et al., 2014; Ozdemir et al., 2021). The worst-case scenario could result in permanent harm to the patient's life. Up to 70% of individuals with epilepsy could live seizure-free if properly diagnosed and treated. Therefore, a timely and accurate diagnosis method for epilepsy is essential for all patients and doctors. In clinical practice, doctors diagnose epilepsy by using patients' medical records, conducting neurological examinations, and employing various clinical tools such as neuroimaging and EEG recording. However, this analysis is considered complex due to the presence of patterns in the EEG that can be challenging to interpret, even for experienced experts. This complexity can lead to different opinions among experts regarding EEG findings, necessitating complementary examinations (Oliva and Rosa, 2019; Oliva and Rosa, 2021). To address the time-consuming nature of visual analysis and errors caused by visual fatigue during the increasing continuous EEG video recordings, numerous automatic methods have been developed.

There have been various methods proposed in the past three decades for the automatic identification of epileptic EEG signals (Ghosh-Dastidar and Adeli, 2009; Sharma et al., 2014; Shanir et al., 2018; Truong et al., 2018). Machine learning (ML) methods can be used to build effective classifiers for automatic epilepsy detection. These automatic seizure detection methods mainly include two steps: feature extraction and classifier construction. The feature extraction includes time domain (T) (Jaiswal and Banka, 2017; Gao et al., 2020; Wijayanto et al., 2020), frequency domain (F) (Altaf and Yoo, 2015; Kaleem et al., 2018; Singh et al., 2020), time-frequency domain (TF) (Tzallas et al., 2007; Abualsaud et al., 2015; Feng et al., 2017; Shen et al., 2017; Goksu, 2018; Sikdar et al., 2018; Yavuz et al., 2018), and a combination of nonlinear approaches (Zeng et al., 2016; Ren and Han, 2019; Sayeed et al., 2019; Wu et al., 2019). In addition, various types of entropy such as fuzzy entropy (Xiang et al., 2015), approximate entropy, sample entropy, and phase entropy (Acharya et al., 2012) have been calculated from the EEG signals to distinguish different epileptic EEG segments. The automatic seizure classifier includes Support Vector Machine (SVM) (Subasi and Ismail Gursoy, 2010; Das et al., 2016; Şengür et al., 2016; Li and Chen, 2021), Convolutional Neural Network (CNN) (Feng et al., 2017; Wijayanto et al., 2020; Ozdemir et al., 2021), Extreme Learning Machine (Yuan et al., 2014), K-Nearest Neighbor (Guo et al., 2011; Tuncer et al., 2021), Deep Neural Network (Sayeed et al., 2019), Recurrent Neural Network (Yavuz et al., 2018).

Gotman (1982) proposed the first widely used new method, which is based on decomposing the EEG into elementary waves and detecting paroxysmal bursts of rhythmic activity with a frequency between 3 and 20 cycles per second. This method was further improved by the same group, who broke down EEG signals into half waves and then extracted features such as peak amplitude, duration, slope, and sharpness to detect seizure activities (Gotman, 1990). Jaiswal and Banka (2017) primarily used time-domain features such as local neighborhood descriptive patterns and one-dimensional local gradient patterns for epilepsy detection. Gao et al. (2020) and Wijayanto et al. (2020) extracted approximate entropy as features and combined with recurrence quantification analysis to detect epilepsy, their method achieved an accuracy of 91.75% in the Bonn dataset (Andrzejak et al., 2001). Wijayanto et al. (2020) used the Higuchi fractal dimension (HFD) to differentiate between ictal and interictal conditions in EEG signals. Many researchers focused on time domain features, while others concentrated on frequency domain, time-frequency domain, and nonlinear approaches. Altaf and Yoo (2015) combined feature extraction with classification engines, implementing multiplex bandpass filter coefficients for feature extraction. Subsequently, a nonlinear SVM was used, achieving a sensitivity of 95.1%. Kaleem et al. (2018) developed a method based on a signal-derived empirical mode decomposition (EMD) dictionary approach.

The integrated time-frequency method has been widely used for feature extraction in various approaches. For instance, Abualsaud et al. (2015) successfully detected epilepsy from compressed and noisy EEG signals using discrete wavelet transformation (DWT), achieving an accuracy of 80% when SNR = 1 dB. Feng et al. (2017) extracted features from three-level Daubechies discrete wavelet transform. Shen et al. (2017) employed a genetic algorithm to select a subset of 980 features subset and used 6 SVMs to classify EEG data into four types, i.e., normal, spike, sharp wave, and seizures. Sikdar et al. (2018) proposed a MultiFractal Detrended Fluctuation Analysis (MFDFA) to address the multifractal behaviors in healthy (Group B), interictal (Group D), and ictal (Group E) patterns. Yavuz et al. (2018) extracted mel frequency cepstral coefficients (MFCCs) as features and applied them in a regression neural network. Goksu (2018) extracted Log Energy Entropy, Norm Entropy, and Energy from wavelet packet analysis (WPA) as features and used multilayer perception (MLP) as a classifier, achieving commendable performance.

Some researchers have used nonlinear or mixed features as classification criteria. Zeng et al. (2016) extracted Sample Entropy and the permutation Entropy, and Hurst Index from EEG segments which were selected through an ANOVA test by four classifiers (Decision Tree, K-Nearest Neighbor Discriminant Analysis, SVM). Ren and Han (2019) extracted both linear and nonlinear features and classified them using an extreme learning machine. Sayeed et al. (2019) employed DWT, Hjorth parameters, statistical features, and a machine learning classifier to differentiate between ictal EEG and interictal EEG patterns.

These methods based on feature extraction are influenced by the intrinsic characteristics of EEG, such as muscle activities and eye movements, which may introduce noise to the original EEG data, potentially altering its actual characteristics (Hussein et al., 2019; Li et al., 2020). To address these challenges, many deep learning models have been developed for automatic epileptic seizure detection.

While other approaches have been proposed in the literature for epilepsy classification (Joshi et al., 2014; Zhu et al., 2014; Hassan et al., 2016; Indira and Krishna, 2021; Qaisar and Hussain, 2021), the prevailing trend involves the application of deep learning techniques (Yuan et al., 2017; Acharya et al., 2018; Tsiouris et al., 2018; Ullah et al., 2018; Covert et al., 2019; Li et al., 2020; Ozdemir et al., 2021) in this domain. However, most traditional methods have focused on specific or local features, resulting in information loss, including time domain features, frequency domain features, time-frequency domain features, and nonlinear features. Deep learning methods have demonstrated strong performance across various fields and have shown promise in epilepsy classification. Therefore, we propose combining FFT feature extraction with a deep learning algorithm.

The structure of this paper is as follows: Section 2 gives a brief overview of the dataset, outlines the proposed method, and introduces the classifier used. Section 3 presents the results and compares them with other methods. Section 4 discusses the proposed approach, while section 5 highlights the main conclusions, contributions, and potential future directions.

2 Materials and methods

2.1 Epilepsy dataset

The EEG dataset used for the epilepsy classification performance is from the University of Bonn (Andrzejak et al., 2001). This comprehensive dataset includes EEG signals from both healthy individuals and those with epilepsy, with recordings taken under various conditions such as eyes opened and closed, intracranial and extracranial potential, and interictal and ictal states. The dataset is divided into five subsets labeled as A, B, C, D, and E, each containing 100 single-channel EEG signal segments. Each signal segment is 23.6 s long and sampled at a rate of 173.61 Hz. Subsets A and B were recorded using surface EEG recordings from five healthy volunteers with eyes open and closed, respectively, follow the standard electrode placement scheme of the International 10–20 System. Subsets C, D, and E consist of intracranial recordings from five epileptic patients, with set D representing recordings from the epileptogenic zone, set C from the hippocampal formation of the opposite hemisphere, and set E exclusively containing seizure recordings. Subsets C and D correspond to epileptic interictal states, while set E captures ictal activity. Further details can be found in Table 1.

Each EEG set in the dataset contains 100 segments, each segment containing 4,096 points. However, since the classifier uses a CNN network, having more segments in the dataset is crucial for influencing the algorithm's performance. To address this issue, we divide each EEG segment into four epochs, each comprising 1,024 points. As a result, the original dataset transforms into one containing five classes (A, B, C, D, and E), with 400 segments each having 1,024 sampling points (Pachori and Patidar, 2014; Figure 1).

In order to determine the performance and accuracy of the epilepsy classification algorithm, 9 classifications are considered to be designed as follows, they are A vs. E, B vs. E, AB vs. E, C vs. E, D vs. E, CD vs. E, AB vs. CD, AB vs. CDE, and ABCD vs. E.

A vs. E and B vs. E can determine if eye closure or opening influences epilepsy detection. AB vs. E, A vs. E, and B vs. E can assess the impact of additional EEG data on epilepsy detection.

C vs. E evaluates the method's performance in distinguishing interictal from ictal patterns. D vs. E examines the method's effectiveness in classifying interictal from ictal patterns and exploring the relationship between brain activity and hippocampal formation in the opposite hemisphere. C vs. E and D vs. E can identify which EEG component (epileptic zone or opposite hemisphere) is more effective in classifying interictal and ictal patterns. C vs. E, D vs. E, and CD vs. E investigate the influence of additional EEG data on interictal-ictal detection.

AB vs. CD tests the method's ability to differentiate healthy volunteers from epileptic interictal patients. AB vs. CDE assesses the method's capability to distinguish healthy volunteers from epileptic patients. ABCD vs. E evaluates the method's capacity to differentiate seizure-free individuals from those experiencing seizures. These binary classification tasks are designed to enhance the effectiveness of the experiments.

All of these binary classification tasks are designed to enhance the effectiveness of the experiments.

2.2 Methods

The proposed automatic system for epilepsy classification is based on FFT feature extraction, CNN, and LSTM.

2.2.1 FFT

Three approaches are selected for comparison to determine an optimal method for binary classification: FFT, wavelet transformation (WT), and EMD features. The discussion section compares the proposed methods with other approaches to assess their performance.

The widely used convolution theorem asserts that circular convolutions in the spatial domain are equivalent to pointwise products in the Fourier domain. Matrix generation plays a crucial role in the proposed framework as a means of quantitatively describing EEG records. The information contained in the EEG record matrix is influenced by fast Fourier transformation (FFT) during classification tasks. The classical FFT comprehensively describes and analyzes EEG traces in the frequency domain (Samiee et al., 2015). To effectively extract valuable features from epilepsy EEG signals, the improved method of FFT is employed to convert an EEG signal into a matrix. The steps involved are outlined below:

Step 1: obtain the Fourier coefficient for a given signal $x(n)$ in the frequency range $[0,2\pi]$ using the discrete Fourier transform algorithm. The discrete Fourier transform is defined as equation (1):

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-i2\pi k\frac{n}{N}} \qquad 0\le k\le N-1, \tag{1}$$

where $X(k)$ are the discrete Fourier transform coefficients, M is the length of the input EEG.

Step 2: calculate the absolute values of the coefficients as $A_r=|X(k)|$.

Step 3: transform the A_k into the $m\times n$. Matrix form according to the sequential order of the sample points. The resulting matrix is then expressed as equation (2):

TABLE 1 Bonn epilepsy dataset.

Class	A	B	C	D	E
Description	Nonepileptic eyes opened	Nonepileptic eyes closed	Epileptic interictal, epileptogenic zone	Epileptic interictal, hippocampal	Epileptic ictal

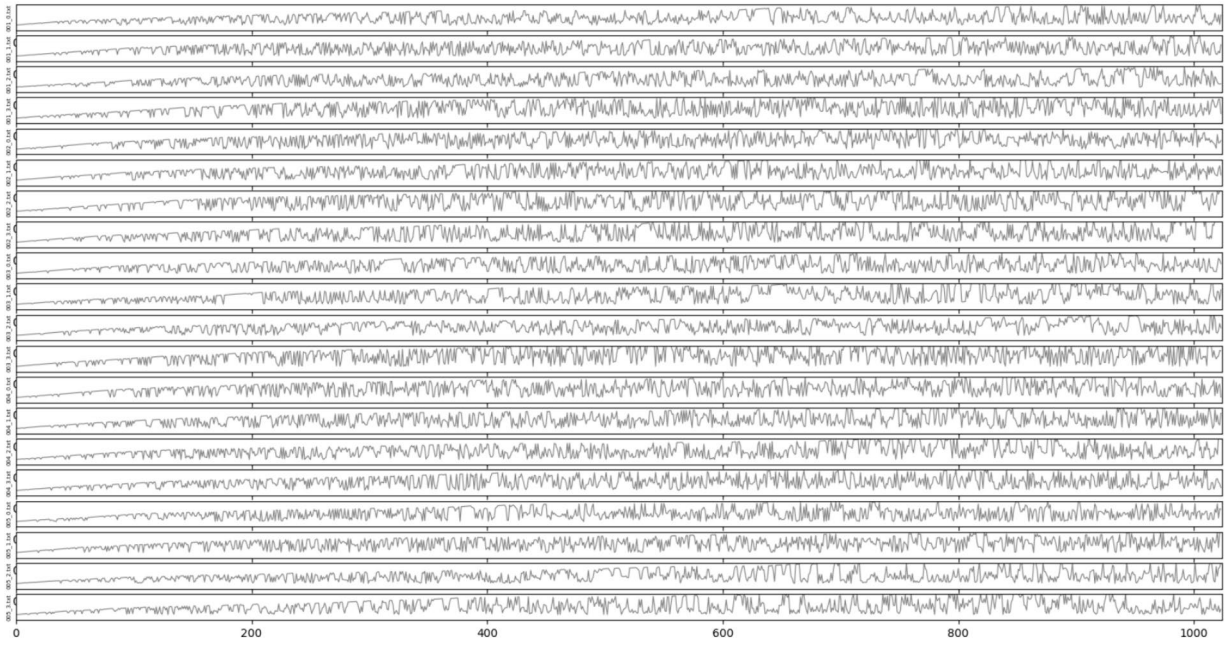


FIGURE 1
Signal display.

$$X = \begin{bmatrix} A_1 A_2 \cdots A_n \\ A_{n+1} A_{n+2} \cdots A_{n+n} \\ \vdots\vdots\vdots \\ A_{(m-1)n+1} A_{(m-1)n+2} \cdots A_{(m-1)n+n} \end{bmatrix} \quad (2)$$

where m and n are the matrix row and matrix column, respectively.

Extracting the FFT features is a crucial step, followed by utilizing these features as training data to train the classifier.

2.2.2 DWT

Wavelets can be defined as small waves with limited duration and an average value of 0. They are mathematical functions that can localize a function or data set in both time and frequency. The concept of wavelets can be traced back to Haar's thesis (Daubechies, 1992; Adeli et al., 2003) in 1909. The wavelet transform is a powerful tool in signal processing, known for its advantageous properties such as time-frequency localization (capturing a signal at specific time and frequency points, or extracting features at different spatial locations and scales) and multi-rate filtering (distinguishing signals with varying frequencies). By leveraging these properties, one can extract specific features from an input signal that exhibit distinct local characteristics in both time and space.

In continuous wavelet transform (CWT), the signal to be analyzed is matched and convolved with the wavelet basis function in a continuous sequence of time and frequency increments. Even in CWT, the data must be digitized. Continuous time and frequency increments mean that data at each digitized point or increment is used. Consequently, the original signal is represented as a weighted integral of the continuous basis wavelet function. In DWT, the basis wavelet function takes the original signal's inner product at discrete points (usually dyadic to ensure orthogonality). The result is a weighted sum of a series of base functions. The wavelet transform is based on the wavelet function, a family of functions that satisfy certain conditions, such as continuity, zero mean amplitude, and finite or near-finite duration.

The CWT of a square integrable function of time, $f(t)$, is defined as equation (3):

$$CWT_{a,b} = \int_{+\infty}^{-\infty} f(t) \frac{1}{\sqrt{|a|}} \psi * \left(\frac{t-b}{a} \right) dt \quad (3)$$

by Chui (1992), where $a, b \in R, a \neq 0$, R is the set of real numbers, the star symbol '*' denotes the complex conjugation. In CWT, the parameters a and b are continuously varying and can have infinite number of values to be taken, but this kind of computation cannot be done in finite time for modern computers. So we take a and b as discrete according to certain rules, which is DWT. If a expands exponentially, we define a as:

$$a = a_0^m$$

Since for wide wavelets we want to translate in larger steps, we can define b as:

$$b = nb_0 a_0^m, where\ b_0 > 0\ is\ fixed\ \ and\ n \in \mathbb{Z}$$

The wavelet function and the transform equation are given by the following two equations, respectively equations (4), (5):

$$\psi_{m,n}(k) = a_0^{-\frac{m}{2}} \psi\left(a_0^{-m}\left(k - nb_0 a_0^m\right)\right) \quad (4)$$

$$DWT_{m,n} = a_0^{-\frac{m}{2}} \sum_{k=-\infty}^{\infty} f(k) \cdot \psi * \left(a_0^{-m} k - nb_0\right) m, n \in \mathbb{Z} \quad (5)$$

2.2.3 EMD

The principle of the EMD technique is to automatically decompose a signal into a set of band-limited functions called Intrinsic Mode Functions (IMFs). Each IMF must satisfy two fundamental conditions (Huang et al., 1998; Bajaj and Pachori, 2012): (1) the number of extreme points and zero crossings in the entire dataset must either be equal or differ by at most one, and (2) the mean value of the envelopes defined by local maxima and minima must be zero at every point (Li et al., 2013).

The EMD is capable of decomposing a segment of EEG signal $x(n)$ into N IMFs: $imf_1, imf_2, \ldots, imf_n$ and a residue signal r. Therefore, $x(n)$ can be reconstructed as a linear combination equation (6):

$$x(n) = \sum_{n=1}^{N} imf_n + r \quad (6)$$

The following describes a systematic method for extracting IMFs:

Given an input signal $x(n), r(n) = x(n), n = 0.$

Step 1: determine the local maximum and local minimum of $x(n)$.

Step 2: determine the upper envelope $e_{\max}(n)$ by connecting all local maximum through cubic spline functions. Repeat the same procedure for the local minima to produce the lower envelope $e_{\min}(n)$.

Step 3: calculate the mean value for each point on the envelopes: $m(n) = \left(e_{\max}(n) + e_{\min}(n)\right)/2$.

Step 4: the equation $h(n) = x(n) - m(n)$, if $h(n)$ satisfies the IMF condition, then $n = n + 1, imf_n = h(n)$, go to step 5, else $x(n) = h(n)$, cycle 1–4.

Step 5: Let $r(n) = r(n) - imf_n$, if $r(n)$ is a monotonic function, end the sifting process, else, $x(n) = r(n)$ and go back to step 1.

The residue contains the lowest frequency. The main features of the ictal EEG are closely related to the first five IMFs. IMF1-IMF5 of each EEG segment is used to extract the EEG features.

2.2.4 CNN + nLSTM

Figure 2 displays the proposed automatic system for epilepsy detection, which is based on the fully-convolutional nested long short-term memory (FC-NLSTM) model.

Each EEG signal is initially segmented into a series of EEG segments, each segment containing M sampling points, by applying a fixed-length window that slides through the entire signal. Then filter the EEG signals using a Chebyshev bandpass filter with a cutoff frequency of 3–40 Hz. These EEG segments are then inputted into a fully convolutional network (FCN) with three convolutional blocks to learn the distinctive seizure characteristics present in the EEG data. The FCN serves as a feature extractor, effectively capturing the hierarchy features and internal structure of EEG signals. Subsequently, the features learned by the FCN are inputted into the NLSTM model to uncover the inherent temporal dependencies within the EEG signals. To extract the output characteristics of all NLSTM time steps, the time-distributed fully connected (FC) layer is used to take the outputs of all NLSTM time steps as inputs, rather than just the output of the last time step. Considering that all EEG segments should contribute equally to the label classification, a one-dimensional average pooling layer is added after the time-distributed fully connected layer. Finally, an FC layer is used for classification, and a softmax layer is employed to compute the probability that the EEG segment belongs to each class and predict the class of the input EEG segment (Li et al., 2020).

Temporal convolutional networks are widely used to analyze time-series signals, enabling the capture of how EEG signals evolve and automatic learning of EEG structures from data. The raw EEG signal comprises low-frequency characteristics with long periods and high-frequency characteristics with short periods (Adeli et al., 2003). It serves as a feature extraction module in the FCN and has been demonstrated as an effective method for time-series analysis problems (Wang et al., 2017). To prevent model overfitting to noise in the training data, this study maintains simplicity and shallowness in the FCN model, which includes three stacked convolutional blocks. Each of the three basic convolutional blocks consists of a convolution layer and a Rectified Linear Unit activation function.

According to the EEG recordings that are close to or even distant from the current EEG epoch, neurologists can determine whether the EEG epoch is a part of a seizure. Recurrent neural

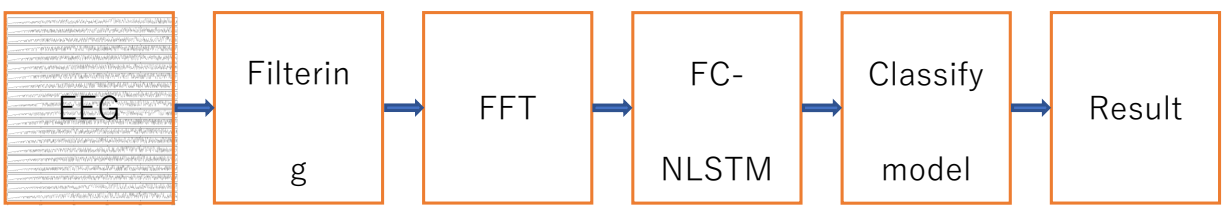


FIGURE 2
Flowchart of proposed method.

networks have made significant progress in emulating this human ability. A more intricate model called LSTM has been proposed based on the simple recurrent neural networks, which incorporates a memory mechanism and addresses the problem of vanishing gradients (Hochreiter and Schmidhuber, 1997). This memory mechanism allows the model to retain previous information from the EEG recordings. In this study, the FC-NLSTM is used to capture the temporal dependencies in EEG signals within the output of the feature extraction module.

2.2.5 Classification

The test data is inputted into the classification model for classification in this step. The 10-fold cross-validation method split the data into 10 parts, using 9 parts to train the model and reserving 1 part as the test set to evaluate the model's performance. This process is repeated 10 times to calculate the average sensitivity, specificity, and accuracy values.

FFT, DWT, and EMD are chosen as features for training and testing, with the results compared in part 3. Subsequently, the best-performing features were selected as the method feature and compared against the performance of existing methods.

2.3 Classifier result estimation

All the experiments results are based on the Bonn University database. The 10-fold cross-validation is used to reduce potential system errors, as well as to assess the stability and reliability of the proposed model.

The EEG data is evenly split into 10 subsets. Nine subsets are designated as training sets, while the remaining one is assigned to test the model. This iterative process is repeated 10 times, and the averaged values across these runs are computed. The performance assessment of the proposed method involves statistical evaluation measures such as sensitivity, specificity, and recognition accuracy.

Before delving into the statistical measures of sensitivity, specificity, and recognition accuracy, let us provide descriptions of four fundamental concepts:

True positive (TP): the number of positive (abnormal) examples classified as positive.

False negative (FN): the number of positive examples classified as negative (normal).

True negative (TN): the number of negative examples classified as negative.

False positive (FP): the number of negative examples classified as positive.

Sensitivity (Sen) is calculated by dividing true positive (TP) by the total number of seizure epochs identified by the experts. TP represents the seizure epochs marked as positive by both the classifier and EEG experts.

$Sen = TP/(TP + FN)$.

Specificity (Spe) is computed by dividing TN by the total number of non-seizure epochs identified by the experts. TN encapsulates the count of non-seizure epochs identified correctly.

$Spe = TN/(TN + FP)$.

Accuracy (Acc) is the number of correctly marked epochs divided by the total number of epochs.

$Acc = (TP + TN)/(TP + TN + FP + FN)$.

TABLE 2 Nine accuracy of three methods in different tasks.

Tasks	FFT	DWT	EMD
A vs. E	0.9962	0.8975	0.7312
B vs. E	0.9900	0.9425	0.5525
AB vs. E	0.9983	0.9658	0.7833
C vs. E	0.9913	0.9338	0.5000
D vs. E	0.9763	0.9400	0.8925
CD vs. E	0.9867	0.9533	0.6667
AB vs. CD	0.9906	0.8631	0.9569
ABCD vs. E	0.9815	0.9655	0.9915
AB vs. CDE	0.9795	0.8505	0.7890

3 Results

All experiments are performed in Python using Keras with TensorFlow backend and are implemented on an NVIDIA GeForce GTX1080-Ti GPU machine. In order to fully evaluate the performance of the proposed method in ideal and real situations, the University of Bonn database is used in this study.

All 9 tasks are tested in three methods. Table 2 shows that FFT and FC-NLSTM obtained the best accuracy in all tasks except ABCD vs. E. EMD performed poorly in every task except ABCD versus E. Therefore, FFT is selected as the optimal feature for comparison with other methods in subsequent sections.

3.1 Normal or interictal or non-ictal vs. ictal classification

Three types of data are used in the experiment. They include non-ictal vs. ictal(A vs. E, B vs. E, AB vs. E, C vs. E, D vs. E, CD vs. E, AB vs. CDE, ABCD vs. E), and normal vs. interictal (AB vs. CD).

The first three experiments compare non-ictal with ictal conditions, including A vs. E, B vs. E, and AB vs. E. The second set of three experiments compare non-ictal with ictal conditions including C vs. E, D vs. E, CD vs. E. The third experiment focuses on distinguishing between non-ictal and ictal states, classifying ABCD as seizure-free and E as seizure epilepsy. These experiments are conducted to validate the effectiveness and reliability of the proposed method.

Table 3 presents the results of the two-class seizure detection problem. As shown in this table, the proposed method demonstrates excellent classification performance across all normal vs. ictal scenarios, achieving nearly 100% sensitivity, specificity, and accuracy in some instances. Although not every fold in the 10-fold cross-validation reaches 100%, the mean sensitivity, specificity, and accuracy values exceed 99%. Notably, the specificity for A vs. E reaches 100%. In the interictal vs. ictal comparison, the proposed method also performs well, achieving 100% sensitivity, specificity, and accuracy in half of the folds in the 10-fold cross-validation. The highest sensitivity of 100% is achieved in the C vs. E experiment, with nearly 100% performance in terms of sensitivity, specificity, and accuracy in multiple folds for C vs. E, D vs. E, and CD vs. E. In the non-ictal vs. ictal

experiments ABCD vs. E, our method achieves a mean accuracy of 98.15%. All classification results exhibit an accuracy rate above 97.63%, demonstrating the robustness of our methods across various classification tasks. Among these experiments, the highest mean accuracy of 99.83% is observed in AB vs. E. Data imbalance is evident in these experiments, with the sensitivity, specificity, and accuracy in ABCD vs. E being lower compared to other experiments. The imbalance of non-ictal data segments in ABCD vs. E is four times greater than A vs. E, B vs. E, C vs. E, D vs. E, and twice as much as AB vs. E and CD vs. E. In this case, the traditional machine learning approaches may struggle to predict the minority classes (Kundu et al., 2013; Hussein et al., 2019). However, our methods continue to perform well under these conditions, without additional operations in our experiment. The 10-fold cross-validation thoroughly validates the method and mitigates the randomness of these experiments.

3.2 Normal vs. epileptic classification

In this section, we discuss two types of epilepsy classification problems to demonstrate the effectiveness and robustness of our proposed method, which includes two experiments comparing normal vs. interictal and normal vs. interictal and ictal cases. The former experiments are AB vs. CD, while the latter compares AB vs. CDE. Table 4 presents the classification results of sensitivity, specificity, and accuracy obtained through 10-fold cross-validation. In our experiment comparing normal vs. interictal (AB vs. CD), our methods achieve mean accuracy, sensitivity and specificity of 99.06, 98.87, and 99.25%, respectively. The comparison between normal vs. interictal and ictal cases yields a mean accuracy of 97.95%, mean sensitivity of 97.58%, and mean specificity of 98.50%.

Every aspect of the AB vs. CD comparison is superior to the AB vs. CDE comparison. The key to this difference lies in the use of

TABLE 3 The results of 10-fold cross-validation for non-ictal vs. ictal based on the Bonn University database.

		*K*1	*K*2	*K*3	*K*4	*K*5	*K*6	*K*7	*K*8	*K*9	*K*10	Mean
A vs. E	Acc	1.0000	1.0000	1.0000	0.9875	0.9875	0.9875	1.0000	1.0000	1.0000	1.0000	0.9962
	Sen	1.0000	1.0000	1.0000	0.9756	0.9756	0.9756	1.0000	1.0000	1.0000	1.0000	0.9927
	Spe	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
B vs. E	Acc	1.0000	0.9875	0.9875	0.9625	0.9875	1.0000	0.9750	1.0000	1.0000	1.0000	0.9900
	Sen	1.0000	1.0000	0.9750	1.0000	0.9750	1.0000	1.0000	1.0000	1.0000	1.0000	0.9950
	Spe	1.0000	0.9750	1.0000	0.9250	1.0000	1.0000	0.9500	1.0000	1.0000	1.0000	0.9850
AB vs. E	Acc	1.0000	1.0000	0.9917	1.0000	0.9917	1.0000	1.0000	1.0000	1.0000	1.0000	0.9983
	Sen	1.0000	1.0000	0.9750	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9975
	Spe	1.0000	1.0000	1.0000	1.0000	0.9875	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988
C vs. E	Acc	1.0000	0.9750	1.0000	1.0000	0.9750	0.9875	0.9875	1.0000	1.0000	1.0000	0.9875
	Sen	1.0000	0.9750	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Spe	1.0000	0.9750	1.0000	1.0000	0.9500	0.9750	0.9750	1.0000	1.0000	1.0000	0.9750
D vs. E	Acc	0.9625	1.0000	0.9750	1.0000	0.9375	0.9875	0.9875	0.9500	0.9750	0.9875	0.9763
	Sen	0.9750	1.0000	0.9750	1.0000	1.0000	1.0000	0.9750	1.0000	1.0000	0.9750	0.9900
	Spe	0.9500	1.0000	0.9750	1.0000	0.8750	0.9750	1.0000	0.9000	0.9500	1.0000	0.9625
CD vs. E	Acc	0.9583	0.9750	1.0000	1.0000	1.0000	0.9833	0.9833	1.0000	0.9833	0.9833	0.9867
	Sen	1.0000	1.0000	1.0000	1.0000	1.0000	0.9500	0.9750	1.0000	0.9750	0.9750	0.9875
	Spe	0.9375	0.9625	1.0000	1.0000	1.0000	1.0000	0.9875	1.0000	0.9875	0.9875	0.9863
ABCD vs. E	Acc	0.9950	0.9950	0.9850	0.9500	1.0000	0.9900	0.9550	0.9900	0.9650	0.9900	0.9815
	Sen	0.9750	0.9750	0.9500	0.8000	1.0000	0.9750	0.8000	0.9500	0.9250	0.9500	0.9300
	Spe	1.0000	1.0000	0.9938	0.9875	1.0000	0.9938	0.9938	1.0000	0.9750	1.0000	0.9944

TABLE 4 Results of 10-fold cross-validation for normal vs. interictal and normal vs. interictal and ictal based on the Bonn University database.

		*K*1	*K*2	*K*3	*K*4	*K*5	*K*6	*K*7	*K*8	*K*9	*K*10	Mean
AB vs. CD	Acc	0.9812	0.9875	0.9875	1.0000	0.9938	0.9875	0.9938	1.0000	0.9938	0.9812	0.9906
	Sen	0.9875	0.9750	0.9875	1.0000	0.9875	0.9875	0.9875	1.0000	1.0000	0.9750	0.9887
	Spe	0.9750	1.0000	0.9875	1.0000	1.0000	0.9875	1.0000	1.0000	0.9875	0.9875	0.9925
AB vs. CDE	Acc	0.9800	0.9650	0.9850	0.9800	0.9900	0.9700	0.9950	0.9800	0.9750	0.9750	0.9795
	Sen	0.9833	0.9583	0.9750	0.9750	0.9833	0.9750	1.0000	0.9667	0.9667	0.9750	0.9758
	Spe	0.9750	0.9750	1.0000	0.9875	1.0000	0.9625	0.9875	1.0000	0.9875	0.9750	0.9850

different data. The combination of ictal and interictal segments and interictal reduces the accuracy, sensitivity and specificity. Conversely, AB vs. E (in Table 2) achieves better results than AB vs. CDE across all evaluation metrics, with accuracy at 99.67%, sensitivity at 99.27%, and specificity at 100.00%. Ictal segments are easier to detect than interictal segments, as evidenced by the superior classification results of the AB vs. E compared to AB vs. CD. These three experiments (AB vs. E, AB vs. CD, AB vs. CDE) demonstrate that ictal segments have greater discriminative power than interictal segments, and the combination of both types makes it more challenging to classify them from normal segments. The experimental results indicate that the proposed method performs well in distinguishing non-ictal from ictal segments and excels in classifying interictal vs. ictal and normal vs. interictal and ictal segments.

4 Discussion

In this study, the deep learning model NLSTM uses FFT as a feature to classify epilepsy segments from normal or interictal segments or a combination of both. The model demonstrates excellent accuracy, sensitivity, and specificity in the Bonn University database. The effectiveness of our approach is validated through 9 experiments presented in Table 2. FFT is employed as a feature within the model and integrated with fully convolutional deep learning and long short-term memory to differentiate between ictal and non-ictal segments. This method uses the FFT features derived from the original EEG data.

The deep learning framework model can effectively learn overall features. The low-level layers of a FCN can capture the internal structure of EEG segments and then transmit them to the higher-level layers of the model for further processing. Subsequently, these EEG features are used to extract the temporal information by being passed to the NLSTM. The NLSTM differs from standard LSTM and the stacked LSTM models in that it enhances the depth of LSTM by nesting to select pertinent information from the EEG segments. In the traditional stacked LSTM architecture, several standard LSTM units are combined into a whole, with the processing outcome of this step serving as the input for the subsequent units. Conversely, the NLSTM structure employs external memory cells to select and process EEG segments, while internal memory cells are responsible for storing and processing them. These two modules are interdependent, with the internal module using the output of the external module as input data. This configuration demonstrates strong performance in capturing the long-term dependencies present in EEG signals.

Most epilepsy detection methods typically involve the extraction or design of features by humans to characterize epilepsy EEG. Subsequently, selection algorithms are applied to identify the most representative features for classification using various classifiers. However, these methods are often complex and time-consuming due to the search for suitable features. In contrast, deep learning frameworks, such as our approach, streamline the process by bypassing feature extraction or automating it, eliminating the need for manual feature selection common in traditional methods. This approach enables the extraction of EEG segment features without human intervention, facilitating the classification of segments into ictal or non-ictal categories. Implementing this method in medical settings alleviates the workload of neurologists by simplifying EEG graph interpretation, thereby reducing the expertise threshold and saving time for healthcare professionals.

Different lengths of EEG segments significantly affect the accuracy of normal vs. interictal vs. ictal problems, which has been demonstrated by Li et al. (2020) that the EEG segment length of 1,024 allows the method to achieve optimal accuracy. This result is verified in the three databases, which include the Bonn University database, the Freiburg Hospital database, and the CHB–MIT database.

There are many methods that have shown good performance in two-class seizure recognition problems. It is necessary and important to compare the accuracy with other research results. The results are compared in Table 5, which consists of three columns containing information on tasks, methods, and the accuracy of the classification experiments. This table includes 9 experiments conducted using the Bonn University database. Our method demonstrates higher accuracy than many other methods across all experiments. Bhattacharyya et al. (2017) used the tunable-Q wavelet transform (TQWT) to extract EEG features, which were then processed using a wrapper-based feature selection method and inputted into an SVM for the identification of ictal EEGs. They achieved 100% accuracy in A vs. E and B vs. E, and 99.5% accuracy in C vs. E. From Table 5, we can see that our method has a good performance in all 9 experiments. Kaya and Ertuğrul (2018) achieved 100% accuracy in A vs. E, but did not perform well in other tasks. Li et al. (2020) achieved 100% accuracy in A vs. E, B vs. E, and CD vs. E. Sharma et al. (2017) and Tuncer et al. (2021) both achieved 100% accuracy in B vs. E. Sharma et al. (2017) also achieved the same accuracy in AB vs. E. Our method demonstrates good performance across all nine classification tasks and achieves a classification accuracy of 99.06% in AB vs. CD.

Table 6 presents the comparative results of statistical differences found in the classification tasks for various small datasets within the Bonn dataset. The performance in A vs. E, AB vs. E, C vs. E, and AB vs. CD is better, while D vs. E and AB vs. CDE show poorer results. The variation in differentiation among these small datasets is influenced by the nature of their data, with some showing greater differentiation and others showing slightly weaker differentiation.

5 Conclusion

In order to promote the application of epilepsy detection in medical practice, the integration of FFT and fully convolutional NLSTM is used in classification. The time domain of the EEG signal transforms into the frequency domain using FFT methods. The data is then divided into training and testing parts, with the former being put into NLSTM to train classification model, and the other parts being put into the classification model to classify them as normal, interictal and ictal categories. Additionally, EMD and WT and FFT are employed as data processing methods to determine the most suitable type for NLSTM, with accuracy, sensitivity and specificity serving as evaluation metrics. Among the 9 experiments conducted, the FFT method yields the best results, confirming the approach as FFT and FC-NLSTM.

In the discussion section, we compare the results with other methods. Our method achieves an accuracy rate exceeding 97.00% across all experiments. The accuracies of 99.62, 99.00, 99.83, 99.13, 97.63, 98.67, 99.06, 98.15 and 97.95% are calculated for the cases A vs. E, B vs. E, AB vs. E, C vs. E, D vs. E, CD vs. E, AB vs. CD, ABCD vs. E and AB vs. CDE, respectively. The accuracy of 6 experiments exceeds 99.00%. These comparative results demonstrate the effectiveness of our method. They indicate its potential for automated epilepsy detection.

TABLE 5 Comparison results for A vs. E, B vs. E, AB vs. E, C vs. E, D vs. E, CD vs. E, AB vs. CDE, ABCD vs. E, AB vs. CD class recognition.

Task	Sample size	Method	10-fold CV	Acc (%)	Our Acc(%)	*p*-value
A vs. E	800	Siuly et al. (2018)	No	99.5	**99.62**	/
		Tuncer et al. (2021)	Yes	99.5		/
		Zhu et al. (2014)	Yes	99		/
		Kaya and Ertuğrul (2018)	Yes	100		/
		Kaya et al. (2014)	Yes	99.5		/
		Fathima et al. (2011)	No	99.75		/
		Das et al. (2016)	No	100		/
		Al Ghayab et al. (2016)	No	99.9		/
		Fu et al. (2014)	Yes	99.13		/
		Bhattacharyya et al. (2017)	Yes	100		/
		Yuan et al. (2014)	Yes	98.63		/
		Tawfik et al. (2016)	Yes	99.5		/
		Ullah et al. (2018)	Yes	99.9		0.2385
		Li et al. (2020)	Yes	100		0.0652
B vs. E	800	Siuly et al. (2018)	No	99	**99.00**	/
		Tuncer et al. (2021)	Yes	100.0		/
		Zhu et al. (2014)	Yes	97		/
		Kaya and Ertuğrul (2018)	Yes	97.5		/
		Wang et al. (2016)	Yes	95		/
		Richhariya and Tanveer (2018)	Yes	95.0		/
		Sharma et al. (2017)	Yes	100		/
		Bhattacharyya et al. (2017)	Yes	100		/
		Swami et al. (2016)	Yes	98.9		/
		Li et al. (2020)	Yes	100		0.0248
		Ullah et al. (2018)	Yes	99		0.9878
AB vs. E	1,200	Sharma et al. (2017)	Yes	100	**99.83**	/
		Ullah et al. (2018)	Yes	99.8		0.0477
		Li et al. (2020)	Yes	100		0.1510
C vs. E	800	Siuly et al. (2018)	No	98.5	**99.13**	/
		Tuncer et al. (2021)	Yes	100.0		/
		Zhu et al. (2014)	Yes	98		/
		Kaya and Ertuğrul (2018)	Yes	97.5		/
		Das et al. (2016)	No	100		/
		Bhattacharyya et al. (2017)	Yes	99.5		/
		Samiee et al. (2015)	No	98.5		/
		Li et al. (2020)	Yes	99.75		0.2457
		Ullah et al. (2018)	Yes	98.1		0.1832
D vs. E	800	Siuly et al. (2018)	No	97.5	**97.63**	/
		Tuncer et al. (2021)	Yes	99.0		/
		Zhu et al. (2014)	Yes	93		/
		Kaya and Ertuğrul (2018)	Yes	94.5		/
		Das et al. (2016)	No	100		/

(Continued)

TABLE 5 (Continued)

Task	Sample size	Method	10-fold CV	Acc (%)	Our Acc(%)	*p*-value
		Nicolaou and Georgiou (2012)	No	83.13		/
		Kumar et al. (2014)	Yes	93		/
		Wang et al. (2013)	No	97.58		/
		Sharma et al. (2017)	Yes	98.5		/
		Ullah et al. (2018)	Yes	99.4		0.8077
		Li et al. (2020)	Yes	99.88		0.0035
CD vs. E	1,200	Sharmila and Geethanjali (2020)	No	98.8	**98.67**	/
		Ullah et al. (2018)	Yes	99.7		0.8850
		Li et al. (2020)	Yes	100		0.0065
ABCD vs. E	2000	Swami et al. (2016)	Yes	95.2	**98.15**	/
		Orhan et al. (2011)	Yes	99.6		/
		Das et al. (2016)	No	100		/
		Hussein et al. (2019)	Yes	100		/
		Li et al. (2020)	Yes	99.9		0.0071
AB vs. CD	1,600	Ullah et al. (2018)	Yes	99.8	**99.06**	0.0471
		Sharma et al. (2017)	Yes	92.5		/
		Li et al. (2020)	Yes	98.44		0.6828
AB vs. CDE	2000	Ullah et al. (2018)	Yes	99.5	**97.95**	0.8109
		Li et al. (2020)	Yes	99.65		0.0014

Bold indicates emphasis on our accuracy.

Furthermore, this model and its framework can be used for EEG signal classification, which offers practical benefits in epilepsy detection. Its performance allows not only the classification of normal vs. ictal states, but also normal vs. interictal and interictal vs. ictal states.

In future work, it is advisable to consider using large datasets, such as the Freiburg hospital database and the CHB-MIT scalp EEG database, to improve the generalizability of the method and facilitate the development of a successful model. The integration of real-time applications has the potential to greatly impact clinical practice. In addition, it is recognized that deep learning approaches have difficulty providing explanations for decisions. Therefore, novel and explainable methods may need to be proposed to effectively address the epilepsy classification problem.

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

JN: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HS: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. FW: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grants 61271312, and in part by the innovation project of Jiangsu Province under grants BZ2023042, BY2022564.

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

TABLE 6 Comparison of differentiation under different datasets.

Number	Group1	Group2	p-Value
1	A vs. E	B vs. E	0.1825
2	A vs. E	AB vs. E	0.3562
3	A vs. E	C vs. E	0.3419
4	A vs. E	D vs. E	0.0091
5	A vs. E	CD vs. E	0.0581
6	A vs. E	AB vs. CD	0.0241
7	A vs. E	ABCD vs. E	0.0659
8	A vs. E	AB vs. CDE	0.0001
9	B vs. E	AB vs. E	0.0642
10	B vs. E	C vs. E	0.641
11	B vs. E	D vs. E	0.0925
12	B vs. E	CD vs. E	0.581
13	B vs. E	AB vs. CD	0.2399
14	B vs. E	ABCD vs. E	0.8928
15	B vs. E	AB vs. CDE	0.0488
16	AB vs. E	C vs. E	0.1137
17	AB vs. E	D vs. E	0.0039
18	AB vs. E	CD vs. E	0.0178
19	AB vs. E	AB vs. CD	0.0093
20	AB vs. E	ABCD vs. E	0.005
21	AB vs. E	AB vs. CDE	0
22	C vs. E	D vs. E	0.0407
23	C vs. E	CD vs. E	0.2994
24	C vs. E	AB vs. CD	0.1121
25	C vs. E	ABCD vs. E	0.6426
26	C vs. E	AB vs. CDE	0.0082
27	D vs. E	CD vs. E	0.2034
28	D vs. E	AB vs. CD	0.5532
29	D vs. E	ABCD vs. E	0.0521
30	D vs. E	AB vs. CDE	0.6553
31	CD vs. E	AB vs. CD	0.4805
32	CD vs. E	ABCD vs. E	0.4219
33	CD vs. E	AB vs. CDE	0.1848
34	AB vs. CD	ABCD vs. E	0.1499
35	AB vs. CD	AB vs. CDE	0.7563
36	ABCD vs. E	AB vs. CDE	0.0057

References

Abualsaud, K., Mahmuddin, M., Saleh, M., and Mohamed, A. (2015). Ensemble classifer for epileptic seizure detection for imperfect EEG data. *Sci. World J.* 2015, 1–15. doi: 10.1155/2015/945689

Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K.-H., and Suri, J. S. (2012). Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control* 7, 401–408. doi: 10.1016/j.bspc.2011.07.007

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* 100, 270–278. doi: 10.1016/j.compbiomed.2017.09.017

Adeli, H., Zhou, Z., and Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform[J]. *J. Neurosci. Methods* 123, 69–87. doi: 10.1016/S0165-0270(02)00340-0

Al Ghayab, H. R., Li, Y., Abdulla, S., Diykh, M., and Wan, X. (2016). Classification of epileptic EEG signals based on simple random sampling and sequential feature selection. *Brain Inform.* 3, 85–91. doi: 10.1007/s40708-016-0039-1

Altaf, M. A. B., and Yoo, J. (2015). A 1.83μ/classification, 8-channel, patient-specific epileptic seizure classification SoC using a non-linear support vector machine, IEEE Transactions on Biomedical Circuits and Systems. *Biomed. Circuits Syst.* 10, 49–60. doi: 10.1109/TBCAS.2014.2386891

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev. E* 64:061907. doi: 10.1103/PhysRevE.64.061907

Bajaj, V., and Pachori, R. B. (2012). Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Trans. Inf. Technol. Biomed.* 16, 1135–1142. doi: 10.1109/TITB.2011.2181403

Bhattacharyya, A., Pachori, R., Upadhyay, A., and Acharya, U. (2017). Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals. *Appl. Sci.* 7:385. doi: 10.3390/app7040385

Chui, C. K. (1992). An introduction to wavelets. San Diego, CA: Academic Press Inc.

Covert, I. C., Krishnan, B., Najm, I., Zhan, J., Shore, M., Hixson, J., et al. (2019). Temporal graph convolutional networks for automatic seizure detection[C]//machine learning for healthcare conference. *PMLR* 106, 160–180. doi: 10.48550/arXiv.1905.01375

Das, A. B., Bhuiyan, M. I. H., and Alam, S. M. S. (2016). Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection. *SIViP* 10, 259–266. doi: 10.1007/s11760-014-0736-2

Daubechies, I. (1992). Ten lectures on wavelets. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Fathima, T., Bedeeuzzaman, M., Farooq, O., and Khan, Y. U. (2011). Wavelet based features for epileptic seizure detection. *MES J. Technol. Manag.* 2, 108–112.

Feng, L., Li, Z., and Wang, Y. (2017). VLSI design of SVM-based seizure detection system with on-chip learning capability. *IEEE Trans. Biomed. Circuits Syst.* 12, 171–181. doi: 10.1109/TBCAS.2017.2762721

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., et al. (2014). ILAE offcial report: a practical clinical definition of epilepsy. *Epilepsia* 55, 475–482. doi: 10.1111/epi.12550

Fu, K., Qu, J., Chai, Y., and Dong, Y. (2014). Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomed. Signal Process. Control* 13, 15–22. doi: 10.1016/j.bspc.2014.03.007

Gao, X., Yan, X., Gao, P., Gao, V., and Zhang, S. (2020). Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks. *Artif. Intell. Med.* 102:101711. doi: 10.1016/j.artmed.2019.101711

Ghosh-Dastidar, S., and Adeli, H. (2009). A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. *Neural Netw.* 22, 1419–1431. doi: 10.1016/j.neunet.2009.04.003

Goksu, H. (2018). EEG based epileptiform pattern recognition inside and outside the seizure states. *Biomed. Signal Process. Control* 43, 204–215. doi: 10.1016/j.bspc.2018.03.004

Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr. Clin. Neurophysiol.* 54, 530–540. doi: 10.1016/0013-4694(82)90038-4

Gotman, J. (1990). Automatic seizure detection: improvements and evaluation. *Electroencephalogr. Clin. Neurophysiol.* 76, 317–324. doi: 10.1016/0013-4694(90)90032-F

Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., and Pazos, A. (2011). Automatic feature extraction using genetic programming: an application to epileptic EEG classification. *Expert Syst. Appl.* 38, 10425–10436. doi: 10.1016/j.eswa.2011.02.118

Hassan, A. R., Siuly, S., and Zhang, Y. (2016). Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. *Comput. Methods Prog. Biomed.* 137, 247–259. doi: 10.1016/j.cmpb.2016.09.008

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London* 454, 903–995. doi: 10.1098/rspa.1998.0193

Hussein, R., Palangi, H., Ward, R. K., and Wang, Z. J. (2019). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clin. Neurophysiol.* 130, 25–37. doi: 10.1016/j.clinph.2018.10.010

Indira, P. B., and Krishna, R. D. (2021). Optimized adaptive neuro fuzzy inference system (OANFIS) based EEG signal analysis for seizure recognition on FPGA. *Biomed. Signal Process. Control* 66:102484. doi: 10.1016/j.bspc.2021.102484

Jaiswal, A. K., and Banka, H. (2017). Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals. *Biomed. Signal Process. Control* 34, 81–92. doi: 10.1016/j.bspc.2017.01.005

Joshi, V., Pachori, R. B., and Vijesh, A. (2014). Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomed. Signal Process. Control* 9, 1–5. doi: 10.1016/j.bspc.2013.08.006

Kaleem, M., Gurve, D., Guergachi, A., and Krishnan, S. (2018). Patient-specific seizure detection in long-term EEG using signal-derived empirical mode decomposition (EMD)- based dictionary approach. *J. Neural Eng.* 15:056004. doi: 10.1088/1741-2552/aaceb1

Kaya, Y., and Ertuğrul, Ö. F. (2018). A stable feature extraction method in classification epileptic EEG signals. *Australas. Phys. Eng. Sci. Med.* 41, 721–730. doi: 10.1007/s13246-018-0669-0

Kaya, Y., Uyar, M., Tekin, R., and Yıldırım, S. (2014). 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. *Appl. Math. Comput.* 243, 209–219. doi: 10.1016/j.amc.2014.05.128

Kumar, Y., Dewal, M., and Anand, R. (2014). Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. *Neurocomputing* 133, 271–279. doi: 10.1016/j.neucom.2013.11.009

Kundu, K., Costa, F., Huber, M., Reth, M., and Backofen, R. (2013). Semi-supervised prediction of SH2-peptide interactions from imbalanced highthroughput data. *PLoS One* 8:e62732. doi: 10.1371/journal.pone.0062732

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. (2020). Automatic seizure detection using fully convolutional nested LSTM[J]. *Int. J. Neural Syst.* 30:2050019. doi: 10.1142/S0129065720500197

Li, S., Zhou, W., Yuan, Q., Geng, S., and Cai, D. (2013). Feature extraction and recognition of ictal EEG using EMD and SVM. *Comput. Biol. Med.* 43, 807–816. doi: 10.1016/j.compbiomed.2013.04.002

Nicolaou, N., and Georgiou, J. (2012). Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst. Appl.* 39, 202–209. doi: 10.1016/j.eswa.2011.07.008

Oliva, J. T., and Rosa, J. L. G. (2019). Classification for EEG report generation and epilepsy detection. *Neurocomputing* 335, 81–95. doi: 10.1016/j.neucom.2019.01.053

Oliva, J. T., and Rosa, J. L. G. (2021). Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. *Biomed. Signal Process. Control* 66:102469. doi: 10.1016/j.bspc.2021.102469

Orhan, U., Hekim, M., and Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst.Appl.* 38, 13475–13481. doi: 10.1016/j.eswa.2011.04.149

Ozdemir, M. A., Cura, O. K., and Akan, A. (2021). Epileptic EEG classification by using time-frequency images for deep learning. *Int. J. Neural Syst.* 31:2150026. doi: 10.1142/S012906572150026X

Pachori, R. B., and Patidar, S. (2014). Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. *Comput. Methods Prog. Biomed.* 113, 494–502. doi: 10.1016/j.cmpb.2013.11.014

Qaisar, S. M., and Hussain, S. F. (2021). Effective epileptic seizure detection by using level-crossing EEG sampling sub-bands statistical features selection and machine learning for mobile healthcare. *Comput. Methods Prog. Biomed.* 203:106034. doi: 10.1016/j.cmpb.2021.106034

Ren, W., and Han, M. (2019). Classification of EEG signals using hybrid feature extraction and ensemble extreme learning machine. *Neural. Process. Lett.* 50, 1281–1301. doi: 10.1007/s11063-018-9919-0

Richhariya, B., and Tanveer, M. (2018). EEG signal classification using universum support vector machine. *Expert Syst. Appl.* 106, 169–182. doi: 10.1016/j.eswa.2018.03.053

Samiee, K., Kovacs, P., and Gabbouj, M. (2015). Epileptic seizure classification of EEG timeseries using rational discrete short-time Fourier transform. *I.E.E.E. Trans. Biomed. Eng.* 62, 541–552. doi: 10.1109/TBME.2014.2360101

Sayeed, M. A., Mohanty, S. P., Kougianos, E., and Zaveri, H. P. (2019). Neuro-detect: a machine learning-based fast and accurate seizure detection system in the IoMT. *IEEE Trans. Consum. Electron.* 65, 359–368. doi: 10.1109/TCE.2019.2917895

Şengür, A., Guo, Y., and Akbulut, Y. (2016). Time–frequency texture descriptors of EEG signals for efficient detection of epileptic seizure. *Brain Inf.* 3, 101–108. doi: 10.1007/s40708-015-0029-8

Shanir, P. M., Khan, K. A., Khan, Y. U., Farooq, O., and Adeli, H. (2018). Automatic seizure detection based on morphological features using one-dimensional local binary pattern on long-term EEG. *Clin. EEG Neurosci.* 49, 351–362. doi: 10.1177/1550059417744890

Sharma, P., Khan, Y. U., Farooq, O., Tripathi, M., and Adeli, H. (2014). A wavelet-statistical features approach for nonconvulsive seizure detection. *Clin. EEG Neurosci.* 45, 274–284. doi: 10.1177/1550059414535465

Sharma, M., Pachori, R. B., and Acharya, U. R. (2017). A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recogn. Lett.* 94, 172–179. doi: 10.1016/j.patrec.2017.03.023

Sharmila, A., and Geethanjali, P. (2020). Evaluation of time domain features on detection of epileptic seizure from EEG signals. *Heal. Technol.* 10, 711–722. doi: 10.1007/s12553-019-00363-y

Shen, C. P., Lin, J. W., Lin, F. S., Lam, A. Y. Y., Chen, W., Zhou, W., et al. (2017). GA-SVM modeling of multiclass seizure detector in epilepsy analysis system using cloud computing. *Soft. Comput.* 21, 2139–2149. doi: 10.1007/s00500-015-1917-9

Sikdar, D., Roy, R., and Mahadevappa, M. (2018). Epilepsy and seizure characterisation by multifractal analysis of EEG subbands. *Biomed. Signal Process. Control* 41, 264–270. doi: 10.1016/j.bspc.2017.12.006

Singh, G., Kaur, M., and Singh, B. (2020). Detection of epileptic seizure EEG signal using multiscale entropies and complete ensemble empirical mode decomposition. *Wirel. Pers. Commun.* 116, 845–864. doi: 10.1007/s11277-020-07742-z

Siuly, S., Alcin OFBajaj, V., Sengur, A., and Zhang, Y. (2018). Exploring Hermite transformation in brain signal analysis for the detection of epileptic seizure. *IET Sci. Meas. Technol.* 13, 35–41. doi: 10.1049/iet-smt.2018.5358

Subasi, A., and Ismail Gursoy, M. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* 37, 8659–8666. doi: 10.1016/j.eswa.2010.06.065

Swami, P., Gandhi, T. K., Panigrahi, B. K., Tripathi, M., and Anand, S. (2016). A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Syst. Appl.* 56, 116–130. doi: 10.1016/j.eswa.2016.02.040

Tawfik, N. S., Youssef, S. M., and Kholief, M. (2016). A hybrid automated detection of epileptic seizures in EEG records. *Comput. Electr. Eng.* 53, 177–190. doi: 10.1016/j.compeleceng.2015.09.001

Truong, N. D., Nguyen, A. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., Ippolito, S., et al. (2018). Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Netw.* 105, 104–111. doi: 10.1016/j.neunet.2018.04.018

Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* 99, 24–37. doi: 10.1016/j.compbiomed.2018.05.019

Tuncer, T., Dogan, S., Naik, G. R., and Pławiak, P. (2021). Epilepsy attacks recognition based on 1D octal pattern, wavelet transform and EEG signals. *Multimed. Tools Appl.* 80, 25197–25218. doi: 10.1007/s11042-021-10882-4

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2007). "A time-frequency based method for the detection of epileptic seizures in EEG recordings" in Twentieth IEEE international symposium on computer-based medical systems (CBMS'07) (Maribor, Slovenia: IEEE), 135–140.

Ullah, I., Hussain, M., and Aboalsamh, H. (2018). An automated system for epilepsy detection using EEG brain signals based on deep learning approach. *Expert Syst. Appl.* 107, 61–71. doi: 10.1016/j.eswa.2018.04.021

Wang, Z., Yan, W., and Oates, T. (2017). Time series classification from scratch with deep neural networks: a strong baseline. *Int. Joint Conf. Neural Networks* 2017, 1578–1585. doi: 10.48550/arXiv.1611.06455

Wang, Y., Zhou, W., Yuan, Q., Li, X., Meng, Q., Zhao, X., et al. (2013). Comparison of ictal and interictal EEG signals using fractal features. *Int. J. Neural Syst.* 23:1350028. doi: 10.1142/S0129065713500287

Wang, H., Zhuo, G., and Zhang, Y. (2016). Analyzing EEG signal data for detection of epileptic seizure: introducing weight on visibility graph with complex network feature. *Austr. Database Conf.* 9877, 56–66. doi: 10.1007/978-3-319-46922-5_5

Wijayanto, I., Hadiyoso, S., Aulia, S., and Atmojo, B. S. (2020). Detecting ictal and Interictal condition of EEG signal using Higuchi fractal dimension and support vector machine. *J. Physics* 1577:012016. doi: 10.1088/1742-6596/1577/1/012016

World Health Organization (2021) Epilepsy. Available at: https://www.who.int/en/news-room/fact-sheets/detail/measles (Accessed March 30, 2022)

Wu, J. M-T., Tsai, M-H., Hsu, C-T., Huang, H-C., and Chen, H-C. (2019). Intelligent signal classifier for brain epileptic EEG based on decision tree, multilayer perceptron and over-sampling approach[C]//future of information and communication conference. Cham: Springer, 11–24.

Xiang, J., Li, C., Li, H., Cao, R., Wang, B., Han, X., et al. (2015). The detection of epileptic seizure signals based on fuzzy entropy. *J. Neurosci. Methods* 243, 18–25. doi: 10.1016/j.jneumeth.2015.01.015

Yavuz, E., Kasapbaşı, M. C., Eyüpoğlu, C., and Yazıcı, R. (2018). An epileptic seizure detection system based on cepstral analysis and generalized regression neural network, Biocybern. *Biomed. Eng.* 38, 201–216. doi: 10.1016/j.bbe.2018.01.002

Yuan, Y., Xun, G., Jia, K., and Zhang, A. (2017). A multi-view deep learning method for epileptic seizure detection using short-time Fourier transform, in Proc. 8th ACM Int. Conf. Bioinformatics, computational biology, and health informatics (ACM, New York)

Yuan, Q., Zhou, W., Yuan, S., Li, X., Wang, J., and Jia, G. (2014). Epileptic EEG classification based on kernel sparse representation[J]. *Int. J. Neural Syst.* 24:1450015. doi: 10.1142/S0129065714500154

Zeng, K., Yan, J., Wang, Y., Sik, A., Ouyang, G., and Li, X. (2016). Automatic detection of absence seizures with compressive sensing EEG. *Neurocomputing* 171, 497–502. doi: 10.1016/j.neucom.2015.06.076

Zhu, G., Li, Y., and Wen, P. P. (2014). Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm. *Comput. Methods Prog. Biomed.* 115, 64–75. doi: 10.1016/j.cmpb.2014.04.001

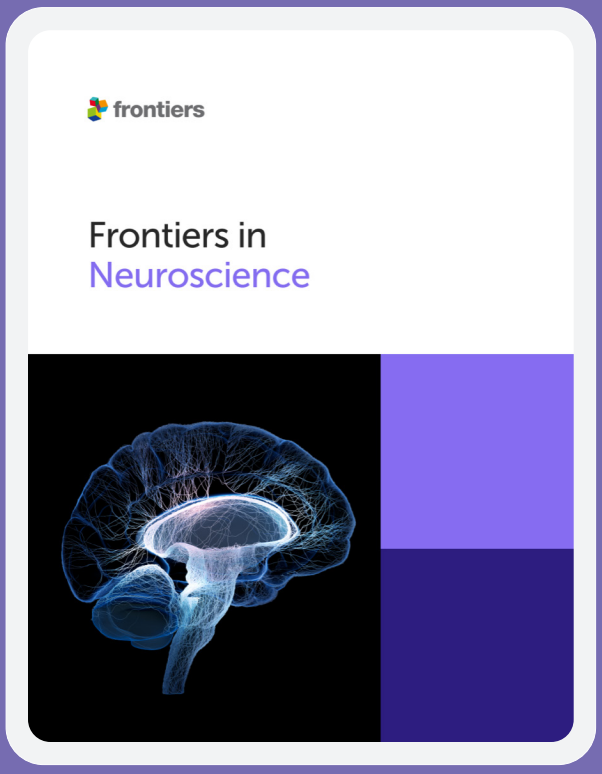
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact