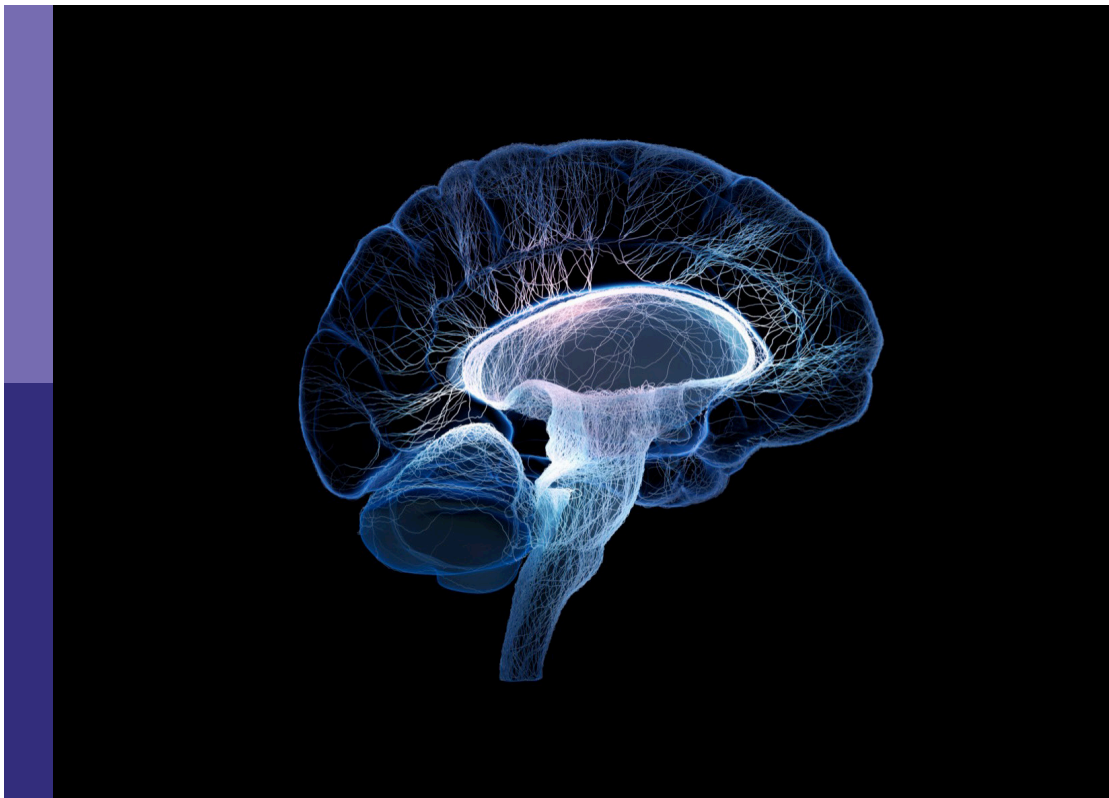
frontiers | Research Topics

Application of multimodal imaging combined with artificial intelligence in eye diseases

Edited by
Xin Huang, Yu-Chen Chen, Vijaya Prakash Krishnan Muthaiah and Zhi Wen

Published in
Frontiers in Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-3610-0
DOI 10.3389/978-2-8325-3610-0

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

Application of multimodal imaging combined with artificial intelligence in eye diseases

Topic editors

Xin Huang — Renmin Hospital of Wuhan University, China
Yu-Chen Chen — Nanjing Medical University, China
Vijaya Prakash Krishnan Muthaiah — University at Buffalo, United States
Zhi Wen — Renmin Hospital of Wuhan University, China

Citation

Huang, X., Chen, Y.-C., Muthaiah, V. P. K., Wen, Z., eds. (2023). *Application of multimodal imaging combined with artificial intelligence in eye diseases*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3610-0

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 19 September 2023
DOI 10.3389/fnins.2023.1287762

OPEN ACCESS

EDITED AND REVIEWED BY
Benjamin Thompson,
University of Waterloo, Canada

*CORRESPONDENCE
Xin Huang
✉ 334966891@qq.com

RECEIVED 02 September 2023
ACCEPTED 05 September 2023
PUBLISHED 19 September 2023

CITATION
Wen Z, Chen Y, Krishnan Muthaiah VP and Huang X (2023) Editorial: Application of multimodal imaging combined with artificial intelligence in eye diseases.
Front. Neurosci. 17:1287762.
doi: 10.3389/fnins.2023.1287762

Editorial: Application of multimodal imaging combined with artificial intelligence in eye diseases

Zhi Wen[1], Yuchen Chen[2], Vijaya Prakash Krishnan Muthaiah[3] and Xin Huang[4]*

[1]Department of Radiology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China, [2]Department of Radiology, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, China, [3]Department of Rehabilitation Sciences, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY, United States, [4]Department of Ophthalmology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, Jiangxi, China

KEYWORDS

eye disease, optic atrophy, visual pathway, OCT, OCTA, AI, fMRI

Editorial on the Research Topic
Application of multimodal imaging combined with artificial intelligence in eye diseases

Introduction

Vision is the most crucial sensory system for human beings to process external information. The retina, optic nerve, and visual pathway (lateral geniculate body, optic radiation, and visual cortex) are all necessary for the complete transmission of visual information. Numerous eye conditions, including glaucoma, diabetic retinopathy, optic neuritis, and hereditary optic neuropathy result in vision loss, and as a result, they remain a key field of study for both clinical and pre-clinical research. Multimodal imaging techniques, such as optical coherence tomography (OCT), optical coherence tomography angiography (OCTA), and functional magnetic resonance imaging (fMRI), give essential biological indicators in the diagnosis of ophthalmic diseases. In fact, combining imaging with artificial intelligence (AI) may enhance diagnostic precision and accuracy, detect imaging biomarkers, develop cutting-edge tools that will impact clinical practice and benefit patient outcomes.

This Research Topic on "*Application of multimodal imaging combined with artificial intelligence in eye diseases*" included 19 articles. It covered a range of AI applications in fundus images, OCT, anterior segment images, infrared videos from eye movement recorders, and steady-state visual evoked potentials (SSVEPs) in order to classify eye diseases, or to find neuroimaging indicators. Other papers reviewed the literature, identified indicators from animal model studies, and determined the prognostic factors of transnasal endoscopic optic decompression. In this editorial, we give an overview of these studies and group them according to the study design.

AI-aided diagnosis

Retinal atrophy is a crucial assessment indicator as it is correlated with the severity of myopia. For 2D fundus images, Chen et al. developed an attention-aware retinal atrophy segmentation network based on UNet structure called ARA-Net. To deal with blurred boundaries and irregular shapes of the retinal atrophic region, they proposed a novel skip self-attention (SSA) connection block. They also proposed a multi-scale feature flow (MSFF) between the SSA connection blocks, greatly enhancing the self-attention mechanism's capacity. The Pathological Myopia (PALM) dataset has been used to validate the proposed method, which performs noticeably better than other methods. A new deep learning model called MyopiaDETR was put up by Li M. et al. for the diagnosis of pathological myopia (PM) using 2D fundus image data. The architecture of Detection Transformer (DETR) allows it to effectively tackle the issue of morphology irregularity. An attentional feature pyramid network (FPN) increases the difference in feature intensity the between foreground and background. The experimental results show that their model outperforms other state-of-the-art object detectors in terms of localization and classification performance on the iChallenge-PM dataset.

Huang et al. introduces GABNet, a novel lightweight classification network model based on the global attention block (GAB) for OCT. By using their method, classification accuracy is increased over the EfficientNetV2B3 network model by 3.7%. Gan, Wu, et al. proposed an AI method based on multifeature fusion to enable automatic macular edema (ME) classification on spectral domain OCT (SD-OCT) images. With an accuracy of 93.8%, the support vector machine (SVM) model performed the best when compared to other classification models.

Gan, Liu, et al. and Li, Huang, Peng developed AI segmentation platforms with a deep transfer-learning algorithm and multi-feature fusion by using anterior segment images for automatic cortical cataract staging and fungal keratitis diagnosis, respectively. One is based on a method of automatic segmentation, whereas the other is based on a method of manual segmentation. While the automatic segmentation platform can stage cataracts and diagnose fungal keratitis more quickly, the manual segmentation platform can do so more accurately. In addition, Gan, Chen, et al. developed an AI model based on ensemble DL that was combined with four benchmark models (the Resnet18, Alexnet, Googlenet, and Vgg11) for identifying pterygium that need to be surgically removed. The ensemble DL model exceed all other models in terms of accuracy and area under the curve (AUC).

Li and Yang demonstrated that torsional nystagmus can be recognized by deep learning networks models. They used convolution neural network to extract the frame features of the infrared video sequence from eye movement recorders, and classified the obtained vector sequence.

For SSVEPs, Wan, Li, et al. proposed a transformer–based EEGformer analysis model to capture the electroencephalogram (EEG) characteristics in a unified manner. Across three EEG datasets [BETA, SJTU emotion EEG dataset (SEED), depressive EEG database (DepEEG)] the EEGformer achieves the best classification performance. This finding suggests that the rationality of model architecture and learning EEG characteristics in a unified manner can improve model classification performance. Another study by Wan, Cheng, et al. propose a deep neural network called GDNet-EEG for SSVEP stimulation frequency recognition that uses group depth-wise convolutional filtering to extract regional characteristics from raw EEG data. The findings show that GDNet-EEG surpasses the existing deep learning models to process EEG data on two publicly available SSVEPs datasets (largescale benchmark and BETA dataset) and their merged dataset.

For resting-state fMRI, Ji, Wang, et al. used the amplitude of low-frequency fluctuation (ALFF) in conjunction with sliding window approach to assess the changes of dynamic neural activity in patients with retinal detachment. Based on dALFF values, the overall accuracies of SVM classification were good under three different time windows. In patients with primary angle-closure glaucoma, Li, Huang, Peng, Liang, et al. identified changes in functional connectivity (FC) with primary visual cortex (V1), and found the increased FC between V1 and calcarine. However, the discrimination of PACG from healthy controls (HC) was poor when utilizing the SVM method and the dFC map as the classification feature.

Neuroimaging indicators

Ji, Huang, et al. explore differences in static FC (sFC) and dynamic FC (dFC) alteration patterns in the V1 among patients with high myopia and HCs via seed-based FC analysis. This disturbance suggests that patients with high myopia could exhibit impaired cognitive and emotional processing functions, top-down control of visual attention, and visual information processing functions. Pang et al. compared retinal OCT and optic nerve diffusion tensor imaging (DTI) parameters in patients with non-functioning pituitary adenoma (NFPA). The degree of adverse changes in OCT and DTI parameters was found to be stronger in the severe compression group than that in the mild compression group. Moreover, the fractional anisotropy (FA) value of the optic chiasma has a high diagnostic ability for visual pathway impairment. A literature review by Wang et al. described the use of DTI technology in glaucoma in humans and animal models, with the advancement of DTI technology and its coupling with artificial intelligence, DTI represents a potential future for MRI technology in glaucoma research.

Indicators from animal models

Accurate axial length (AL) measurement is crucial for developing animal models of myopia. The accuracy of Quantel A-B scan, OD-1 A scan, and vernier caliper were compared by Wu et al. for measuring AL in Sprague Dawley rats. They found that Quantel A-B scan might be more accurate than OD-1 A scan. AL and refractive error (RE) are both

influenced by gender. Duan et al. employed OCT to assess the changes in retina thickness after shape deprivation in myopic mice, and found considerably thinner retina, nerve fiber layer, inner nuclear layer, and outer nuclear layer. In addition, Shi et al. found epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) may have a role in the choroidal thickness regulation by using guinea pig model for myopia.

Last but not least, Tu et al. investigate the clinical effectiveness and prognostic variables of trans nasal endoscopic optic decompression in the treatment of traumatic optic neuropathy. The prognosis is highly dependent on the presence of residual light perception and the timing of surgery within 7 days.

As we draw to a close, we wish to emphasize the need of making more efforts to enhance the technical, clinical, and preclinical advances described in this Research Topic. In future, with the aid of AI, it will be feasible to provide a comprehensive description of eye diseases by integrating data from multidimensional modalities, as well as those from patients and animal models. Furthermore, it is hoped that the existing large-scale image dataset will be useful in this regard.

Author contributions

ZW: Writing—original draft. YC: Writing—review and editing. VK: Writing—review and editing. XH: Writing—review and editing.

Funding

This work was supported by the Natural Science Foundation of Jiangxi Province (grant no. 20212BAB216058), Jiangxi Provincial Health Technology Project (grant nos. 202210012 and 202310114), Jiangxi Provincial Traditional Chinese Technology Project (grant no. 2022B840), and Fundamental Research Funds for the Central Universities (grant no. 2042018kf0178).

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

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fnins.2022.1084118

OPEN ACCESS

EDITED BY
Zhi Wen,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Chen-Xing Qi,
Renmin Hospital of Wuhan University, China
Tianming Huo,
Wuhan University, China

*CORRESPONDENCE
Yu-Lin Zhong
✉ 804722489@qq.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 30 October 2022
ACCEPTED 02 December 2022
PUBLISHED 20 December 2022

CITATION
Gan F, Chen W-Y, Liu H and Zhong Y-L (2022) Application of artificial intelligence models for detecting the pterygium that requires surgical treatment based on anterior segment images. *Front. Neurosci.* 16:1084118. doi: 10.3389/fnins.2022.1084118

Application of artificial intelligence models for detecting the pterygium that requires surgical treatment based on anterior segment images

Fan Gan[1,2], Wan-Yun Chen[2], Hui Liu[2] and Yu-Lin Zhong[2]*

[1]Medical College of Nanchang University, Nanchang, China, [2]Department of Ophthalmology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, China

Background and aim: A pterygium is a common ocular surface disease, which not only affects facial appearance but can also grow into the tissue layer, causing astigmatism and vision loss. In this study, an artificial intelligence model was developed for detecting the pterygium that requires surgical treatment. The model was designed using ensemble deep learning (DL).

Methods: A total of 172 anterior segment images of pterygia were obtained from the Jiangxi Provincial People's Hospital (China) between 2017 and 2022. They were divided by a senior ophthalmologist into the non-surgery group and the surgery group. An artificial intelligence model was then developed based on ensemble DL, which was integrated with four benchmark models: the Resnet18, Alexnet, Googlenet, and Vgg11 model, for detecting the pterygium that requires surgical treatment, and Grad-CAM was used to visualize the DL process. Finally, the performance of the ensemble DL model was compared with the classical Resnet18 model, Alexnet model, Googlenet model, and Vgg11 model.

Results: The accuracy and area under the curve (AUC) of the ensemble DL model was higher than all of the other models. In the training set, the accuracy and AUC of the ensemble model was 94.20% and 0.978, respectively. In the testing set, the accuracy and AUC of the ensemble model was 94.12% and 0.980, respectively.

Conclusion: This study indicates that this ensemble DL model, coupled with the anterior segment images in our study, might be an automated and cost-saving alternative for detection of the pterygia that require surgery.

KEYWORDS

anterior segment images, artificial intelligence, deep learning, pterygium, surgery

Introduction

Pterygium is a common ocular surface lesion characterized by wing-shaped, fibrovascular conjunctival outgrowth that invades the clear cornea (Xue et al., 2014). The pathogenesis of pterygia is still not completely understood, several factors including ultraviolet radiation, immunoinflammatory process, virus infection, and genetic factors have been reported to be related to pterygial formation (Chen et al., 2010). A pterygium not only affects facial appearance but can also grow into the tissue layer, causing astigmatism and vision loss. A pterygium is divided commonly into active and stationary phases. In stationary phase, it is relatively flat, the color is light red or white, congestion is not obvious, and growth is slow. In active period, it is thickened, congestion is obvious, the color is red, and growth is faster. Surgical treatment should be administered in a timely manner for an active pterygium to prevent damage to the ocular surface caused by cytokines from pterygium (Liu et al., 2020).

In a clinical setting, ocular surface diseases are diagnosed by professional ophthalmologists based on anterior segment images. In primary medical institutions, such as community hospital, there are few professional ophthalmologists, so it is difficult to diagnose a pterygium, let alone identify a pterygium that requires surgical treatment. This might result in a delay to surgical intervention, and the optimal treatment time being missed. Unfortunately, surgical removal of an advanced pterygium also carries higher risk of post-operative complications, such as corneal scarring, post-operative complication-induced astigmatism, higher rates of recurrence, and so poorer prognosis (Fang et al., 2021). Therefore, it is necessary to detect and refer pterygium requiring surgical treatment timely. For this reason, an automated approach for detecting a pterygium that requires surgical treatment is needed.

With the development of artificial intelligence (AI), it is being applied increasingly in various healthcare disciplines, especially in fields, such as ophthalmology, in which medical image assessment has a key role (Tang et al., 2022). DL algorithms have been used widely to diagnose fundus diseases such as diabetic retinopathy (DR), diabetic macular edema (DME), central serous retinopathy (CSR), and age-related macular degeneration (AMD) (Gulshan et al., 2016; Syed et al., 2016; Chen et al., 2021; Tang et al., 2021) and has good results. Few previous studies using AI to diagnose ocular surface diseases. Pterygium is one of the most common ocular surface diseases. Zaki et al. (2018) used support vector machine (SVM) to detect pterygium. This study on pterygium detection was based on the two-class detection of pterygium and traditional machine learning. It has not been further determined whether the pterygium requires surgical treatment. Hung et al. (2022) used a multi-layer perception (MLP) model to perform pterygium grading and further predict surgical prognosis. Zheng et al. (2021) used MobileNet, AlexNet, VGG16, and the ResNet18 model for pterygium grading. These studies all based on the traditional DL models. The performance of these models still needs to be improved.

Ensemble learning is a kind of machine-learning paradigm in which multiple models, such as decision trees, neural networks, and support vector machine (SVM), are combined together to solve a particular problem (Wang et al., 2015). An ensemble of various machine-learning models could help to reduce the bias in a single machine-learning algorithm to provide a much better prediction performance than single models (Kong et al., 2021).

Therefore, in this study, an ensemble scheme consisting of four DL model and multilayer perceptron (MLP) classifier was designed with aim to detect a pterygium that requires surgical treatment. In anticipation of detecting and referring pterygium requiring surgical treatment timely in the clinic work applying in treatment selection in personalized precision therapy of the pterygium. Thus, we hypothesized that an ensemble of various machine-learning models could help to provide a much better prediction performance than single models.

Materials and methods

Dataset preparation

The total of 172 anterior segment images of pterygium were obtained from the Jiangxi Provincial People's Hospital between 2017 and 2022. All the images selected had high quality and were obtained from the same slit lamp digital microscopy. They were divided by a senior ophthalmologist into the non-surgery group and the surgery group. The classifying standard was as follows (Zheng et al., 2021). The anterior segment images of the non-surgery group was characterized by the horizontal length of the pterygium head tissue invading the limbus of the cornea <3 mm. The anterior segment images of the surgery group was characterized by the horizontal length of the pterygium head tissue invading the limbus of the cornea ≥ 3 mm. LabelMe software was applied to label the regions of interest (ROIs) of the anterior segment images. The non-surgery group was labeled "label 0" and the surgery group was labeled "label 1." To avoid over fitting problems, the dataset was established and divided by stratified sampling into training set (n = 136) and testing set (n = 34) at a ratio of 8:2 by referring to the previous research (Lu et al., 2022), ensuring that there was no overlap between the same person's data in the development and internal test sets. The anterior segment images of two group were shown in **Figure 1**.

Model development

This study developed five prediction models based on deep learning neural network models, Which were resnet18 model,

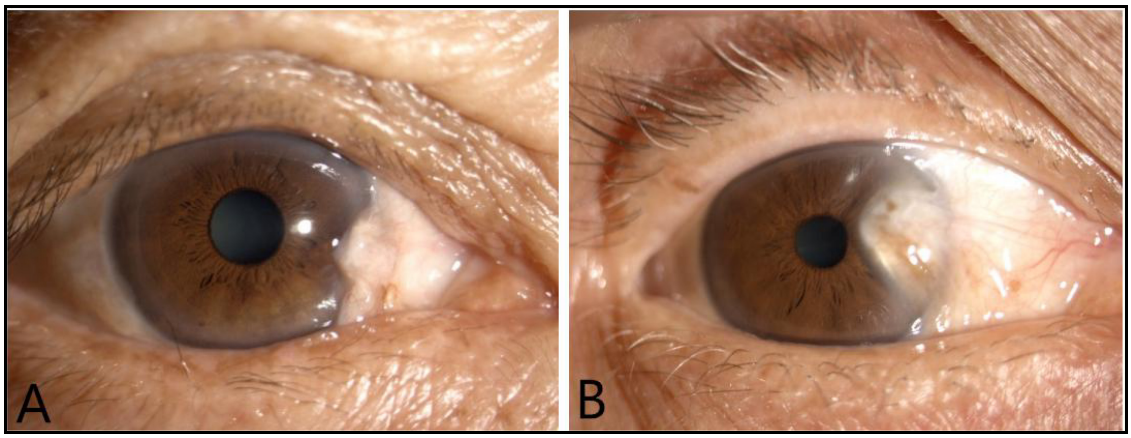


FIGURE 1
The anterior segment images of two group: The anterior segment image of the non-surgery group **(A)**. The anterior segment image of the surgery group **(B)**.

alexnet model, googlenet model, vgg11 model, and the ensemble deep learning (DL) model, respectively. The ensemble DL model was integrated with the alexnet, googlenet, vgg11, and resnet18 models. The flowchart of the ensemble DL model as shown in **Figure 2**.

During the training process of deep learning, the anterior segment images on the training set and the classification labels of pterygium was entered into the five deep learning models, separately. Convolutional neural networks (CNNs) coupled with batch normalization layers were used and the convolutional layer weights were initialized based on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset pre-trained models (Fang et al., 2021). And then these CNNS were used to extract features. Usually, lower CNN layers are used to extract abstract features like edges, and deeper CNN layers are used to find features that are informative for the target task (Shen et al., 2017). Deep feature on "avgpool" layer before last FC layers were extracted. These extracted features from each model were screened by Least absolute shrinkage selection operator (LASSO) method separately. Finally, these selected features were then used to build classification model through a multilayer perceptron (MLP) neural network. The LASSO method and MLP neural network were also a part of deep learning models. And then the models were applied to the test set.

Model visualization

The gradient-weighted class activation mapping (Grad-CAM) is a widely used method to interpret which features are responsible for determining outputs. It functions by capturing a specific class's vital features from the last Conv layer of a CNN model to localize its important regions (Chen et al., 2010; Kwon et al., 2020; Montalbo, 2021).

To understand which areas of the anterior segment images of our dataset were most likely to be used by the model to predict whether the pterygium requires surgical treatment. We used Grad-CAM for visualizing the filters of the penultimate layer of the deep learning process, that can highlight which parts of an image contribute to the deep learning models.

Statistical analysis

The accuracy, sensitivity, specificity, F1-score, confusion matrix and the area under the receiver operating characteristic curve (AUC) of the prediction models on the training set and test set were calculated, separately. And the predictive performances of the ensemble DL, resnet18, googlenet, alexnet, and vgg11 models were compared based on the above indicators.

Software

The anterior segment images labeled using the LabelMe Open Annotation Tool.[1] Deep learning techniques were completed using Python version 3.9.

Results

A total of 172 anterior segment images of pterygia are in our dataset. Of these, 95 images were labeled as the surgery group and 77 images were labeled as the non-surgery group. The dataset was divided by stratified sampling into training and

1 http://labelme.csail.mit.edu

testing sets at a ratio of 8:2. The 136 anterior segment images of pterygia in the training set were used to train the ensemble DL model, Resnet18 model, Googlenet model, Alexnet model, and Vgg11 model. The 35 images in the testing set were used to test the models.

The LASSO model was used to screen independent predicting features of ensemble DL model in training set. Each feature has a coefficient as its weight provided by LASSO, when the binomial deviance was minimized, features with non-zero coefficients were selected by optimal lambda. The optimal lambda was 0.068665, as shown in Figure 3.

Compared with other models, the ensemble DL model achieved the highest accuracy in both the training and testing sets. The accuracy was 94.20 and 94.12% in the training set and testing set, respectively. The accuracies of the Alexnet, Googlenet, Resnet18, and Vgg11 model in the training set were only 85.40, 84.67, 80.29, and 84.67%, respectively; whereas their accuracies were only 85.71, 80.00, 77.14, and 85.71% in the testing set. All results were shown in Table 1.

All the models had good results with AUC (Table 1 and Figure 4), of which the ensemble DL model had the highest AUC in both the training and testing set, of 0.978 and 0.980,

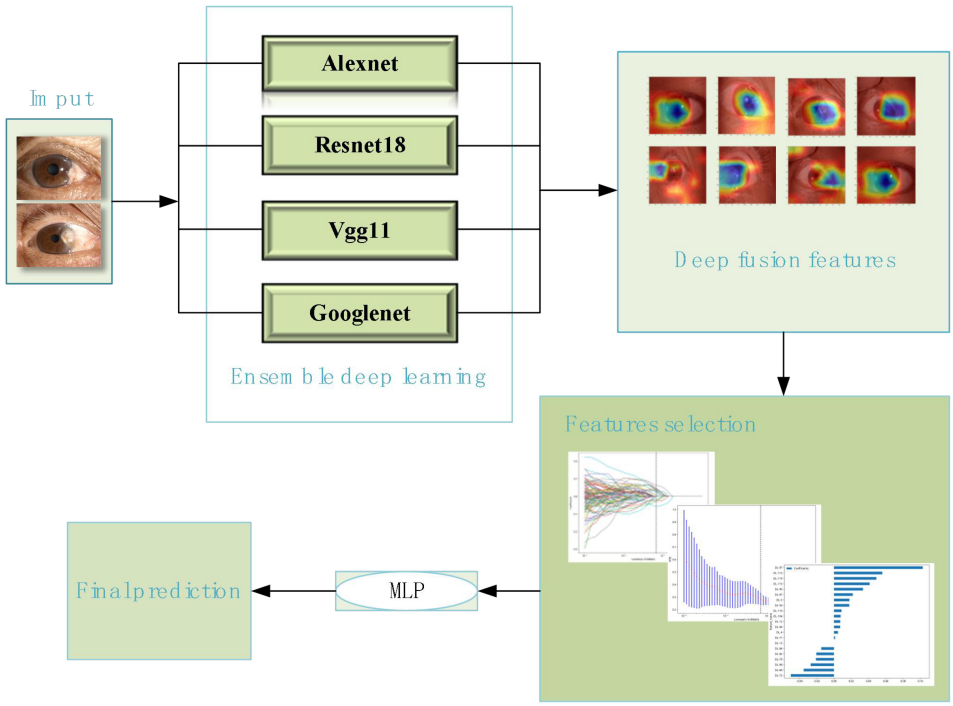


FIGURE 2
The flowchart of the ensemble deep learning model.

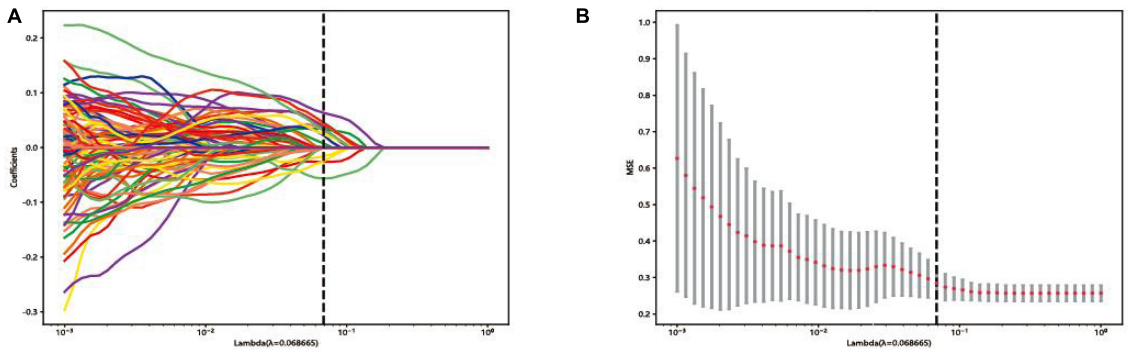


FIGURE 3
Feature selection in the least absolute shrinkage selection operator (LASSO) model: **(A)** LASSO coefficient profiles of the features. **(B)** Selection of tuning parameters in the Lasso regression analysis by 10-fold cross-validation.

TABLE 1 The performances of the models of the models.

Models	Train/Test	Accuracy	AUC	95% CI	Sensitivity	Specificity	F1-score
Ensemble DL	Train	94.20%	0.978	(0.958–0.998)	90.79%	98.39%	94.52%
	Test	94.12%	0.980	(0.968–1.000)	89.47%	100.00%	94.44%
Alexnet	Train	85.40%	0.900	(0.847–0.953)	90.79%	78.69%	87.34%
	Test	85.71%	0.878	(0.754–1.000)	89.47%	81.25%	87.18%
Googlenet	Train	84.67%	0.912	(0.867–0.958)	77.63%	93.44%	84.89%
	Test	80.00%	0.872	(0.759–0.985)	68.42%	93.75%	78.79%
Resnet18	Train	80.29%	0.835	(0.765–0.906)	82.89%	77.05%	82.35%
	Test	77.14%	0.816	(0.673–0.959)	63.16%	93.75%	75.00%
Vgg11	Train	84.67%	0.920	(0.878–0.963)	85.52%	83.61%	86.10%
	Test	85.71%	0.921	(0.836–1.000)	94.74%	75.00%	87.80%

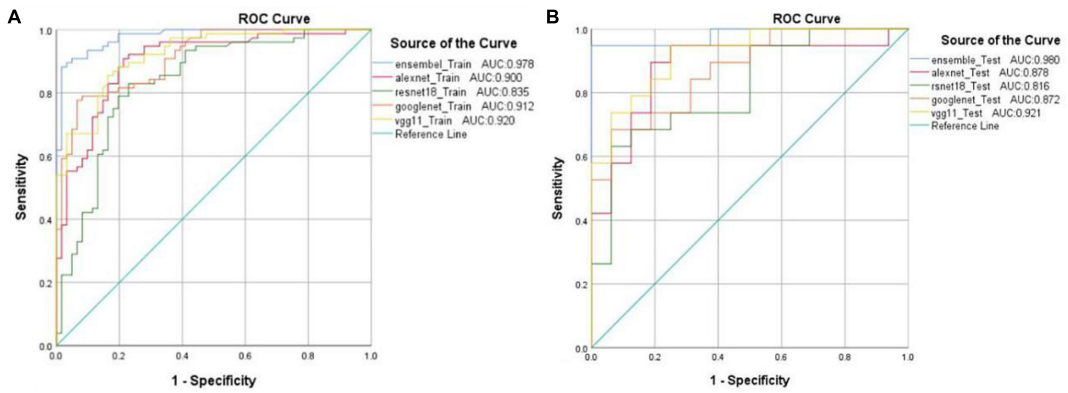


FIGURE 4
Receiver operating characteristic (ROC) curves and the area under the receiver operating characteristic (AUC) values from five different models in the training set **(A)** and testing set **(B)**.

respectively. The AUC of the other models in the training and testing sets were all <0.950.

The heatmaps of Grad-CAM highlighted areas of the anterior segment images where the DL models probably focused on detecting a pterygium that required surgical treatment. The areas highlighted were the actual sites that correspond well with the pterygium (Figure 5).

Discussion

In this study, we developed an AI model based on ensemble DL that was integrated with the alexnet, googlenet, vgg11, and resnet18 models for detecting the pterygium that requires surgical treatment and visualizes the DL process by Grad-CAM.

Compared with the classical Resnet18, Alexnet, Googlenet, and Vgg11 models, respectively, the results showed that the ensemble DL model achieved the highest accuracy and AUC than the four classical DL models in both the training and testing sets. The accuracy of the ensemble DL model was up to 94%. The maximum accuracy of the classical DL models was only 85%. The ensemble DL model outperformed classical DL models with an improvement in accuracy of 9%. The AUC of the ensemble DL model was up to 98% and was at least 6% higher than that of the other four classical models. Thus, our results indicated that this ensemble DL model coupled with the anterior segment images might be an automated and cost-saving alternative for detecting the pterygium that require surgery.

In the previous study, Zaki et al. (2018) used an artificial neural network (ANN) and a SVM to detect a pterygium. The average accuracy reached 91.27% and the AUC reached 95.6%. Hung et al. (2022) developed a DL system in pterygium grading based on the multilayer perceptron (MLP) and the accuracy was 86.67 to 91.67%. Xu et al. (2021) used MobileNet, AlexNet, VGG16, and ResNet18 models to diagnose pterygia. The highest accuracy of the models was 88.30% and the highest AUC of the models was 0.872. However, the accuracy and AUC of our ensemble DL model was up to 94 and 98%. In comparison, the performance of our model was better. It could be related to the method that they used, which was classical machine learning or

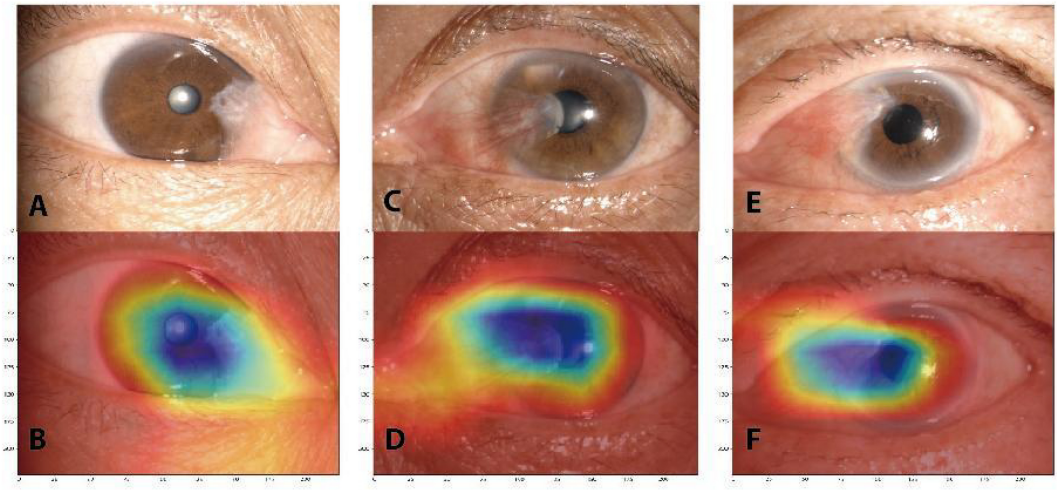


FIGURE 5
Gradient-weighted class activation mapping (Grad-CAM) heatmaps that highlight important regions for the deep learning model predicted for pterygium classification. The original anterior segment images **(A,C,E)**. The corresponding heatmaps **(B,D,F)**.

deep learning method. Compared with the traditional method, DL is a mainly data-driven feature extraction, which does not require much feature extraction of specific domain knowledge, and can extract deep abstract features that are difficult to extract by the traditional method. Its expression of data sets is more efficient and accurate, and the extracted abstract features are more robust and have better generalization ability. Also, compared with the classical DL models, ensemble learning has the advantages of improving prediction performance, directly cascading different models, easy implementation, and fewer parameters.

Precision medicine is a medical model for prevention, diagnosis, and treatment that aims to achieve an optimal therapeutic regimen for an individual (Yan et al., 2021). It has become a focal area of interest and development in medicine of the 21st century (Iacobas and Xi, 2022). Previous researchers such as Zaki et al. (2018), Zulkifley et al. (2019), and Zamani et al. (2020) was based on the two-class detection of pterygium. Unlike their studies, the model of our study was developed to further detect the pterygium that requires surgical treatment. This more closely fits with the current strategy for precision medicine.

Another key strength of this study was the use of Grad-CAM. The disadvantage of ensemble learning is that the prediction results are as uninterpretable as the DL model, i.e., black-box system (Yamashita et al., 2021). Therefore, the Grad-CAM was introduced in our study for visualizing the filters of the penultimate layer of the DL process. The result of Grad-CAM heatmaps highlighted important regions for the DL model predicted for pterygium classification and these regions were consistent with the actual location of the pterygium. It indicated that the model was making predictions based on clinical features of the pterygium.

Of course, our study has some shortcomings. Our sample size was small and might limit the generalizability of our findings. Our study also did not combine clinical features in models. The majority of studies showed that imaging features combined with clinical features have a high value in predicting and diagnosing (Stubblefield et al., 2020; Zhou et al., 2021; Huang et al., 2022). In future research, we will continue to enroll more cases and use the data augmentations method to addresses issues of small sample sizes. And we will also combine with clinical features to detect pterygia that require surgery to improve the performance of the model.

Conclusion

We developed an AI model based on ensemble DL method to classify pterygium. The results indicated that ensemble DL model based on the anterior segment images might be an automated and cost-saving alternative for detection of the pterygium that require surgery. We also used the Grad-CAM to visualize the DL process. The highlighted important regions in the Grad-CAM heatmaps were consistent with the actual location of the pterygium. It indicated that the model was making predictions based on clinical features of the pterygium.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by institutional review board of Jiangxi Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

FG, W-YC, HL, and Y-LZ: statistical analyses and wrote the manuscript. All authors read and approved the final manuscript, contributed to data collection and article, and approved the submitted version.

Funding

We acknowledge the assistance provided by the Natural Science Foundation of Jiangxi Province (20212BAB216058), Jiangxi Provincial Health Technology Project (202210012 and 202310114), and Jiangxi Provincial traditional Chinese Technology Project (2022B840).

Conflict of interest

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

Publisher's note

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

References

Chen, P., Yeh, K., Tsai, Y., Koeh, H., Liu, Y., Lee, H., et al. (2010). XRCC1, but not APE1 and hOGG1 gene polymorphisms is a risk factor for pterygium. *Mol. Vis.* 16, 991–996.

Chen, Y., Huang, W., Ho, W., and Tsai, J. (2021). Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning. *BMC Bioinformatics* 22(Suppl 5):99. doi: 10.1186/s12859-021-04001-1

Fang, X., Deshmukh, M., Chee, M. L., Soh, Z.-D., Teo, Z. L., Thakur, S., et al. (2021). Deep learning algorithms for automatic detection of pterygium using anterior segment photographs from slit-lamp and hand-held cameras. *Br. J. Ophthalmol.* 106, 1642–1647. doi: 10.1136/bjophthalmol-2021-318866

Gulshan, V., Peng, L., Coram, M., Stumpe, M., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410.

Huang, Y., Wang, T., Chen, M., Lin, C. C., Chang, C. W., Tai, H. C., et al. (2022). Radiomics-based nomogram as predictive model for prognosis of hepatocellular carcinoma with portal vein tumor thrombosis receiving radiotherapy. *Front. Oncol.* 12:906498. doi: 10.3389/fonc.2022.906498

Hung, K., Lin, C., Roan, J., Kuo, C., Hsiao, C., Tan, H., et al. (2022). Application of a deep learning system in pterygium grading and further prediction of recurrence with slit lamp photographs. *Diagn. Basel Switz.* 12:888. doi: 10.3390/diagnostics12040888

Iacobas, D. A., and Xi, L. (2022). Theory and applications of the (Cardio) genomic fabric approach to post-ischemic and hypoxia-induced heart failure. *J. Pers. Med* 12:1246. doi: 10.3390/jpm12081246

Kong, G., Wu, J., Chu, H., Yang, C., Lin, Y., Lin, K., et al. (2021). Predicting prolonged length of hospital stay for peritoneal dialysis–treated patients using stacked generalization: Model development and validation study. *JMIR Med. Inform.* 9:e17886. doi: 10.2196/17886

Kwon, Y., Shin, W., Ko, J., and Lee, J. (2020). AK-score: Accurate protein-ligand binding affinity prediction using an ensemble of 3D-convolutional neural networks. *Int. J. Mol. Sci.* 21:8424. doi: 10.3390/ijms21228424

Liu, C., Song, Y., Wang, X., Lai, Z., Li, C., Wan, P., et al. (2020). The key role of vegf in the cross talk between pterygium and dry eye and its clinical significance. *Ophthalmic Res.* 63, 320–331. doi: 10.1159/000503636

Lu, C., Song, J., Li, H., Yu, W., Hao, Y., Xu, K., et al. (2022). Predicting venous thrombosis in osteoarthritis using a machine learning algorithm: A population-based cohort study. *J. Pers. Med.* 12:114. doi: 10.3390/jpm12010114

Montalbo, F. J. (2021). Diagnosing Covid-19 chest x-rays with a lightweight truncated DenseNet with partial layer freezing and feature fusion. *Biomed. Signal Process. Control* 68:102583. doi: 10.1016/j.bspc.2021.102583

Shen, D., Wu, G., and Suk, H. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248.

Stubblefield, J., Hervert, M., Causey, J. L., Qualls, J. A., Dong, W., Cai, L., et al. (2020). Transfer learning with chest X-rays for ER patient classification. *Sci. Rep.* 10:20900. doi: 10.1038/s41598-020-78060-4

Syed, A. M., Hassan, T., Akram, M. U., Naz, S., and Khalid, S. (2016). Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces. *Comput. Methods Programs Biomed.* 137, 1–10. doi: 10.1016/j.cmpb.2016.09.004

Tang, F., Wang, X., Ran, A., Chan, C. K., Ho, M., Yip, W., et al. (2021). A multitask deep-learning system to classify diabetic macular edema for different optical coherence tomography devices: A multicenter analysis. *Diabetes Care* 44, 2078–2088. doi: 10.2337/dc20-3064

Tang, J., Yuan, M., Tian, K., Wang, Y., Wang, D., Yang, J., et al. (2022). An artificial-intelligence–based automated grading and lesions segmentation system for myopic maculopathy based on color fundus photographs. *Transl. Vis. Sci. Technol.* 11:16. doi: 10.1167/tvst.11.6.16

Wang, H., Xu, Q., and Zhou, L. (2015). Large unbalanced credit scoring using lasso-logistic regression ensemble. *PLoS One* 10:e0117844. doi: 10.1371/journal.pone.0117844

Xu, W., Jin, L., Zhu, P., He, K., Yang, W., and Wu, M. (2021). Implementation and application of an intelligent pterygium diagnosis system based on deep learning. *Front. Psychol.* 12:759229. doi: 10.3389/fpsyg.2021.759229

Xue, C., Chen, Y., Huang, Z., Ge, Y., Wang, H., and Wang, J. (2014). EphB4 expression in pterygium is associated with microvessel density. *Int. J. Clin. Exp. Med.* 7, 4008–4015.

Yamashita, R., Long, J., Saleem, A., Rubin, D. L., and Shen, J. (2021). Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Sci. Rep.* 11:2047. doi: 10.1038/s41598-021-81506-y

Yan, J., Yan, Y., Young, A., Yan, Z., and Yan, Z. (2021). Effectiveness and safety of chinese medicine decoctions for behcet's disease: A systematic review and meta-analysis. *Evid. Based Complement. Altern. Med. ECAM* 2021:8202512. doi: 10.1155/2021/8202512

Zaki, W. M., Daud, M. M., Abdani, S. R., Hussain, A., and Mutalib, H. A. (2018). Automated pterygium detection method of anterior segment photographed images. *Comput. Methods Programs Biomed.* 154, 71–78.

Zamani, N. S., Zaki, W. M., Huddin, A. B., Hussain, A., Mutalib, H. A., and Ali, A. (2020). Automated pterygium detection using deep neural network. *IEEE Access* 8, 191659–191672.

Zheng, B., Liu, Y., He, K., Wu, M., Jin, L., Jiang, Q., et al. (2021). Research on an intelligent lightweight-assisted pterygium diagnosis model based on anterior segment images. *Dis. Markers* 2021:7651462. doi: 10.1155/2021/7651462

Zhou, Y., Yang, R., Wang, Y., Zhou, M., Zhou, X., Xing, J., et al. (2021). Histogram analysis based on multi-parameter MR imaging as a biomarker to predict lymph node metastasis in T3 stage rectal cancer. *BMC Med. Imaging* 21:176. doi: 10.1186/s12880-021-00706-0

Zulkifley, M. A., Abdani, S. R., and Zulkifley, N. H. (2019). Pterygium-Net: A deep learning approach to pterygium detection and localization. *Multimed. Tools Appl.* 78, 34563–34584.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 06 January 2023
DOI 10.3389/fnins.2022.1106904

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Jiayi Yang,
Renmin Hospital of Wuhan University, China
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China

*CORRESPONDENCE
Wensheng Li
✉ drlws@qq.com
Xiaobo Cen
✉ xbcen@glpcd.com

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 24 November 2022
ACCEPTED 12 December 2022
PUBLISHED 06 January 2023

CITATION
Wu Y, Luo X, Feng Y, Yang J, Fan H, Cen X and Li W (2023) Comparison of the accuracy of axial length measurement by different imaging methods in Sprague Dawley rats. *Front. Neurosci.* 16:1106904. doi: 10.3389/fnins.2022.1106904

Comparison of the accuracy of axial length measurement by different imaging methods in Sprague Dawley rats

Yajun Wu[1,2,3†], Xiangdong Luo[4†], Yuliang Feng[1,2,3], Jiasong Yang[1,2,3], Hua Fan[2,3], Xiaobo Cen[5*] and Wensheng Li[2,3,1*]

[1]Aier School of Ophthalmology, Central South University, Changsha, Hunan, China, [2]Department of Ophthalmology, Shanghai Aier Eye Ophthalmology Hospital, Shanghai, China, [3]Shanghai Aier Eye Institute, Shanghai, China, [4]Department of Ophthalmology, Xiamen Eye Center of Xiamen University, Xiamen, Fujian, China, [5]WestChina-Frontier PharmaTech Co., Ltd., Chengdu, Sichuan, China

Background: Obtaining accurate axial length (AL) is very important for the establishment of animal models of myopia. The purpose of this study is to compare the accuracy of Quantel A-B scan, OD-1 A scan, and vernier caliper in measuring AL in Sprague Dawley (SD) rats.

Methods: In total, 60 5-week-old SD rats were divided into female rat group (n = 30) and male rat group (n = 30). Quantel A-B scan and OD-1 A scan were, respectively, used to measure the AL of both eyes of each living rat, and vernier caliper was used to measure the anterior-posterior diameter of each rat's eyeball. Besides, the correlation between refractive error (RE) and AL measured by different instruments was evaluated, and the accuracy of the three measurement methods was compared according to gender and left/right eyes.

Results: There were significant differences in AL and diopter of SD rats at the same age ($p < 0.05$). the AL of male rats was greater than that of female rats, while diopter (D) was the opposite; There was no significant difference in AL and D between left and right eyes in the same SD rats ($p > 0.05$); There were statistical differences among the three measurement methods ($p < 0.05$), AL measured by vernier caliper was the largest, followed by Quantel A-B scan, OD-1 A scan; Difference in AL between male and female was not statistically significant between the results obtained by Quantel A-B scan and vernier caliper ($p > 0.05$), but there were statistically significant differences between the other two measurement methods ($p < 0.05$).

Conclusion: Sex is the influencing factor of AL and RE. Imaging measurement can accurately measure the AL in living small rodents. Compared with OD-1 A scan, Quantel A-B scan may be more accurate.

KEYWORDS

Quantel A-B scan, OD-1 A scan, vernier caliper, axial length, measuring method

Introduction

Myopia is now a puzzle for public health worldwide (Ramamurthy et al., 2015) with a prevalence that continues to rise. It is predicted that, by 2050, the number of people with myopia will reach as high as 4.95 billion, nearly half of the global population (Holden et al., 2016). At present, the characteristics of myopia include low age and advanced development, with more and more young people suffering from myopia. In East Asia in particular (He et al., 2015), myopia is more common, and, indeed, 80–90% of young people aged 14–35 in China suffer from myopia (Ma et al., 2021). According to statistics, the myopia rate of 12-year-old children in Hong Kong is as high as 61%, and that of adults is as high as 41.1% (Mak et al., 2018). The myopia rate in Taiwan's 18–24-year-old youth conscription group is as high as 86% (Lee et al., 2013). Moreover, high myopia (HM) often causes myopic choroidal neovascularization (mCNV), complicated glaucoma, cataract, macular hole (MH), and other blinding complications (Wu et al., 2016), which result in a heavy economic burden to families and society. Statistically, the annual productivity loss as a result of myopia in China is as high as US $26.3 billion (Ma et al., 2022), whereas the global productivity loss caused by myopia alone was as high as US $244 billion in 2015 (Naidoo et al., 2019). These values will increase as the prevalence of myopia increases in the population. Therefore, myopia prevention and control is urgent.

Obviously, understanding the mechanisms of myopia occurrence and development is the key to slowing down the rapid rise of its global prevalence effectively (Baird et al., 2020), and the emergence of animal models of myopia is very important to the research of myopia-related mechanisms. At present, the animals used widely in myopia research include monkeys (Zhu et al., 2013), chickens (Liu et al., 2020), guinea pigs (Zhou et al., 2020), rats (Chen et al., 2021), and mice (Lin et al., 2021). The two primary animal models are form-deprivation myopia (FDM) and lens-induced myopia (LIM), respectively. The refractive status of the eye is determined mainly by the axial length (AL) or refractive error (RE) and the matching relationship between them. In fact, the AL is related closely to the RE (Chamberlain et al., 2021). Myopia is not a simple RE, but is often accompanied by axial growth, retinal and scleral thinning, and other pathological changes (Baird et al., 2020), and the animal models of myopia confirmed these above pathological changes clearly (Troilo et al., 2019). Therefore, AL is also one of the important parameters to evaluate myopia, which is why animal models usually judge whether myopia induction is successful according to the results of diopter (D) and AL at the same time.

Although there are many kinds of animals that can be selected for myopia models, there are differences in the physiological structure of the eyes of different animals, which leads to different choices of different animals in establishing myopia models (Schaeffel and Feldkaemper, 2015). The structure of the chicken eyeball is quite different from that of human eye (Glasser and Howland, 1995), and the accuracy of the conclusion of myopia research in chicken models is worthy of further confirmation. The eyes of monkeys have macular fovea, the physiological structure of which is similar to that of humans, and the instruments available for human eyes can be applied directly to monkeys, which therefore makes the monkey eye an ideal model for myopia. However, monkeys are expensive, have a long experimental period, and are difficult to domesticate, therefore, few research teams can use monkeys to conduct large-sample myopia studies. Small rodents, rats, mice, and guinea pigs, are cheap and reproduce quickly and in large numbers. Their eyeball development process is also similar to that of humans. They all have hyperopia reserves in rodents, and gradually face up with age, which can be better used in myopia research.

However, although small rodents have been used widely in myopia research, their eyeballs are small and it is difficult to measure ophthalmic parameters. Although, at present, optical coherence tomography (OCT) can be used successfully to measure the living eyeball parameters of small animals, such as mice (Zhou et al., 2008), the cost of measuring AL is high and it OCT is inconvenient to use. For animals with AL $>$ 15 mm that can be measured by conventional ophthalmology A-scan, it is difficult to measure small animals, such as rats, and guinea pigs, with small eyeballs accurately. Even if measurements can be made, manual mode should be used, which has certain measurement errors. A-B scan (Quantel, Les Ulis, France) has been used to measure the AL of guinea pigs (Dong et al., 2019). Compared with OCT or B scan, it is simple to operate and is a portable instrument, but it cannot be measured automatically on small animals directly. The manual measurement mode is required, and there is error. OD-1 A scan (Kaixin, China) developed an A-scan measurement mode independently, which

can manually measure the AL of small animals such as, rats, and guinea pigs, manually.

So far, there has been no comparative study on the AL measurement methods of the AL of small animals such as rats. Therefore, the purpose of this study is to compare the AL of Sprague Dawley (SD) rats measured by Quantel A-B scan and Kaixin OD-1 A scan with the AL measured by vernier caliper after anatomy, and also to explore if there are differences in the AL of SD rats with different of sexes and eyes, so as to provide a reliable basis for the study of myopia in small rodents.

Materials and methods

Animals

A total of 60 SD rats [Purchased from Beijing Weitong Lihua Laboratory Animal Technology Co., Ltd., China. Production License No: SCXK (Beijing) 2021-0011] were included in this experiment, male/female.

Inclusion criteria: (1) 5-week old SD rats, with age difference $\leq$ 1 week; (2) The body weight of females was 180–220 g and males 300–400 g, and the individual weight was within the mean $\pm$ 20%; (3) Eye condition is good, cornea, lens, vitreous eye refractive system is normal, no turbidity, inflammation. Exclusion criteria: (1) SD rats with poor ocular and systemic conditions, such as eye trauma, corneal opacity, and poor mental state; (2) SD rats with irritable temperament and difficult to accept A-scan and retinoscopy under non-general anesthesia; (3) Rats whose weight and age exceed the inclusion criteria.

All SD rats were adapted to the environment for 3–5 days before the experiment, and healthy rats were selected as the test animals. According to gender, All SD rats were divided into female group (n = 30) and male group (n = 30). All animals were housed in a PP-4 mouse cage (L $\times$ W $\times$ H: 400 mm $\times$ 250 mm $\times$ 200 mm) in WestChina-Frontier PharmaTech Co., Ltd., Chengdu, Sichuan, China, with five animals per cage. If the animals were abnormal, they were housed in a single cage. The feeding environment was in accordance with the national standard of the People's Republic of China GB14925-2010, with room temperature of 20–26°C (daily temperature difference $\leq$ 4°C), relative humidity of 40–70%, artificial lighting, alternating light and dark day and night for 12/12 h, and all animals were free to eat and drink.

This study has passed the ethical approval of the Committee of WestChina-Frontier PharmaTech Co., Ltd. with the ethical approval number of IACUC- SW-S2022007-P001-01.

Retinoscopy

Fix the rats, completely expose the cornea, and measure the refractive state with the red strip Optometry (Suzhou 66 Vision Technology Co., Ltd., China). Under dark room conditions, SD rats were dripped with 5 g/l compound tropicamide eye drops to dilate their pupils and paralyze the ciliary muscles, once every 5 min, at least three consecutive times. After 30 min, the experimenter grabbed the rats to expose the examined eyes. An experienced optometrist performed retinoscopy and optometry on all rats at a working distance of 50 cm, and performed retinoscopy on the horizontal and vertical meridians at an interval of 0.5 D, respectively. The astigmatism will be calculated by half equivalent spherical lens (Zhou et al., 2007).

Quantel A-B scan (Quantel, Les Ulis, France) measures AL

All SD rats were subjected to superficial anesthesia by dropping 4 g/l obucaine hydrochloride eye drops on the eye surface. The operation was repeated 2–3 times with an interval of 5 min each time, and the corneal reflex disappeared as the standard. Open the instrument, input the rat number, select the eye type (right then left), select the A-scan manual measurement mode, anterior chamber and vitreous set to 1,557.5 ms^{-1}, lens set to 17,233.3 ms^{-1} (Jiang et al., 2014). The probe was pointed at the pupil area of the rat, and the cornea was vertically touched (without compression). The clear and stable waveform was taken as the determined image. The value of the waveform of the anterior interface of the retina is read as AL.

OD-1 A scan (Kaixin, China) measures AL

All SD rats were surface anesthetized with 4 g/l obucaine hydrochloride eye drops. The operation was repeated 2–3 times with an interval of 5 min. The disappearance of corneal reflex was taken as the standard. Turn on the instrument, enter the rat number, select the eye type (right first, then left), select the mode of manual measurement of AL in small animals. aim the probe at the center of the rat pupil, touch the cornea vertically (without pressing the cornea), take a clear and stable waveform as the determined image, read and record the scale value of the anterior interface of the retina wave peak, measure three times for each eye, and take the average value, accurate to 0.01 mm.

The vernier caliper (Shanghai Measuring Tools Co., Ltd., China) measures the AL

Sprague Dawley (SD) rats were anesthetized with 3% sodium pentobarbital sodium (50 mg/kg, intraperitoneal injection) and euthanized by bleeding of abdominal aorta. quickly remove the complete eyeballs in the super clean workbench, measure the anterior and posterior diameters of the binocular axes

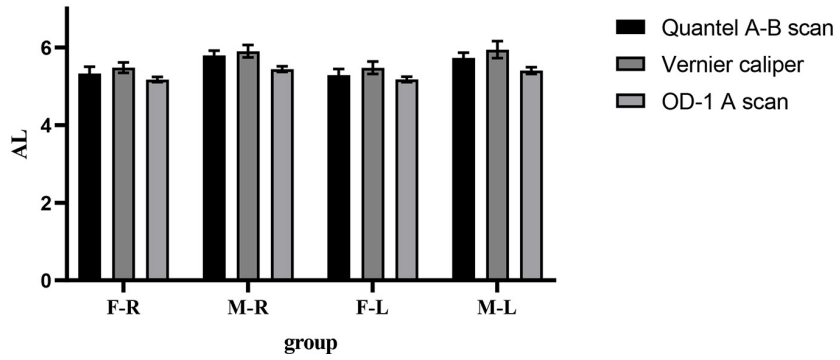


FIGURE 1
Comparison results of the mean values of three AL measurement methods. F-R, female group right eye; M-R, male group right eye; F-L, female group left eye; M-L, male group left eye; AL, axial length.

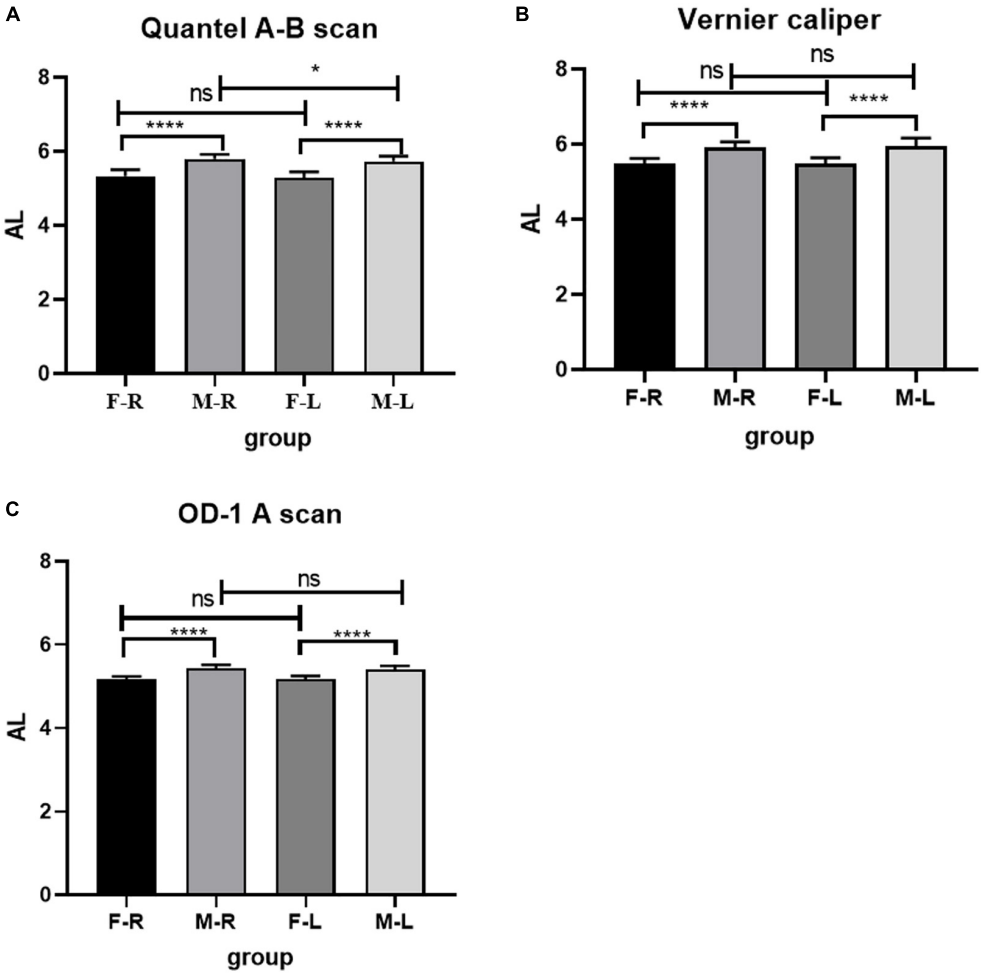


FIGURE 2
Comparison of AL in Sprague Dawley (SD) rats of different sex and eye types. **(A)** AL results from Quantel A-B Scan. **(B)** AL results from Vernier caliper. **(C)** AL results from OD-1 A scan. F-R, female group right eye; M-R, male group right eye; F-L, female group left eye; M-L, male group left eye; AL, axial length; ns indicates no statistical significance; *$p < 0.05$; ****$p < 0.01$.

TABLE 1 Axial length (AL) of left and right eyes in each group.

Group	Male (n = 30)		T	P	Female (n = 30)		T	P
	R (mm)	L (mm)			R (mm)	L (mm)		
Quantel A-B scan	5.798 ± 0.122	5.729 ± 0.140	0.879	0.383	5.324 ± 0.176	5.286 ± 0.158	2.050	0.045
Vernier caliper	5.902 ± 0.160	5.942 ± 0.218	0.123	0.903	5.481 ± 0.135	5.476 ± 0.159	−0.816	0.418
OD-1 A scan	5.441 ± 0.073	5.403 ± 0.088	−0.257	0.798	5.173 ± 0.068	5.178 ± 0.072	1.831	0.072
F	114.629	88.522			39.555	37.116		
P	0.000	0.000			0.000	0.000		

R, right eye; L, left eye; T, t-value; P, p-value; F, f-value ($p < 0.05$) difference was statistically significant.

with a vernier caliper (Distance from corneal apex to optic disc of eyeball with clean connective tissue and optic nerve removed), accurate to 0.02 mm. The axial values were read and recorded, and the average values were measured three times for each eye.

Statistical analysis

Statistical software SPSS25.0 (IBM Corp., Armonk, NY, USA) was used for analysis. Measurement data in this study were described as mean ± SD. Levene test was used to test the homogeneity of variance for axial index. When the variance was homogeneity ($P > 0.05$), one-way analysis of variance (ANOVA) was used for statistical test. When the variance was uneven ($P \leq 0.05$), Kruskal-Wallis H rank sum test was used for statistical analysis. Paired-t-test was used for comparison of AL between left and right eyes, independent sample-t test was used for pairwise comparison between different groups, and ANOVA was used for comparison among three groups. $P < 0.05$ was considered statistically significant. Pearson correlation analysis was used to analyze the correlation between AL and D.

Results

Comparison of AL results from three different measurement methods

In the male and female rat groups, the AL measured by vernier caliper of the left and right eyes was greater than that measured by Quantel A-B scan, and the result value of OD-1 A scan was the smallest (as shown in Figure 1). The AL measured by the three methods was statistically significant ($p < 0.05$). Taking the eye type as the variable, the Quantel A-B scan measurement results in the female rat group showed that there was a statistical difference in the AL of the left and right eyes ($t = 2.050$, $p = 0.045$; as shown in Figure 2A), and the other measurement results showed that there was no statistical difference in the AL of the left and right eyes of SD rats of the same sex ($p < 0.05$; as shown in Figures 2B, C). In the same eye, the AL of male rats was significantly higher than that of female rats. See Table 1 for the details.

Results of RE

The RE of the female group was higher than that of the male group (F-R: 5.083 ± 0.506D; M-R: 4.183 ± 0.328D; F-L: 4.833 ± 0.610D; M-L: 4.208 ± 0.322D), the difference was statistically significant ($p < 0.05$), and there was no significant difference in RE between the left and right eyes ($p > 0.05$). See Figure 3 for details.

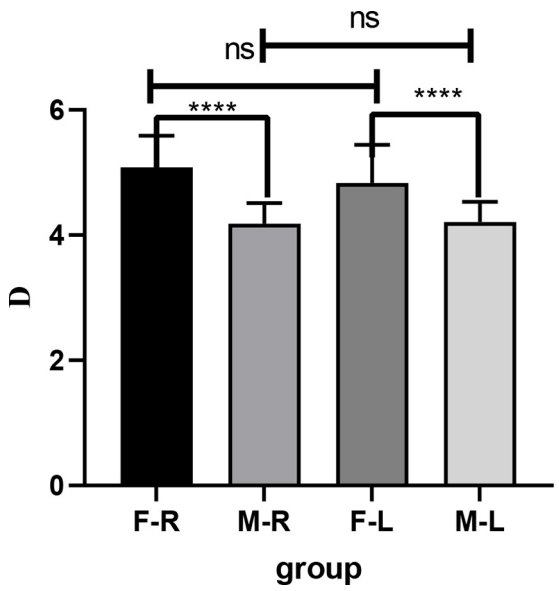


FIGURE 3
Comparison of RE in Sprague Dawley (SD) rats of different sexes and eye types. F-R, female group right eye; M-R, male group right eye; F-L, female group left eye; M-L, male group left eye; RE, refractive errors; D, diopter; ns indicates no statistical significance; ****$p < 0.01$.

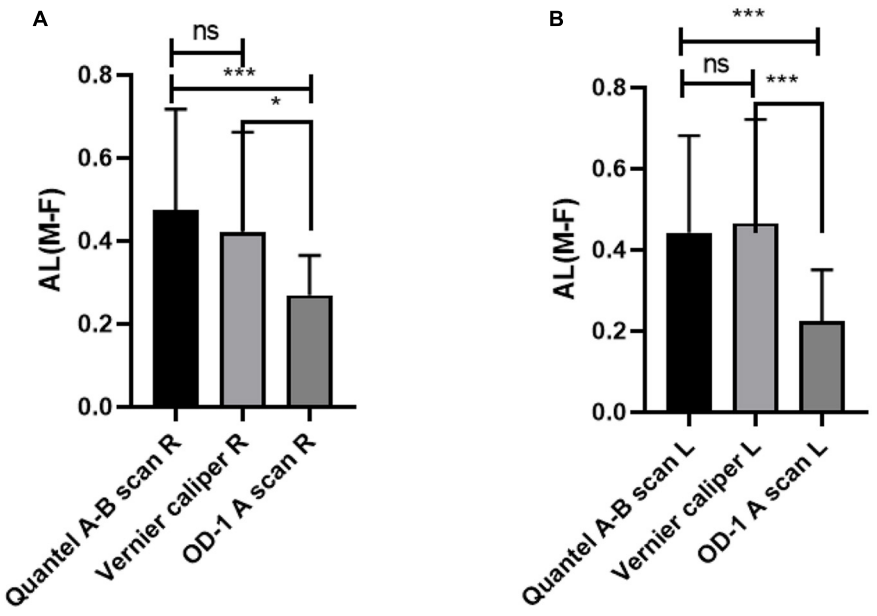


FIGURE 4
Comparison of AL difference between male and female rats by three different measurement methods. **(A)** AL difference in right eye between male and female rats. **(B)** AL difference in left eye between male and female rats. AL, axial length; R, right eye; L, left eye; M-F, AL difference between male and female rats; ns indicates no statistical significance; *$p < 0.05$; ***$p < 0.01$.

Comparison of AL difference between male and female rats by three different measurement methods

The comparison of AL difference between male and female rats shows that there is no significant difference in the measurement results between Quantel A-B scan and Vernier caliper, while there are statistical differences between Quantel A-B scan and OD-1 A scan, Vernier caliper and OD-1 A scan. The details are shown in **Figure 4**.

Correlation analysis between AL and D

The AL results measured in three different ways have negative correlation with D ($p < 0.0001$). The Pearson correlation coefficient r^2 shows that Quantel A-B scan is 0.2100, Vernier caliper is 0.2622, and OD-1 A scan is 0.2973. The details are shown in **Figure 5**.

Discussion

The commonly used AL measurement methods in clinical practice mainly include mainly the optical method and ultrasonic bioassay, the former is mainly intraocular lens-master (IOL-master) (Roy et al., 2012) and OCT, whereas the latter is mainly A-scan. However, the optical method is used more widely because of its higher measurement accuracy than the ultrasonic method, and is considered as the gold standard of AL measurement (Goto et al., 2020a). However, the optical method has high requirements on the refractive stroma. Once the lens, cornea, and other refractive stroma, are seriously cloudy, AL cannot be obtained accurately, whereas the ultrasound method is not subject to this limitation. Thus, at present, both methods are often used for AL measurement. In addition, AL measured by the optical method refers to the distance from corneal apex to retinal pigment epithelium (RPE) cells, whereas AL measured by the ultrasonic method refers to the distance from corneal apex to inner limiting membrane (ILM), so the result of the latter is slightly smaller than that of the former (Goto et al., 2020b).

However, although the accuracy of AL measurement by A-scan is not as high as that by the optical method, it is more convenient to apply to small animals. Both OCT and IOL Master are difficult to coordinate with the measurement of animals, which not only needs to be performed under general anesthesia, but also tends to increase the mortality of experimental animals, and it is difficult to support microaxial detection using this instrument. Therefore, A-scan is more suitable for the measurement of AL in small animals. In fact, in most studies of small rodent myopia models, AL was measured by A-scan (Tian L. et al., 2021; Yang et al., 2021; Yu et al., 2021). This is the reason why two kinds of A-scan were selected to measure the AL of rats in this study.

Axial length (AL) is defined as the length of the anterior and posterior diameters of the eyeball. For this

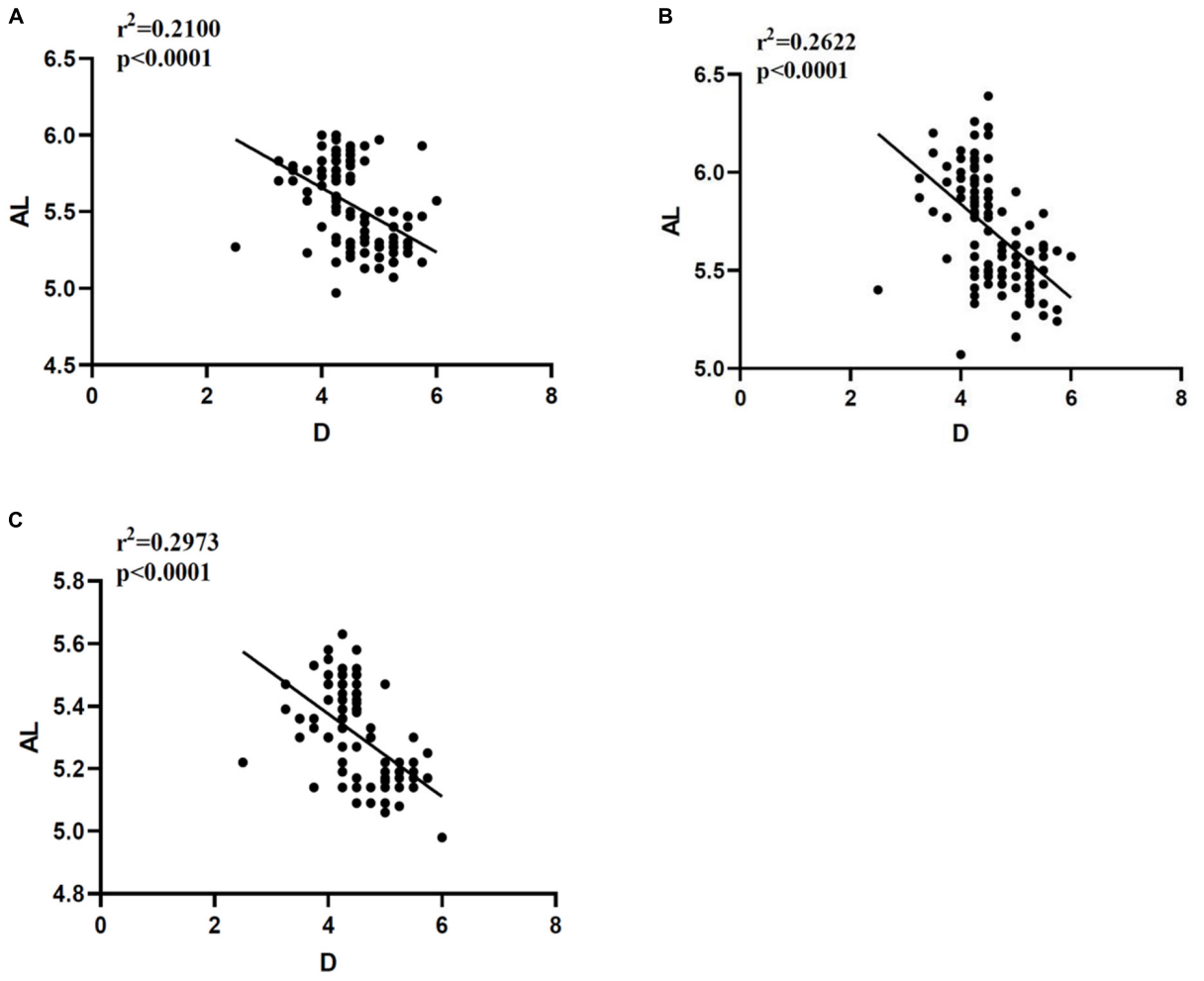


FIGURE 5
Correlation analysis between AL and D measured in different ways. **(A)** Correlation between AL and D measured by Quantel A-B scan. **(B)** Correlation between AL and D measured by Vernier caliper. **(C)** Correlation between AL and D measured by OD-1 A scan. AL, axial length; D, diopter; r2, Pearson correlation coefficient; $p < 0.0001$ indicates that there is great statistical significance.

reason, we chose vernier caliper to measure the anterior posterior diameter of the eyeball of the intact rat after removing the conjunctival tissue as the gold standard of AL. Because the thickness of all tissues (choroid and sclera) after retinal ILM is included, the measurement result will be greater than that of A-scan. Consistent with our results, the AL results measured by the vernier calipers of all rats were greater than those of the other two groups of A-scan.

Moreover, because the results of RE tend to be synchronized with the increase of AL (Tian T. et al., 2021), there is a negative correlation between them (Jiang et al., 2009), and the AL of infant female rats is smaller than that of male rats of the same age, therefore we examined the AL and RE results of male and female rats at 5 weeks of age. Finally, we found that the AL of male rats was greater than that of female rats, and the difference was significant.

Furthermore, the correlation between AL and D in each group was analyzed and all three AL values had significant negative correlation with D. In addition, we compared the AL difference between male and female rats measured by three different ocular axis measurement methods, and found that there was no significant difference between Quantel A-B scan and vernier caliper, although there was a statistical difference between OD-1 A scan and the other two methods. These indicate that Quantel A-B scan might be more accurate for AL measurement of living small rodents, although the accuracy of this instrument is only 0.1 mm, whereas OD-1 A scan can be accurate to 0.01 mm. In addition, the price of the latter is only 1/10 of the former, which might therefore be a better choice for laboratories under poor economic conditions. Although the image measurement results are not as accurate as the anatomical results, they are sufficient for measuring the AL of small experimental animals.

As far as we know, this is the first study comparing the AL values of SD rats measured by an anatomical method and an imaging method, and the results have certain significance for the study of myopia in small animals. Unlike human eye measurements, the AL of small animals is too small to be measured accurately, although, admittedly, the AL measurement methods for human eyes are improving constantly. Among them, swept source optical coherence tomography (SS-OCT) is considered to be the most accurate AL measurement method at present (Vounotrypidis et al., 2019). Huang et al. (2019) compared the AL values measured by three kinds of SS-OCT with those measured by partial coherence interferometry (PCI), and found that the success rate of AL measurement of SS-OCT biometrics was significantly higher than that of PCI, and so they believed that SS-OCT might become the gold standard for AL measurement. For animal studies, the myopic animal model can meet the acquisition of real AL (anterior posterior diameter of ocular axis), if it is not required to obtain AL data of living animals, we believe that AL measured by vernier caliper can be used as the gold standard of rat AL, which can be accurate to 0.02 mm. However, we recommend A-scan for the acquisition of AL in small living animals.

Our study has the following limitations. First, we did not measure the tissue thickness of the eyeball retina after ILM, so we could not judge accurately the difference between the AL value measured by the two A-scan and the actual eyeball anteroposterior diameter, which requires the assistance of SS-OCT. Unfortunately, we are unable currently to meet this condition; Second, we did not detect the changes of AL and D from young to adult rats to judge the specific impact of changes in the ocular axis on RE, so as to further detect the accuracy of Quantel A-B scan and OD-1 A scan. Third, we chose SD rats as the representative of small animals because they can be used as myopia models. Compared with other common models, such as, mice, chickens, their eyeballs are larger and easier to measure. In fact, we also used these two A-ultrasound to measure AL in guinea pigs, and successfully obtained AL. However, we have not tested mice and chickens, so we cannot determine whether their AL can be obtained by A-scan. We aim to address these limitations in future studies.

Conclusion

From what we have discussed here, we have confirmed that A-scan can satisfy the acquisition of AL in small animals. Our results of Quantel A-B scan for AL measurement of living small animals might be more reliable, whereas the vernier caliper can obtain the actual anterior posterior diameter of the ocular axis, however, it needs to be obtained after the animals are dissected. With the development of technology, we believe that accurate and economical AL measuring instruments for small animals can be developed in the future, which will provide strong support to relevant myopia research.

Data availability statement

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

Ethics statement

The animal study was reviewed and approved by the Committee of WestChina-Frontier PharmaTech Co., Ltd. with the ethical approval number of IACUC- SW-S2022007-p001-01.

Author contributions

YW and XL were responsible for writing and modifying the manuscripts. YF and JY were responsible for proofreading and revising the language of the manuscript. HF was responsible for data statistics and chart making. WL and XC were responsible for the design of the study and guiding the revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the Science Research Foundation of Aier Eye Hospital Group (Grant Nos. AF2101D9 and AR2001D1).

Acknowledgments

We thank the Aier Eye Group for its funding support. We also thank the WestChina-Frontier PharmaTech Co., Ltd. for providing laboratory facilities and materials for this study.

Conflict of interest

XC was employed by WestChina-Frontier PharmaTech Co., Ltd.

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

Publisher's note

References

Baird, P. N., Saw, S. M., Lanca, C., et al. (2020). Myopia. *Nat Rev Dis Primers.* 6, 99. doi: 10.1038/s41572-020-00231-4

Chamberlain, P., Lazon de la Jara, P., Arumugam, B., and Bullimore, M. A. (2021). Axial length targets for myopia control. *Ophthalmic Physiol Opt.* 41, 523–531. doi: 10.1111/opo.12812

Chen, L., Ouyang, J., Li, X., et al. (2021). DNAH17 is essential for rat spermatogenesis and fertility. *J Genet.* 100, 14. doi: 10.1007/s12041-021-01264-8

Dong, L., Shi, X. H., Kang, Y. K., et al. (2019). Amphiregulin and ocular axial length. *Acta Ophthalmol.* 97, e460–e470. doi: 10.1111/aos.14080

Glasser, A., and Howland, H. C. (1995). In vitro changes in back vertex distance of chick and pigeon lenses: species differences and the effects of aging. *Vision Res.* 35, 1813–1824. doi: 10.1016/0042-6989(94)00292-T

Goto, S., Maeda, N., Noda, T., et al. (2020a). Change in optical axial length after cataract surgery. *Journal of Cataract & Refractive Surgery.* 46, 710–715. doi: 10.1097/j.jcrs.0000000000000149

Goto, S., Maeda, N., Noda, T., et al. (2020b). Comparison of composite and segmental methods for acquiring optical axial length with swept-source optical coherence tomography. *Sci Rep.* 10, 4474. doi: 10.1038/s41598-020-61391-7

He, M., Xiang, F., Zeng, Y., et al. (2015). Effect of Time Spent Outdoors at School on the Development of Myopia Among Children in China: A Randomized Clinical Trial. *JAMA.* 314, 1142–1148. doi: 10.1001/jama.2015.10803

Holden, B. A., Fricke, T. R., Wilson, D. A., et al. (2016). Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. *Ophthalmology.* 123, 1036–1042. doi: 10.1016/j.ophtha.2016.01.006

Huang, J., Chen, H., Li, Y., et al. (2019). Comprehensive Comparison of Axial Length Measurement With Three Swept-Source OCT-Based Biometers and Partial Coherence Interferometry. *J Refract Surg.* 35, 115–120. doi: 10.3928/1081597X-20190109-01

Jiang, L., Schaeffel, F., Zhou, X., et al. (2009). Spontaneous axial myopia and emmetropization in a strain of wild-type guinea pig (Cavia porcellus). *Invest Ophthalmol Vis Sci.* 50, 1013–1019. doi: 10.1167/iovs.08-2463

Jiang, L., Zhang, S., Schaeffel, F., et al. (2014). Interactions of chromatic and lens-induced defocus during visual control of eye growth in guinea pigs (Cavia porcellus). *Vision Res.* 94, 24–32. doi: 10.1016/j.visres.2013.10.020

Lee, Y. Y., Lo, C. T., Sheu, S. J., and Lin, J. L. (2013). What factors are associated with myopia in young adults? A survey study in Taiwan Military Conscripts. *Invest Ophthalmol Vis Sci.* 54, 1026–1033. doi: 10.1167/iovs.12-10480

Lin, M. Y., Lin, I. T., Wu, Y. C., et al. (2021). Stepwise candidate drug screening for myopia control by using zebrafish, mouse, and Golden Syrian Hamster myopia models. *EBioMedicine.* 65, 103263. doi: 10.1016/j.ebiom.2021.103263

Liu, H., Schaeffel, F., Trier, K., and Feldkaemper, M. (2020). Effects of 7-Methylxanthine on Deprivation Myopia and Retinal Dopamine Release in Chickens. *Ophthalmic Res.* 63, 347–357. doi: 10.1159/000502529

Ma, M., Xiong, S., Zhao, S., Zheng, Z., Sun, T., and Li, C. (2021). COVID-19 Home Quarantine Accelerated the Progression of Myopia in Children Aged 7 to 12 Years in China. *Invest Ophthalmol Vis Sci.* 62, 37. doi: 10.1167/iovs.62.10.37

Ma, Y., Wen, Y., Zhong, H., et al. (2022). Healthcare utilization and economic burden of myopia in urban China: A nationwide cost-of-illness study. *J Glob Health.* 12, 11003. doi: 10.7189/jogh.12.11003

Mak, C. Y., Yam, J. C., Chen, L. J., Lee, S. M., and Young, A. L. (2018). Epidemiology of myopia and prevention of myopia progression in children in East Asia: a review. *Hong Kong Med J.* 24, 602–609. doi: 10.12809/hkmj187513

Naidoo, K. S., Fricke, T. R., Frick, K. D., et al. (2019). Potential Lost Productivity Resulting from the Global Burden of Myopia: Systematic Review, Meta-analysis, and Modeling. *Ophthalmology.* 126, 338–346. doi: 10.1016/j.ophtha.2018.10.029

Ramamurthy, D., Lin Chua, S. Y., and Saw, S. M. (2015). A review of environmental risk factors for myopia during early life, childhood and adolescence. *Clin Exp Optom.* 98, 497–506. doi: 10.1111/cxo.12346

Roy, A., Das, S., Sahu, S. K., and Rath, S. (2012). Ultrasound biometry vs. IOL Master. *Ophthalmology* 119, .e1–.e1937. doi: 10.1016/j.ophtha.2012.06.006

Schaeffel, F., and Feldkaemper, M. (2015). Animal models in myopia research. *Clin Exp Optom.* 98, 507–517. doi: 10.1111/cxo.12312

Tian, L., Guo, Y. T., Ying, M., Liu, Y. C., Li, X., and Wang, Y. (2021). Co-existence of myopia and amblyopia in a guinea pig model with monocular form deprivation. *Ann Transl Med.* 9, 110. doi: 10.21037/atm-20-5433

Tian, T., Zou, L., Wang, S., Liu, R., and Liu, H. (2021). The Role of Dopamine in Emmetropization Modulated by Wavelength and Temporal Frequency in Guinea Pigs. *Invest Ophthalmol Vis Sci.* 62, 20. doi: 10.1167/iovs.62.12.20

Troilo, D., Smith, E. L. III, Nickla, D. L., et al. (2019). IMI - Report on Experimental Models of Emmetropization and Myopia. *Invest Ophthalmol Vis Sci.* 60, M31–M88. doi: 10.1167/iovs.18-25967

Vounotrypidis, E., Haralanova, V., Muth, D. R., et al. (2019). Accuracy of SS-OCT biometry compared with partial coherence interferometry biometry for combined phacovitrectomy with internal limiting membrane peeling. *J Cataract Refract Surg.* 45, 48–53. doi: 10.1016/j.jcrs.2018.08.027

Wu, P. C., Huang, H. M., Yu, H. J., Fang, P. C., and Chen, C. T. (2016). Epidemiology of Myopia. *Asia Pac J Ophthalmol (Phila).* 5, 386–393. doi: 10.1097/APO.0000000000000236

Yang, Y., Wu, J., Wu, D., et al. (2021). Intravitreal brimonidine inhibits form-deprivation myopia in guinea pigs. *Eye and Vision** 8, doi: 10.1186/s40662-021-00248-0

Yu, T., Xie, X., Wei, H., et al. (2021). Choroidal changes in lens-induced myopia in guinea pigs. *Microvasc Res.* 138, 104213. doi: 10.1016/j.mvr.2021.104213

Zhou, X., Lu, F., Xie, R., et al. (2007). Recovery from axial myopia induced by a monocularly deprived facemask in adolescent (7-week-old) guinea pigs. *Vision Res.* 47, 1103–1111. doi: 10.1016/j.visres.2007.01.002

Zhou, X., Xie, J., Shen, M., et al. (2008). Biometric measurement of the mouse eye using optical coherence tomography with focal plane advancement. *Vision Res.* 48, 1137–1143. doi: 10.1016/j.visres.2008.01.030

Zhou, X., Zhang, S., Zhang, G., et al. (2020). Increased Choroidal Blood Perfusion Can Inhibit Form Deprivation Myopia in Guinea Pigs. *Invest Ophthalmol Vis Sci.* 61, 25. doi: 10.1167/iovs.61.13.25

Zhu, X., McBrien, N. A., Smith, E. L. III, Troilo, D., and Wallman, J. (2013). Eyes in various species can shorten to compensate for myopic defocus. *Invest Ophthalmol Vis Sci.* 54, 2634–2644. doi: 10.1167/iovs.12-10514

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 30 January 2023
DOI 10.3389/fnins.2023.1097291

OPEN ACCESS

EDITED BY
Zhi Wen,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Xiang Liu,
The First Affiliated Hospital of Nanchang University, China
Fei Chen,
The First Affiliated Hospital of Zhengzhou University, China

*CORRESPONDENCE
Yu-Lin Zhong
804722489@qq.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 13 November 2022
ACCEPTED 09 January 2023
PUBLISHED 30 January 2023

CITATION
Gan F, Wu F-P and Zhong Y-L (2023) Artificial intelligence method based on multi-feature fusion for automatic macular edema (ME) classification on spectral-domain optical coherence tomography (SD-OCT) images. *Front. Neurosci.* 17:1097291. doi: 10.3389/fnins.2023.1097291

Artificial intelligence method based on multi-feature fusion for automatic macular edema (ME) classification on spectral-domain optical coherence tomography (SD-OCT) images

Fan Gan[1,2], Fei-Peng Wu[2] and Yu-Lin Zhong[2]*

[1]Medical College of Nanchang University, Nanchang, China, [2]Department of Ophthalmology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, China

Purpose: A common ocular manifestation, macular edema (ME) is the primary cause of visual deterioration. In this study, an artificial intelligence method based on multi-feature fusion was introduced to enable automatic ME classification on spectral-domain optical coherence tomography (SD-OCT) images, to provide a convenient method of clinical diagnosis.

Methods: First, 1,213 two-dimensional (2D) cross-sectional OCT images of ME were collected from the Jiangxi Provincial People's Hospital between 2016 and 2021. According to OCT reports of senior ophthalmologists, there were 300 images with diabetic (DME), 303 images with age-related macular degeneration (AMD), 304 images with retinal-vein occlusion (RVO), and 306 images with central serous chorioretinopathy (CSC). Then, traditional omics features of the images were extracted based on the first-order statistics, shape, size, and texture. After extraction by the alexnet, inception_v3, resnet34, and vgg13 models and selected by dimensionality reduction using principal components analysis (PCA), the deep-learning features were fused. Next, the gradient-weighted class-activation map (Grad-CAM) was used to visualize the-deep-learning process. Finally, the fusion features set, which was fused from the traditional omics features and the deep-fusion features, was used to establish the final classification models. The performance of the final models was evaluated by accuracy, confusion matrix, and the receiver operating characteristic (ROC) curve.

Results: Compared with other classification models, the performance of the support vector machine (SVM) model was best, with an accuracy of 93.8%. The area under curves AUC of micro- and macro-averages were 99%, and the AUC of the AMD, DME, RVO, and CSC groups were 100, 99, 98, and 100%, respectively.

Conclusion: The artificial intelligence model in this study could be used to classify DME, AME, RVO, and CSC accurately from SD-OCT images.

KEYWORDS

artificial intelligence, macular edema, multi-feature fusion, SD-OCT images, classification models

Introduction

Macular edema (ME) is a common ocular manifestation of fluid infiltration or inflammation in the sensitive macular area of the retina, and is an important cause of visual deterioration (Song et al., 2022). There are several ME-related eye diseases, including diabetic ME (DME), retinal-vein occlusion (RVO), age-related macular degeneration (AMD), and central serous chorioretinopathy (CSC). Chronic hyperglycemia in diabetes mellitus (DM) causes damage to capillaries, resulting in retinal ischemia and increased vascular permeability, which leads to DME (Kim et al., 2020). Wet AMD is a result of subretinal choroidal neovascularization, resulting in fragile and leaky blood vessels that penetrate through Bruch's membrane and cause edema (McGill et al., 2017). Excessive angiogenic growth factors, caused by hypoxia secondary to RVO, leads to vascular leakage and ME (Narayanan et al., 2021). In CSC, choroidal congestion, thickening, and hyperpermeability are considered to cause leakage through the retinal pigment epithelium (RPE) (Schellevis et al., 2019).

Optical coherence tomography (OCT) is a high-resolution, non-contact, and non-invasive biomedical imaging technique and is often used in eye clinics for macular disease. The basic principle of OCT imaging is that a beam of light is emitted into the tissues to be examined and detects the reflected or back-scattered light from the tissues; this reflected light will interfere with light that originated from the same source and the reflectivity profile along the light beam can be derived from the interference signal and used to generate an A-scan (Liu et al., 2019). The combination of multiple A-scans along the horizontal axis produces a brightness scan (B-scan) (Khalid et al., 2017). Compared with Fundus cameras, OCT systems has the high contrast and depth sectioning capability (LaRocca et al., 2014). And high-quality cross-sectional images of the neurosensory retina can be acquired without pupil dilatation in a matter of seconds (Ouyang et al., 2013). Therefore the sensitivity of OCT for detection of a variety of retinal irregularities was higher.

Macular edema diagnosis by OCT is based on the visualization of the retinal structure. However, spectral-domain optical coherence tomography (SD-OCT) can better delineate the different retinal layers so that the histological changes of ME can be shown in more detail. In DME patients, SD-OCT shows mild retinal edema with cystic spaces located only in the outer plexiform layer (OPL), whereas, when edema worsens, they involved both the OPL and the outer nuclear layer (ONL) (Leung et al., 2008). SD-OCT image analysis was also more sensitive than FAF for identifying geographic atrophy GA in patients treated for exudative AMD (Massamba et al., 2019). For CSC patients, SD-OCT can show shallow serous detachments and provided precise information about the amount and localization of subretinal fluid and RPE abnormalities (Murthy et al., 2016). SD-OCT also can quantify retinal thickness changes in eyes with cystoid macular edema (CME) from central retinal vein occlusion (CRVO) and is superior to contact lens–assisted biomicroscopy to identify foveal edema (Decroos et al., 2013).

Currently, ME diagnosis depends on the subjective evaluation of OCT and the clinical experience of ophthalmologists. Not only does this process take a lot of time, energy, and requires training, but the ability of ophthalmologists at different levels to diagnose diseases ranges widely. With the application of artificial intelligence in ophthalmology, a large number of machine learning-based computer-aided diagnosis (CAD) models have been developed for the quantitative analysis of OCT images to achieve the automatic diagnosis of macular diseases. Alsaih et al. (2017) applied machine-learning techniques for DME classification on SD-OCT images, both the sensitivity (SE) and specificity (SP) of the best result were 87.5%. Chen Y. et al. (2021) applied convolutional-neural-network-based transfer learning to classify AMD, the CNN models with appropriate algorithm hyperparameters had excellent capability and performance in classifying OCT images of AMD and DME. However, their studies made only binary classification, which limits the application of machine-learning algorithms in the diagnosis of many diseases. Wang et al. (2016) proposed a CAD model to discriminate AMD, DME, and healthy macula on OCT images, the best model based on the sequential minimal optimization (SMO) algorithm achieved 99.3% in the overall accuracy for the three classes of samples. However, the coverage of disease types was still inadequate and their studies were all based on single features.

Currently, the signal fusion methods have attracted the attention of many researchers for solving pattern recognition problems, and that were divided into three categories which are early fusion, intermediate fusion, and late fusion (Verma and Tiwary, 2014). Early fusion is also named as feature level fusion which emphasizes the data combination before the classification (Zhang et al., 2017). It was defined as performing merge and splitting operations on existing feature sets to generate new feature sets. Using the feature fusion approach of deep learning and machine learning, the complementary information of abstract features of deep learning and detailed features of machine learning can be realized (Wang et al., 2022). The accuracy of models could be improved (Khan and Hasan, 2020). Therefore, we introduced an artificial intelligence method of fusion of traditional features and deep features, aiming to automate the classification of DME, AME, RVO, and CSC from DM based on SD-OCT images (**Figure 1**).

Materials and methods

Image collection and pre-processing

A total of 1,213 two-dimensional (2D) cross-sectional OCT images of ME were collected from the Jiangxi Provincial People's Hospital (China) between 2016 and 2021. According to OCT reports of a senior ophthalmologist, 300 images with DME, 303 images with AMD, 304 images with RVO, and 306 images with CSC were included. And the set was randomly divided into a training set and a test set, at a ratio of 8:2. To protect patient privacy, patient images were all anonymized prior to analysis. All the OCT images were acquired by the same experienced ophthalmologist on the same machine, i.e., Heidelberg Spectralis OCT (Heidelberg Engineering, Dossenheim, Germany). Then, the macular area was outlined manually with ITK-SNAP software, which was the mask region of interest (ROI) images. The obtained mask files were used for traditional omics features extraction. Next, the image was cropped to the ROI specifications and the segments of ROI were used for DL features extraction.

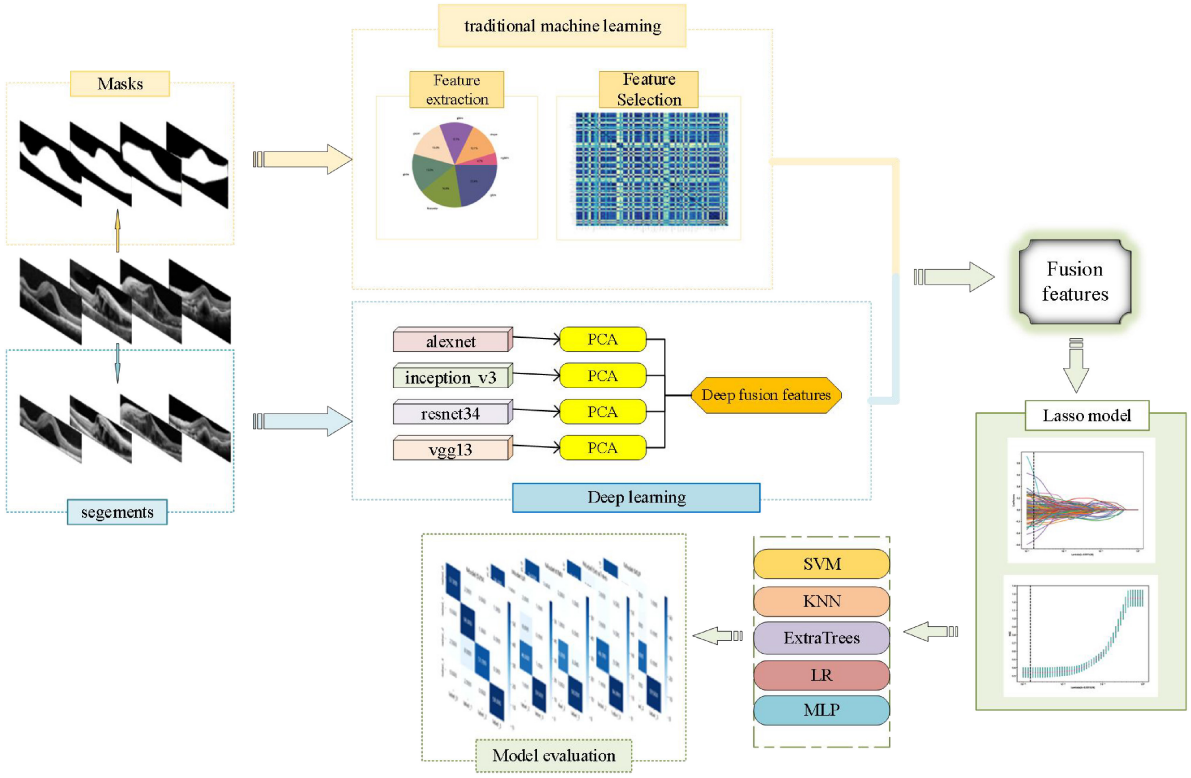


FIGURE 1
The flowchart of this study.

Traditional omics features extraction and selection

Based on the mask files of the original image files, traditional omics features of images were extracted based on the first-order statistics, shape, size, and texture. The *Z*-score standardization method was then used to normalize the extracted features and the Spearman correlation coefficients were used to select the normalized feature.

DL features extraction and model visualization

The segments of ROI were input into alexnet, inception_v3, resnet34, and vgg13 models, respectively, which were initialized using the pre-trained weights from ImageNet, and the DL features were obtained. And then the DL features were selected by dimensionality reduction using principal components analysis (PCA). Finally, the selected features of the four DL models were fused.

In order to evaluate the deep learning-focused regions, the gradient-weighted class activation map (Grad-CAM) method was used. In this method, gradient information flowing from input layers to the last convolution layer of a convolutional neural network (CNN) is used, and coarse heat maps of important regions in the input images are generated (Chen T. et al., 2021). Based on the coarse heat maps, we can understand which areas of the segments are most likely to be focused by the DL models.

Early fusion and lasso model established

Feature fusion was performed after the pooling layer in the model. The traditional omics features and the deep-fusion features were fused into a composite feature vector. Then in the training set, the composite feature vector was input into further fused as fusion features set. A *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm was used to visualize the features vectors from feature space of high dimensions into 2D space. Then, the fusion features set was divided into a training set and a test set, at a ratio of 7:3. The lasso model which was established to further select features. We chosed the optimal λ based on the minimum criteria according to fivefold cross validation.

Classification models established

The support vector machine (SVM), K-nearest neighbor (KNN), ExtraTrees, logistic regression (LR), and multilayer perceptron (MLP) were used to establish the classification models in the training set and the performance of the final classification models was evaluated in the test set. Finally, the classification performance of the different models was assessed and compared.

Statistical analysis

The accuracy, confusion matrix and the receiver operating characteristic (ROC) curve of the classification models were used

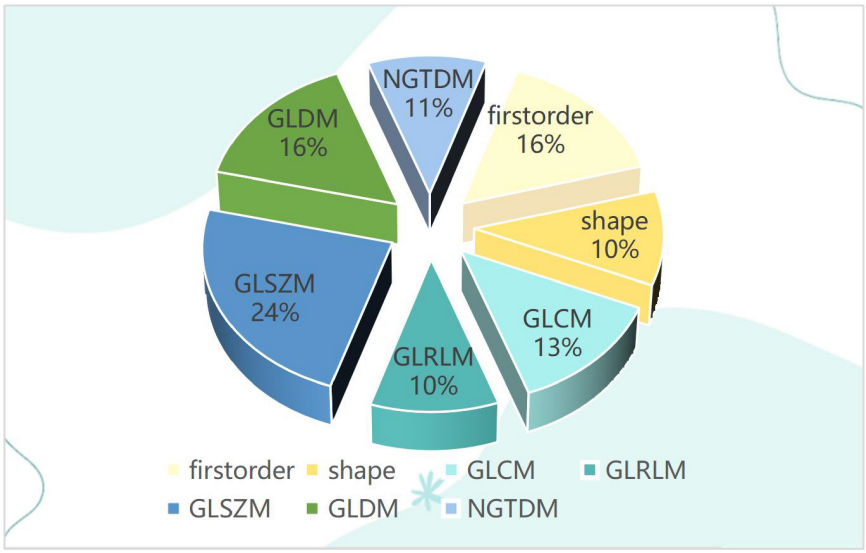


FIGURE 2
The pie chart for traditional omics features distribution.

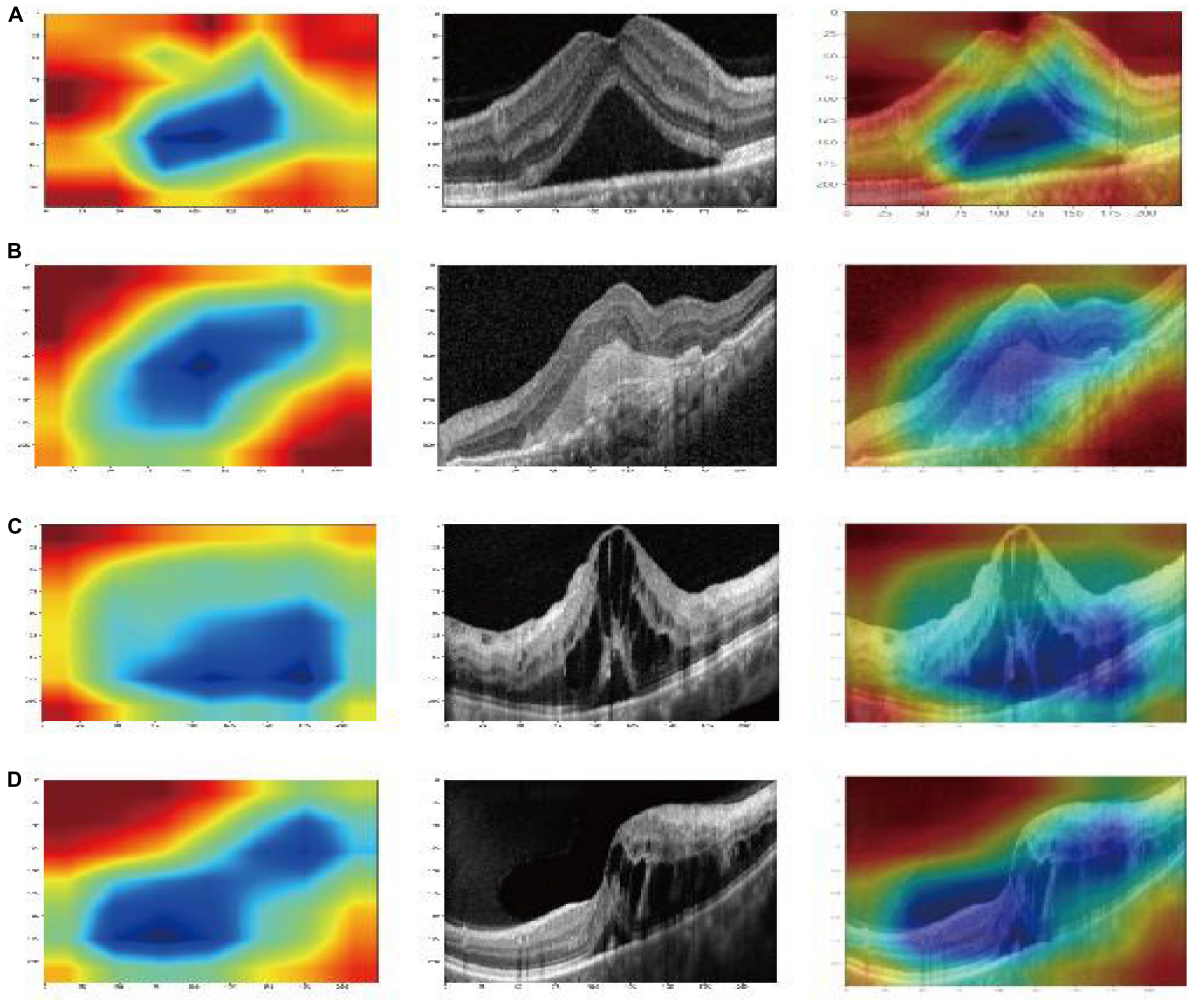


FIGURE 3
Gradient-weighted class-activation map (Grad-CAM) visualization of deep learning feature extraction: CSC **(A)**; AMD **(B)**; DME **(C)**; RVO **(D)**. The blue part that gathers inward from the red part is active, indicating that the model pays particular attention to this area (Huang et al., 2022).

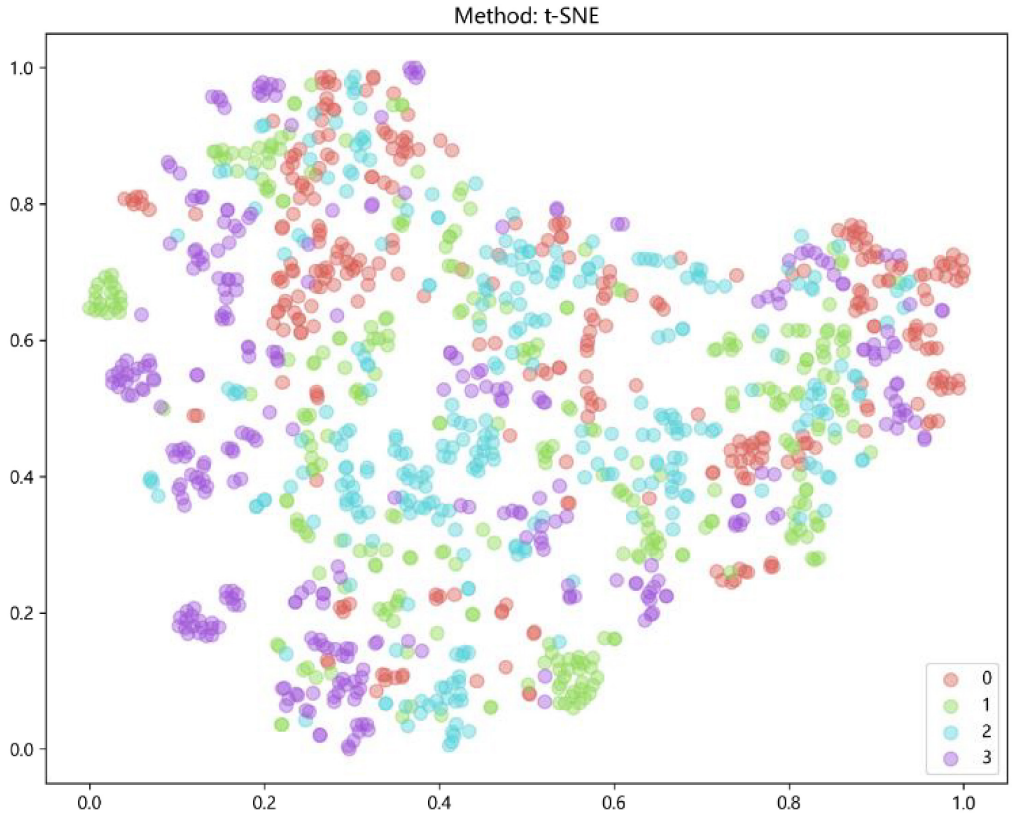


FIGURE 4
Feature visualization by *t*-distributed stochastic neighbor embedding (*t*-SNE): AMD (red); DME (green); RVO (blue); CSC (purple).

to evaluate the performance of models. All statistical analyses were performed and visualized in Python (version 3.9.7).

Results

Characteristics of OCT images

The total of 1,213 original images of ME were collected, included DME (n = 300), AMD (n = 303), RVO (n = 304), and CSC (n = 306). The training set was consisted of 849 images, included DME (n = 240), AMD (n = 243), RVO (n = 243), and CSC (n = 245). The test set was consisted of 364 images, included DME (n = 60), AMD (n = 60), RVO (n = 61), and CSC (n = 61). And then, the original image files and corresponding mask files were obtained to use for traditional omics features extraction. The segments based on maximum ROI mask were used for deep learning features extraction.

Characteristics of traditional omics features

For each ROI, a total of 107 features of each image were extracted, and after the Spearman correlation coefficients, the final 38 features of each image were selected. Including 6 first-order features, 4 shape-based features, and 28 textural features. The textural features were composed of 5 Gray Level Co-occurrence Matrix (GLCM), 4 Gray Level Run Length Matrix (GLRLM), 9 Gray Level Size Zone Matrix (GLSZM), 6 Gray Level Dependence Matrix (GLDM), and 4 Neighboring Gray Tone Difference Matrix (NGTDM) as shown in Figure 2.

Characteristics of deep learning features and model visualization

There were 9,216, 2,048, 512, and 16,383 deep learning features of each image were obtained from the alxnet, inception_v3, resnet34, and vgg13, respectively, which were on "avgpool" layer before last FC layers. Dimension reduction with PCA compressed features into 31. Finally, a deep fusion feature subset containing 124 compression features were obtained. And the heatmaps of Grad-CAM highlighted areas which the deep learning models likely focused on as shown in Figure 3.

Characteristics of fusion features

After the early fusion, a subset of each image that contains 162 features was got. Feature was visualized by *t*-SNE for an intuitive perception of how well these features can distinguish different types of ME, was shown in Figure 4.

Lasso model evaluation

The LASSO was used for automated feature selection in this study. 53 features were selected to build the final classification models

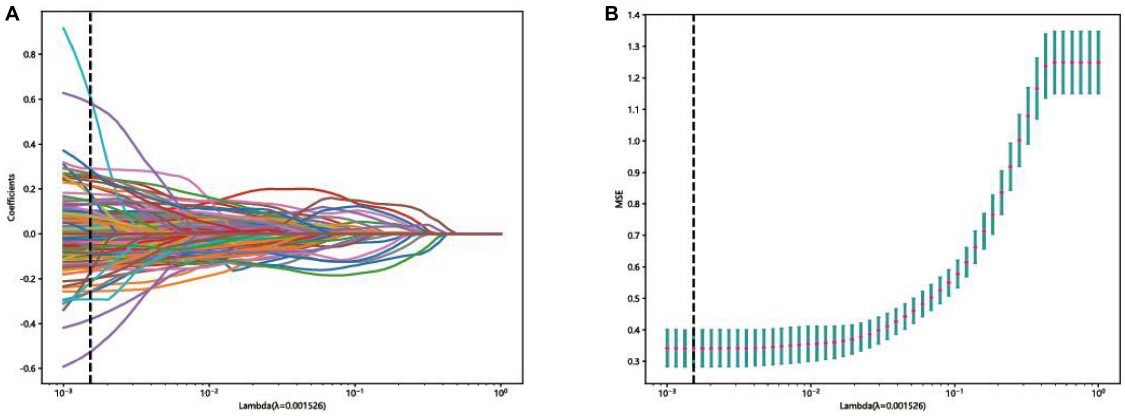


FIGURE 5
Feature selection in the lasso model: **(A)** Lasso coefficient profiles of the 162 fusion features, where each curve corresponds to one feature, the vertical black line indicates an optimal λ. **(B)** Curve of binomial deviation varied with parameter λ, where value of the optimal log (λ) is marked by vertical dashed lines.

based on the optimal lambda value and the corresponding coefficients in the training set as shown in Figure 5.

Classification models evaluation

The accuracy of SVM model was highest than other models, up to 93.8% in the test set. And the KNN, ExtraTrees, MLP, and LR models in the test set were only 90.08, 82.23, 90.50, and 86.77%, respectivily, as shown in Table 1.

Since our problem is a multiclass classification, AUC of binary class classification cannot be considered. So, micro- and macro-averages (Sokolova and Lapalme, 2009) were calculated from ROC curves, the macro-average could give equal weight to the classification of each label, whereas the micro-average incorporates the frequency of the labels into the label weighting. In the test set, the area under curves (AUCs) of micro- and macro-averages of the SVM and MLP models both were 99%, which was highest than the other models. The ROC curve of the test set for each group compared with that of the other groups, each group were clearly distinguished from other groups in the SVM model and the AUC of the AMD, DME, RVO, and CSC groups were 100, 99, 98, and 100%, respectively. While, the AUC of the AMD, DME, RVO, and CSC groups compared with that of other groups in the MLP model only were 99, 97, 97, and 100%, respectively. It could be seen that in terms of ROC curve results, the SVM model has the best performance.

The test set was distributed in a 4 × 4 matrix according to the labeled labels and the classification results. It could be seen that the recognition performance of SVM and MLP models was better than others. For example, the recognition rates of RVO were relatively high in the SVM and MLP model, while RVO was easily misrecognized in the other three models. However, there were also some differences. In the SVM model, the recognition rates of AMD, DME, and RVO were relatively high. While, in the MLP model, the recognition rates of CSC were relatively high. As shown in Figure 6.

Discussion

The current study used a multi-feature fusion method for automatic ME classification on SD-OCT images. It fused the features of traditional omics and four DL models, which were the alxnet, inception_v3, resnet34, and vgg13. The Grad-CAM was used to visualize the explanation of DL black-box model. Finally, after the fusion features were screened by the lasso model, the non-zero coefficients features were used to developed six classification models. According the accuracy results, as well as the ROC curve and confusion matrix, the performance of the SVM model was the best, and could be used to classify the DME, AMD, RVO, and CSC accurately from DM SD-OCT images.

Early intelligent diagnosis mainly relies on artificially designed feature templates or uses single traditional machine-learning methods (Turchin et al., 2009), treating intelligent diagnosis as a classification problem (Srinivasan et al., 2014; Alsaih et al., 2017). Because a single feature is usually sensitive to the changes of part of the image features and is not sensitive to the changes of other features, when the difference between two kinds of images is not big in the sensitive features of a certain feature, the classifier based on the training of a single feature cannot output the correct classification. In addition, the complex background noise in the image will also lead to the deterioration of feature data quality, which not only increases the difficulty of classifier training, but also reduces the accuracy of classification. Our proposed fusion features method, by contrast,

TABLE 1 The accuracy of classification models in the training set and test set.

Model	Accuracy	Task
SVM	97.22%	train
	93.80%	test
KNN	90.32%	train
	90.08%	test
ExtraTrees	100%	train
	82.23%	test
MLP	98.35%	train
	90.50%	test
LR	90.83%	train
	86.77%	test

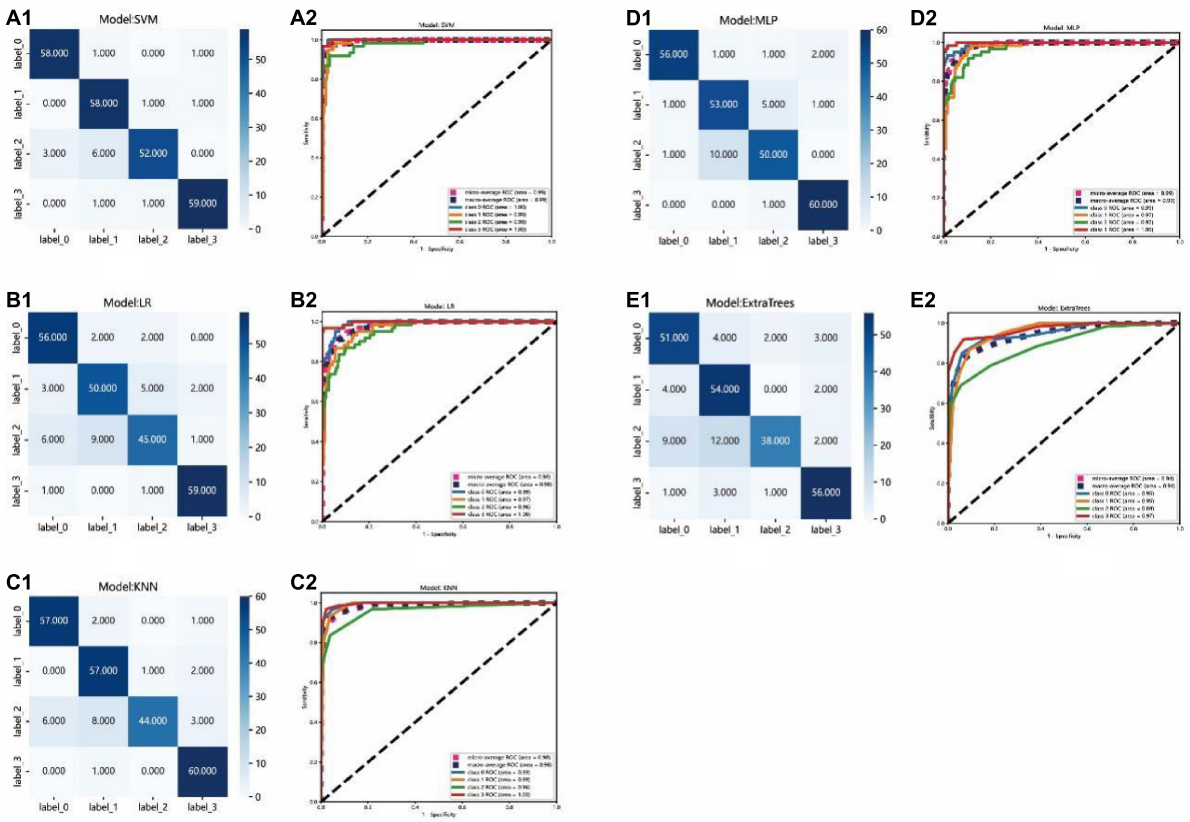


FIGURE 6
Confusion matrix and ROC curve of the different models: **(A1–E1)**: Confusion matrix of the SVM, LR, KNN, MLP, and ExtraTrees model, respectively. Each row of the matrix represents the actual class and each column indicates the predicted class. **(A2–E2)**: ROC curve of SVM, LR, KNN, MLP, and ExtraTrees model, respectively. Label 0 for AMD, label 1 for DME, label 2 for RVO, label 3 for CSC.

realized feature complementarity and reduced the influence of single feature inherent defects.

In previous studies, Lu et al. (2018) used ResNet to detect normal images, cystoid ME, serous macular detachment, epiretinal membrane, and macular hole based on the single deep learning feature extraction method. The accuracy of their method for detecting cystoid ME cases was 84.8% which was much lower than our result. This also confirmed that the feature fusion method can improve the accuracy of the model compared with the single feature extraction method. Chen T. et al. (2021) used a convolutional-neural-network to classify AMD. Chan et al. (2018) used information from AlexNet, VggNet, and GoogleNet to design a decision model for automatic classification of normal ME and DME. Although these models have performed well, they lack the interpretation capability. The Grad-CAM was introduced in our study to overcome the common drawback of DL models. It uses the gradient of the target class and propagates to the final convolutional layer to generate a rough positioning map, which is used to visualize the features (Yang et al., 2021). The Grad-CAM could address the mechanism by which the CAM approach requires changes to the model architecture. Compared with other interpretation methods, the computational complexity is reduced and the interpretability of the model is increased. It also combines the advantages of fine-grained detection (unable to locate the image) and image positioning (unable to improve the positioning resolution). The result of Grad-CAM heatmaps in our study highlighted important areas that the DL models probably focused on extracting features. This is the same area in which our eyes recognize ME. This is a good example of the Grad-CAM identifying the pathologic region of an OCT image correctly.

Of course, there were also shortcomings to this study. First, we just collected the OCT images from a single-center study so the sample does not represent the entire patient population. Second, single-omics methods were used in this study. For multi-classification, using multi-omics data can obtain better accuracy (Lin et al., 2020). Third, the accuracy of our study needs to be improved. Therefore, in future studies, we will try to incorporate multicenter data to reinforce the conclusion of our study and combined multiomics techniques to automate classification of DM based on SD-OCT images and the color fundus pictures.

Conclusion

In this study, an artificial intelligence method based on multi-feature fusion was introduced for automatic ME classification on SD-OCT images. The results showed that the model could be used to classify the DME, AMD, RVO, and CSC accurately from SD-OCT images. The result of Grad-CAM heatmaps in our study highlighted important areas that the DL models probably focused on extracting features. The results of Grad-CAM heatmap highlighted that the important areas for the DL model to extract features was the same as the areas in which our eyes recognize ME.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the Jiangxi Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

FG, F-PW, and Y-LZ contributed to data collection, statistical analyses, and wrote the manuscript. All authors read and approved the final manuscript, contributed to the manuscript, and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

Alsaih, K., Lemaitre, G., Rastgoo, M., Massich, J., Sidib, D., and Meriaudeau, F. (2017). Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomed. Eng. Online* 16:68. doi: 10.1186/s12938-017-0352-9

Chan, G. C., Kamble, R., Muller, H., Shah, S. A., Tang, T. B., and Meriaudeau, F. (2018). "Fusing results of several deep learning architectures for automatic classification of normal and diabetic macular edema in optical coherence tomography," in *Proceedings of the 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, Vol. 2018, (Honolulu, HI: IEEE), 670–673. doi: 10.1109/EMBC.2018.8512371

Chen, T., Lim, W. S., Wang, V. Y., Ko, M., Chiu, S., Huang, Y., et al. (2021). Artificial intelligence–assisted early detection of retinitis pigmentosa–the most common inherited retinal degeneration. *J. Digit. Imaging* 34, 948–958. doi: 10.1007/s10278-021-00479-6

Chen, Y., Huang, W., Ho, W., and Tsai, J. (2021). Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning. *BMC Bioinform.* 22(Suppl. 5):99. doi: 10.1186/s12859-021-04001-1

Decroos, F. C., Stinnett, S. S., Heydary, C. S., Burns, R. E., and Jaffe, G. J. (2013). Reading center characterization of central retinal vein occlusion using optical coherence tomography during the COPERNICUS trial. *Transl. Vis. Sci. Technol.* 2:7. doi: 10.1167/tvst.2.7.7

Huang, T., Yang, R., Shen, L., Feng, A., Li, L., He, N., et al. (2022). Deep transfer learning to quantify pleural effusion severity in chest X-rays. *BMC Med. Imaging* 22:100. doi: 10.1186/s12880-022-00827-0

Khalid, S., Akram, M. U., Hassan, T., Nasim, A., and Jameel, A. (2017). Fully automated robust system to detect retinal edema, central serous chorioretinopathy, and age related macular degeneration from optical coherence tomography images. *Biomed Res. Int.* 2017:7148245. doi: 10.1155/2017/7148245

Khan, M. U., and Hasan, M. A. (2020). Hybrid EEG-fNIRS BCI fusion using multi-resolution singular value decomposition (MSVD). *Front. Hum. Neurosci.* 14:599802. doi: 10.3389/fnhum.2020.599802

Kim, J., Park, I. W., and Kwon, S. (2020). Factors predicting final visual outcome in quiescent proliferative diabetic retinopathy. *Sci. Rep.* 10:17233. doi: 10.1038/s41598-020-74184-9

LaRocca, F., Nankivil, D., Farsiu, S., and Izatt, J. A. (2014). True color scanning laser ophthalmoscopy and optical coherence tomography handheld probe. *Biomed. Opt. Express* 5, 3204–3216. doi: 10.1364/BOE.5.003204

Leung, C. K., Cheung, C. Y., Weinreb, R. N., Lee, G., Lin, D., Pang, C. P., et al. (2008). Comparison of macular thickness measurements between time domain and spectral domain optical coherence tomography. *Invest. Ophthalmol. Vis. Sci.* 49, 4893–4897. doi: 10.1167/iovs.07-1326

Lin, Y., Zhang, W., Cao, H., Li, G., and Du, W. (2020). Classifying breast cancer subtypes using deep neural networks based on multi-Omics data. *Genes* 11:888. doi: 10.3390/genes11080888

Liu, Y., Carass, A., He, Y., Antony, B. J., Filippatou, A., Saidha, S., et al. (2019). Layer boundary evolution method for macular OCT layer segmentation. *Biomed. Opt. Express* 10, 1064–1080. doi: 10.1364/BOE.10.001064

Lu, W., Tong, Y., Yu, Y., Xing, Y., Chen, C., and Shen, Y. (2018). Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl. Vis. Sci. Technol.* 7:41. doi: 10.1167/tvst.7.6.41

Massamba, N., Sellam, A., Butel, N., Skondra, D., Caillaux, V., and Bodaghi, B. (2019). Use of fundus autofluorescence combined with optical coherence tomography for diagnose of geographic atrophy in age-related macular degeneration. *Med. Hypothesis Discov. Innov. Ophthalmol.* 8, 298–305.

McGill, T. J., Bohana-Kashtan, O., Stoddard, J. W., Andrews, M. D., Pandit, N., Rosenberg-Belmaker, L. R., et al. (2017). Long-term efficacy of GMP grade Xeno-Free hESC-derived RPE cells following transplantation. *Transl. Vis. Sci. Technol.* 6:17. doi: 10.1167/tvst.6.3.17

Murthy, R. K., Haji, S., Sambhav, K., Grover, S., and Chalam, K. V. (2016). Clinical applications of spectral domain optical coherence tomography in retinal diseases. *Biomed. J.* 39, 107–120. doi: 10.1016/j.bj.2016.04.003

Narayanan, R., Kelkar, A., Abbas, Z., Goel, N., Soman, M., Naik, N., et al. (2021). Sub-optimal gain in vision in retinal vein occlusion due to under-treatment in the real world: Results from an open-label prospective study of Intravitreal Ranibizumab. *BMC Ophthalmol.* 21:33. doi: 10.1186/s12886-020-01757-7

Ouyang, Y., Heussen, F. M., Keane, P. A., Sadda, S. R., and Walsh, A. C. (2013). The retinal disease screening study: Prospective comparison of nonmydriatic fundus photography and optical coherence tomography for detection of retinal irregularities. *Invest. Ophthalmol. Vis. Sci.* 54, 1460–1468. doi: 10.1167/iovs.12-10727

Schellevis, R. L., Dijk, E. H., Breukink, M. B., Keunen, J. E., Santen, G. W., Hoyng, C. B., et al. (2019). Exome sequencing in families with chronic central serous chorioretinopathy. *Mol. Genet. Genomic Med.* 7:e00576. doi: 10.1002/mgg3.576

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 45, 427–437. doi: 10.1016/j.ipm.2009.03.002

Song, R., Jiang, J., and Wang, H. (2022). Macular edema and visual acuity observation after cataract surgery in patients with diabetic retinopathy. *J. Healthc. Eng.* 2022:3311751. doi: 10.1155/2022/3311751

Srinivasan, P. P., Kim, L. A., Mettu, P. S., Cousins, S. W., Comer, G. M., Izatt, J. A., et al. (2014). Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* 5, 3568–3577. doi: 10.1364/BOE.5.003568

Turchin, A., Shubina, M., Breydo, E., Pendergrass, M., and Einbinder, J. (2009). Comparison of information content of structured and narrative text data sources on the example of medication intensification. *J. Am. Med. Inform. Assoc.* 16, 362–370. doi: 10.1197/jamia.M2777

Verma, G. K., and Tiwary, U. S. (2014). Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *Neuroimage* 102 Pt 1, 162–172. doi: 10.1016/j.neuroimage.2013.11.007

Wang, H., Zheng, L., Pan, S., Yan, T., and Su, Q. (2022). Image recognition of pediatric pneumonia based on fusion of texture features and depth features. *Comput. Math. Methods Med.* 2022:1973508. doi: 10.1155/2022/1973508

Wang, Y., Zhang, Y., Yao, Z., Zhao, R., and Zhou, F. (2016). Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images. *Biomed. Opt. Express* 7, 4928–4940. doi: 10.1364/BOE.7.004928

Yang, B., Zhu, Y., and Zhou, S. (2021). Accurate wheat lodging extraction from multi-channel UAV images using a lightweight network model. *Sensors* 21:6826. doi: 10.3390/s21206826

Zhang, P., Wang, X., Chen, J., and You, W. (2017). Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload. *Sensors* 17:2315. doi: 10.3390/s17102315

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fnins.2023.1131247

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Yu Ji,
The First Affiliated Hospital of Nanchang University, China
Chen-Xing Qi,
Renmin Hospital of Wuhan University, China
Tianming Huo,
Wuhan University, China

*CORRESPONDENCE
Bing-Lin Huang
✉ huangbinglin2009@163.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 24 December 2022
ACCEPTED 16 January 2023
PUBLISHED 02 February 2023

CITATION
Li D-J, Huang B-L, Peng Y, Liang L-Y and Liu H (2023) Altered dynamic functional connectivity in the primary visual cortex in patients with primary angle-closure glaucoma. *Front. Neurosci.* 17:1131247. doi: 10.3389/fnins.2023.1131247

Altered dynamic functional connectivity in the primary visual cortex in patients with primary angle-closure glaucoma

Dong-Jin Li[1], Bing-Lin Huang[2]*, Yuan Peng[3], Ling-Yan Liang[4] and Hui Liu[5]

[1]Health Management Center, The First People's Hospital of Jiujiang City, Jiujiang, Jiangxi, China, [2]College of Clinical Medicine, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China, [3]Department of Ophthalmology, Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China, [4]College of Graduate, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China, [5]Department of Ophthalmology, Jiangxi Provincial People's Hospital, Nanchang, Jiangxi, China

Purpose: Glaucoma is the main blindness-causing disease in the world. Previous neuroimaging studies demonstrated that glaucoma not only causes the loss of optic ganglion cells but also leads to the abnormal function of the optic nerve pathway and the visual cortex. However, previous studies also reported that patients with glaucoma have dysfunction in the visual cortex in a static state. Whether or not patients with primary angle-closure glaucoma (PACG) were accompanied by dynamic functional connectivity (FC) changes in the primary visual cortex (V1) remains unknown.

Methods: A total of 34 patients with PACG (23 men and 11 women) and 34 well-matched healthy controls (HCs) were enrolled in the study. The dynamic functional connectivity (dFC) with the sliding window method was applied to investigate the dynamic functional connectivity changes in the V1.

Results: Compared with HCs, patients with PACG showed increased dFC values between left V1 and bilateral calcarine (CAL). Meanwhile, patients with PACG showed increased dFC values between right V1 and bilateral CAL.

Conclusion: Our study demonstrated that patients with PACG showed increased dFC within the visual network, which might indicate the increased variability FC in the V1 in patients with PACG.

KEYWORDS

dynamic functional connectivity, functional magnetic resonance imaging, primary angle-closure glaucoma, primary visual cortex, machine learning

Introduction

Primary angle-closure glaucoma (PACG) is a serious irreversible blinding eye disease. PACG leads to increased intraocular pressure, which in turn leads to the apoptosis of optic ganglion cells. PACG is characterized by ocular pain and visual field defects, which might lead to blindness in advanced cases. PACG is more common in Asians than Europeans and Africans, with >80% of PACG cases worldwide occurring in Asia. At present, the prevalence of PACG is 0.75% in adult Asians worldwide (Cheng et al., 2014). Glaucoma not only leads to the loss of retinal ganglion

cells but also leads to abnormal changes in the optic nerve pathway and visual center function. Retinal ganglion cell death could leads to anterograde and retrograde in visual pathway (Lawlor et al., 2018).

With the development of functional magnetic resonance imaging (fMRI) technology, multimodal MRI techniques, including diffusion tensor imaging (DTI) and fMRI, have been used widely to investigate the functional changes of the visual cortex in glaucoma. Numerous studies have shown that patients with glaucoma have primary visual cortex (V1) and V1-related dysfunction. Li et al. (2017) demonstrated that patients with PACG have decreased functional connectivity (FC) between the left V1 and right V2 and increased FC between the left V1 and higher cognitive cortex.

Meanwhile, Jiang et al. (2020) reported that patients with PACG showed abnormal effective connectivity between V1 and the higher visual area, motor cortices, somatosensory cortices, and frontal lobe. Graham et al. (2021) found that dogs with unilateral PACG showed reduced fractional anisotropy (FA) in the visual pathway. Nuzzi et al. (2018) demonstrated that patients with glaucoma showed abnormal FC changes in the visual pathway. In addition, Pankowska et al. (2022) reported that the patients with advanced glaucoma showed an increase in the gray matter thickness in the V1 region. Meanwhile, early glaucoma was associated with reduced thickness in the right lateral occipital gyrus and the left lingual gyrus (Pankowska et al., 2022). Haykal et al. (2022) reported that patients with primary open-angle glaucoma (POAG) showed a loss of both axonal coherence and density in the pregeniculate visual pathways, while the postgeniculate pathways exhibited a loss of axonal coherence. Meanwhile, patients with POAG showed a decrease in fiber density and fiber-bundle cross-section in the pregeniculate optic tracts, whereas the postgeniculate optic radiation showed a decrease in fiber density (Haykal et al., 2019).

Thus, combined with our findings, previous studies demonstrated that patients with glaucoma were along with visual pathway and visual cortex dysfunction. However, the abovementioned studies focused mainly on static brain activity changes in the visual cortex. It is recognized increasingly that the functional activity changes in the brain not only are constant over time but also show dynamic changes over time.

The human brain is a complex dynamic system capable of non-stationary neural activity. The human brain showed inherently dynamic activity, which is related to the functional ability of neural networks. The dynamics of neural activity in the brain is closely related to neurophysiological activity. Growing evidence demonstrated that the fluctuations in FC are of neural origin and are also even more prominent during the resting period when mental activity is unconstrained. Temporal variability of the dynamic brain activity is related closely to various neurophysiological activities, including working memory (Shunkai et al., 2021), vision (Di and Biswal, 2020), and cognition (Patil et al., 2021). Therefore, recent studies have shown that the dynamics of brain activity can better reflect neural activity. Dynamic FC with the sliding window method is applied widely to investigate the dynamic neural activity changes in neuroimaging studies (Yuan et al., 2019; Li et al., 2020; Tsurugizawa and Yoshimaru, 2021). Previous neuroimaging studies demonstrated that patients with PACG showed abnormal neural activity changes in the visual cortex and visual-related cortex. However, there have been no studies on the dynamic FC changes in the V1 in patients with PACG. Thus, we hypothesized that patients with PACG might be associated with dynamic FC changes in the V1.

Based on these assumptions, our study is the first to determine whether or not patients with PACG are associated with the dynamic FC changes in the V1. Moreover, we chose the dFC maps as a feature of machine-learning classification. Machine learning for MRI-related neuroimaging has been widely applied in the diagnosis of nervous system diseases. In addition, the support vector machine (SVM) method is the most commonly used supervised machine learning algorithm for MRI classification related to random forests, decision trees, and convolutional neural networks.

Materials and methods

Participants

In total, 34 patients with PACG and 34 healthy controls (matched for sex and age) were recruited in this study. The inclusion criteria for individuals with PACG were as follows: (1) the intraocular pressure is higher than 21 mmHg; (2) without other eye diseases (cataract, optic neuritis, high myopia, etc.); (3) no medical treatment of PACG; (4) eliminate patients with primary open-angle glaucoma (POAG) and normal tension glaucoma (NTG); and (5) without brain lesions including cerebral hemorrhage and cerebral infarction.

The inclusion criteria for healthy controls were as follows: (1) normal visual acuity [>1.0] in both eyes; (2) no ophthalmic diseases (i.e., optic neuritis, cataract, keratitis, etc.). This study was performed in accordance with the tenets of the Declaration of Helsinki. Each participant provided written informed consent before inclusion in the study.

MRI acquisition and experimental procedure

Magnetic resonance imaging scanning was performed on a 3-tesla magnetic resonance scanner (Discovery MR 750W system; GE Healthcare, Milwaukee, WI, USA) with the eight-channel head coil. Functional images were obtained using a gradient-echo-planar imaging sequence. fMRI scanning parameter:repetition time = 2,000 ms, echo time = 25 ms, thickness = 3.0 mm, gap = 1.2 mm, acquisition matrix = 64 × 64, field of view = 240 mm^2 × 240 mm^2, flip angle = 90°, voxel size = 3.6 mm^3 × 3.6 mm^3 × 3.6 mm^3, and 35 axial slices. Before fMRI scanning, all study subjects were informed of all the caveats of the experiment and were asked to remove any metal objects. Then, during the fMRI scanning, all subjects were asked to keep their eyes closed and relaxed without falling asleep.

fMRI data analysis

All preprocessing was performed using the toolbox for Data Processing and Analysis of Brain Imaging (DPABI)[1] (Yan et al., 2016), which is based on the Statistical Parametric Mapping (SPM12)[2] implemented in MATLAB 2013a (MathWorks, Natick,

1 http://www.rfmri.org/dpabi

2 http://www.fil.ion.ucl.ac.uk

TABLE 1 Clinical characteristics for health controls (HCs) and patients with primary angle-closure glaucoma (PACG).

	PACG group	HC group	*T*-Value	*P*-Value
Sex (male/female)	23/11	23/11	N/A	0.780
Age (years)	45.15 ± 14.95	45.30 ± 13.87	−0.038	0.970
BCVA-OD	0.44 ± 0.27	1.16 ± 0.16	−11.474	< 0.001*
BCVA-OS	0.43 ± 0.37	1.19 ± 0.16	−9.352	< 0.001*

Chi-square test for sex. Independent *t*-test was used for other normally distributed continuous data.
Data are presented as mean ± standard deviation.
HC, healthy control; BCVA, best-corrected visual acuity; OD, oculus dexter; OS, oculus sinister; N/A, not applicable; R, right.
*$p < 0.001$.

MA, USA) and the details of steps according to a previous study (Chen et al., 2022).

dFC analysis

The dFC method was performed using DPABI software. Based on previous studies (Wen et al., 2018; Qi et al., 2021; Tong et al., 2021), left V1(x = −8, y = −76, z = 10) and right V1 (x = 7, y = −76, z = 10) with 6 mm radius. Specifically, the R-fMRI indices mentioned above were computed with hamming windows (window length = 30 TR, window step = 1 TR). To avoid the introduction of spurious fluctuations during the sliding window select, the minimum window length should be larger than 1/$f$min, where $f$min is the minimum frequency of the time series.

Support vector machine (SVM) analysis

The support vector machine (SVM) algorithm for binary classification is implemented on the Pattern Recognition for Neuroimaging Toolbox (PRoNTo) software[3] (Schrouff et al., 2013). The details of the steps for the SVM method were as follows: (1) The dFC maps were selected as a classification feature. (2) The leave-one-out cross-validation (LOOCV) was applied to perform the SVM classifier validation. (3) The permutation test was applied to assess the statistical significance of the total accuracy of this classification. (4) The total accuracy, specificity, sensitivity, and area under the receiver

3 http://www.mlnl.cs.ucl.ac.uk/pronto/prtsoftware.html

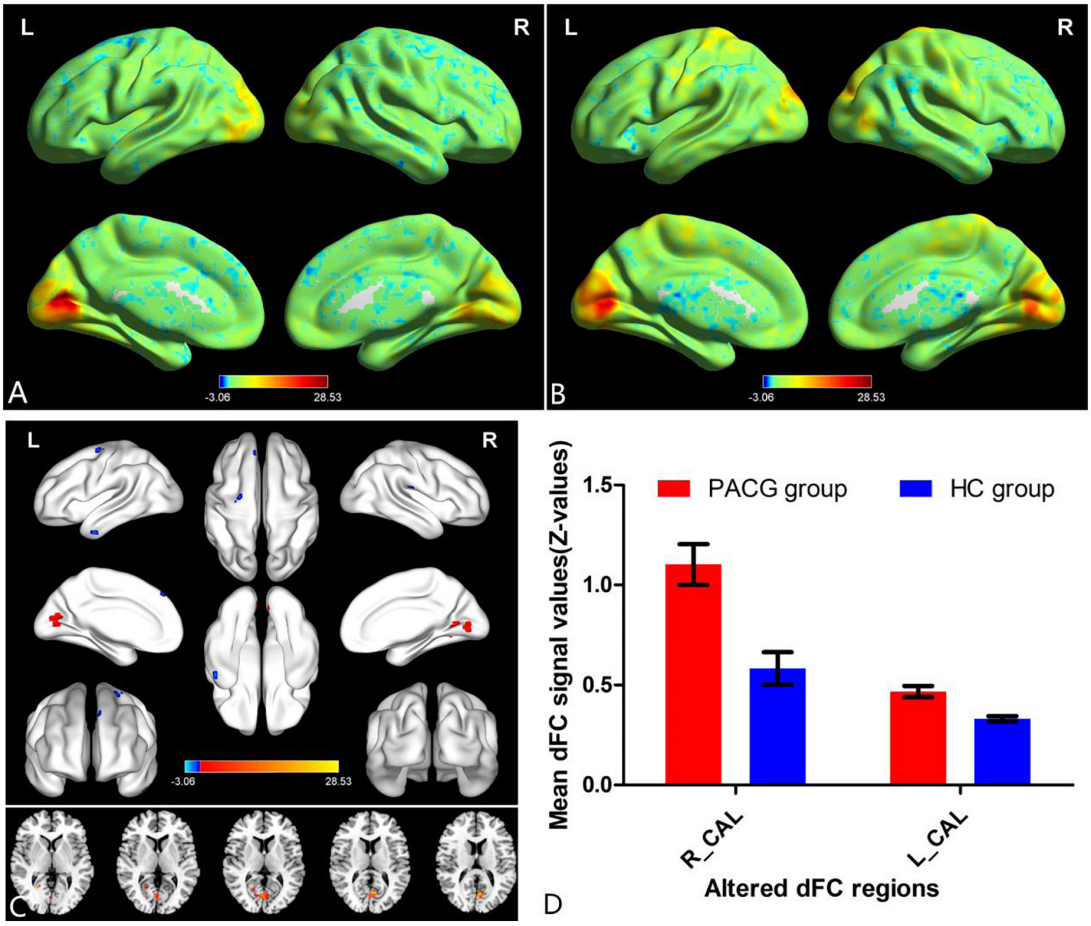


FIGURE 1
One-sample *t*-test results of dynamic functional connectivity (dFC) in left V1 maps within primary angle-closure glaucoma (PACG) group **(A)** and HC group **(B)**; significant dFC differences in left V1 between two groups **(C)**; the mean of altered dFC values of left V1 between the patients with PACG and HCs **(D)**. PACG, primary angle-closure glaucoma; HC, health control; V1, primary visual cortex; dFC, dynamic functional connectivity; CAL, calcarine; R, right; L, left.

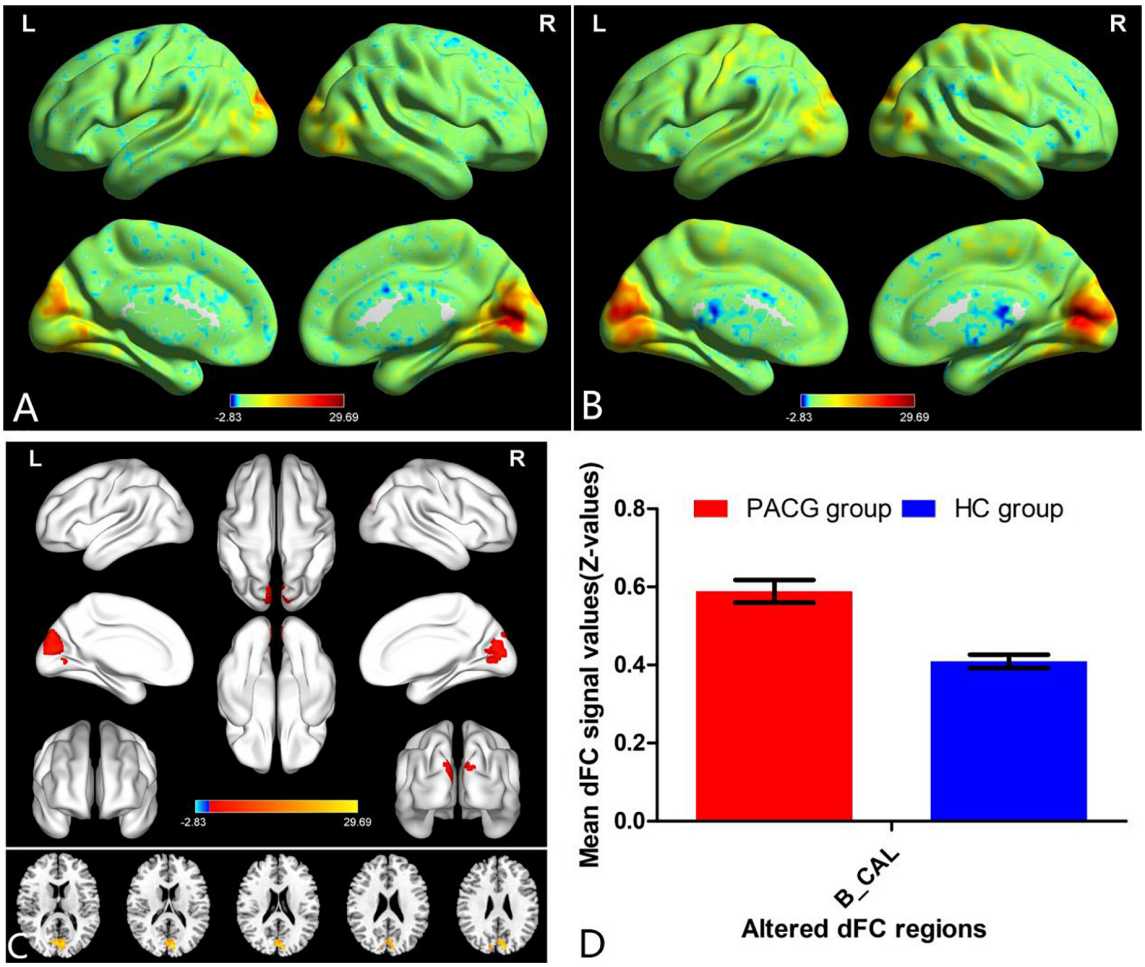


FIGURE 2
One-sample *t*-test results of dFC in right V1 maps within PACG group **(A)** and HC group **(B)**; significant dFC differences in right V1 between two groups **(C)**; the mean of altered dFC values of right V1 between the patients with PACG and HCs **(D)**. PACG, primary angle-closure glaucoma; HC, health control; V1, primary visual cortex; dFC, dynamic functional connectivity; CAL, calcarine; B, bilateral; R, right; L, left.

TABLE 2 Different dynamic functional connectivity (dFC) values in the left V1 between two groups.

Condition	Brain regions	BA	Peak T-scores	MNI coordinates (x, y, z)	Cluster size (voxels)
PACG > HC	Calcarine_R	–	3.5784	24, −57, 3	10
PACG > HC	Calcarine_L	–	4.0202	−6, −69, 12	51

Different dFC values in left V1 between two groups (voxel-level $p < 0.01$, GRF correction, cluster-level $p < 0.05$).
PACG, primary angle-closure glaucoma; HC, health control; MNI, Montreal Neurological Institute; dFC, dynamic functional connectivity; R, right; L, left.

TABLE 3 Different dFC values in the right V1 between two groups.

Condition	Brain regions	BA	Peak T-scores	MNI coordinates (x, y, z)	Cluster size (voxels)
PACG > HC	Calcarine_B	–	4.5846	21, −75, 3	237

Different dFC values in right V1 between two groups (voxel-level $p < 0.01$, GRF correction, cluster-level $p < 0.05$).
PACG, primary angle-closure glaucoma; HC, health control; MNI, Montreal Neurological Institute; dFC, dynamic functional connectivity; B, bilateral.

operating characteristic curve (AUC) were determined to assess and classify the patients with PACG and the HCs.

Statistical analysis

The independent sample *t*-test was used to investigate the clinical features between the two groups. In this study, a one-sample *t*-test was applied to assess the group the mean of dFC maps within two groups, and the two-sample *t*-test was used to compare the two group differences in the dFC maps between the two groups using the Gaussian random field (GRF) method (two-tailed, voxel-level $p < 0.01$, GRF correction, cluster-level $p < 0.05$).

Results

Analysis of ophthalmic clinical data

There were no statistically significant differences between the PACG and HC groups in gender or age. However, there were

10.3389/fnins.2023.1131247

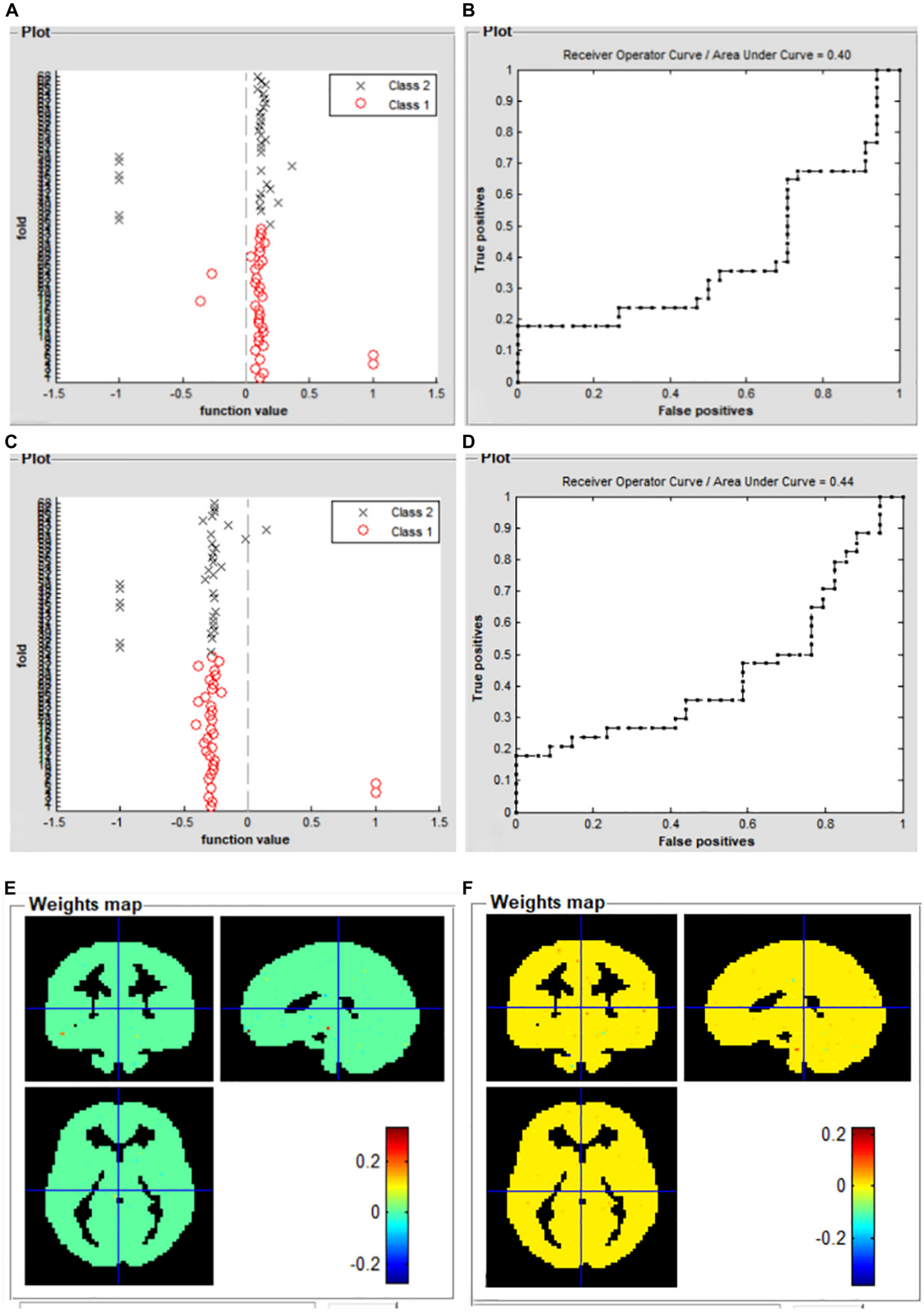


FIGURE 3
Support vector machine classification results based on the mean dFC maps. Function dFC in left V1 value of SVM algorithm for two groups (class 1: PACG group; class 2: HC group) **(A)**; the receive operating characteristic curve of the SVM classifier **(B)**; function dFC in left V1 value of SVM algorithm for two groups (class 1: PACG group; class 2: HC group) **(C)**; the receive operating characteristic curve of the SVM classifier **(D)**; weight maps for SVM models for left V1 **(E)**; weight maps for SVM models for right V1 **(F)**. PACG, primary angle-closure glaucoma; HC, health control; V1, primary visual cortex; dFC, dynamic functional connectivity; SVM, support vector machine.

significant differences between the two groups in visual acuity. More details of the results are shown in Table 1.

Different dFC values between two groups

The group means of dFC maps of the PACG and HC are shown in Figures 1A, B, 2A, B. Compared with HCs, patients with PACG showed increased dFC values between left V1 and bilateral calcarine (CAL) (Figure 1C and Table 2). The mean values of altered dFC in left V1 values were shown with a histogram (Figure 1D). Meanwhile, patients with PACG showed increased dFC values between right V1 and bilateral calcarine (CAL) (Figure 2C and Table 3). The mean values of dFC in right V1 values were shown with a histogram (Figure 2D).

Support vector machine (SVM) result

For the left V1 map, the total classification accuracy was 55.88%. Function value of SVM algorithm for two groups (in scatter diagram. class 1: PACG group; class 2: HC group) (Figure 3A). A receive operating characteristic curve of the binary classifiers was generated to evaluate the system's performance in distinguishing individuals with PACG from HCs and the AUC was 0.40 (Figure 3B).

For the right V1 map, total classification accuracy was 51.47%. Function value of SVM algorithm for two groups (in scatter diagram. class 1: PACG group; class 2: HC group) (Figure 3C). A receive operating characteristic curve of the binary classifiers was generated to evaluate the system's performance in distinguishing individuals with PACG from HCs and the AUC was 0.44 (Figure 3D). Weight maps for SVM models for left V1 (Figure 3E), and weight maps for SVM models for right V1 (Figure 3F).

Discussion

Our study is the first to investigate the dFC alterations of the V1 in patients with PACG. Patients with PACG showed increased dFC values between the left V1 and bilateral CAL. Meanwhile, patients with PACG showed increased dFC values between the right V1 and bilateral CAL related to the HC group.

Our most interesting finding is that the patients with PACG showed an increased dFC within the visual network. As we all know, the most important pathological mechanism of glaucoma is the loss of retinal ganglion cells. Meanwhile, retinal ganglion cell death can lead to anterograde and retrograde retinal ganglion cell degeneration (Lawlor et al., 2018). Thus, the damage to the optic nerve leads to abnormal visual signaling in the visual cortex. Thus, we speculated that the reduced afferent visual signals may lead to increased variability and decreased stability of primary visual cortex neural activity. Hernowo et al. (2011) reported that reduced volume in the visual pathway was found in patients with glaucoma. Chen et al. (2013) found that patients with glaucoma showed decreased gray matter (GM) volume in the visual cortex. Frezzotti et al. (2014) also demonstrated that patients with POAG had brain atrophy in both the visual cortex and other distant GM regions (the frontoparietal cortex, the hippocampi, and the cerebellar cortex). Wang et al. (2016) found that patients with glaucoma had significantly reduced cortical thickness in the right frontal pole and decreased GM volume in LGN, the right V1, and the left amygdala. Thus, brain structural atrophy in VI may lead to increased dFC values of V1 in patients with glaucoma. We speculated that the abnormal brain structure may lead to local neural activity flexibility changes. Consistent with these findings, our study found that patients with PACG showed increased dFC values within the visual network. Increased dFC might reflect increased neural activity flexibility in the visual cortex, which might indicate visual function compensation in patients with PACG.

In addition, in our study, the SVM method was used to investigate the predictive values of classifying individual patient populations with PACG versus HCs. For the left V1 map, the total classification accuracy was 55.88%. A receiver operating characteristic (ROC) curve of the binary classifiers was generated to evaluate the system's performance in distinguishing individuals with PACG from HCs and the area under the ROC curve (AUC) was 0.40. For the right V1 map, the total classification accuracy was 51.47%. A ROC curve of the binary classifiers was generated to evaluate the system's performance in distinguishing individuals with PACG from HCs and the AUC was 0.44. The SVM algorithm was trained with input data labeled previously by PACG or HCs to predict the desired outcome. This powerful multivariate analysis enabled us to make clinical predictions at the individual subject level.

However, there were some limitations to the study. First, the sample size of our study was small. Second, blood oxygen level-dependent (BOLD) signals would be influenced by a variety of physiological noises such as breathing and heartbeat. Third, the machine noise of MRI can also have an effect on the BOLD signal. Finally, a different time window length might affect the results. We would select a different window length in the future study for the stability of the results.

In conclusion, our study demonstrated that patients with PACG showed an increased dFC within the visual network, which might be indicative of an increased variability of FC in the primary visual cortex. Our study also provides an important imaging reference to aid the understanding of the mechanism of nerve damage in the visual center of patients with glaucoma.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by tenets of the Declaration of Helsinki of Jiangxi Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

D-JL, B-LH, YP, L-YL, and HL contributed to the data collection, statistical analyses, and wrote the manuscript. B-LH, YP, L-YL, and

HL designed the protocol and contributed to the MRI analysis and designed the study, oversaw all clinical aspects of study conduct, and manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the National Natural Science Foundation of China Youth Science Fund Project (No. 821049352), the Natural Science Foundation of Jiangxi Province (No. 20202BAB2060743), and the Science and Technology Research Project of Jiangxi Provincial Department of Education (No. GJJ190647).

Conflict of interest

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

Publisher's note

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

References

Chen, J., Chen, W. Y., and Huang, X. (2022). More than static: Altered dynamic functional connectivity of primary visual cortex in the comitant exotropia patients. *Neuroreport* 33, 812–818. doi: 10.1097/WNR.0000000000001849

Chen, W. W., Wang, N., Cai, S., Fang, Z., Yu, M., Wu, Q., et al. (2013). Structural brain abnormalities in patients with primary open-angle glaucoma: A study with 3T MR imaging. *Invest. Ophthalmol. Vis. Sci.* 54, 545–554. doi: 10.1167/iovs.12-9893

Cheng, J., Zong, Y., Zeng, Y., and Wei, R. (2014). The prevalence of primary angle closure glaucoma in adult Asians: A systematic review and meta-analysis. *PLoS One* 9:e103222. doi: 10.1371/journal.pone.0103222

Di, X., and Biswal, B. B. (2020). Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. *Neuroimage* 216:116698. doi: 10.1016/j.neuroimage.2020.116698

Frezzotti, P., Giorgio, A., Motolese, I., Leucio, A. D., Iester, M., Motolese, E., et al. (2014). Structural and functional brain changes beyond visual system in patients with advanced glaucoma. *PLoS One* 9:e105931. doi: 10.1371/journal.pone.0105931

Graham, K. L., Johnson, P. J., Barry, E. F., Orrico, M. P., Soligo, D. J., Lawlor, M., et al. (2021). Diffusion tensor imaging of the visual pathway in dogs with primary angle-closure glaucoma. *Vet. Ophthalmol.* 24, 63–74. doi: 10.1111/vop.12824

Haykal, S., Curcic-Blake, B., Jansonius, N., and Cornelissen, F. (2019). Fixel-Based analysis of visual pathway white matter in primary open-angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* 60, 3803–3812. doi: 10.1167/iovs.19-27447

Haykal, S., Invernizzi, A., Carvalho, J., Jansonius, N., and Cornelissen, F. (2022). Microstructural visual pathway white matter alterations in primary open-angle glaucoma: A neurite orientation dispersion and density imaging study. *AJNR Am. J. Neuroradiol.* 43, 756–763. doi: 10.3174/ajnr.A7495

Hernowo, A. T., Boucard, C. C., Jansonius, N. M., Hooymans, J. M., and Cornelissen, F. W. (2011). Automated morphometry of the visual pathway in primary open-angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* 52, 2758–2766. doi: 10.1167/iovs.10-5682

Jiang, F., Fang, J., Ye, Y., Tian, Y., Zeng, X., and Zhong, Y. (2020). Altered effective connectivity of primary visual cortex in primary angle closure glaucoma using Granger causality analysis. *Acta Radiol.* 61, 508–519. doi: 10.1177/0284185119867644

Lawlor, M., Danesh-Meyer, H., Levin, L. A., Davagnanam, I., Vita, E. D., and Plant, G. T. (2018). Glaucoma and the brain: Trans-synaptic degeneration, structural change, and implications for neuroprotection. *Surv. Ophthalmol.* 63, 296–306. doi: 10.1016/j.survophthal.2017.09.010

Li, S., Li, P., Gong, H., Jiang, F., Liu, D., Cai, F., et al. (2017). Intrinsic functional connectivity alterations of the primary visual cortex in primary angle-closure glaucoma patients before and after surgery: A resting-state fMRI study. *PLoS One* 12:e170598. doi: 10.1371/journal.pone.0170598

Li, Y., Zhu, Y., Nguchu, B. A., Wang, Y., Wang, H., Qiu, B., et al. (2020). Dynamic functional connectivity reveals abnormal variability and hyper-connected pattern in autism spectrum disorder. *Autism Res.* 13, 230–243. doi: 10.1002/aur.2212

Nuzzi, R., Dallorto, L., and Rolle, T. (2018). Changes of visual pathway and brain connectivity in glaucoma: A systematic review. *Front. Neurosci.* 12:363. doi: 10.3389/fnins.2018.00363

Pankowska, A., Matwiejczuk, S., Koziol, P., Żarnowski, T., Pietura, R., Kosior-Jarecka, E., et al. (2022). Visual tract degradation in bilateral normal-tension glaucoma-cortical thickness maps and volumetric study of visual pathway areas. *J. Clin. Med.* 11:1907. doi: 10.3390/jcm11071907

Patil, A. U., Ghate, S., Madathil, D., Tzeng, O. J., Huang, H., and Huang, C. (2021). Static and dynamic functional connectivity supports the configuration of brain networks associated with creative cognition. *Sci. Rep.* 11:165. doi: 10.1038/s41598-020-80293-2

Qi, C., Huang, X., Tong, Y., and Shen, Y. (2021). Altered functional connectivity strength of primary visual cortex in subjects with diabetic retinopathy. *Diabetes Metab. Syndr. Obes.* 14, 3209–3219. doi: 10.2147/DMSO.S311009

Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., et al. (2013). PRoNTo: Pattern recognition for neuroimaging toolbox. *Neuroinformatics* 11, 319–337. doi: 10.1007/s12021-013-9178-1

Shunkai, L., Su, T., Zhong, S., Chen, G., Zhang, Y., Zhao, H., et al. (2021). Abnormal dynamic functional connectivity of hippocampal subregions associated with working memory impairment in melancholic depression. *Psychol. Med.* 1–13. doi: 10.1017/S0033291721004906

Tong, Y., Huang, X., Qi, C., and Shen, Y. (2021). Altered functional connectivity of the primary visual cortex in patients with iridocyclitis and assessment of its predictive value using machine learning. *Front. Immunol.* 12:660554. doi: 10.3389/fimmu.2021.660554

Tsurugizawa, T., and Yoshimaru, D. (2021). Impact of anesthesia on static and dynamic functional connectivity in mice. *Neuroimage* 241:118413. doi: 10.1016/j.neuroimage.2021.118413

Wang, J., Li, T., Sabel, B. A., Chen, Z., Wen, H., Li, J., et al. (2016). Structural brain alterations in primary open angle glaucoma: A 3T MRI study. *Sci. Rep.* 6:18969. doi: 10.1038/srep18969

Wen, Z., Zhou, F., Huang, X., Dan, H. D., Xie, B., and Shen, Y. (2018). Altered functional connectivity of primary visual cortex in late blindness. *Neuropsychiatr. Dis. Treat.* 14, 3317–3327. doi: 10.2147/NDT.S183751

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

Yuan, Y., Zhang, L., Li, L., Huang, G., Anter, A., Liang, Z., et al. (2019). Distinct dynamic functional connectivity patterns of pain and touch thresholds: A resting-state fMRI study. *Behav. Brain Res.* 375:112142. doi: 10.1016/j.bbr.2019.112142

frontiers | Frontiers in Neuroscience

TYPE Review
PUBLISHED 02 February 2023
DOI 10.3389/fnins.2023.1125638

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Hua-Tao Xie,
Huazhong University of Science and Technology, China
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China

*CORRESPONDENCE
Gang Liu
✉ liugangdr@163.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 16 December 2022
ACCEPTED 11 January 2023
PUBLISHED 02 February 2023

CITATION
Wang J, Zhang Y, Meng X and Liu G (2023) Application of diffusion tensor imaging technology in glaucoma diagnosis. *Front. Neurosci.* 17:1125638. doi: 10.3389/fnins.2023.1125638

Application of diffusion tensor imaging technology in glaucoma diagnosis

Jiaqi Wang, Yaqiong Zhang, Xueyu Meng and Gang Liu*

Department of Ophthalmology, Xiangyang No. 1 People's Hospital, Hubei University of Medicine, Xiangyang, Hubei, China

Glaucoma is the first major category of irreversible blinding eye illnesses worldwide. Its leading cause is the death of retinal ganglion cells and their axons, which results in the loss of vision. Research indicates that glaucoma affects the optic nerve and the whole visual pathway. It also reveals that degenerative lesions caused by glaucoma can be found outside the visual pathway. Diffusion tensor imaging (DTI) is a magnetic resonance imaging (MRI) technique that can investigate the complete visual system, including alterations in the optic nerve, optic chiasm, optic tract, lateral geniculate nuclear, and optic radiation. In order to provide a more solid foundation for the degenerative characteristics of glaucoma, this paper will discuss the standard diagnostic techniques for glaucoma through a review of the literature, describe the use of DTI technology in glaucoma in humans and animal models, and introduce these techniques. With the advancement of DTI technology and its coupling with artificial intelligence, DTI represents a potential future for MRI technology in glaucoma research.

KEYWORDS

glaucoma, diffusion tensor imaging, visual field, OCT, OCTA

1. Introduction

Glaucoma is a diverse group of diseases characterized by damage to the retinal ganglion cells (RGC) and their axons, the retinal nerve fiber layer (RNFL), which leads to progressive vision loss (Stein et al., 2021). If left untreated, this condition can result in permanent blindness. Glaucoma is thought to affect more than 76 million people worldwide; by 2040, that number is expected to increase to 112 million (Tham et al., 2014). High intraocular pressure (IOP) is a significant risk factor for glaucoma. In some instances, however, reducing the IOP to normal or even below normal can cause visual impairment. IOP is the only clinical risk factor that may be modified. In addition, risk factors for glaucoma include advanced age (Kühn et al., 2021), race (Cheng and Tanna, 2022), myopia (Haarman et al., 2020), and glaucoma-positive family history (Bhandari et al., 2021); nevertheless, the exact reason is unknown (Schuster et al., 2020). Epidemiology studies show that the pathophysiological pathways of glaucoma are similar to those of Alzheimer's disease and Parkinson's disease. They suggested that the brain may be involved in the development of glaucoma (Zhang et al., 2019; Saccà et al., 2020). Historically, glaucoma referred to illnesses resulting in optic nerve (ON) atrophy and visual field (VF) abnormalities. However, an increasing amount of evidence indicates that intracranial vision-related regions and visual pathways, such as the lateral geniculate nuclear (LGN), optic tract (OT), and optic radiation (OR), are also affected (Zhou et al., 2017; Schmidt et al., 2018; Song et al., 2018). In addition, some investigations have found a decrease in the volume of brain structures outside the visual pathway (Chen et al., 2013). Visual pathways are made up

of white matter (WM) tracts, so studying glaucomatous WM degeneration may be crucial to understanding how glaucomatous neurodegeneration spreads through the visual system.

Diffusion tensor imaging (DTI) is now the most prevalent method for investigating WM degeneration in glaucoma patients (Hanekamp et al., 2021; Haykal et al., 2022). In addition, the DTI technique is frequently utilized to detect glaucoma changes in various glaucoma types (Zhang et al., 2019). In this review, we explored the application of DTI in glaucoma and assessed it as a recently created technique.

2. Conventional glaucoma screening methods

Routinely used methods for glaucoma evaluation include gonioscopy, IOP measurement, corneal thickness measurement, fundus photography, VF inspection, and optical coherence tomography (OCT) (Stein et al., 2021). Moreover, optical coherence tomography angiography (OCTA) identifies glaucoma by detecting alterations in the retinal blood vessels (Rao et al., 2020). Due to irreversible glaucomatous vision damage, early detection is essential for preventing its progression. Through screening and identifying high-risk patients, ophthalmologists can diagnose glaucoma in its early stages. Given that the morphology of a healthy ON varies considerably and various eye disorders can present with glaucomatous alterations in the ON, early identification of glaucoma can be difficult. For glaucoma diagnosis, direct visual evaluation of optic papilla morphology has a high mistake rate. Consequently, early glaucoma is identified by longitudinally assessing structural changes in the ocular papilla.

The term "visual field" refers to the area the human eye can perceive, and the "VF examination" primarily refers to the perimeter measuring this region. VF examination is a subjective test used to diagnose glaucomatous damage and quantify the degree and severity of visual impairment in glaucoma patients (Hashimoto et al., 2018). The automated static visual field test effectively detects and monitors visual function loss caused by glaucoma. Even in the earliest phases of glaucoma, more and more macular abnormalities are detected (Arai et al., 2018). Considering the significance of central visual function, some have recently proposed that regular targeted central VF testing, such as the 10-2 approach on Humphrey Field Analyzer (HFA), can detect glaucomatous alterations in their early phases (Grillo et al., 2016). In addition, the 10-2 strategy on the HFA can detect abnormal outcomes in glaucoma patients that the 24-2 strategy cannot detect (Grillo et al., 2016). Therefore, suitable parameter values may facilitate earlier diagnosis of glaucoma. In order to sample the same retinal position, the patient needs to maintain a steady position throughout standard VF testing for SAP. Thus, the researchers utilized a fundus-tracked visual field test technique to lower measurement variability and to be able to quickly detect VF advancement when patient participation declines over time (Wu et al., 2016; Rao et al., 2017). Furthermore, implementing a new visual field threshold technique and an upgraded visual field progression analysis offer promise for the early identification of glaucoma (Aoki et al., 2017; Wild et al., 2017).

Optical coherence tomography is a non-invasive imaging technology that permits the quantitative measurement of changes in the ON, RGC axon, and RGC body layer, RNFL (Kang and Tanna, 2021). Since its inception in 1991 (Huang et al., 1991), OCT has undergone several modifications, including time-domain OCT (TD-OCT), spectral-domain OCT (SD-OCT), and swept-source OCT (SS-OCT) (Geevarghese et al., 2021). These enhancements have substantially increased scanning resolution and speed. OCT has transformed the ability to analyze the anatomical characteristics of an optic papilla afflicted by glaucoma. In addition, OCT can detect glaucoma before VF changes (Swaminathan et al., 2021). OCT has transformed glaucoma from a disease that could only be diagnosed subjectively to one that can now be evaluated objectively. This significant development has revolutionized the diagnosis, monitoring, and treatment of glaucoma. As a result, it has become the most common method for detecting and monitoring glaucoma among clinicians. According to the study, RNFL-based SD-OCT and TD-OCT have the same ability to distinguish glaucoma from other eye diseases. However, SD-OCT is superior for determining RNFL advancement (Bengtsson et al., 2012). Due to the SD-OCT rapid scan speed and excellent images, the retina may be analyzed more precisely, and there is less variance in the results. Finally, it increased the precision of glaucoma diagnosis.

Optical coherence tomography angiography is a non-invasive, dye-free imaging technique that can measure the retina, optic papilla, and choroidal blood vessels in qualitative and quantitative ways. OCTA can provide information on the extent of perfusion damage at various depths in the retina and choroid. Because advanced glaucoma has a lower vascular density (Yarmohammadi et al., 2018), OCTA can be utilized to track the progression of ocular damage in this condition. Patients at risk for a more rapid glaucoma progression are also included in the OCTA database. OCTA works with VF and OCT tests (Rao et al., 2020) to diagnose glaucoma, identify development, and assess progression risk.

Conventional glaucoma diagnostic technologies, including fundus photography, OCT, OCTA, and VF, focus exclusively on the retinal region and disregard information regarding the WM of the brain, which transmits and systematically processes vision. The visual pathway consists mainly of the WM of the brain, and DTI is the advanced imaging technology currently available for evaluating the WM of the brain (Cheng et al., 2021); consequently, DTI is essential for detecting glaucoma.

3. Application of DTI technology in glaucoma

Diffusion-weighted imaging (DWI) is a magnetic resonance imaging (MRI) technique used to assess water diffusion (Graham et al., 2021). Due to the presence of water diffusion in numerous types of biological tissues, pathophysiological alterations impact typical cellular structures, resulting in variations in water diffusion. DWI can detect aberrant changes in diffusion and can be used to evaluate brain tissue's integrity, connectivity, and structure (Martinez-Heras et al., 2021). Glaucoma is considered a multifactorial neurodegenerative disease affecting the visual pathway rather than just an ophthalmic condition with VF abnormalities and optic neuropathy. DTI, which was first introduced by Basser et al. (1994) in 1994, is a magnetic resonance technique based on DWI technology that can analyze the water molecule dispersion characteristics in tissue in three dimensions at regular intervals and quantitatively and has significant applications in neuroimaging. DTI enables a sensitive evaluation of potential microstructural changes that may occur prior to brain tissue shrinkage. Therefore, this technique is up-and-coming for the study

of intracerebral damage in glaucoma (Le Bihan and Iima, 2015). DTI measures several parameters, including fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD). FA measures the propensity of water to diffuse in one direction, with values ranging from 0 (isotropic) to 1 (anisotropic) (Li and Zhang, 2020). MD describes the average mobility of water molecules within a voxel. In contrast, AD measures diffusivity along the horizontal axes, and RD offers an average diffusion measurement along the two vertical axes (Altobelli et al., 2015). Changes in these variables may signify alterations in axonal integrity and demyelination. Recently, DTI has been extensively utilized to evaluate glaucoma-induced impairments to the intracranial visual system in humans and animals.

4. DTI in glaucoma *in vivo*

Changes in DTI parameters of the glaucomatous ON and beyond the ON are hot areas of current research (Sidek et al., 2014; Zhang Q. J. et al., 2015). The following section focuses on developing DTI investigations of the visual pathway in animal and human populations with glaucoma.

4.1. DTI in animal models of glaucoma

Diffusion tensor imaging has been examined extensively and thoroughly in glaucoma-affected animal models with altered visual pathways in experimental animal models. Numerous investigations on rat glaucoma models have documented an increase in MD and a decrease in FA of the ON and OT (Hui et al., 2007; Ho et al., 2015; Yang et al., 2018). Moreover, *in vivo* DTI was demonstrated to detect RGC axonal damage earlier than immunohistochemistry in the early stages of the disease in a mouse model of optic nerve crush-induced glaucoma (Zhang et al., 2011). DTI can therefore be utilized to detect FA of OR and OT and aid in the early identification of POAG. DTI was used to test the visual system integrity in five animal models of glaucoma (Colbert et al., 2021). The evaluation of three glaucoma models that are caused by genes and two that are induced by experiments. Significantly less FA and more RD were found in the visual pathways of DBA/2J mice and LTBP2-mutant cats; also, AD was slower in DBA/2J mice. Chronic high IOP is linked to less FA and more RD along the ON or OT, which suggests that microstructural integrity has been compromised. In addition, Graham et al. (2021) use DTI to describe measurements of structures associated with canine primary angle closure glaucoma (PACG) in the retina and visual pathways beyond the optic papilla. Quantitative measurements of the ON, optic chiasma (OC), OT, and LGN were taken in dogs with and without PACG. A tendency toward a disease-related decline in FA was seen for all structures evaluated. *In vivo* evaluations of axonal, myelin, and trans-synaptic degeneration in canine PACG can be done using DTI. Research on DTI is currently being done primarily in rat eyes, and future research on glaucoma in large animals, including dogs, pigs, and monkeys, will also be done.

4.2. DTI in humans with glaucoma

Glaucoma-related modifications to DTI parameters in the visual pathway constitute a significant research focus. Multiple studies have found that glaucoma patients had lower FA and higher MD on DTI (Chen et al., 2013; Murai et al., 2013). Moreover, studies have established that FA is a more sensitive and reliable glaucoma detection indicator than MD (Garaci et al., 2009; Chen et al., 2013). These works of literature focus on the various types of glaucoma and the associated changes in the intracranial visual and extra-visual pathways involved in glaucoma (Giorgio et al., 2018; Schmidt et al., 2018; Wang et al., 2018; Qu et al., 2019). It is well-known that VF examination and OCT measurement of the RNFL thickness are the accepted methods for diagnosing primary open angle glaucoma (POAG) and ocular hypertension (OHT). And Song et al. (2018) indicate that DTI distinguishes POAG from OHT by evaluating the FA and MD of the OT, LGN, and OR in the visual pathway. DTI characteristics can quantify the evolution of POAG.

Schmidt et al. (2018) studied the potential advantages of volumetric LGN and DTI evaluation techniques. Normal tension glaucoma (NTG) was observed to considerably diminish the size of the LGN compared to healthy controls. In addition, FA of the OT and OR are lowered in NTG. Although RNFL thickness was related to LGN volume, FA was not correlated with LGN volume. Likewise, Li et al. (2019) examined the diagnostic value of DTI parameters and LGN size in POAG. FA values of the OT may be a sensitive and accurate biomarker for glaucoma assessment, even though MD is not connected with this condition. Zhou et al. (2017) discovered that the DTI parameters FA and RD correspond with monocular right and left visual fields, although there is no significant correlation between FA and RNFL thickness. The connection between FA and contralateral VF scores for OR is significantly positive. By splitting the left and right VF, FA readings for OR can be used to evaluate the degree of glaucomatous visual field loss. The results of the OT evaluation revealed no correlation. The results of the OT evaluation revealed no correlation. The OT is relatively narrow, has fewer neuronal bundles, and has poorer sensitivity, which may cause the absence of association. Engelhorn et al. (2011) used DTI to assess the pathological abnormalities in glaucoma patients and found that it can demonstrate the sparsity of OR. The study included 50 glaucoma patients and 50 age-matched healthy controls. Twenty-two glaucoma patients (44%) were found to have significantly lower OR volumes than the control group (67.16%). As a result, glaucomatous ON atrophy and OR thinning brought on by DTI coexist. Furthermore, compared to controls, glaucoma patients had significantly higher incidences of cerebral microangiopathy affecting OR.

Researchers are particularly interested in determining whether abnormalities occur outside the visual pathway in glaucoma patients. In contrast to nature controls, NTG patients show decreased FA of OR and forceps major in the occipital lobe, according to a DTI study by Boucard et al. (2016). The same changes were observed in non-visual regions, including the corpus callosum and parietal lobe. Moreover, Giorgio et al. (2018) discovered aberrant WM in the lingual gyrus and lateral occipital cortex of the occipital lobe in the NTG compared to the NC group, demonstrating a decrease in FA and an increase in AD. Moreover, Zikou et al. (2012) used DTI to identify decreases in FA in the inferior frontal-occipital fasciculus, the longitudinal and inferior frontal fasciculi, the putamen, the caudate nucleus, the anterior and posterior thalamic radiations, and the anterior and posterior limbs of the internal capsule.

As a result, we can now consider glaucoma a neurodegenerative disorder affecting the visual and extra-visual pathways owing to analyzing various types of glaucoma utilizing DTI. Additionally, DTI

offers objective and quantifiable changes in the visual pathway and improves understanding of the central visual pathway's degenerative process in human and animal glaucoma.

5. Correlation of DTI parameters and glaucoma severity

Multiple investigations have demonstrated that as the disease progresses, FA and MD values in the ON and OR decrease and increase, respectively (Garaci et al., 2009; Chen et al., 2013; Li et al., 2014). In addition, Sidek et al. (2014) found substantial variations between mild and severe glaucoma in the FA and MD of the ON and OR. FA values of either the ON or OR showed greater sensitivity and specificity in differentiating between mild and severe glaucoma than MD values. Similarly, Li et al. (2014) found that the heterogeneity analysis shows that FA may have a link with the severity of glaucoma. This study suggests that DTI may be helpful for glaucoma diagnosis and management. A meta-analysis investigated the connection between structural WM alterations and glaucoma severity. Typically, the severity of glaucoma increases the prominence of cerebral WM loss (Sidek et al., 2014). Therefore, based on the extent of parameter measurement in the visual pathway, DTI can differentiate the severity of glaucoma and offer some clinical support.

6. Limitations of DTI in glaucoma research

Application DTI has some limitations. The human ON is small, only 3–4 mm in diameter. There may be specific areas throughout the inspection when artifacts lead to examination errors (Wheeler-Kingshott et al., 2006). This restriction can be removed by utilizing advanced DTI and higher MRI scanners. It is important to remember that DTI parameters are voxel-based values that the fiber image located around the same voxel can change. Significant noise is detected for the DTI image, and the noise image leads to measurement inaccuracies.

Moreover, the DTI signal acquisition time is extended, which could cause motion distortions if the patient moves a lot during the examination. DTI should be enhanced and modified prior to its widespread usage in clinical glaucoma testing, which is now employed primarily for research. The accuracy and reproducibility of DTI results can be impacted by oculomotor interference, large magnetic sensitivity differences in the orbit caused by gas and bone in the sinus combined with the thin diameter of the optic nerve, encapsulation by cerebrospinal fluid in the sphincter cavity, and surrounding orbital fat (Techavipoo et al., 2009). The final results may need to be interpreted as a result of inaccurate diffusion measurements that need to adequately reflect microstructure information due to methodological and artificial factors (Concha, 2014). However, DTI measurements need a more precise biological meaning, and the method has many technical limitations. The DTI-based analysis is still an impenetrable technology that relies on complex data gathering and geometrical models that are predicated on many different assumptions (Pujol et al., 2015).

In this review, we only discuss the alterations of DTI in the visual and extra-visual pathways because not only WM alterations but also gray matter alterations are present in the central visual pathways of glaucoma patients, and intracranial gray matter alterations are typically detected by functional MRI (Garaci et al., 2008; Zhang et al., 2015), which is not discussed in this paper and is its limitation. Future research will delve deeper into glaucoma sufferers' gray matter changes.

7. Future prospects

The area of glaucoma and ophthalmology, in general, is primarily image-based, and AI is situated to solve many of these problems (Mayro et al., 2020). Deep learning (DL) is a subset of AI. Using multilayer neural networks modeled after the mammalian visual cortex, DL in AI may generate images that modify the glaucoma field. Autonomous DL algorithms can maximize information in digital fundus pictures, optical coherence tomography, and visual fields (Ting et al., 2019; Girard and Schmetterer, 2020). Suppose DL technology can be merged with DTI to evaluate changes in images and parameters. In that case, it will be possible to explore the characteristics of glaucoma as a neurodegenerative disease and distinguish glaucoma patients from those without the problem. In the future, the combination of AI and DTI will significantly impact outpatient glaucoma screening, management, and the exploration of glaucoma's relationship to other neurodegenerative illnesses. In addition, as technology improves and is combined with other MRI technologies (Lanzafame et al., 2016; Wang et al., 2018), glaucoma will be detected earlier and can be studied as a whole to find out how glaucoma starts, cease it from getting worse, improve the prognosis of glaucoma, and take glaucoma research to a new level, all of which will help glaucoma patients in the long run.

8. Conclusion

In summary, DTI often showed a rise in MD and a decrease in FA, strongly linked with increasing disease severity in glaucoma patients or animal models. DTI is a promising non-invasive technique for assessing the severity and prognosis of glaucoma. As clinical and scientific uses of DTI continue to develop, practitioners in the area will engage with ophthalmologists to overcome its limits and enhance patient care.

Author contributions

All authors contributed to draft and revise the manuscript, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

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

Altobelli, S., Toschi, N., Mancino, R., Nucci, C., Schillaci, O., Floris, R., et al. (2015). Brain imaging in glaucoma from clinical studie s to clinical practice. *Progr. Brain Res.* 221, 159–175. doi: 10.1016/bs.pbr.2015.06.004

Aoki, S., Murata, H., Fujino, Y., Matsuura, M., Miki, A., Tanito, M., et al. (2017). Investigating the usefulness of a cluster-based trend analysis to detect visual field progression in patients with open-angle glaucoma. *Br. J. Ophthalmol.* 101, 1658–1665. doi: 10.1136/bjophthalmol-2016-310069

Arai, T., Murata, H., Matsuura, M., Usui, T., and Asaoka, R. (2018). The association between ocular surface measurements with visual field reliability indices and gaze tracking results in preperimetric glaucoma. *Br. J. Ophthalmol.* 102, 525–530. doi: 10.1136/bjophthalmol-2017-310309

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. doi: 10.1016/S0006-3495(94)80775-1

Bengtsson, B., Andersson, S., and Heijl, A. (2012). Performance of time-domain and spectral-domain optical coherence tomography for glaucoma screening. *Acta Ophthalmol.* 90, 310–315. doi: 10.1111/j.1755-3768.2010.01977.x

Bhandari, R. D., Khanal, B., Poudel, M., Shrestha, M. K., and Thapa, S. S. (2021). Glaucoma screening in family members of glaucoma patients at a tertiary eye hospital in eastern region of Nepal. *Nepal. J. Ophthalmol.* 13, 128–136. doi: 10.3126/nepjoph.v13i2.34673

Boucard, C. C., Hanekamp, S., Ćurčić-Blake, B., Ida, M., Yoshida, M., and Cornelissen, F. W. (2016). Neurodegeneration beyond the primary visual pathways in a population with a high incidence of normal-pressure glaucoma. *Ophthalmic Physiol. Opt.* 36, 344–353. doi: 10.1111/opo.12297

Chen, W. W., Wang, N., Cai, S., Fang, Z., Yu, M., Wu, Q., et al. (2013). Structural brain abnormalities in patients with primary open-angle glaucoma: A study with 3T MR imaging. *Invest. Ophthalmol. Vis. Sci.* 54, 545–554. doi: 10.1167/iovs.12-9893

Chen, Z., Lin, F., Wang, J., Li, Z., Dai, H., Mu, K., et al. (2013). Diffusion tensor magnetic resonance imaging reveals visual pathway damage that correlates with clinical severity in glaucoma. *Clin. Exp. Ophthalmol.* 41, 43–49. doi: 10.1111/j.1442-9071.2012.02832.x

Cheng, B. T., and Tanna, A. P. (2022). Association of race and ethnicity with the frequency of outpatient glaucoma care. *Ophthalmology* 129, 585–587. doi: 10.1016/j.ophtha.2021.11.026

Cheng, S. J., Tsai, P. H., Lee, Y. T., Li, Y. T., Chung, H. W., and Chen, C. Y. (2021). Diffusion tensor imaging of the spinal cord. *Magn. Reson. Imaging Clin. North Am.* 29, 195–204. doi: 10.1016/j.mric.2021.02.002

Colbert, M. K., Ho, L. C., van der Merwe, Y., Yang, X., McLellan, G. J., Hurley, S. A., et al. (2021). Diffusion tensor imaging of visual pathway abnormalities in five glaucoma animal models. *Invest. Ophthalmol. Vis. Sci.* 62:21. doi: 10.1167/iovs.62.10.21

Concha, L. (2014). A macroscopic view of microstructure: Using diffusion-weighted images to infer damage, repair, and plasticity of white matter. *Neuroscience* 276, 14–28. doi: 10.1016/j.neuroscience.2013.09.004

Engelhorn, T., Michelson, G., Waerntges, S., Struffert, T., Haider, S., and Doerfler, A. (2011). Diffusion tensor imaging detects rarefaction of optic radiation in glaucoma patients. *Acad. Radiol.* 18, 764–769. doi: 10.1016/j.acra.2011.01.014

Garaci, F. G., Bolacchi, F., Cerulli, A., Melis, M., Spanò, A., Cedrone, C., et al. (2009). Optic nerve and optic radiation neurodegeneration in patients with glaucoma: In vivo analysis with 3-T diffusion-tensor MR imaging. *Radiology* 252, 496–501. doi: 10.1148/radiol.2522081240

Garaci, F. G., Cozzolino, V., Nucci, C., Gaudiello, F., Ludovici, A., Lupattelli, T., et al. (2008). A in neuroimaging of the visual pathways and their use in glaucoma. *Prog. Brain Res.* 173, 165–177. doi: 10.1016/S0079-6123(08)01112-6

Geevarghese, A., Wollstein, G., Ishikawa, H., and Schuman, J. S. (2021). Optical coherence tomography and glaucoma. *Annu. Rev. Vision Sci.* 7, 693–726. doi: 10.1146/annurev-vision-100419-111350

Giorgio, A., Zhang, J., Costantino, F., De Stefano, N., and Frezzotti, P. (2018). Diffuse brain damage in normal tension glaucoma. *Hum. Brain Mapp.* 39, 532–541. doi: 10.1002/hbm.23862

Girard, M. J. A., and Schmetterer, L. (2020). Artificial intelligence and deep learning in glaucoma: Current state and future prospects. *Prog. Brain Res.* 257, 37–64. doi: 10.1016/bs.pbr.2020.07.002

Graham, K. L., Johnson, P. J., Barry, E. F., érez Orrico, M. P., Soligo, D. J., Lawlor, M., et al. (2021). Diffusion tensor imaging of the visual pathway in dogs with primary angle-closure glaucoma. *Vet. Ophthalmol.* 24(Suppl. 1), 63–74. doi: 10.1111/vop.12824

Grillo, L. M., Wang, D. L., Ramachandran, R., Ehrlich, A. C., De Moraes, C. G., Ritch, R., et al. (2016). The 24-2 visual field test misses central macular damage confirmed by the 10-2 visual field test and optical coherence tomography. *Transl. Vision Sci. Technol.* 5:15. doi: 10.1167/tvst.5.2.15

Haarman, A. E. G., Enthoven, C. A., Tideman, J. W. L., Tedja, M. S., Verhoeven, V. J. M., and Klaver, C. C. W. (2020). The complications of myopia: A review and meta-analysis. *Invest. Ophthalmol. Vis. Sci.* 61:49. doi: 10.1167/iovs.61.4.49

Hanekamp, S., Ćurčić-Blake, B., Caron, B., McPherson, B., Timmer, A., Prins, D., et al. (2021). White matter alterations in glaucoma and monocular blindness differ outside the visual system. *Sci. Rep.* 11:6866. doi: 10.1038/s41598-021-85602-x

Hashimoto, S., Matsumoto, C., Eura, M., Okuyama, S., Nomoto, H., Tanabe, F., et al. (2018). Distribution and progression of visual field defects with binocular vision in glaucoma. *J. Glaucoma* 27, 519–524. doi: 10.1097/IJG.0000000000000949

Haykal, S., Invernizzi, A., Carvalho, J., Jansonius, N. M., and Cornelissen, F. W. (2022). Microstructural visual pathway white matter alterations in primary open-angle glaucoma: A neurite orientation dispersion and density imaging study. *Am. J. Neuroradiol.* 43, 756–763. doi: 10.3174/ajnr.A7495

Ho, L. C., Wang, B., Conner, I. P., van der Merwe, Y., Bilonick, R. A., Kim, S. G., et al. (2015). In vivo evaluation of white matter integrity and anterograde transport in visual system after excitotoxic retinal injury with multimodal MRI and OCT. *Invest. Ophthalmol. Vis. Sci.* 56, 3788–3800. doi: 10.1167/iovs.14-15552

Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., et al. (1991). Optical coherence tomography. *Science* 254, 1178–1181. doi: 10.1126/science.1957169

Hui, E. S., Fu, Q. L., So, K. F., and Wu, E. X. (2007). Diffusion tensor MR study of optic nerve degeneration in glaucoma. *IEEE Eng. Med. Biol. Soc.* 2007, 4312–4315. doi: 10.1109/IEMBS.2007.4353290

Kang, J. M., and Tanna, A. P. (2021). Glaucoma. *Med. Clin. North Am.* 105, 493–510. doi: 10.1016/j.mcna.2021.01.004

Kühn, T., Rohrmann, S., Karavasiloglou, N., Friedman, D. S., Cassidy, A., Bärnighausen, T., et al. (2021). Glaucoma and mortality risk: Findings from a prospective population-based study. *Sci. Rep.* 11:11771. doi: 10.1038/s41598-021-91194-3

Lanzafame, S., Giannelli, M., Garaci, F., Floris, R., Duggento, A., Guerrisi, M., et al. (2016). Differences in Gaussian diffusion tensor imaging and non-Gaussian diffusion kurtosis imaging model-based estimates of diffusion tensor invariants in the human brain. *Med. Phys.* 43:2464. doi: 10.1118/1.4946819

Le Bihan, D., and Iima, M. (2015). Diffusion magnetic resonance imaging: What water tells us about biological tissues. *PLoS Biol.* 13:e1002203. doi: 10.1371/journal.pbio.1002203

Li, K., Lu, C., Huang, Y., Yuan, L., Zeng, D., and Wu, K. (2014). Alteration of fractional anisotropy and mean diffusivity in glaucoma: Novel results of a meta-analysis of diffusion tensor imaging studies. *PLoS One* 9:e97445. doi: 10.1371/journal.pone.0097445

Li, M., Ke, M., Song, Y., Mu, K., Zhang, H., and Chen, Z. (2019). Diagnostic utility of central damage determination in glaucoma by magnetic resonance imaging: An observational study. *Exp. Ther. Med.* 17, 1891–1895. doi: 10.3892/etm.2018.7134

Li, Y., and Zhang, W. (2020). Quantitative evaluation of diffusion tensor imaging for clinical management of glioma. *Neurosurg. Rev.* 43, 881–891. doi: 10.1007/s10143-018-1050-1

Martinez-Heras, E., Grussu, F., Prados, F., Solana, E., and Llufriu, S. (2021). Diffusion-weighted imaging: Recent advances and applications. *Semin. Ultrasound CT MR* 42, 490–506. doi: 10.1053/j.sult.2021.07.006

Mayro, E. L., Wang, M., Elze, T., and Pasquale, L. R. (2020). The impact of artificial intelligence in the diagnosis and management of glaucoma. *Eye (Lond)* 34, 1–11. doi: 10.1038/s41433-019-0577-x

Murai, H., Suzuki, Y., Kiyosawa, M., Tokumaru, A. M., Ishii, K., and Mochizuki, M. (2013). Positive correlation between the degree of visual field defect and optic radiation damage in glaucoma patients. *Jpn. J. Ophthalmol.* 57, 257–262. doi: 10.1007/s10384-013-0233-0

Pujol, S., Wells, W., Pierpaoli, C., Brun, C., Gee, J., Cheng, G., et al. (2015). DTI challenge: Toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. *J. Neuroimaging* 25, 875–882. doi: 10.1111/jon.12283

Qu, X., Wang, Q., Chen, W., Li, T., Guo, J., Wang, H., et al. (2019). Combined machine learning and diffusion tensor imaging reveals altered anatomic fiber connectivity of the brain in primary open-angle glaucoma. *Brain Res.* 1718, 83–90. doi: 10.1016/j.brainres.2019.05.006

Rao, H. L., Pradhan, Z. S., Suh, M. H., Moghimi, S., Mansouri, K., and Weinreb, R. N. (2020). Optical coherence tomography angiography in glaucoma. *J. Glaucoma* 29, 312–321. doi: 10.1097/IJG.0000000000001463

Rao, H. L., Raveendran, S., James, V., Dasari, S., Palakurthy, M., Reddy, H. B., et al. (2017). Comparing the performance of compass perimetry with humphrey field analyzer in eyes with glaucoma. *J. Glaucoma* 26, 292–297. doi: 10.1097/IJG.0000000000000609

Saccà, S. C., Paluan, F., Gandolfi, S., Manni, G., Cutolo, C. A., and Izzotti, A. (2020). Common aspects between glaucoma and brain neurodegeneration. *Mutat. Res. Rev. Mutat. Res.* 786:108323. doi: 10.1016/j.mrrev.2020.108323

Schmidt, M. A., Knott, M., Heidemann, R., Michelson, G., Kober, T., Dörfler, A., et al. (2018). Investigation of lateral geniculate nucleus volume and diffusion tensor imaging in patients with normal tension glaucoma using 7 tesla magnetic resonance imaging. *PLoS One* 13:e0198830. doi: 10.1371/journal.pone.0198830

Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., and Pfeiffer, N. (2020). The diagnosis and treatment of glaucoma. *Dtsch. Arztebl. Int.* 117, 225–234. doi: 10.3238/arztebl.2020.0225

Sidek, S., Ramli, N., Rahmat, K., Ramli, N. M., Abdulrahman, F., and Tan, L. K. (2014). Glaucoma severity affects diffusion tensor imaging (DTI) parameters of the optic nerve and optic radiation. *Eur. J. Radiol.* 83, 1437–1441. doi: 10.1016/j.ejrad.2014.05.014

Song, X. Y., Puyang, Z., Chen, A. H., Zhao, J., Li, X. J., Chen, Y. Y., et al. (2018). Diffusion tensor imaging detects microstructural differences of visual pathway in patients with primary open-angle glaucoma and ocular hypertension. *Front. Hum. Neurosci.* 12:426. doi: 10.3389/fnhum.2018.00426

Stein, J. D., Khawaja, A. P., and Weizer, J. S. (2021). Glaucoma in adults-screening, diagnosis, and management: A review. *JAMA* 325, 164–174. doi: 10.1001/jama.2020.21899

Swaminathan, S. S., Jammal, A. A., Berchuck, S. I., and Medeiros, F. A. (2021). Rapid initial OCT RNFL thinning is predictive of faster visual field loss during extended follow-up in glaucoma. *Am. J. Ophthalmol.* 229, 100–107. doi: 10.1016/j.ajo.2021.03.019

Techavipoo, U., Okai, A. F., Lackey, J., Shi, J., Dresner, M. A., Leist, T. P., et al. (2009). Toward a practical protocol for human optic nerve DTI with EPI geometric distortion correction. *J. Magn. Reson. Imaging* 30, 699–707. doi: 10.1002/jmri.21836

Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., and Cheng, C. Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* 121, 2081–2090. doi: 10.1016/j.ophtha.2014.05.013

Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., et al. (2019). Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* 103, 167–175. doi: 10.1136/bjophthalmol-2018-313173

Wang, R., Tang, Z., Sun, X., Wu, L., Wang, J., Zhong, Y., et al. (2018). White matter abnormalities and correlation with severity in normal tension glaucoma: A whole brain atlas-based diffusion tensor study. *Invest. Ophthalmol. Vis. Sci.* 59, 1313–1322. doi: 10.1167/iovs.17-23597

Wheeler-Kingshott, C. A., Trip, S. A., Symms, M. R., Parker, G. J., Barker, G. J., and Miller, D. H. (2006). In vivo diffusion tensor imaging of the human optic nerve: Pilot study in normal controls. *Magn. Reson. Med.* 56, 446–451. doi: 10.1002/mrm.20964

Wild, D., Kucur, S. S., and Sznitman, R. (2017). Spatial entropy pursuit for fast and accurate perimetry testing. *Invest. Ophthalmol. Vis. Sci.* 58, 3414–3424. doi: 10.1167/iovs.16-21144

Wu, Z., McKendrick, A. M., Hadoux, X., Gaskin, J. C., Ang, G. S., Sarossy, M. G., et al. (2016). Test-retest variability of fundus-tracked perimetry at the peripapillary region in open angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* 57, 3619–3625. doi: 10.1167/iovs.15-19002

Yang, X. L., van der Merwe, Y., Sims, J., Parra, C., Ho, L. C., Schuman, J. S., et al. (2018). Age-related changes in eye, brain and visuomotor behavior in the DBA/2J mouse model of chronic glaucoma. *Sci. Rep.* 8:4643. doi: 10.1038/s41598-018-22850-4

Yarmohammadi, A., Zangwill, L. M., Manalastas, P. I. C., Fuller, N. J., Diniz-Filho, A., Saunders, L. J., et al. (2018). Peripapillary and macular vessel density in patients with primary open-angle glaucoma and unilateral visual field loss. *Ophthalmology* 125, 578–587. doi: 10.1016/j.ophtha.2017.10.029

Zhang, H. J., Mi, X. S., and So, K. F. (2019). Normal tension glaucoma: From the brain to the eye or the inverse? *Neural Regen. Res.* 14, 1845–1850. doi: 10.4103/1673-5374.259600

Zhang, Q. J., Wang, D., Bai, Z. L., Ren, B. C., and Li, X. H. (2015). Diffusion tensor imaging of optic nerve and optic radiation in primary chronic angle-closure glaucoma using 3T magnetic resonance imaging. *Int. J. Ophthalmol.* 8, 975–979.

Zhang, S., Wang, B., Xie, Y., Zhu, S., Thomas, R., Qing, G., et al. (2015). Retinotopic changes in the gray matter volume and cerebral blood flow in the primary visual cortex of patients with primary open-angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* 56, 6171–6178. doi: 10.1167/iovs.15-17286

Zhang, X., Sun, P., Wang, J., Wang, Q., and Song, S. K. (2011). Diffusion tensor imaging detects retinal ganglion cell axon damage in the mouse model of optic nerve crush. *Invest. Ophthalmol. Vis. Sci.* 52, 7001–7006. doi: 10.1167/iovs.11-7619

Zhou, W., Muir, E. R., Chalfin, S., Nagi, K. S., and Duong, T. Q. (2017). MRI study of the posterior visual pathways in primary open angle glaucoma. *J. Glaucoma* 26, 173–181. doi: 10.1097/IJG.0000000000000558

Zikou, A. K., Kitsos, G., Tzarouchi, L. C., Astrakas, L., Alexiou, G. A., and Argyropoulou, M. I. (2012). Voxel-based morphometry and diffusion tensor imaging of the optic pathway in primary open-angle glaucoma: A preliminary study. *Am. J. Neuroradiol.* 33, 128–134. doi: 10.3174/ajnr.A2714

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fnins.2023.1126262

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Chen-Xing Qi,
Renmin Hospital of Wuhan University, China
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China

*CORRESPONDENCE
Xiao-rong Wu
✉ wxr98021@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 17 December 2022
ACCEPTED 09 January 2023
PUBLISHED 02 February 2023

CITATION
Ji Y, Huang S-q, Cheng Q, Fu W-w, Zhong P-p, Chen X-l, Shu B-l, Wei B, Huang Q-y and Wu X-r (2023) Exploration of static functional connectivity and dynamic functional connectivity alterations in the primary visual cortex among patients with high myopia *via* seed-based functional connectivity analysis. *Front. Neurosci.* 17:1126262. doi: 10.3389/fnins.2023.1126262

Exploration of static functional connectivity and dynamic functional connectivity alterations in the primary visual cortex among patients with high myopia *via* seed-based functional connectivity analysis

Yu Ji†, Shui-qin Huang†, Qi Cheng, Wen-wen Fu, Pei-pei Zhong, Xiao-lin Chen, Ben-liang Shu, Bin Wei, Qin-yi Huang and Xiao-rong Wu*

Department of Ophthalmology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

Aim: This study was conducted to explore differences in static functional connectivity (sFC) and dynamic functional connectivity (dFC) alteration patterns in the primary visual area (V1) among high myopia (HM) patients and healthy controls (HCs) *via* seed-based functional connectivity (FC) analysis.

Methods: Resting-state functional magnetic resonance imaging (fMRI) scans were performed on 82 HM patients and 59 HCs who were closely matched for age, sex, and weight. Seed-based FC analysis was performed to identify alterations in the sFC and dFC patterns of the V1 in HM patients and HCs. Associations between mean sFC and dFC signal values and clinical symptoms in distinct brain areas among HM patients were identified *via* correlation analysis. Static and dynamic changes in brain activity in HM patients were investigated by assessments of sFC and dFC *via* calculation of the total time series mean and sliding-window analysis.

Results: In the left anterior cingulate gyrus (L-ACG)/left superior parietal gyrus (L-SPG) and left V1, sFC values were significantly greater in HM patients than in HCs. In the L-ACG and right V1, sFC values were also significantly greater in HM patients than in HCs [two-tailed, voxel-level $P < 0.01$, Gaussian random field (GRF) correction, cluster-level $P < 0.05$]. In the left calcarine cortex (L-CAL) and left V1, dFC values were significantly lower in HM patients than in HCs. In the right lingual gyrus (R-LING) and right V1, dFC values were also significantly lower in HM patients than in HCs (two-tailed, voxel-level $P < 0.01$, GRF correction, cluster-level $P < 0.05$).

Conclusion: Patients with HM exhibited significantly disturbed FC between the V1 and various brain regions, including L-ACG, L-SPG, L-CAL, and R-LING. This disturbance suggests that patients with HM could exhibit impaired cognitive and emotional processing functions, top-down control of visual attention,

and visual information processing functions. HM patients and HCs could be distinguished from each other with high accuracy using sFC and dFC variabilities. These findings may help to identify the neural mechanism of decreased visual performance in HM patients.

KEYWORDS

high myopia, brain function, brain region, resting-state functional magnetic resonance imaging, static functional connectivity, dynamic functional connectivity, seed-based functional connectivity analysis

1. Introduction

High myopia (HM) is a type of ametropia characterized by a need for more than −6.00 diopters of refractive correction or the presence of an axial length ≥ 26 mm[1] (Ucak et al., 2020). The typical clinical manifestations of HM include decline in distance vision, normal near vision, and decline in dark adaptive function; these manifestations are often accompanied by visual fatigue. Patients with HM can exhibit degenerative changes in the fundus related to excessive elongation of the ocular axis; such changes include patchy or diffuse chorioretinal atrophy, retinal nerve fiber layer thinning, macular atrophy, macular hole formation, and altered retinal vessel morphology (Li et al., 2011; Silva, 2012; Kamal Salah et al., 2015). In East Asia, 80–90% of 18-year-olds exhibit myopia, and approximately 10–20% of these individuals have HM (Jonas and Panda-Jonas, 2019). The incidence rate of HM among Chinese students in primary school does not exceed 1%, but it is > 2% in junior high school (Wang et al., 2018). There are multiple risk factors for increased axial length in patients with HM, such as sex, best-corrected visual acuity, axial length, type of myopic maculopathy, and choroidal neovascularization status (Du et al., 2021). Jonas et al. (2020) demonstrated that the prevalence of glaucoma-like optic neuropathy increases with longer axial length in eyes with HM; other investigations have shown that glaucoma is both an eye disease and a degenerative illness of the central nervous system (Nucci et al., 2015; Mancino et al., 2018). Therefore, a connection may exist between HM and the central nervous system.

Recently, resting-state functional magnetic resonance imaging (fMRI) has emerged as an important method for non-invasive analyses of changes in brain function; it has been used to investigate various clinical diseases. Resting-state fMRI is based on the paramagnetic effect of deoxygenated hemoglobin. Increased oxygen consumption in local tissues leads to an increase in deoxygenated hemoglobin, which indirectly indicates the degree of local neural activity in the brain (Glover, 2011). Patients with HM reportedly exhibit significantly decreased voxel-mirror homotopic connectivity between the putamen and fusiform gyrus, suggesting that the visual and recognition functions are affected in HM patients (Cheng et al., 2022). Huang et al. (2018a) revealed that the whole-brain gray matter volume of the right cuneus gyrus was decreased in patients with HM, indicating potential visual cortex functional impairment. Our previous study demonstrated that HM patients had altered dynamic regional homogeneity values in the left fusiform gyrus, right inferior temporal gyrus, right Rolandic operculum, right postcentral gyrus, and right precentral gyrus (Ji et al., 2022). Thus far, research has mainly focused on changes in static brain activity in HM patients; and it is believed that the functional interaction of brain regions remains unchanged in time during the whole MRI scan, which is obviously not objective. Recent research has revealed time-dependent characteristics of brain activity; even when fully at rest, the brain experiences brief spontaneous oscillations that are strongly associated with its activities. In some respects, a larger value may indicate a higher degree of adaptability (Vincent et al., 2007; Allen et al., 2014). Therefore, dynamic functional connectivity (dFC) could serve as a new indicator of complex brain functional structure through the acquisition of time-dependent connections over short durations of time (Du et al., 2018). Additionally, Preti et al. (2017) found that resting-state functional connectivity (FC) dynamically changed throughout the scan cycle. The above findings suggest that analyses of changes in brain activity in HM patients are not sufficiently comprehensive when conducted solely on the basis of static functional connectivity (sFC) or dFC. To our knowledge, previous studies have not combined static and dynamic analysis methods to characterize altered brain function in HM patients. We suspect that such analyses can be used to improve the broader understanding of altered neural mechanisms in HM patients.

The primary visual area (V1; Brodmann 17), located around the calcarine cortex of the occipital lobe (Ding et al., 2013), is the main source of feed forward visual stimuli in higher-level visual cortices. The lateral geniculate nucleus of the thalamus receives visual stimuli from the retina and sends it to the V1 (Mock et al., 2018). Recent studies have revealed two pathways of visual transmission (ventral and dorsal), both of which originate from the retina and project to the V1 (Yamasaki and Tobimatsu, 2018). The ventral pathway is mainly involved in the perception of shape and color, whereas the dorsal pathway is mainly involved in the perception of motion (Tobimatsu and Celesia, 2006). Visual stimuli received by the V1 are then projected to the secondary visual cortex (V2), which is involved in the perception of color and orientation (Tootell et al., 2003). Because the V1 is the first stage of visual information cortex processing visual signals, V1 impairment results in vision loss. Wu et al. (2020) found that the cortical surface thickness of the right V1 was decreased in patients with HM, suggesting that visual and speech functions were affected in those patients. The aforementioned neuroimaging investigations confirmed that patients with HM exhibit unique functional and structural changes in the V1. However, few studies have examined whether individuals with HM exhibit specific altered FC patterns in the V1. Here, we hypothesized that patients with HM would show characteristic alterations in FC patterns in the V1.

2. Participants and methods

2.1. Participants

From August 2021 to December 2021, 82 HM patients and 59 healthy controls (HCs) were examined in the Department of Ophthalmology at Nanchang University's First Affiliated Hospital. For each participant, age, sex, and educational background were all met. Individuals with brain illnesses were excluded on the basis of a clinical examination and physical assessment. All eligible individuals were examined in the same clinic and provided written informed consent to participate in the study. All procedures were conducted in accordance with the Declaration of Helsinki, and the study protocol was approved by Nanchang University's First Affiliated Hospital's Medical Ethics Committee (Jiangxi Province, China).

The inclusion criteria for HM patients were binocular visual acuity of −6 diopters or worse; corrected visual acuity of better than 1.0; and the completion of MRI-related tests, optical coherence tomography, ultrasonography, and other ophthalmic examinations. The exclusion criteria for HM patients were binocular visual acuity of better than −6 diopters; presence of retinal detachment, maculopathy, choroidal neovascularization, and/or retinal pigment epithelium disease; and/or history of ocular trauma or ophthalmic surgery, neurological diseases, and/or cerebral infarction.

According to age, sex, and educational background, HCs were chosen at random from Nanchang City. The inclusion criteria for HCs were the absence of eye diseases and major illnesses (e.g., neurological illness or cerebral infarction); the presence of uncorrected vision or visual acuity better than 1.0; and the completion of MRI-related tests, optical coherence tomography, ultrasonography, and other ophthalmic examinations.

2.2. fMRI data acquisition

All MRI data were obtained using a 3-TeslaTrio magnetic resonance imaging scanning system (Trio Tim, Siemens Medical Systems, Erlangen, Germany). During image acquisition, we asked the participants to close their eyes, minimize movement, and avoid falling asleep. We also asked the participants to use earplugs to reduce the impacts of head movement and machine noise during scanning. The following three-dimensional high-resolution T1-weighted imaging parameters were used in this study: repetition time = 1,900 ms, echo time = 2.26 ms, thickness = 1, no intersection gap, acquisition matrix = 256 × 256, field of view = 240 × 240 mm^2, and flip angle = 12°.

2.3. fMRI data preprocessing analysis

The Statistical Parametric Mapping (SPM12) and Data Processing and Analysis for Brain Imaging (DPABI) toolboxes running on MATLAB 2013b were used for data preprocessing. The following preprocessing steps were performed: (1) conversion of DICOM format to NIFTI format; (2) removal of the first 10 volumes of functional images to eliminate erratic data related to machine initialization; (3) analysis of functional volumes; (4) time correction; (5) head motion correction; (6) spatial normalization; (7) spatial smoothing; and (8) removal of linear data, interference noise, and low-frequency filtering.

TABLE 1 Montreal Neurological Institute coordinates for region of interest.

Region of interest	X	Y	Z
Right V1 (Brodmann 17)	8	−76	10
Left V1 (Brodmann 17)	−8	−76	10

MNI, Montreal Neurological Institute.

TABLE 2 Demographic characteristics of HM patients and HCs.

Characteristic	HM patients	HCs
Men/women	43/39	24/35
Age (years)	26.53 ± 5.291	25.67 ± 3.102

HM, high myopia; HCs, healthy controls.

2.4. Definition of region of interest (ROI)

The ROI for this investigation was defined as the V1 (Brodmann 17). Computation of FC was conducted with the center of the V1 as the seed point. The Montreal Neurological Institute (MNI) coordinates for the right V1 and left V1 were (8, −76, 10) and (−8, −76, 10), respectively (Table 1). The radius of the ROI was set to 6 mm. The mean time course of the ROI was compared with the time courses of all other regions to generate a Pearson correlation coefficient. Fisher's z-transform analysis was applied to the Pearson correlation coefficient to evaluate data normality and obtain an approximate normal distribution for further statistical analyses.

2.5. Seed-based FC analysis

We used seed-based FC analysis to identify the FC of the V1. First, we established the sFC analysis parameters. In the DPABI toolbox, we created a sphere with a radius of 6 mm, then calculated Pearson correlation coefficients for each participant's seed region and all voxels in the brain in sequence, using Fisher's z-transform analysis to improve normality. The results were used for sFC analysis. Next, we established dFC analysis parameters. In the DPABI toolbox, we used the sliding window method within the "Dynamic and Stability Analyses" module to measure dFC. We extracted the time series signal of the ROI for each participant, then selected a time window with a width of 30 repetition time and a length of 1 repetition time. The Pearson correlation coefficients of the mean time series signals of all voxels in the whole brain were calculated in a sequential manner; this allowed acquisition of the correlation coefficient of the whole-brain voxel sliding window for the ROI of each participant. Finally, the results of multiple sliding window correlation coefficients for each participant were normalized, and FC variability was represented by calculating the standard deviation of the *z*-value for each voxel correlation coefficient. The results were used for dFC analysis.

2.6. Statistical analysis

SPSS 8.0 software was used for analyses of aggregated clinical and demographic data. The chi-squared test was used to analyze

10.3389/fnins.2023.1126262

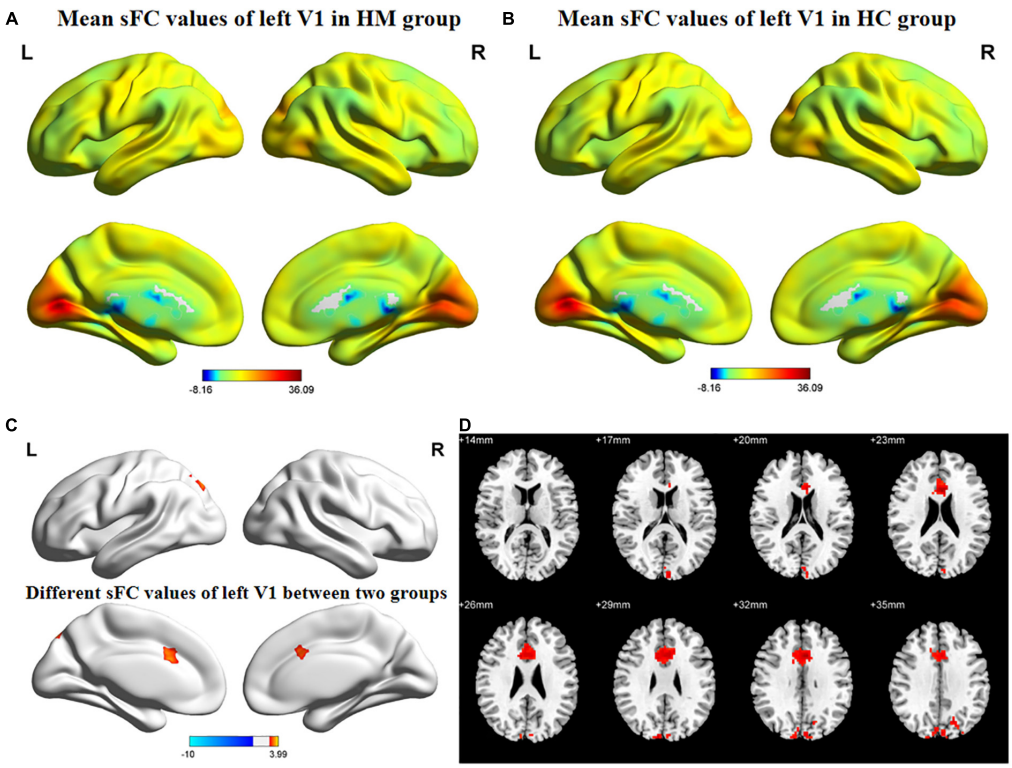


FIGURE 1
Spatial of distributions of sFC patterns of the left V1 in HM patients and HCs. **(A)** Mean sFC values of left V1 in HM group; **(B)** mean sFC values of left V1 in HC group; **(C)** different sFC values of left V1 between two groups and **(D)** significant zsFC maps of left V1 differences among two groups. HCs, healthy controls; HM, high myopia; sFC, static functional connectivity; zsFC, z-values static functional connectivity; L, left; R, right.

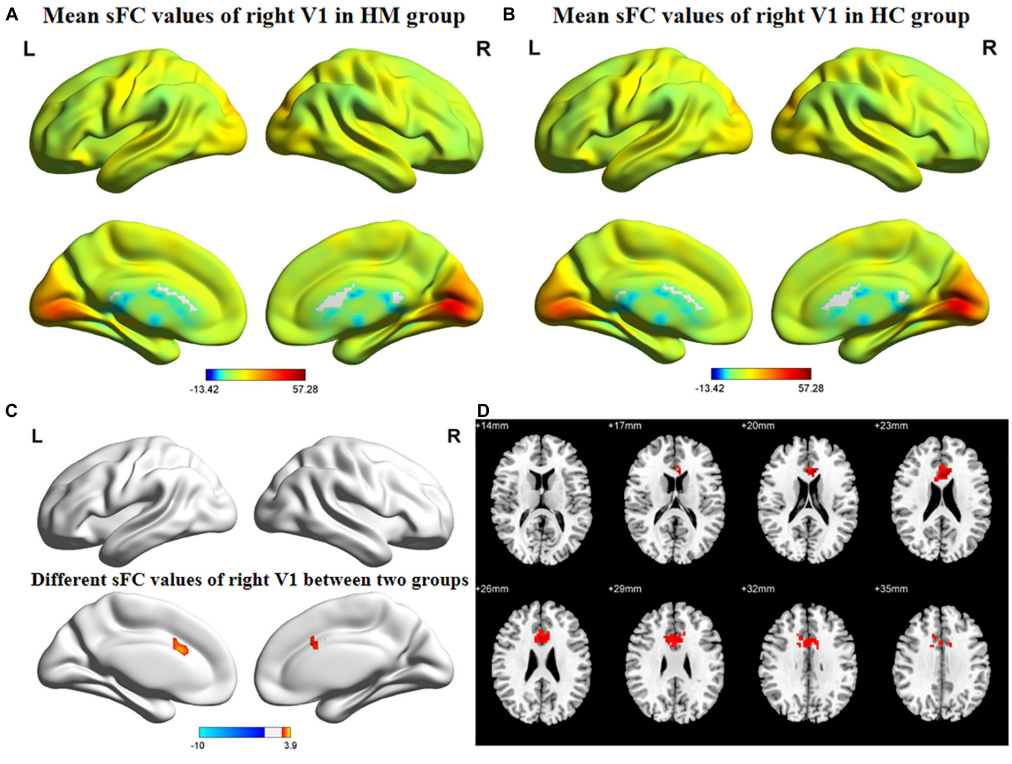


FIGURE 2
Spatial of distributions of sFC patterns of the right V1 in HM patients and HCs. **(A)** Mean sFC values of right V1 in HM group; **(B)** mean sFC values of right V1 in HC group; **(C)** different sFC values of right V1 between two groups and **(D)** significant zsFC maps of right V1 differences among two groups. HCs, healthy controls; HM, high myopia; sFC, static functional connectivity; zsFC, z-values static functional connectivity; L, left; R, right.

TABLE 3 Significant differences in sFC values of the left V1 between HM patients and HCs.

Brain region	BA	Peak *t*-score	MNI coordinates (x, y, z)	Cluster size (voxels)
L-ACG	32	3.5628	−3, 15, 33	200
L-SPG		3.9879	−18, −81, 48	214

sFC, static functional connectivity; V1, primary visual area; HCs, healthy controls; HM, high myopia; BA, Brodmann area; L-ACG, left anterior cingulate gyrus; L-SPG, left superior parietal gyrus; MNI, Montreal Neurological Institute.

TABLE 4 Significant differences in sFC values of the right V1 between HM patients and HCs.

Brain region	BA	Peak *t*-score	MNI coordinates (x, y, z)	Cluster size (voxels)
L-ACG		3.8977	−9, 21, 21	158

sFC, static functional connectivity; V1, primary visual area; HCs, healthy controls; HM, high myopia; BA, Brodmann area; L-ACG, left anterior cingulate gyrus; MNI, Montreal Neurological Institute.

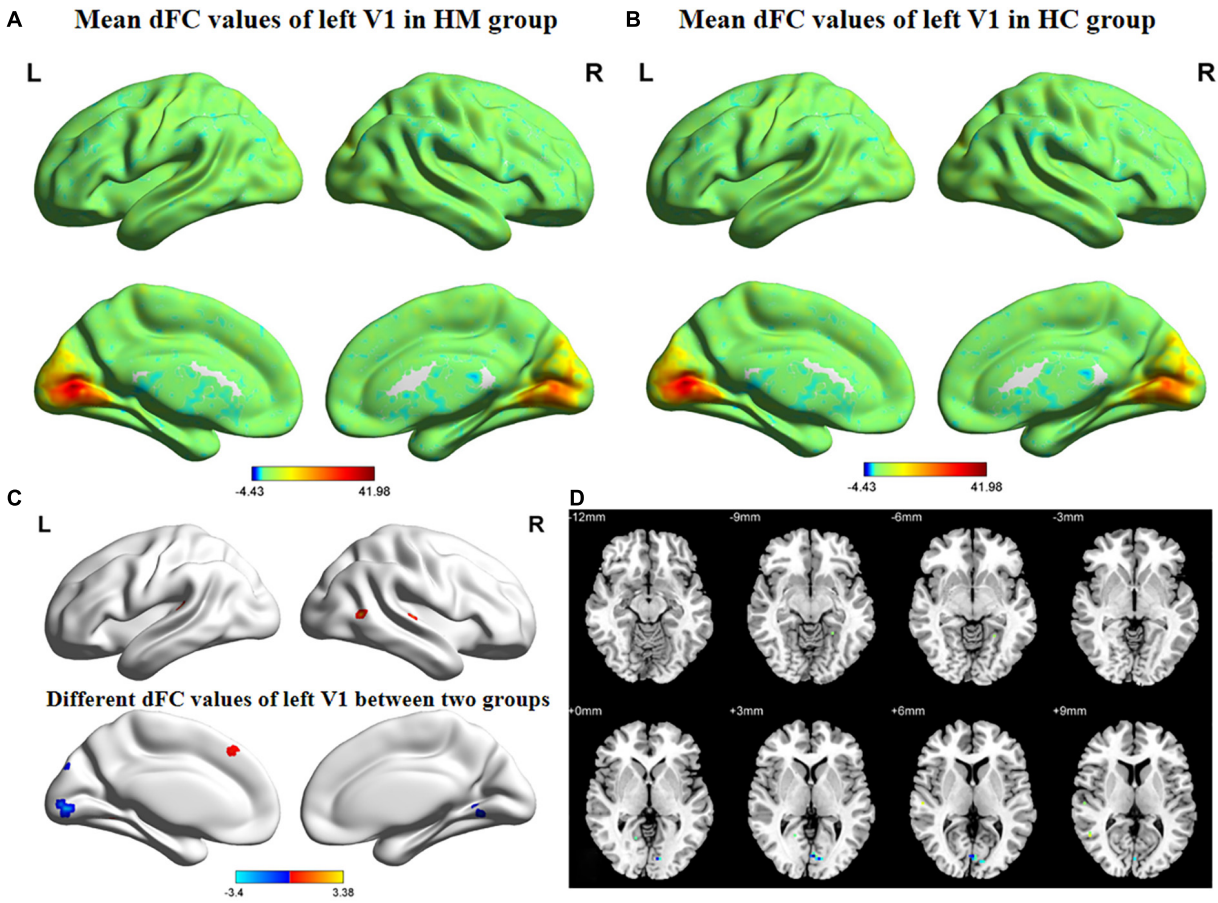


FIGURE 3
Spatial of distributions of dFC patterns of the left V1 in HM patients and HCs. **(A)** Mean dFC values of left V1 in HM group; **(B)** mean dFC values of left V1 in HC group; **(C)** different dFC values of left V1 between two groups and **(D)** significant zdFC maps of left V1 differences among two groups. HCs, healthy controls; HM, high myopia; sFC, static functional connectivity; zdFC, z-values dynamic functional connectivity; L, left; R, right.

proportions, whereas independent two-sample *t*-tests were used to assess continuous variables (*P*-values < 0.05 were considered indicative of statistical significance). Using SPM12 software, one-sample *t*-tests were conducted to assess intragroup *z*-value FC patterns, and two-sample *t*-tests were conducted to investigate differences in *z*-value FC patterns between two groups [voxel-level $P < 0.01$, Gaussian random field (GRF) correction, cluster-level $P < 0.05$]. Additionally, Pearson correlation coefficients were used to examine associations between mean FC signal values in various brain locations and clinical characteristics in patients with HM (*P*-values < 0.05 were considered indicative of statistical significance).

3. Results

3.1. Demographics

This study included 82 HM patients (43 men and 39 women; mean age, 26.53 ± 5.291 years) and 59 HCs (24 men and 35 women;

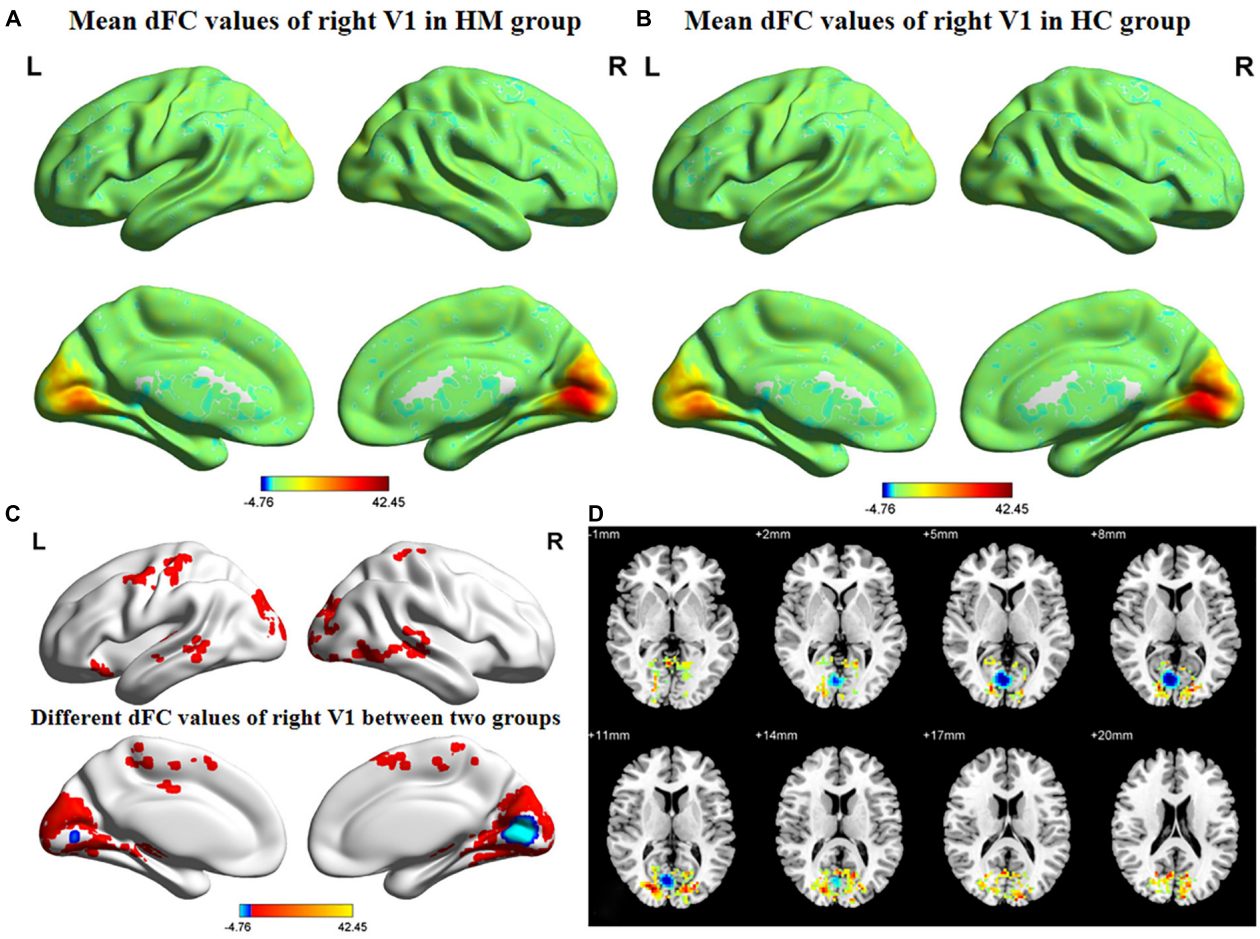


FIGURE 4
Spatial of distributions of dFC patterns of the right V1 in HM patients and HCs. **(A)** Mean dFC values of right V1 in HM group; **(B)** mean dFC values of right V1 in HC group; **(C)** different dFC values of right V1 between two groups and **(D)** significant zdFC maps of right V1 differences among two groups. HCs, healthy controls; HM, high myopia; sFC, static functional connectivity; zdFC, z-values dynamic functional connectivity; L, left; R, right.

mean age, 25.67 ± 3.102 years). Demographic characteristics are shown in Table 2.

3.2. Group differences in sFC

Figures 1, 2 show the spatial distributions of sFC values of the bilateral V1 in HM patients and HCs. In the left anterior cingulate gyrus (L-ACG)/left superior parietal gyrus (L-SPG) and left V1, sFC values were significantly greater in HM patients than in HCs (Figure 1 and Table 3) (voxel-level $P < 0.01$, GRF correction, cluster-level $P < 0.05$). In the L-ACG and right V1, sFC values were also significantly greater in HM patients than in HCs (Figure 2 and Table 4) (voxel-level $P < 0.01$, GRF correction, cluster-level $P < 0.05$).

3.3. Group differences in dFC

Figures 3, 4 show the spatial distributions of dFC values of the bilateral V1 in HM patients and HCs. In the left calcarine cortex (L-CAL) and left V1, dFC values were significantly lower in HM patients than in HCs (Figure 3 and Table 5) (voxel-level $P < 0.01$, GRF correction, cluster-level $P < 0.05$). In the right lingual gyrus (R-LING) and right V1, dFC values were also significantly lower in HM patients than in HCs (Figure 4 and Table 6) (voxel-level $P < 0.01$, GRF correction, cluster-level $P < 0.05$).

4. Discussion

A reliable and efficient approach to assess the correlation coefficients of blood oxygen level-dependent signal time series between different brain areas and V1 is the seed-based FC technique used in the present study. This method has been used multiple times for analyses of individuals with ophthalmic or other systemic diseases (Table 7). To our knowledge, the present study is the first to investigate the sFC and dFC between the V1 and other brain regions in patients with HM; the findings may facilitate greater understanding of V1 HCs in such patients. Notably, we found that, in the L-ACG/L-SPG and left V1, sFC values were significantly greater in HM patients than in HCs; in the L-ACG and right V1, sFC values were also significantly greater in HM patients than in HCs. However, in the L-CAL and left V1, dFC values were significantly lower in HM patients than in HCs; in the R-LING and right V1, dFC values were also

TABLE 5 Significant differences in dFC values of the left V1 between HM patients and HCs.

Brain region	BA	Peak *t*-score	MNI coordinates (x, y, z)	Cluster size (voxels)
L-CAL		−3.3452	−3, −81, 3	15

dFC, dynamic functional connectivity; V1, primary visual area; HCs, healthy controls; HM, high myopia; BA, Brodmann area; L-CAL, left calcarine cortex; MNI, Montreal Neurological Institute.

TABLE 6 Significant differences in dFC values of the right V1 between HM patients and HCs.

Brain region	BA	Peak t-score	MNI coordinates (x, y, z)	Cluster size (voxels)
R-LING	23	−11.017	6, −75, 9	1,489

dFC, dynamic functional connectivity; V1, primary visual area; HCs, healthy controls; HM, high myopia; BA, Brodmann area; R-LING, right lingual gyrus; MNI, Montreal Neurological Institute.

significantly lower in HM patients than in HCs. These results may provide some insights regarding the neural mechanisms involved in HM; they may also be useful in identifying potential neurological causes of decreased visual performance in HM patients.

4.1. sFC alterations and their significance

Within the bilateral V1 and L-ACG in HM patients, significantly increased sFC was observed. The anterior cingulate gyrus, which plays a key role in regulating cognitive and emotional processing (Takayanagi et al., 2013), is an important anatomical component of the salience network (Christopher et al., 2015). The salience network acts as a mediator within the brain, continuously monitoring the external environment and assessing how other brain networks respond to new information and stimuli. Previous studies have shown that the salience network is primarily responsible for regulating the switch between the default mode network and the central execution network, avoiding simultaneous excitation of both networks (Goulden et al., 2014). Philippi et al. (2018) found that major depressive disorder patients with greater negative self-focus thoughts showed significantly increased sFC within the anterior cingulate gyrus. Hafkemeijer et al. (2013) demonstrated that Subjective memory complaints (SMC) patients had significantly increased sFC values in the anterior cingulate gyrus, implying that the increase in sFC is related to compensation for the loss of cognitive function and maintenance of task performance. Sacca et al. (2022) reported that the treatment of migraine with 20-Hz transcutaneous auricular vagus nerve stimulation led to a significant increase in sFC within the anterior cingulate gyrus. In a separate study, Sun et al. (2020) revealed that divergent thinking training intervention resulted in a significant increase in sFC within the anterior cingulate gyrus. Similarly, we found a higher sFC between the bilateral V1 and L-ACG in the present study. Our findings suggest that patients with HM have difficulty switching between brain networks, which leads to impairments in cognitive and emotional processing functions. Furthermore, our findings suggest that neural hyperactivity between the bilateral V1 and L-ACG is a compensatory response to the loss of cognitive and emotional processing functions in HM patients.

Additionally, we discovered that patients with HM had considerably greater sFC values between the left V1 and L-SPG. The superior parietal gyrus is a component of the apical region of the parietal lobe, surrounded by the postcentral gyrus, precuneus, and inferior parietal lobules of the parietal lobe. The superior parietal gyrus is within the anatomical area of the dorsal attention network (He et al., 2007), which participates in top-down control of visual attention (Rajan et al., 2021). As a brain network that focuses human attention, the dorsal attention network directs attention to the most salient and active brain networks. Previous studies have shown that the dorsal attention network maintains top-down attention control, allowing focus toward and away from external noise or environmental changes; conversely, the ventral attention network interrupts ongoing cognitive activities (Corbetta et al., 2008) and causes changes *via* bottom-up attention control (Corbetta and Shulman, 2002). Zou et al. (2021) found that the sFC of the superior parietal gyrus was increased in patients with chronic migraine, suggesting that the saliency of painful input increases in response to aural stimuli. Li et al. (2014) revealed that patients with primary insomnia exhibit significantly increased sFC in the superior parietal gyrus, thereby offering additional insights into the neurobiological mechanism of working memory deficiency caused by primary insomnia. Ye et al. (2019) reported that the white matter hyperintensities (WMH) with cognitive impairment (CI) group led to a significant increase in sFC within the superior parietal gyrus, which may reflect a compensatory functional enhancement. Furthermore, Huang et al. (2019) demonstrated that comatose patients had significantly increased sFC values in the superior parietal gyrus, implying that an increased sFC is associated with compensatory remodeling. Consistent with the previous findings, we demonstrated that patients with HM had significantly increased sFC values between the left V1 and L-SPG. Thus, patients with HM may have an inability to focus their attention to the network that is currently most active, leading to impaired top-down control of visual attention. Additionally, hyperactive neuronal connections between the left V1 and L-SPG may be a compensatory mechanism that protects against the loss of top-down control of visual attention in HM patients.

TABLE 7 Use of seed-based FC technique for analysis of individuals with ophthalmic or other systemic diseases.

References	Disease	Year
Wu et al. (2021)	Bronchial asthma	2021
Tang et al. (2021)	Alzheimer's disease and mild cognitive impairment	2021
Qi et al. (2021)	Diabetic retinopathy	2021
Yu et al. (2020)	Proliferative diabetic retinopathy	2020
Ke et al. (2020)	Migraine without aura	2020
Nie et al. (2015)	Primary insomnia	2015

FC, functional connectivity.

4.2. dFC alterations and their significance

The calcarine cortex and lingual gyrus are anatomical regions in the V1 that receive visual signals from the visual pathway, then transmit those signals to the higher visual cortex (Beckmann et al., 2005). These rich dynamic properties exhibited by early visual neurons suggest that V1 does not encode the environment in a static manner; it exhibits rich spatial and temporal dynamic features (Lazar et al., 2021). Huang et al. (2018b) found that amplitude of low-frequency fluctuation values in the bilateral lingual gyrus were considerably lower in retinitis pigmentosa patients than in HCs. Tong et al. (2021) demonstrated that patients with iridocyclitis displayed significantly lower FC between the V1 and both the bilateral calcarine. Dan et al. (2019) revealed that regional homogeneity values in the lingual gyrus were significantly lower in retinitis pigmentosa patients than in HCs. Furthermore, Wen et al. (2018) also found that late blindness patients showed a decreased FC between the V1 and the bilateral calcarine cortex/lingual gyrus. Additionally, Huang et al. (2020) demonstrated that the dynamic amplitude of low-frequency fluctuation/dFC of the bilateral calcarine cortex/lingual gyrus was lower in patients with late blindness than in HCs. Consistent with the previous findings, the present study showed that HM patients exhibited significantly decreased dFC values between the L-CAL and left V1 and the R-LING and right V1. The dFC is an indicator of the degree of spontaneous neural activity, which represents temporal variation in energy consumption and reflects neural network adaptability. Compared with sFC, dFC can better reflect the dynamic involvement of different brain regions in the actual brain and is considered a more accurate representation of functional brain networks (Zhao et al., 2020). Therefore, when the dFC value between the L-CAL and left V1 and the R-LING and right V1 in HM patients decreases significantly, it may be difficult to accept visual stimulation. This leads to impairment of visual information processing functions.

This study had some limitations. First, the number of HM patients was limited. Second, the data were frequently affected by unavoidable factors in the fMRI environment (e.g., heartbeat, muscle beat, and respiratory motion). Third, there remains debate regarding the selection of sliding window length because no standardized criteria have been established. An inadequate window length results in insufficient time points in each window to generate a stable dFC, but an excessive window length may reduce temporal variation in FC and thus fail to detect valid connections. Finally, the duration of HM varied among patients, which may have impacted the reliability of the results. In future studies, we plan to increase the sample size, improve the testing environment, and expand the size of the sliding window.

5. Conclusion

The results of this study indicated that, compared with HCs, patients with HM have altered sFC and dFC values in various brain regions, implying that HM causes extensive changes in static and dynamic spontaneous brain activity; these changes presumably lead to the corresponding clinical manifestations. Our findings improve the broader understanding of altered neural mechanisms in HM patients and provide new insights into potential neural causes of vision loss in those patients.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Nanchang University's First Affiliated Hospital's Medical Ethics Committee (Jiangxi Province, China). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YJ was responsible for writing the manuscript. S-QH was in charge of proofreading and refining the manuscript's wording. QC, W-WF, P-PZ, X-LC, B-LS, BW, and Q-YH contributed to data collection and statistical analyses. YJ and S-QH designed the protocol and contributed to the MRI analysis. YJ, S-QH, and X-RW designed the study, oversaw all clinical aspects of study conduct, and prepared the manuscript. All authors contributed to the article and approved the submitted version.

Funding

We acknowledge the assistance provided by the National Nature Science Foundation of China (grant no.82160207), Key projects of Jiangxi Youth Science Fund (No.20202ACBL216008), Science and Technology Plan of Jiangxi Provincial Health and Health Commission(202130156), and Postgraduate Innovation Special Fund Project in Jiangxi Province(YC2022—s198).

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

Allen, E., Damaraju, E., Plis, S., Erhardt, E., Eichele, T., and Calhoun, V. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676.

Beckmann, C., DeLuca, M., Devlin, J., and Smith, S. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1001–1013.

Cheng, Y., Chen, X., Shi, L., Li, S., Huang, H., Zhong, P., et al. (2022). Abnormal functional connectivity between cerebral hemispheres in patients with high myopia: A Resting FMRI Study based on voxel-mirrored homotopic connectivity. *Front. Hum. Neurosci.* 16:910846. doi: 10.3389/fnhum.2022.910846

Christopher, L., Duff-Canning, S., Koshimori, Y., Segura, B., Boileau, I., Chen, R., et al. (2015). Salience network and parahippocampal dopamine dysfunction in memory-impaired Parkinson disease. *Ann. Neurol.* 77, 269–280. doi: 10.1002/ana.24323

Corbetta, M., and Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Corbetta, M., Patel, G., and Shulman, G. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58, 306–324.

Dan, H., Zhou, F., Huang, X., Xing, Y., and Shen, Y. (2019). Altered intra- and inter-regional functional connectivity of the visual cortex in individuals with peripheral vision loss due to retinitis pigmentosa. *Vis. Res.* 159, 68–75. doi: 10.1016/j.visres.2019.02.013

Ding, K., Liu, Y., Yan, X., Lin, X., and Jiang, T. (2013). Altered functional connectivity of the primary visual cortex in subjects with amblyopia. *Neural. Plast* 2013 :612086.

Du, R., Xie, S., Igarashi-Yokoi, T., Watanabe, T., Uramoto, K., Takahashi, H., et al. (2021). Continued increase of axial length and its risk factors in adults with high myopia. *JAMA Ophthalmol.* 139, 1096–1103.

Du, Y., Fryer, S., Fu, Z., Lin, D., Sui, J., Chen, J., et al. (2018). Dynamic functional connectivity impairments in early schizophrenia and clinical high-risk for psychosis. *Neuroimage* 180(Pt B), 632–645.

Glover, G. (2011). Overview of functional magnetic resonance imaging. *Neurosurg. Clin. N. Am.* 22, 133–139,vii.

Goulden, N., Khusnulina, A., Davis, N., Bracewell, R., Bokde, A., McNulty, J., et al. (2014). The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *Neuroimage* 99, 180–190. doi: 10.1016/j.neuroimage.2014.05.052

Hafkemeijer, A., Altmann-Schneider, I., Oleksik, A., van de Wiel, L., Middelkoop, H., van Buchem, M., et al. (2013). Increased functional connectivity and brain atrophy in elderly with subjective memory complaints. *Brain Connect.* 3, 353–362.

He, B., Snyder, A., Vincent, J., Epstein, A., Shulman, G., and Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron* 53, 905–918.

Huang, L., Zheng, Y., Zeng, Z., Li, M., Zhang, L., and Gao, Y. (2019). Fractional amplitude of low-frequency fluctuations and functional connectivity in comatose patients subjected to resting-state functional magnetic resonance imaging. *Ann. Indian Acad. Neurol.* 22, 203–209. doi: 10.4103/aian.AIAN_420_17

Huang, X., Hu, Y., Zhou, F., Xu, X., Wu, Y., Jay, R., et al. (2018a). Altered whole-brain gray matter volume in high myopia patients: A voxel-based morphometry study. *Neuroreport* 29, 760–767. doi: 10.1097/WNR.0000000000001028

Huang, X., Wen, Z., Qi, C., Tong, Y., Dan, H., Xie, B., et al. (2020). Altered temporal dynamic intrinsic brain activity in late blindness. *Biomed. Res. Int.* 2020:1913805. doi: 10.1155/2020/1913805

Huang, X., Zhou, F., Dan, H., and Shen, Y. (2018b). Abnormal intrinsic brain activity in individuals with peripheral vision loss because of retinitis pigmentosa using amplitude of low-frequency fluctuations. *Neuroreport* 29, 1323–1332. doi: 10.1097/WNR.0000000000001116

Ji, Y., Cheng, Q., Fu, W., Zhong, P., Huang, S., Chen, X., et al. (2022). Exploration of abnormal dynamic spontaneous brain activity in patients with high myopia via dynamic regional homogeneity analysis. *Front. Hum. Neurosci.* 16:959523. doi: 10.3389/fnhum.2022.959523

Jonas, J., and Panda-Jonas, S. (2019). [Epidemiology and anatomy of myopia]. *Ophthalmologe* 116, 499–508.

Jonas, J., Wang, Y., Dong, L., and Panda-Jonas, S. (2020). High myopia and glaucoma-like optic neuropathy. *Asia Pac. J. Ophthalmol. (Phila)* 9, 234–238.

Kamal Salah, R., Morillo-Sánchez, M., García-Ben, A., Rius-Diaz, F., Cilveti-Puche, Á, Figueroa-Ortiz, L., et al. (2015). The effect of peripapillary detachment on retinal nerve fiber layer measurement by spectral domain optical coherence tomography in high myopia. *Ophthalmologica* 233, 209–215. doi: 10.1159/000371903

Ke, J., Yu, Y., Zhang, X., Su, Y., Wang, X., Hu, S., et al. (2020). Functional alterations in the posterior insula and cerebellum in migraine without aura: A resting-state MRI study. *Front. Behav. Neurosci.* 14:567588. doi: 10.3389/fnbeh.2020.567588

Lazar, A., Lewis, C., Fries, P., Singer, W., and Nikolic, D. (2021). Visual exposure enhances stimulus encoding and persistence in primary cortex. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2105276118. doi: 10.1073/pnas.2105276118

Li, H., Mitchell, P., Rochtchina, E., Burlutsky, G., Wong, T., and Wang, J. (2011). Retinal vessel caliber and myopic retinopathy: The blue mountains eye study. *Ophthalmic Epidemiol.* 18, 275–280. doi: 10.3109/09286586.2011.602508

Li, Y., Wang, E., Zhang, H., Dou, S., Liu, L., Tong, L., et al. (2014). Functional connectivity changes between parietal and prefrontal cortices in primary insomnia patients: Evidence from resting-state fMRI. *Eur. J. Med. Res.* 19:32. doi: 10.1186/2047-783X-19-32

Mancino, R., Martucci, A., Cesareo, M., Giannini, C., Corasaniti, M., Bagetta, G., et al. (2018). Glaucoma and alzheimer disease: One age-related neurodegenerative disease of the brain. *Curr. Neuropharmacol.* 16, 971–977.

Mock, V., Luke, K., Hembrook-Short, J., and Briggs, F. (2018). Dynamic communication of attention signals between the LGN and V1. *J. Neurophysiol.* 120, 1625–1639. doi: 10.1152/jn.00224.2018

Nie, X., Shao, Y., Liu, S., Li, H., Wan, A., Nie, S., et al. (2015). Functional connectivity of paired default mode network subregions in primary insomnia. *Neuropsychiatr. Dis. Treat.* 11, 3085–3093. doi: 10.2147/NDT.S95224

Nucci, C., Martucci, A., Cesareo, M., Garaci, F., Morrone, L., Russo, R., et al. (2015). Links among glaucoma, neurodegenerative, and vascular diseases of the central nervous system. *Prog. Brain Res.* 221, 49–65.

Philippi, C., Cornejo, M., Frost, C., Walsh, E., Hoks, R., Birn, R., et al. (2018). Neural and behavioral correlates of negative self-focused thought associated with depression. *Hum. Brain Mapp.* 39, 2246–2257.

Preti, M., Bolton, T., and Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage* 160, 41–54.

Qi, C., Huang, X., Tong, Y., and Shen, Y. (2021). Altered functional connectivity strength of primary visual cortex in subjects with diabetic retinopathy. *Diabetes Metab. Syndr. Obes.* 14, 3209–3219. doi: 10.2147/DMSO.S311009

Rajan, A., Meyyappan, S., Liu, Y., Samuel, I., Nandi, B., Mangun, G., et al. (2021). The microstructure of attentional control in the dorsal attention network. *J. Cogn. Neurosci.* 33, 965–983.

Sacca, V., Zhang, Y., Cao, J., Li, H., Yan, Z., Ye, Y., et al. (2022). Evaluation of the modulation effects evoked by different transcutaneous auricular vagus nerve stimulation frequencies along the central vagus nerve pathway in migraines: A functional magnetic resonance imaging study. *Neuromodulation* S1094–7159(22)01256–9. doi: 10.1016/j.neurom.2022.08.459 [Epub ahead of print].

Silva, R. (2012). Myopic maculopathy: A review. *Ophthalmologica* 228, 197–213.

Sun, J., Zhang, Q., Li, Y., Meng, J., Chen, Q., Yang, W., et al. (2020). Plasticity of the resting-state brain: Static and dynamic functional connectivity change induced by divergent thinking training. *Brain Imaging Behav.* 14, 1498–1506. doi: 10.1007/s11682-019-00077-9

Takayanagi, M., Wentz, J., Takayanagi, Y., Schretlen, D., Ceyhan, E., Wang, L., et al. (2013). Reduced anterior cingulate gray matter volume and thickness in subjects with deficit schizophrenia. *Schizophr. Res.* 150, 484–490. doi: 10.1016/j.schres.2013.07.036

Tang, F., Zhu, D., Ma, W., Yao, Q., Li, Q., and Shi, J. (2021). Differences changes in cerebellar functional connectivity between mild cognitive impairment and Alzheimer's Disease: A seed-based approach. *Front. Neurol.* 12:645171. doi: 10.3389/fneur.2021.645171

Tobimatsu, S., and Celesia, G. (2006). Studies of human visual pathophysiology with visual evoked potentials. *Clin. Neurophysiol.* 117, 1414–1433.

Tong, Y., Huang, X., Qi, C., and Shen, Y. (2021). Altered functional connectivity of the primary visual cortex in patients with iridocyclitis and assessment of its predictive value using machine learning. *Front. Immunol.* 12:660554. doi: 10.3389/fimmu.2021.660554

Tootell, R., Tsao, D., and Vanduffel, W. (2003). Neuroimaging weighs in: Humans meet macaques in "primate" visual cortex. *J. Neurosci.* 23, 3981–3989. doi: 10.1523/JNEUROSCI.23-10-03981.2003

Ucak, T., Icel, E., Yilmaz, H., Karakurt, Y., Tasli, G., Ugurlu, A., et al. (2020). Alterations in optical coherence tomography angiography findings in patients with high myopia. *Eye (Lond)* 34, 1129–1135.

Vincent, J., Patel, G., Fox, M., Snyder, A., Baker, J., Van Essen, D., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447, 83–86.

Wang, S., Guo, Y., Liao, C., Chen, Y., Su, G., Zhang, G., et al. (2018). Incidence of and factors associated with myopia and high myopia in chinese children, based on refraction without cycloplegia. *JAMA Ophthalmol.* 136, 1017–1024.

Wen, Z., Zhou, F., Huang, X., Dan, H., Xie, B., and Shen, Y. (2018). Altered functional connectivity of primary visual cortex in late blindness. *Neuropsychiatr. Dis. Treat.* 14, 3317–3327.

Wu, Y., Rao, J., Huang, X., Wu, N., Shi, L., Huang, H., et al. (2021). Impaired interhemispheric synchrony in bronchial asthma. *Int. J. Gen. Med.* 14, 10315–10325. doi: 10.2147/IJGM.S343269

Wu, Y., Wu, N., Huang, X., Rao, J., Yan, L., Shi, L., et al. (2020). Evidence of cortical thickness reduction and disconnection in high myopia. *Sci. Rep.* 10:16239. doi: 10.1038/s41598-020-73415-3

Yamasaki, T., and Tobimatsu, S. (2018). Driving ability in alzheimer disease spectrum: Neural basis, assessment, and potential use of optic flow event-related potentials. *Front. Neurol.* 9:750. doi: 10.3389/fneur.2018.00750

Ye, Q., Chen, X., Qin, R., Huang, L., Yang, D., Liu, R., et al. (2019). Enhanced regional homogeneity and functional connectivity in subjects with white matter hyperintensities and cognitive impairment. *Front. Neurosci.* 13:695. doi: 10.3389/fnins.2019.00695

Yu, Y., Lan, D., Tang, L., Su, T., Li, B., Jiang, N., et al. (2020). Intrinsic functional connectivity alterations of the primary visual cortex in patients with proliferative diabetic retinopathy: A seed-based resting-state fMRI study. *Ther. Adv. Endocrinol. Metab.* 11:2042018820960296. doi: 10.1177/2042018820960296

Zhao, L., Zeng, W., Shi, Y., Nie, W., and Yang, J. (2020). Dynamic visual cortical connectivity analysis based on functional magnetic resonance imaging. *Brain Behav.* 10:e01698.

Zou, Y., Tang, W., Qiao, X., and Li, J. (2021). Aberrant modulations of static functional connectivity and dynamic functional network connectivity in chronic migraine. *Quant. Imaging Med. Surg.* 11, 2253–2264. doi: 10.21037/qims-20-588

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fnins.2023.1130609

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China
Yanggang Feng,
Beihang University, China
Dongsheng Wang,
Jiangsu University of Science and Technology, China

*CORRESPONDENCE
Renping Zhu
✉ xgczrp@ncu.edu.cn
Zhijiang Wan
✉ wandndn@gmail.com

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

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 23 December 2022
ACCEPTED 23 January 2023
PUBLISHED 07 February 2023

CITATION
Li M, Liu S, Wang Z, Li X, Yan Z, Zhu R and Wan Z (2023) MyopiaDETR: End-to-end pathological myopia detection based on transformer using 2D fundus images. *Front. Neurosci.* 17:1130609. doi: 10.3389/fnins.2023.1130609

MyopiaDETR: End-to-end pathological myopia detection based on transformer using 2D fundus images

Manyu Li[1†], Shichang Liu[2†], Zihan Wang[1], Xin Li[2], Zezhong Yan[1], Renping Zhu[1,3,4*] and Zhijiang Wan[1,3*]

[1]School of Information Engineering, Nanchang University, Jiangxi, China, [2]School of Computer Science, Shaanxi Normal University, Xi'an, China, [3]Industrial Institute of Artificial Intelligence, Nanchang University, Jiangxi, China, [4]School of Information Management, Wuhan University, Hubei, China

Background: Automated diagnosis of various retinal diseases based on fundus images can serve as an important clinical decision aid for curing vision loss. However, developing such an automated diagnostic solution is challenged by the characteristics of lesion area in 2D fundus images, such as morphology irregularity, imaging angle, and insufficient data.

Methods: To overcome those challenges, we propose a novel deep learning model named MyopiaDETR to detect the lesion area of normal myopia (NM), high myopia (HM) and pathological myopia (PM) using 2D fundus images provided by the iChallenge-PM dataset. To solve the challenge of morphology irregularity, we present a novel attentional FPN architecture and generate multi-scale feature maps to a traditional Detection Transformer (DETR) for detecting irregular lesion more accurate. Then, we choose the DETR structure to view the lesion from the perspective of set prediction and capture better global information. Several data augmentation methods are used on the iChallenge-PM dataset to solve the challenge of insufficient data.

Results: The experimental results demonstrate that our model achieves excellent localization and classification performance on the iChallenge-PM dataset, reaching AP_{50} of 86.32%.

Conclusion: Our model is effective to detect lesion areas in 2D fundus images. The model not only achieves a significant improvement in capturing small objects, but also a significant improvement in convergence speed during training.

KEYWORDS

myopia detection, fundus images, attentional FPN, detection transformer (DETR), dichotomous graph matching

Introduction

Retinal diseases are one of the main causes of vision loss, and severe retinal diseases can also cause irreversible damage to vision. Medical research has found that the deformation of the front of the eyeball varies with the degree of myopia (Wong et al., 2014), these changes may be related to the complications of ocular diseases, the complications of pathological myopia (PM) are considered to be the main cause of visual impairment and blindness. Due to changes in the environment and lifestyle, the incidence of high myopia-related diseases has been increasing year by year (Hsu et al., 2004; Iwase et al., 2006; Yamada et al., 2010; You et al., 2011;

Furtado et al., 2012). As a common eye disease, it affects 20 to 40% of adults (Iwase et al., 2006) and has become a global burden of public health, 35% of myopic patients are high myopia (HM) (Yamada et al., 2010), which will develop into pathological myopia. The PM is characterized by excessive and progressive elongation of the globe, which is now considered to be the most visually impaired and blind cause. Therefore, timely diagnosis and regular review for PM are very important.

Nowadays, people pay more attention to their health, and the demand for medical services is also increasing. Although the number of ophthalmologists in the developed countries is growing (Hsu et al., 2004; You et al., 2011; Furtado et al., 2012; Sakaguchi et al., 2019), there is still a big gap in the demand for ophthalmologists. Due to the long training time for cultivating doctors, the underdeveloped regions will still face the problem of shortage of medical resources in the next few decades. With the development of imaging technology, myopia-related complications have been identified (Lu et al., 2018; Peng et al., 2019; Nazir et al., 2020; Cui et al., 2021; Zhang et al., 2021; Muthukannan and Glaret Subin, 2022). At present, fundus imaging is an important basis for the diagnosis of various ophthalmic diseases. Most retinal diseases can be avoided with early and timely treatment. Therefore, early detection and early treatment are of great significance for the cure of retinal diseases. However, analysing medical images relies on the extensive medical experience of doctors, which is laborious and time-consuming. Thus, designing a reliable and accurate automatic detection method for fundus images is crucial to the prevention and treatment of diseases.

Many studies utilize deep learning techniques to diagnose eye diseases. For instance, Nazir et al. (2020) proposed a FRCNN algorithm with fuzzy k-means for automatic detecting three types of retinal diseases at early stage. Vyas et al. used common convolutional neural network for dry eye disease (DED) detection based on Tear Film Breakup Time (TBUT) videos, the approach shows high performance in classifying TBUT frames and detecting DED. Muthukannan and Glaret Subin (2022) designed a CNN that optimized by flower pollination for feature extraction, increased the speed and the accuracy of the network for detecting four types of eye diseases. However, most efforts in the existed deep learning focused on applying existing techniques to the myopia detection task rather than proposing new ones specifically suited to the domain. The standard well-known network architectures were designed for the data collected in natural scenes (e.g., natural images) and do not take the peculiarities of the myopia images' characteristics into account. Therefore, research is necessary to understand how these architectures can be optimized for myopia data.

Detection Transformer (DETR) is a new paradigm for end-to-end object detection. DETR always failed in detecting small object and it has a slow convergence speed. Since DETR only utilizes the feature maps (32 × down sampling) from the last layer of backbone, which leads to a large semantic loss of small objects, thus DETR performs poorly on small object detection. Additionally, in the decoder structure of DETR, self-attention is computed for all input pixels, so the model is presented with computational complexity in square level, which further results in slow convergence speed. To solve the problems of the two aspects, Deformable DETR (Zhu et al., 2020) improves the performance of small object detection and accelerates the convergence speed by limiting the range of computed attention using multi-scale feature maps. Conditional DETR (Meng et al., 2021) uses the conditional spatial query explicitly to find the extremity region of the object to reduce the searching range of object and accelerate the convergence.

Figure 1 shows some examples selected from the iChallenge-PM dataset. The figure shows the typical image characteristics of the fundus images, the green background indicates normal myopia (NM), the purple background represents high myopia (HM), and the yellow background delegates pathological myopia (PM). The black mask on the right side of myopia image is atrophy area, which various a lot in shape or to some extend very similar. The white oval area in yellow background is eye's optical disk region and the lesion area appears randomly. These characteristics challenge the model performance of the deep learning methods. The specific challenges are illustrated as follows:

(1) **Morphology Irregularity:** As for the NM and HM shown in Figure 1, the lesion area which in green and purple background is irregular and similar, its area only occupies a small part of the location, which makes the model troubling to learn its morphological features and the most of the rest area is background.

(2) **Imaging Angle:** From all images in Figure 1, it is obvious that the optical disk region hava a tendency to the left of the image, this man-made imaging method may mislead our model, so the differences brought by the imaging angle require the learning ability of the model for location correlation demanding.

(3) **Insufficient Data:** The iChallenge-PM dataset only contains 1,200 images, around 600 images for PM and 600 images for Non-PM (NM+HM). Fewer images and the strong fitting ability of neural network make it easy to overfit, which will reduce the generalization ability of the model.

In this paper, a novel deep learning model named MyopiaDETR is proposed for detecting the lesion of NM, HM, and PM using 2D fundus images. Our base model adopts scalable Swin Transformer as backbone, which is flexible in depth. When it comes to morphology irregularity, it is worth noting that the lesion tissue is not only in irregular shape, but also distributed in a small area of the whole fundus image, most of the pixels are redundant, and an impure background will adversely affect the prediction results. To address those problems, we propose a novel Attentional feature pyramid networks (FPN) architecture that can purify the feature representation during the aggregation of feature maps, specifically, object queries are added to FPN (Lin et al., 2017) levels with positional encoding to execute multi-head self-attention, which are used to give more activation weight to the object regions, produce a larger gap between object and background. Our attentional FPN solves the problem that traditional DETR cannot utilize multi-scale feature maps, resulting in poor performance in detecting small objects in the fundus image. Since most of the regions in the fundus image are background or useless regions, we want our model to focus on the ophthalmic disease regions of interest so as to reduce the computational complexity.

As for the imaging angle, unlike the traditional object detection method, which first generates many candidate boxes on the image, and then adjusts the offset of the boxes according to the calculated loss between the predicted results and the real labels. This method has strong traces of artificial design, which is different from the way humans observe objects, and must require post-processing methods like Non-maximum suppression (NMS) (Neubeck and Van Gool, 2006), whose drawbacks are inevitable in the face of large overlapping areas between real labels. The DETR (Carion et al., 2020) structure, on the other hand, views the object detection from the perspective of set prediction, calculates the loss through the dichotomous graph matching method, and the Transformer structure has global information, which is more in line with the way humans

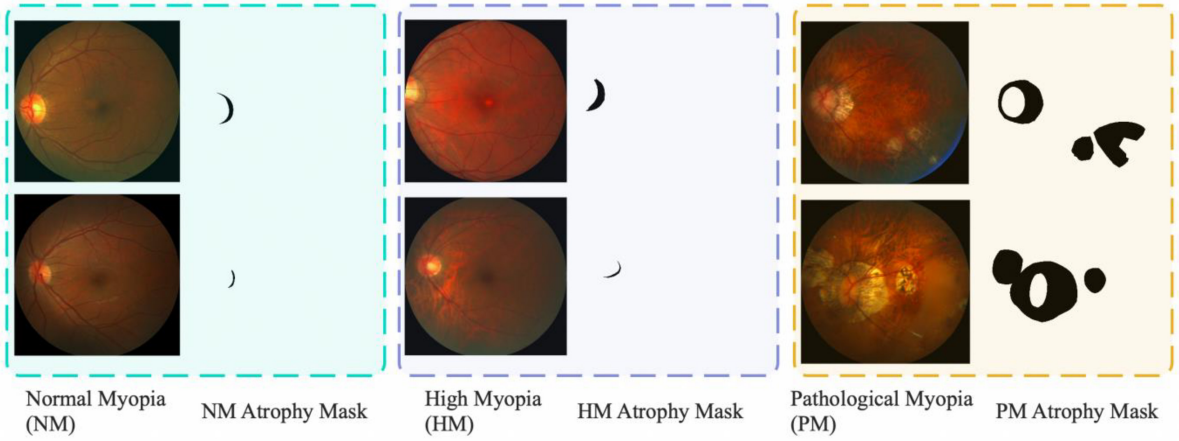


FIGURE 1
Fundus images selected from iChallenge-PM dataset. The dataset contains three types of 2D fundus images (NM, HM, PM).

observe objects. Therefore, we choose the DETR structure for further optimization in the face of sensitive location information.

Finally for the insufficient data mentioned above, Ghiasi et al. (2021) used data augmentation to address the insufficiency of the training data for classifying PM. Almahairi et al. (2018) used CycleGAN with cycle consistency to generate more realistic and reliable images for training, a res-guided U-Net is constructed for segmentation, they achieved superior result on PM detection. We adopt variety of strong data augmentation to enrich iChallenge-PM dataset while training our model. Unlike the common two-branch head design, inspired by DETR, we use a feed forward network that takes the output from Transformer decoder, which will produce the box coordinates and a matrix for classification if the query has an object, no post-processing is needed.

The main contributions of this paper are summarized as follows:

(1) Inspired by DETR, we present a novel post-processing free object detector that uses fundus image data for pathological myopia diagnosis. For specific, we design an attentional FPN that uses object queries on feature maps of each level of FPN, the self-attention mechanism increases the feature intensity gap between foreground and background. Due to the architecture of DETR, it can well solve the challenge of morphology irregularity. To the best of our knowledge, this is the first work using object detection for pathological myopia diagnosis based on iChallenge-PM dataset.

(2) Several data augmentation methods are used on the iChallenge-PM dataset to accelerate model convergence and enhance model robustness.

(3) Extensive experiments are conducted on iChallenge-PM dataset for discriminating NM, HM, and PM. The results demonstrate the superiority of our method than other state-of-the-art (SOTA) object detectors.

The rest of this paper is organized as follows. Section "Related works" introduces related works of the deep learning based retinal disease analysis methods. Section "Materials and methods" illustrates the details of our proposed MyopiaDETR model, which comprizes of Swin Transformer (Liu et al., 2021) backbone, attentional FPN, Transformer encoder, and decoder, shared feed forward network for specific retinal disease analysis tasks. Section "Experimental results" describes the experimental results of ablation studies and comparison studies. Finally, section "Discussion" has a discussion about our method and section "Conclusion" presents the conclusions of this paper and expounds ideas of future work.

Related works

Deep learning based object detection

Object detection is a popular task in computer vision and is widely applied in many real-world scenes such as autonomous driving, video surveillance, remote sensing, and medical diagnosis. The main task of object detection is to locate and classify the target of interest from an image. In the context of the rapid development of computing power, deep learning has been researched and applied as never before. In the trajectory of vision model, AlexNet (Krizhevsky et al., 2017) opened a new era of computer vision by using convolutional neural network years ago, making CNN architecture the mainstream approach of deep learning for many years. Object detectors can be divided into anchor-based and anchor-free model based on whether or not anchor is used during the detection pipeline. The anchor based models can be further divided into two-stage and one-stage detector. One-stage model predicts bounding boxes on grid while two-stage model uses a proposal network to generate candidate boxes, and then uses a second network to refine the result. The advantage of one-stage detectors is that it can complete localization and classification by going through the network once, hence the one-stage detectors can offer significant advantages in terms of speed, such as SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018; Bochkovskiy et al., 2020) series. Two-stage models sacrifice speed for obtaining high accuracy, most of the mainstream detectors with high performance adopt two-stage methods, such as Faster R-CNN (Ren et al., 2015) and Cascade R-CNN (Cai and Vasconcelos, 2018). With the trend of Transformer architecture gradually unifying natural language processing (NLP) and computer vision (CV), Vision Transformer (Dosovitskiy et al., 2020) (ViT) are gradually dominating visual tasks. The excellent relational modeling capability of self-attention mechanism is bringing feature extraction to a new era. Pyramid Vision Transformer brought pyramid structure

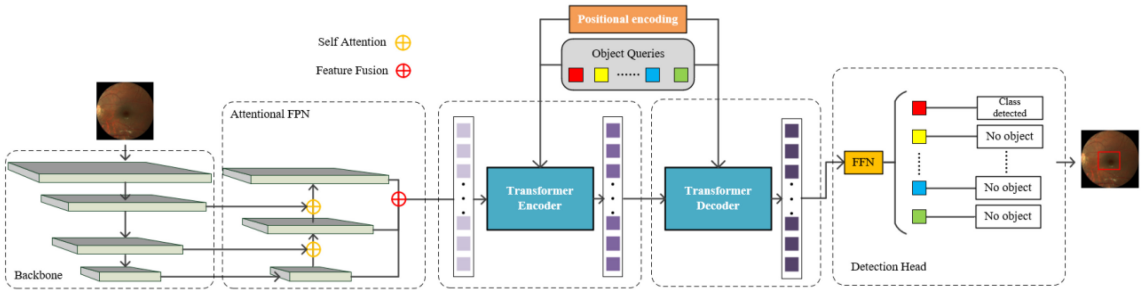


FIGURE 2
Overall architecture of our MyopiaDETR.

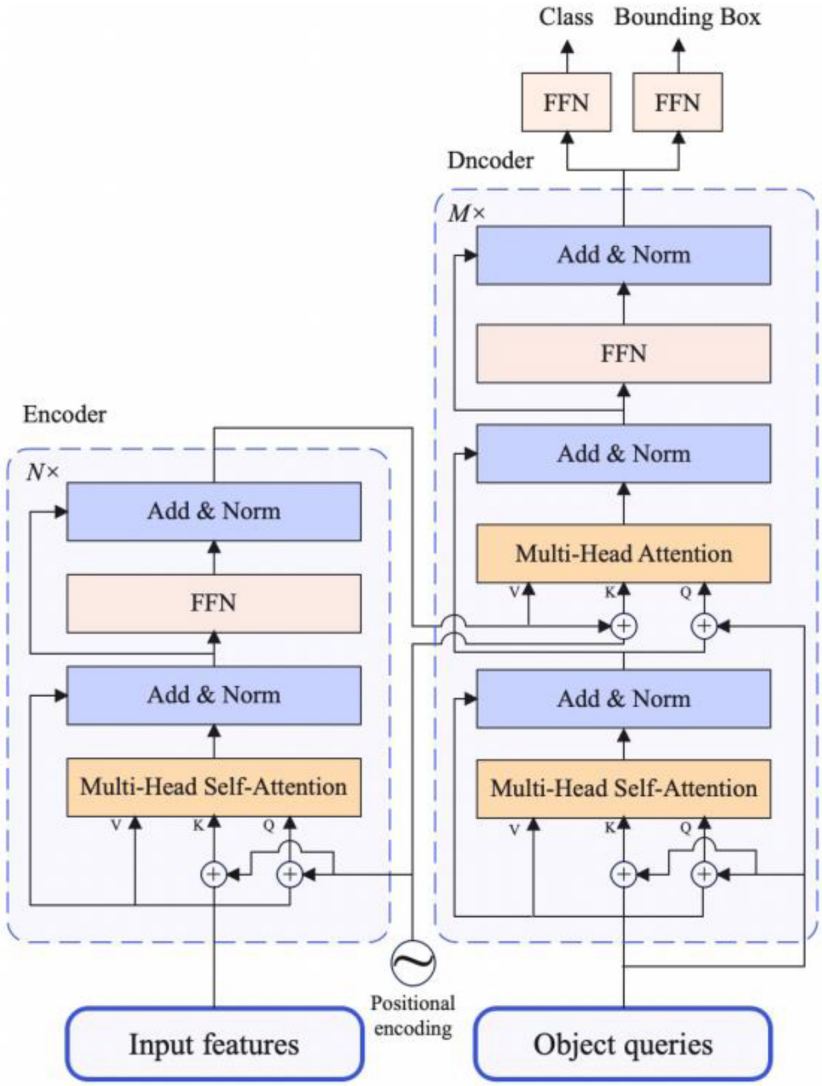


FIGURE 3
The architecture of transformer encoder and decoder.

in to Transformer, making it seamlessly accessible to a variety of downstream tasks (e.g., object detection, semantic segmentation). Swin Transformer (Swin-T) proposed a vision transformer with sliding window operation and hierarchical design, achieved state-of-the-art in many tasks. The hierarchical design makes feature fusion easier. Swin Transformer is an improved version based on ViT, which has a similar structure of CNN. The hierarchical structure of Swin Transformer is more suitable to be applied to many downstream tasks. While ViT is a straight structure, it does not change the dimension of input feature map. In addition, the resolution of fundus image is generally large, and many rich semantic features will be lost if the image shape is resized to a range that is acceptable to ViT.

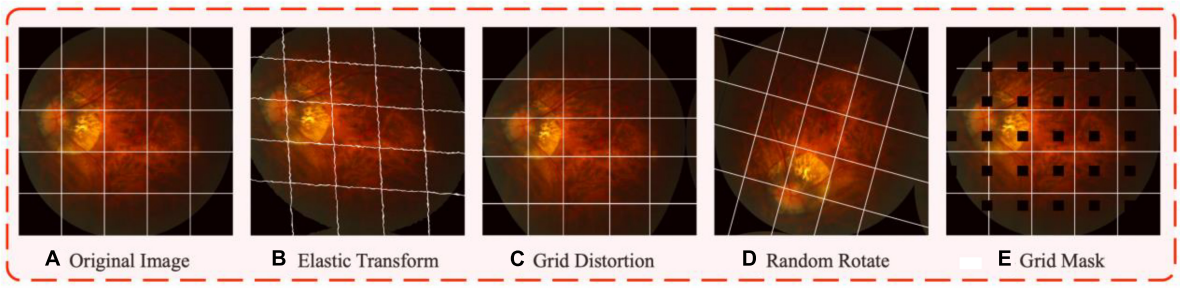


FIGURE 4
The visualized results after data augmentations. Subplot **(A)** is the sample of fundus images in iChallenge-PM dataset, the white grid line added to the original graph is to better show the effect of different augmentation. Subplot **(B–E)** represents elastic transform, grid distortion, random rotate, and grid mask, respectively.

TABLE 1 The effects of attentional FPN.

Model	AP_{50}	AP_S	AP_M	AP_L	Epochs	FPS	MS/SS
DETR 120e	62.53	29.52	65.93	96.18	120	19.31	SS
DETR 200e	71.21	39.12	74.89	97.62	200	19.31	SS
DETR 300e	71.87	39.69	75.47	97.81	300	19.31	SS
DETR + FPN	OOM	–	–	–	–	–	SS
UP-DETR	76.08	44.13	81.77	98.03	300	16.21	SS
Deformable DETR	80.53	58.92	82.55	98.76	120	17.86	MS
Conditional DETR	84.29	65.35	**83.73**	99.10	120	16.60	MS
Ours	**85.18**	**68.71**	83.66	**99.34**	120	8.73	MS

Bold values represent the best metric values achieved by our method and other comparison methods.

Swin Transformer can receive larger image resolution, which means the Swin Transformer can do better in processing images with large resolution than ViT model. Thus, we choose Swin Transformer as our feature extraction backbone network.

Deep learning based eye diseases detection

The object in medical images usually have small sizes and certain morphological features. Many studies utilize deep learning techniques to diagnose eye diseases. Early work such as (Liu et al., 2010) developed a system called PAMELA (Pathological Myopia Detection by Peripapillary Atrophy) that automatically identifies pathological myopia in retinal fundus images. (Wen et al., 2020) placed the key research on the distinction between pathological myopia and high myopia, a two-branch network is proposed, where the first branch distinguishes between normal and abnormal, while the other branch classifies pathological myopia and high myopia. Specifically, the previous studies on iChallenge-PM dataset have been related to image classification and instance segmentation. Cui et al. (2021) used data augmentation to address the insufficiency of the training data for classifying PM. Zhang et al. (2021) used CycleGAN with cycle consistency to generate more realistic and reliable images for training. A res-guided U-Net is also constructed for segmentation, they achieved superior result on PM detection. Our work first use an object detection method for classifying and locating the atrophy based on retinal fundus images. From the previous studies, we can know that the majority of automatic diagnose method uses CNN architecture for feature extraction, and to the best of our knowledge, our work is the first that uses Transformer architecture as backbone and also the first post-processing-free end-to-end detector in myopia diagnosis.

Attention mechanism

With the trend of Transformer architecture gradually unifying NLP and CV (Gumbs et al., 2021), Vision Transformer is gradually dominating the visual tasks. In the evolution of vision attention mechanism, common method can be embedded into CNN for building more relevant feature, such as soft-attention. Soft-attention is a continuous distribution problem, focusing more on spatial or channel, it can be divided into spatial attention and channel attention. Non-local first utilized the idea of Transformer in computer vision model, it takes the approach of doing attention on the feature maps of the intermediate layers, which greatly avoids the computational cost. Self-attention based model, such as ViT and Swin Transformer, has excellent capabilities in extracting relationships between image patches, building the connections of those that are mostly relevant to each other. The pioneer work DETR is worth noting, it applies Transformer architecture as an end-to-end object detector, the post-processing-free design is more compatible with human visual patterns and also avoid the unstable performance of NMS. In the retinal fundus image scenario, the cluttered tissue makes it more difficult to extract the feature of atrophy, so we believe that the self-attention mechanism can help the model to better capture the difference between foreground and background.

TABLE 2 Performance improvement on various augmentations.

Original image	Elastic transform	Grid distortion	Random rotate	Grid mask	AP_{50}
√					80.45
√	√				81.81
√	√	√			82.33
√	√	√	√		85.11
√	√	√	√	√	**85.18**

Bold values represent the best metric values achieved by our method and other comparison methods.

Materials and methods

Diagnosing myopia by detecting lesions based on fundus images requires sufficient data for the deep learning model to have a steady performance (Peng et al., 2019; Virmani et al., 2019). The iChallenge-PM dataset released by Baidu encourage the data-driven methods to automatically detect fundus lesion, it contains three types of fundus diseases images and lesion masks, as illustrate in **Figure 1**, previous studies try to design general deep learning based methods to address the fundus disease identification and localization problems, which contain image classification and segmentation that are insufficient and slow, respectively. In order to implement object detection on this dataset, we transform the mask of lesions into bounding box by obtaining the length and width of the mask. It can be clearly found from the sample images that the region occupied by the disease lesion is only a small part of the fundus image, and the rest of the tissue without lesions can be considered as redundant information, which will be detrimental to the feature extraction and representation of the model. That's why we choose DETR as our baseline model, detail improvements are as follows.

Overall architecture

Inspired by DETR, the overall architecture is illustrated as **Figure 2**, the main components are: backbone network, attentional FPN, Transformer Encoder-Decoder, and a detection head, in which the backbone is responsible for feature extraction, the feature maps of three stages with different resolutions are fed into attentional FPN for feature aggregation. The outputs of attentional FPN are further sent to the encoder-decoder architecture, which consists of multi-head self-attention mechanism, layer normalization and a feed forward neural network, the details are similar to the DETR. Positional encoding is also adopted for retaining positional information of the feature blocks, while the object queries are used for information aggregation, give more attention to the positions where objects are likely to appear. The detection head is in charge of classifying the output of the decoder and finally getting the detection result. In particular, the yellow circles in the **Figure 2** represent self-attention module, which calculates the similarity while fusing feature maps. The red circles represent multiple layers of features for fusion. Small squares of different colors in object queries represent different query objects. As the feature map flows from encoder to decoder, the feature representation becomes clearer and the purple squares become darker. The final feed forward network (FFN) is shared to calculate the object box position and category attributes of the object query. During the training phase, the image batch is fed into the backbone to obtain feature maps of four stages and reduce the computational overhead, three relatively small feature maps are selected and fed into the detection neck for integration. In the upsampling process of the neck network, self-attention operation is added to obtain features with better feature representation ability. Then the aggregation of the neck feature is sent to the following encoder part. In the detection head, FFN takes the output of decoder as input and utilizes bipartite matching loss (i.e., Hungarian Maximum Matching algorithms) to calculate the corresponding loss values. During inference phase, the learned object query generates box candidates through Transformer decoder to select boxes with larger confidence as the final prediction result.

Model backbone

In the overall model architecture, the backbone is in charge of feature extraction of the image. The mainstream architecture of backbone are mainly divided into CNN [AlexNet, ResNet (He et al., 2016), Res2Net (Gao et al., 2019), ResNeXt (Xie et al., 2017), ConvNeXt (Liu et al., 2022)] and Transformer (e.g., ViT, Swin Transformer). The translational invariance and localization of CNNs provide inductive bias, makes CNN models converge faster, however, the fixed receptive field limits the global view of convolution operation. The positional encoding enables Transformer based network to obtain better capabilities in learning global dependencies. Swin Transformer first splits the image into small patches and then feed each patch as a token into Transformer encoder, the core idea is to calculate the similarity between patches for training an attention-intensive network without convolution operations. Compare with the Swin Transformer, the CNNs possess inductive bias, and their convergence speed is relatively fast. The advantages of inductive bias are reflected at two aspects: (1) the convolutional kernel size is generally fixed which result in high local correlation existed in feature maps; (2) the feature maps generated by the CNNs have characteristics of translational invariance, which indicates the output of the convolutional layer does not change no matter where the object appears in the image.

Since our neck network also uses the architecture of the self-attention mechanism, Swin Transformer is selected as our backbone so as to keep the consistency of the features representation. Moreover, the Transformer architecture has good parallel computing capability and global view characteristics. When we input a fundus image with the shape of $H \times W \times C$, it will first pass through a patch partition module with the purpose of descending and chunking the image. The output gets a sequence of $N \times (P2 \times C)$ spreading 2D image blocks, where N is the number of image blocks, P2 is the area of each patch, H and W are the height and width of the image, respectively, and C is the number of image channels. Here we set P to 4, and N is computed by $H/4 \times W/4$. In summary, an input image with the shape $H \times W \times C$ passes through the patch partition module and outputs a tensor of $H/4 \times W/4 \times 48$, which can be understood as a total of H/4*W/4 image patches, each of which is composed of a 48-dimensional token. The W-MSA and SW-MSA modules help the Swin Transformer to improve its ability of extracting the global features in the fundus image. The W-MSA module restricts the receptive field of the model by only applying the self-attentive mechanism within each patch. The SW-MSA model adds a cyclic shift operation to the W-MSA for extracting the features between patches.

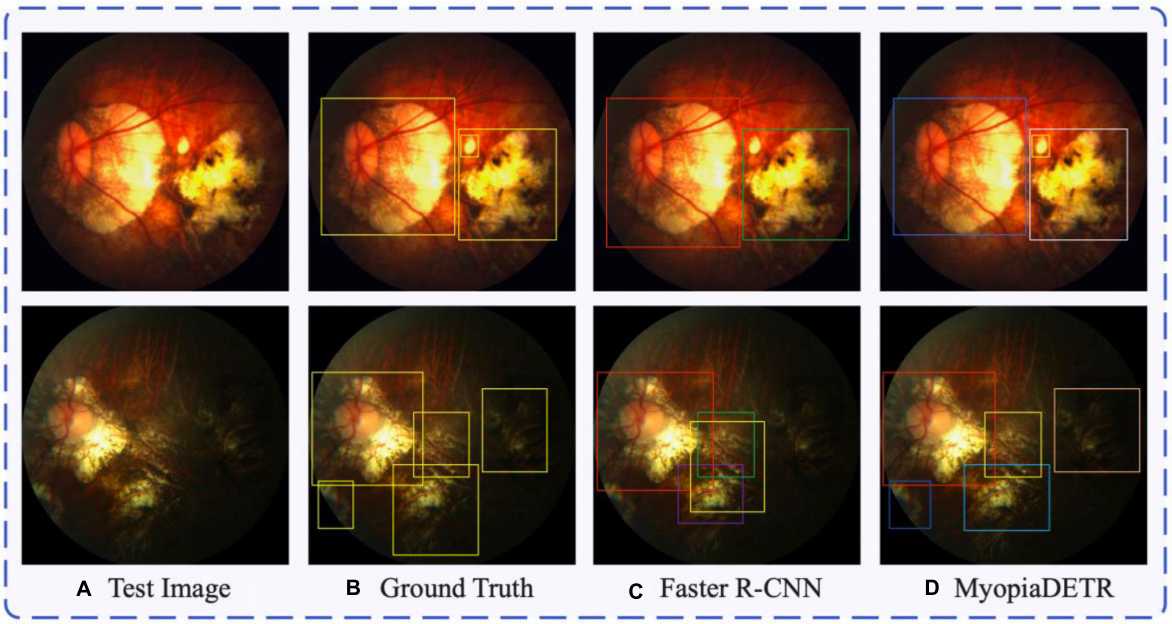


FIGURE 5
The comparison between traditional object detection algorithm Faster R-CNN and MyopiaDETR. Subplot **(A)** shows the test image, subplot **(B)** shows the PM location labels, subplot **(C)** shows the output of the Faster R-CNN algorithm, and subplot **(D)** shows the output of our MyopiaDETR.

As illustrated by Figure 2. The Swin Transformer backbone network has four stages corresponding to four feature maps of different sizes. The feature map sizes are H/4 × W/4 × C, H/8 × W/8 × 2C, H/16 × W/16 × 4C, H/32 × W/32 × 8C. The feature map of each layer will be fed to the subsequent attentional FPN structure for further processing.

Attentional FPN

Due to the high computational complexity of the Transformer architecture, which is of O(n2) level, only the feature map of the last layer (32 times down sampling) is utilized in the original DETR, resulting in the loss of small object feature information and has the limitation of single scale in feature representation. To address the above issues, we added the attentional FPN architecture to utilize all the features that output from the backbone network. The output of the first layer of the attentional FPN, which is also the last stage output from the backbone, is down sampled by 32 times. And then, for each neck level, the feature size is upsampled by 2 times. As the structure illustrated in Figure 2, the feature size of each neck level is 16 × 16, 32 × 32, and 64 × 64 when the input size is 512 × 512. Self-attention is aggregate to the FPN to focus on the regions of interest in each layer of the feature map. The feature map of each layer will be sliced into 8 times × 8 times pixels blocks, and the results computed between the blocks are used as weights for the output, the final computation is used to activate the part of interest in the feature map. Specifically, we add an auxiliary head after the attentional FPN to distinguish the foreground from the background, and only the foreground region will be input to the follow-up Transformer structure. The attentional FPN outputs the final blocks while recording the sparse encoding matrix of them, which is fed to the Transformer Decoder to map the features back to the original image.

Since each feature map is partitioned into 8 times × 8 times pixels blocks, there will be a total of 64 pixel blocks after the slice. Let's define the input $x = (x_1, x_2,..., x_{64})$. The input elements will be passed through an embedding layer W to obtain a multiset of one-dimensional vectors, denoted as $a = (a_1, a_2,..., a_{64})$. Meanwhile, three learnable matrices are also set as W^Q, W^K, W^V, represent the query matrix, the key matrix, and the value matrix, respectively. In particular for a single input x_i, the proportion of its weights is calculated by the following formula:

$$q_i = W^Q W x_i = W^Q a_i \tag{1}$$

$$k_i = W^K W x_i = W^K a_i \tag{2}$$

$$v_i = W^V W x_i = W^V a_i \tag{3}$$

$$\alpha_{i,\,64} = \frac{q_i^T \cdot k_i}{\sqrt{d_{q,\,k}}},\ \widetilde{\alpha}_{i,\,64} = Softmax(\alpha_{i,\,64}) \tag{4}$$

$$b_i = \sum_{i=1}^{64} \widetilde{\alpha}_{i,\,64} \cdot v_i \tag{5}$$

The subscript (i, 64) of α in the above equation (4), represents that the similarity of the i-th input patch is currently being calculated, and there are a total of 64 patches to be calculated. The purpose of dividing by root $d_{q,k}$ when calculating α is for normalization to avoid gradient vanish and $d_{q,k}$ means the dimension of q and k vector. The softmax function is to map the sum of the weight ratios to 1, which is convenient for calculating b_i.

Transformer encoder-decoder

The structure of Transformer encoder-decoder is shown in Figure 3, details are as follows:

Transformer encoder

The input requirement of Transformer encoder module is a sequence, so when getting the feature map output by attentional FPN, let the original feature map shape be C × H × W, we need to descend the channel dimension of the feature map first, and then flatten the H and W dimension. Finally, we get the feature map with the shape of C × L, where L equals to H × W. Each encoder block has a unified structure: multi-head self-attention, Add and Norm and feed forward network. As for Add and Norm structure, Add stands for residual connection to prevent network degradation, Norm is the layer normalization, which normalizes the activation values of each layer. Because of the invariance of Transformer Architecture, we add the fixed spatial positional encoding to the attention layer to complement the location information.

Transformer decoder

The architecture of the decoder part is the same as the traditional Transformer, but the difference is that each decoder module of our model decodes N inputs from encoder in parallel. Because the decoder structure is also permutation-invariant, the N inputs should be different so that different results can be generated. The meaning of object queries is similar to that of anchor in traditional object detection methods, and it is learnable. We input them to each multi-head self-attention module, which will eventually be decoded into object boxes location information and category information by the FFN structure, which is described in section "Detection head".

To sum up, the features extracted from the backbone are passed through the multi-head self-attention module in the encoder structure along with the spatial positional encoding. After that, the N outputs of encoder and object queries are fed to the decoder part. Finally, the final object boxes and category information is output by the multiple multi-head self-attention and decoder-encoder attention and the FNN structure.

Detection head

The role of detection head in DETR is to predicting a fixed set of object detections for each input image, rather than using a sliding window or anchor-based approach in traditional object detection models. After obtaining the output of the decoder, the final result is processed by a 3-layer perceptron and a linear projection layer, where the perceptron is comprized of a ReLU activation function. FNN computes the position of the box. The linear layer computes the category to which it belongs by softmax function. Since our predicted set is composed of N boxes, but in fact N is much larger than the actual number of objects present in the image, we mark the object query with no detected object as a background class. In particular, our FFN share the same weights and are calculated equally for all N object queries.

Loss function

The loss function calculation is performed in two steps. First, finding the optimal pairwise loss between the ensemble predictio

TABLE 3 Comparisons between different backbones.

Backbone	AP_{50}	AP_S	AP_M	AP_L
Swin-S	85.18	68.71	83.66	99.34
Swin-B	86.23	69.76	84.65	99.35
Swin-L	85.62	69.08	83.72	99.34
ResNet-50	82.25	65.91	80.34	98.72
ResNet-101	84.67	67.21	82.76	99.12
ResNet-152	86.32	69.65	84.78	99.21

and the ground-truth label in the Hungarian algorithm alignment, the index of the set of solutions is set to $\widehat{\sigma}$, as illustrated in formula (6), where $y_i$is the set of prediction, $\hat{y}_i$is the ground-truth, both of them need to be stretched to a fixed length by adding None value, where length = max(len(y_i), len($\hat{y}_i$)). The L_{MATCH} is defined as the gap between the predicted set and the ground-truth labels in the case of the first group pairing.

$$\widehat{\sigma} = argmin \sum L_{MATCH}(y_i, \hat{y}_i). \tag{6}$$

Second, the index value $\widehat{\sigma}$ that calculated in the first steps is used to calculate the classification loss and the predicted bounding box loss, c_i is the class label, and the Hungarian loss is calculated as:

$$L_{Hungarian}(\mathrm{y}, \hat{y}) = \sum_{i=1}^{N} [-log_{\hat{p}_{\widehat{\sigma}(i)}}(c_i) + L_{box}(b_i, \hat{b}_{\widehat{\sigma}}(i))] \tag{7}$$

where bounding box loss uses weighted IoU and L1 loss, λ_{iou} and λ_{L1} are the weight of IoU loss and L1 loss, respectively. The box loss is calculated as:

$$L_{box}(b_i, \hat{b}_{\widehat{\sigma}}(i)) = \lambda_{iou} L_{iou}(b_i, \hat{b}_{\widehat{\sigma}}(i)) + \lambda_{L1} ||b_i - \hat{b}_{\widehat{\sigma}}(i)||_1 \tag{8}$$

Experimental results

Dataset description

iChallenge-PM

Myopia has become a global public health burden. As the refractive error of myopia increases, high myopia will progress to pathological myopia, causing irreversible visual damage to the patient. Therefore, early diagnosis and regular follow-up are very important. With this challenge, the iChallenge competition jointly organized by Baidu Brain and Zhongshan Eye Center of Sun Yat-sen University provides iChallenge-PM, a dataset on pathological myopia, which includes 1,200 annotated retinal fundus images from non-pathological myopic subjects and pathological myopic patients (about 50%). There are 400 training data, validation data, and test data sets each.

Data augmentation

We adopt several data augmentation methods to address the challenges of imaging angle and insufficient data. Subplot (b) in Figure 4 shows the elastic transform, which was proposed by Simard et al. (2003) and made great progress on the MNIST handwritten dataset, and the method is gradually applied to medical image processing and has been widely used [e.g., Mottl et al. (2002), Gulshad et al. (2021)], for the parameter settings we set the alpha to 50 and the

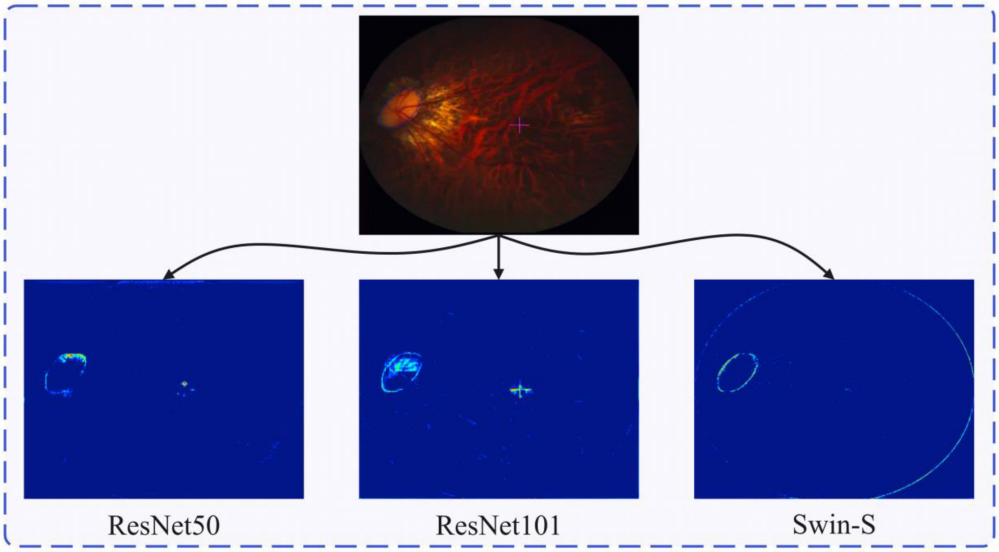


FIGURE 6
Retinal fundus image and the corresponding feature heatmap generated by different backbones, the heatmap of the feature learned by ResNet50, ResNet101, and Swin-S from left to right.

sigma to 5. Subplot (c) is grid distortion, its effect is similar to that of the elastic transform, which is a non-rigid body transformation. Submap (d) is random rotation data augmentation, which aims to increase the diversity of imaging angles to solve the challenge of imaging angle, and the rotation angle is set from −180 to 180 degrees. Subplot (e) is the grid mask data augmentation proposed by Chen et al. (2020), which randomly masks a number of block locations on the image and fills them with 0 pixel values, this data augmentation can mask part of the positive samples with certain probability, thus preventing the model from overfitting to simple local features, for the parameter settings we set the ratio to 0.3 and the x_holes and y_holes to be set randomly between 5 and 7. The gains from each enhancement will be presented in the ablation study part.

Evaluation metrics

For evaluating the performance of our detection method on iChallenge-PM dataset, we use mAP as the evaluation metrics, which is the mean average precision of all categories. AP θ is calculated as the area enclosed by the Precision (P) and Recall (R) and the coordinate axis at an IoU (Intersection over Union of predicted box and ground-truth) threshold of θ, as illustrated below:

$$IoU(A,\ B) = \frac{A \bigcap B}{A \bigcup B} \tag{9}$$

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

where TP (True Positive) is the number of IoU between the predicted box and the ground-truth label that is greater than or equal to the threshold θ, while FP (False Positive) means the number of IoU between the predicted box and the ground-truth label that is less than the threshold θ. FN (False Negative) means no positive object is detected. Finally, we can get mAP by averaging AP θ at different thresholds.

Implementation details

Our model is implemented using MMDetection object detection algorithm library based on Pytorch1.8 deep learning framework using four NVIDIA RTX 3090 GPUs. We pre-trained our model on COCO dataset for 36 epochs and fine-tuned on iChallenge-PM for 100 epochs with a mini-batch size of 16 due to the limited data amount. The learning rate is initiated to 0.001, and we use CosineAnnealingLR to decay the learning rate with 5 epochs warm-up. The AdamW optimizer is used to optimize the hyper-parameters. For evaluating model performance, the IoU threshold and confidence threshold are set to 0.5 and 0.01, respectively. NMS is adopted as post-process method. As for the data augmentation implementation approach, we adopt the image processing algorithms based on opencv-python and Albumentations library (Buslaev et al., 2020).

Ablation study

To verify the effectiveness of our attentional FPN, we set up a set of ablation experiments. The backbone network used as the base model was Swin-Small, DETR and its variants are selected to test the gains we obtained by adding attentional FPN. In object detection, taking COCO object definition as an example (Kisantal et al., 2019), we define a small object as an object which pixel number is less than 32 × 32, a medium object is an object which pixel number is between 32*32 and 96*96, and a large object is an object which pixel number is larger than 96*96. AP_S means object area smaller than 32 × 32 pixel points, AP_M means object area between 32 × 32 and 96 × 96 pixel points, AP_L means object area bigger than 96 × 96 pixel points. As can be seen from **Table 1**, the DETR works well for large target detection, but suffers from a major shortcoming in small object detection. If the FPN is added directly to the DETR structure,

since the Tranformer has an $O(n^2)$ computational complexity and limited computational resources, this method will result in an out of memory (OOM) error. In contrast, by adding a self-attention mechanism to the FPN, only a small number of features need to be fed into the Transformer structure, and our model achieves 29.02% improvement in detecting small objects, and obtains 8.19 and 1.53% improvement in detecting medium and large object, respectively. It is worth noting that Deformable DETR, Conditional DETR and our method achieve more significant improvement with multi-scale (MS) than the other models with single-scale (SS) design. The result indicates that multi-scale learning allows greater variety of features. With Attentional FPN, our model not only achieves a significant improvement in capturing small objects, but also a significant improvement in convergence speed during training. In a nutshell, our model not only achieves a significant improvement in capturing small objects, but also a significant improvement in convergence speed during training. Note that all experiments adopt the same data augmentation methods.

Furthermore, it is important to carefully evaluate the impact of data augmentation on model performance, which can provide insights into the factors that contribute to model performance. In Table 2, we focus on the AP boost from the augmentation method, presenting in a cumulative manner. As can be seen evidently, Random Rotate brings the most significant performance gains because it simulates a situation that is similar to the characteristics of the original data (i.e., the lesion sites tend to appear in different directions), this greatly expands the training sample. The other augmentation methods also achieved considerable improvements, proving that the increment of data volume by the data augmentation methods is effective to improving model performance.

Comparison study

Subplot (a) in Figure 5 shows the test image, subplot (b) shows the PM location labels, subplot (c) shows the output of the Faster R-CNN algorithm, and subplot (d) shows the output of our MyopiaDETR. The detection result in the first row of subplot (c) fails to detect the small object compared to the ground truth, and the traditional object detection algorithm has limited learning of complex morphological features. The yellow box in the second row of subplot (c) is the model's false detection, and if the threshold of post-processing of NMS is adjusted down, it will lead to the purple or green box being removed by the post-processing algorithm, and the detection ability is still very poor for irregular PM regions. Our MyopiaDETR does not have these problems, and not only can handle the detection of irregular PM, but also do not have to worry about the false removal of detection boxes caused by post-processing.

In addition, we compare the effect of different backbone networks on the performance of our model. As shown in Table 3, in the small to medium sized network architecture with a similar number of parameters, Swin Transformer shows better performance, we believe this is because the feature extracted by Swin Transformer is consistent with the operations in attentional FPN. Due to the morphology irregularity, we need a feature extraction network with strong capability to capturing global context feature. Swin Transformer has better feature extraction capability than ResNet because of its global field of view. Thus, the Swin Transformer avoids the problem of morphology irregularity and shows a better detection performance than the ResNet. However, because the Transformer architecture lacks inductive bias, as Swin Transformer becomes deeper, it requires a lot more data and the model performance is somewhat weakened. Compared to the CNN structure, the global feature representation capability of Transformer is more prominent, and Swin Transformer performance has a significant advantage over the CNN structure in the case of small model structures.

Figure 6 shows the 8 × down sampled feature maps of ResNet-50, ResNet-101, and Swin-S. As the CNN is more localized, it will to some extent activate non-focal regions, such as the cross-focused symbols in the figure. The Swin Transformer structure, on the other hand, has a global field of view and can focus on more important information (e.g., lesion borders as well as slice edge contours). As shown in Figure 6, Swin Transformer has a better understanding of the global information and can pay more attention to the global contour information. ResNet, on the other hand, has a strong feature extraction capability, but it only pays attention to the local contour information and has more activation within the local contour information. In contrast, we do not need to pay attention to all the information within the local contour information when dealing with the ophthalmic disease region, thus highlighting the superiority of Swin Transformer in feature extraction.

Discussion

The sources of novelty in our work are: (1) Ordinary deep learning-based object detection methods utilize many modules with obvious traces of artificial design, that is, a paradigm that does not match the way human vision finds objects. (2) DETR proposes a new object detection paradigm based on ensemble prediction, which directly predicts all objects in an image without post-processing, which is more in line with human visual habits. (3) Due to the high computational complexity of Transformer and the fact that DETR only uses the output features of the last layer of the model, which contains rich semantic information but is weak in the representation of local information, the performance in small object detection is poor, so we propose attentional FPN for feature aggregation, which uses all the output features of the backbone network, significantly improve the performance in small object detection.

However, although our proposed new method solves the problem of small object detection as well as purifying the background features, there are still some drawbacks. The use of Swin Transformer as backbone makes the training time much longer than traditional DETR model. Specifically, in DETR, each query is responsible for a part of the location and size of the object in an image, which requires all objects in all images in the training set to be well apportioned to different queries, so more epochs are needed. The data amount is very important for deep Transformer architecture model. Our follow-up research is considering to replace the backbone network with an optimized architecture to adapt to scenarios with small data volumes.

Conclusion

In this paper, we propose attentional FPN and use a new paradigm for object detection to solve the eye disease detection challenges based on 2D fundus images. The experimental results show that our attentional FPN can be adapted to other deep learning architectures with only a small increment in computational cost to achieve a significant accuracy improvement. Several augmentation

methods are utilized to improve the data volume and make our model achieve considerable performance improvements, proving that the increment of data volume by the data augmentation methods is effective to detecting lesion area in 2D fundus images. Our model not only achieves a significant improvement in capturing small objects, but also a significant improvement in convergence speed during training.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://aistudio.baidu.com/aistudio/datasetdetail/177172.

Author contributions

ML, SL, and ZJW contributed to the statistical analyses and wrote the manuscript. ZHW, ZY, and XL contributed to the data collection and figures plot. RZ provided the technique and wrote the guidance. All authors read and approved the final manuscript, contributed to data collection and article, and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

Almahairi, A., Rajeshwar, S., Sordoni, A., Bachman, P., and Courville, A. (2018). "Augmented cyclegan: Learning many-to-many mappings from unpaired data," in *Proceedings of the international conference on machine learning (PMLR)*, Stockholm, 195–204.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv* [Preprint]. arXiv:2004.10934.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information* 11:125. doi: 10.3390/info11020125

Cai, Z., and Vasconcelos, N. (2018). "Cascade R-CNN: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City: IEEE), 6154–6162. doi: 10.1109/CVPR.2018.00644

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers," in *Proceedings of the European conference on computer vision*, (Cham: Springer), 213–229. doi: 10.1007/978-3-030-58452-8_13

Chen, P., Liu, S., Zhao, H., and Jia, J. (2020). Gridmask data augmentation. *arXiv* [Preprint]. arXiv:2001.04086.

Cui, J., Zhang, X., Xiong, F., and Chen, C.-L. (2021). Pathological myopia image recognition strategy based on data augmentation and model fusion. *J. Healthc. Eng.* 2021:5549779. doi: 10.1155/2021/5549779

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* [Preprint]. arXiv:2010.11929.

Furtado, J. M., Lansingh, V. C., Carter, M. J., Milanese, M. F., Peña, B. N., Ghersi, H. A., et al. (2012). Causes of blindness and visual impairment in Latin America. *Surv. Ophthalmol.* 57, 149–177.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. (2019). Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 652–662. doi: 10.1109/TPAMI.2019.2938758

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., et al. (2021). "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Nashville, TN, 2918–2928. doi: 10.1109/CVPR46437.2021.00294

Gulshad, S., Sosnovik, I., and Smeulders, A. (2021). Built-in elastic transformations for improved robustness. *arXiv* [Preprint]. arXiv:2107.09391.

Gumbs, A. A., Frigerio, I., Spolverato, G., Croner, R., Illanes, A., Chouillard, E., et al. (2021). Artificial intelligence surgery: How do we get to autonomous actions in surgery? *Sensors* 21:5526. doi: 10.3390/s21165526

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, 770–778.

Hsu, W.-M., Cheng, C.-Y., Liu, J.-H., Tsai, S.-Y., and Chou, P. (2004). Prevalence and causes of visual impairment in an elderly Chinese population in Taiwan: The Shihpai eye study. *Ophthalmology* 111, 62–69. doi: 10.1016/j.ophtha.2003.05.011

Iwase, A., Araie, M., Tomidokoro, A., Yamamoto, T., Shimizu, H., Kitazawa, Y., et al. (2006). Prevalence and causes of low vision and blindness in a Japanese adult population: The Tajimi study. *Ophthalmology* 113, 1354–1362. doi: 10.1016/j.ophtha.2006.04.022

Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., and Cho, K. (2019). Augmentation for small object detection. *arXiv* [Preprint]. arXiv:1902.07296.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 2117–2125. doi: 10.1109/CVPR.2017.106

Liu, J., Wong, D. W., Lim, J. H., Tan, N. M., Zhang, Z., Li, H., et al. (2010). Detection of pathological myopia by Pamela with texture-based features through an SVM approach. *J. Healthc. Eng.* 1, 1–11.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "SSD: Single shot multibox detector," in *Proceedings of the European conference on computer vision*, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer), 21–37.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, Nashville, TN, 10012–10022. doi: 10.7717/peerj-cs.1093

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, New Orleans, LA, 11976–11986. doi: 10.1109/CVPR52688.2022.01167

Lu, W., Tong, Y., Yu, Y., Xing, Y., Chen, C., and Shen, Y. (2018). Applications of artificial intelligence in ophthalmology: General overview. *J. Ophthalmol.* 2018:5278196.

Meng, D., Chen, X., Fan, Z., Zeng, G., Li, H., Yuan, Y., et al. (2021). "Conditional detr for fast training convergence," in *Proceedings of the IEEE/CVF international conference on computer vision*, Montreal, QC, 3651–3660.

Mottl, V., Kopylov, A., Kostin, A., Yermakov, A., and Kittler, J. (2002). "Elastic transformation of the image pixel grid for similarity based face identification," in *Proceedings of the 2002 international conference on pattern recognition*, Vol. 3, (Manhattan, NY: IEEE), 549–552.

Muthukannan, P., and Glaret Subin, P. (2022). Optimized convolution neural network based multiple eye disease detection. *Comput. Biol. Med.* 146:105648. doi: 10.1016/j.compbiomed.2022.105648

Nazir, T., Irtaza, A., Javed, A., Malik, H., Hussain, D., and Naqvi, R. A. (2020). Retinal image analysis for diabetes-based eye disease detection using deep learning. *Appl. Sci.* 10:6185. doi: 10.3390/app10186185

Neubeck, A., and Van Gool, L. (2006). "Efficient non-maximum suppression," in *In Proceedings of the 18th international conference on pattern recognition (ICPR'06)*, Vol. 3, (Piscataway, NJ: IEEE), 850–855. doi: 10.1109/ICPR.2006.479

Peng, Y., Dharssi, S., Chen, Q., Keenan, T. D., Agron, E., Wong, W. T., et al. (2019). Deepseenet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology* 126, 565–575. doi: 10.1016/j.ophtha.2018.11.015

Redmon, J., and Farhadi, A. (2017). "YOLO9000: Better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 7263–7271. doi: 10.1109/CVPR.2017.690

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv* [Preprint]. arXiv:1804.02767.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 1:28. doi: 10.1109/TPAMI.2016.2577031

Sakaguchi, A., Wu, R., and Kamata, S.-I. (2019). "Fundus image classification for diabetic retinopathy using disease severity grading," in *Proceedings of the 2019 9th international conference on biomedical engineering and technology*, Tokyo, 190–196.

Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the 7th international conference on document analysis and recognition*, Edinburgh, 3.

Virmani, J., Singh, G. P., Singh, Y., and Kriti. (2019). PNN-based classification of retinal diseases using fundus images. *Sens. Health Monit.* 5, 215–242.

Wen, Y., Chen, L., Qiao, L., Deng, Y., Dai, S., Chen, J., et al. (2020). "On automatic detection of central serous chorioretinopathy and central exudative chorioretinopathy in fundus images," in *Proceedings of the 2020 IEEE international conference on bioinformatics and biomedicine (BIBM)*, (Manhattan, NY: IEEE), 1161–1165. doi: 10.1109/BIBM49941.2020.9313274

Wong, T. Y., Ferreira, A., Hughes, R., Carter, G., and Mitchell, P. (2014). Epidemiology and disease burden of pathologic myopia and myopic choroidal neovascularization: An evidence-based systematic review. *Am. J. Ophthalmol.* 157, 9–25. doi: 10.1016/j.ajo.2013.08.010

Xie, S., Girshick, R., Dollar, P., Tu, Z., and He, K. (2017). "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Manhattan, NY, 1492–1500. doi: 10.1109/CVPR.2017.634

Yamada, M., Hiratsuka, Y., Roberts, C. B., Pezzullo, M. L., Yates, K., Takano, S., et al. (2010). Prevalence of visual impairment in the adult Japanese population by cause and severity and future projections. *Ophthalmic Epidemiol.* 17, 50–57. doi: 10.3109/09286580903450346

You, Q. S., Xu, L., Yang, H., Wang, Y. X., and Jonas, J. B. (2011). Five-year incidence of visual impairment and blindness in adult Chinese: The Beijing eye study. *Ophthalmology* 118, 1069–1075. doi: 10.1016/j.ophtha.2010.09.032

Zhang, Z., Ji, Z., Chen, Q., Yuan, S., and Fan, W. (2021). Joint optimization of cyclegan and cnn classifier for detection and localization of retinal pathologies on color fundus photographs. *IEEE J. Biomed. Health Inf.* 26, 115–126. doi: 10.1109/JBHI.2021.3092339

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., and Dai, J. (2020). Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv* [Preprint]. arXiv:2010.04159.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fnins.2023.1057781

OPEN ACCESS

EDITED BY
Vijaya Prakash Krishnan Muthaiah,
University at Buffalo, United States

REVIEWED BY
Pinaki Dutta,
Post Graduate Institute of Medical Education and Research (PGIMER), India
Alhusain Nagm,
Al-Azhar University, Egypt

*CORRESPONDENCE
Jingxiang Zhong
✉ zjx85221206@163.com
Wei Mo
✉ waysonsci@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 30 September 2022
ACCEPTED 17 January 2023
PUBLISHED 09 February 2023

CITATION
Pang YH, Tan Z, Chen XX, Liao ZH, Yang X, Zhong Q, Huang BQ, Zhong QS, Zhong JX and Mo W (2023) Evaluation of preoperative visual pathway impairment in patients with non-functioning pituitary adenoma using diffusion tensor imaging coupled with optical coherence tomography. *Front. Neurosci.* 17:1057781. doi: 10.3389/fnins.2023.1057781

Evaluation of preoperative visual pathway impairment in patients with non-functioning pituitary adenoma using diffusion tensor imaging coupled with optical coherence tomography

Yanhua Pang[1†], Zhi Tan[2†], Xinxin Chen[3], Zhihui Liao[1], Xin Yang[1], Qin Zhong[2], Baqi Huang[4], Qianshuo Zhong[4], Jingxiang Zhong[5,6*] and Wei Mo[4*]

[1]Department of Ophthalmology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China, [2]Department of Radiology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China, [3]Teaching and Research Center of Medical Communication Science, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China, [4]Department of Neurosurgery, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China, [5]Department of Ophthalmology, The First Affiliated Hospital of Jinan University, Guangzhou, China, [6]Department of Ophthalmology, The Sixth Affiliated Hospital of Jinan University, Guangzhou, China

Objective: Optic chiasma compression and associated visual impairment induced by a non-functioning pituitary adenoma (NFPA) is commonly assessed by the optic disk and retina but is inadequate to understand the entire visual pathway impairment. We aim to evaluate the use of optical coherence tomography (OCT) coupled with diffusion tensor imaging (DTI) for the preoperative evaluation of visual pathway impairment.

Methods: Fifty-three patients with NFPA (categorized into mild and heavy compression subgroups) were subjected to OCT to calculate the thickness of the circumpapillary retinal nerve fiber layer (CP-RNFL), macular ganglion cell complex (GCC), macular ganglion cell layer (GCL), and macular inner plexus layer (IPL), as well as to DTI to calculate the fractional anisotropy (FA) and apparent diffusion coefficient (ADC) values.

Results: Compared to mild compression, heavy compression caused decreased FA value, increased ADC value of several segments of the visual pathway, thin temporal CP-RNFL, and quadrant macular GCC, IPL, and GCL. Average CP-RNFL thickness, inferior-macular inner-ring IPL and GCC thicknesses, inferior CP-RNFL thickness, and superior CP-RNFL thickness were the best indicators of the impairment of the optic nerve, optic chiasma, optic tract, and optic radiation, respectively.

Conclusion: DTI and OCT parameters can effectively evaluate visual pathway impairment and are beneficial for the objective preoperative evaluation of visual pathway impairment in patients with NFPA.

KEYWORDS

non-functioning pituitary adenoma, diffusion tensor imaging (DTI), visual pathway injury, optical coherence tomography (OCT), circumpapillary retinal nerve fiber layer (CP-RNFL)

Introduction

Non-functional pituitary adenoma (NFPA) is the most common type of pituitary adenoma (PA) and is characterized by a lack of hormone-related clinical symptoms and signs (Melmed et al., 2022). Delayed diagnosis of NFPA caused by the absence of indications generates giant adenomas and subsequently extends to the surrounding structures. Compressed optic chiasma and the resulting visual impairment patterns (i.e., vision loss and visual field defects) occur in 78% of patients with NFPA (Subramanian et al., 2021). Ophthalmic examinations and the assessment of visual pathway impairment are critical for early disease diagnosis and surgery decisions for NFPA. Although visual field examination and visual evoked potentials can be used to identify impaired visual fields or pathways, the results are relatively subjective and time-consuming. Given these limitations, many studies have explored more objective and quantitative methods for assessing visual impairment in patients with NFPA.

Optical coherence tomography (OCT) is a new technique for predicting outcomes after the surgical decompression of PA (Chung et al., 2020; Agarwal et al., 2021; Santorini et al., 2022). In addition, diffusion tensor imaging (DTI) is a new type of functional magnetic resonance used for objectively assessing white matter fiber connectivity and integrity in the central system tissue by detecting the diffusion differences between the parallel and vertical motions of white matter water molecules (Yamada et al., 2016), as well as for objectively predicting the post-surgery visual recovery of patients with PA. A strategy that combines OCT and DTI is increasingly used to predict visual impairment in degenerative neuropathy (Hubers et al., 2016; Alves et al., 2018) and visual pathways characteristic of amblyopia (Altintas et al., 2017). In our previous study, fractional anisotropy (FA) values of the visual pathway were found to be positively correlated with the thickness of the circumpapillary retinal nerve fiber layer (CP-RNFL) in patients with PA, suggesting the feasibility of combining OCT and DTI in evaluating the impairment of the entire visual pathway (Pang et al., 2022). The current study compared retinal OCT parameters and optic nerve DTI parameters in 53 patients with NFPA (classified into mild and heavy optic chiasma compression subgroups). The degree of adverse changes in OCT and DTI parameters was found to be stronger in the severe compression group than that in the mild compression group, suggesting the potential application of these parameters in assessing optic nerve damage in patients with NFPA.

Materials and methods

Patients

We conducted a retrospective clinical study involving 57 patients who were first diagnosed with NFPA and received tumor resection by endoscopic sphenoidal sinus surgery at the Affiliated Hospital of Guangdong Medical University from January 2020 to April 2022. All surgeries were done by the same experienced neurosurgeon and reasonable optic apparatus decompression was accomplished following tumor removal. Intraoperatively, the structure of the ultra care was paid to preserve relevant sellar/suprasellar neurovascular structures, and achieve adequate hemostasis to avoid postoperative compressing hematoma. Overfilling with skull base reconstructive materials was avoided to prevent optic apparatus compression. Fifty-three healthy subjects with matching average gender and age as the controls. All subjects underwent examination of best-corrected visual acuity (BCVA) and the visual field, OCT of the optical disk and macular, and DTI of the visual pathway.

The inclusion and exclusion criteria for the NFPA group were as follows. I. PA was indicated by plain MRI and enhanced examination of the brain. II. PA was the first complete resection obtained by endoscopic sphenoidal sinus surgery by the same brain surgeon without additional optic nerve damage and was confirmed by histopathological examination. III. NFPA was clinically diagnosed in patients aged between 18 and 60 years. IV. Non-contact intraocular pressure was $\leq$21 mmHg (1 mmHg = 0.133 kPa). V. There was no previous history of intracranial diseases and trauma, intracranial surgery, ocular trauma, glaucoma, neuroretinal disease, or internal eye surgery. VI. Previous refractive errors were $<\pm$6.0D (spherical mirror) and $<$3.00D (column mirror). VII. The OCT images were clear, and the DTI images were of good quality.

The inclusion and exclusion criteria for the control group were as follows. I. Non-contact intraocular pressure was $\leq$21 mmHg. II. Visual acuity or corrected visual acuity was $\geq$ 0.6, and refractive errors were $<\pm$ 6.0D (spherical mirror) and $<$3.00D (column mirror). III. There was no previous history of intracranial diseases, trauma, or intracranial surgery. IV. There was no history of ocular trauma, glaucoma, neuroretinal diseases, or internal eye surgery. V. The subjects' age and sex-matched those of the NFPA group. VII. The OCT image was clear, and the DTI images were of good quality.

This study was conducted in accordance with the principles of the Helsinki Declaration and was approved by the Ethics Committee of the Affiliated Hospital of Guangdong Medical University (Approval Document No. PJ2020-006 and VJ2020-006-03). All subjects signed informed consent forms.

Visual field examination

The patients with NFPA underwent a visual field examination after a corrected refractive error (KowaAP7000 precision visual field meter, Kowa, Japan) before pituitary tumor resection. The visual field test was the center 30°. If solid vision disappeared and false-negative or false-positive errors exceeded 20%, the test was considered unreliable and hence repeated. Two reliable visual field examinations were performed for each patient, and mean defect (MD) was used to assess the overall visual field defect.

Magnetic resonance examination of the tumor

All patients were subjected to a preoperative head MRI scan plus enhancement (Discovery MR750 3.0T, GE, USA) to measure their tumor sizes. The horizontal diameter line, anteroposterior diameter line, the epitaxial height of the sella turcica diameter line, the vertical diameter line of the tumor, and the thickness of the optic chiasma were recorded (**Figures 1A–D**). The maximum height of the vertical diameter line was measured in the sagittal view and used to determine the anteroposterior diameter line of the tumor. The epitaxial height of the sella turcica diameter line was measured as the height over the horizontal line between two points on the anterior and posterior

tuberosity of the diaphragmatic saddle. The thickest position of the optic chiasma was selected for measurement. The tumor and optic chiasma were distinguished according to the degree of enhancement. Each value was measured three times and averaged. The degree of tumor compression on the optic chiasma was divided into five grades (Figure 1E). Grade 0 was defined as no contact between the tumor and optic chiasma. Grade 1 was defined as a tumor in contact with the optic chiasma but without surface deformation of the optic chiasma. Grade 2 was defined as a tumor in contact with the optic chiasma with surface deformation of the optic chiasma but visible suprachiasmatic cisterns. Grade 3 was defined as a tumor in contact with the optic chiasma, with the superior surface of the optic chiasma malformation and the superior chiasma cisterns invisible but without brain malformation. Grade 4 was defined as brain malformation, in addition to these changes. Four grade 0 patients who met the inclusion and exclusion criteria were excluded from this study because their tumors had no contact with the optic chiasma.

OCT measures the thicknesses of RNFL, macular ganglion cell complex, inner plexus layer, and ganglion cell layer

The optic disk and macular area were scanned using 3D-OCT (Heideberg Engineering Spectralis, Germany). The OCT parameters included CP-RNFL of average thickness and four quadrants (nasal, supra, temporal, and inferior) and ganglion cell layer (GCL) and inner plexus layer (IPL) thicknesses of 1 mm macular center of the four quadrants (nasal, supra, temporal, and inferior) of the inner ring (1–3 mm from the macular fovea) and outer ring (within 3–6 mm from the macular fovea). Each parameter was evaluated automatically using Spectralis mapping software. GCC thickness was superimposed by the thicknesses of the macular RNFL, GCL, and IPL.

DTI examination and image processing

T1WI and DTI scanning were performed using a GE3.0T Optima MR360 imaging system with a head 16-channel phased-array coil. The T1WI scanning parameters were set as an axial 3D BRAVO sequence, 12.3/5.1 ms TR/TE, 256 × 256 matrix, 240 mm × 240 mm FOV, 1.4 mm layer thickness, 0 mm interval, and NEX 1. The DTI scanning parameters were set as a single-excitation DW-SE-EPI sequence, 9000/100.1 ms TR/TE, 128 × 128 matrix, 240 mm × 240 mm FOV, one acquisition, 25 diffusion-sensitive gradient directions, b value = 1000 s/mm^2, layer thickness and layer spacing 2/0 mm, and axial scanning. The scanning results were presented in the form of a color-coded tensor FA graph and an ADC graph (Figure 2). In DTI data processing, these two graphs were set as green in the front and back directions, red in the left and right directions, and blue in the top and bottom directions. The anterior, middle, and posterior of the optic nerve, optic tract, and optic radiation, and the left, middle, and right of the optic chiasma were measured, analyzed, and recorded using the GE3.0 NMR machine software (ADW 4.2 Function Tool). Three regions of interest (ROIs) were selected at the clearest locations on the bilateral optic nerve, optic chiasma, bilateral optic tract, and optic radiation to measure the FA and ADC values. According to the classic neuroanatomical description and relevant literature (Li et al., 2019), the ROI was delineated and measured as an area of 8–12 mm^2. In particular, when measuring an ultra-thin optic chiasm resulting from severe compression, FA and ADC signals were detected at the tumor

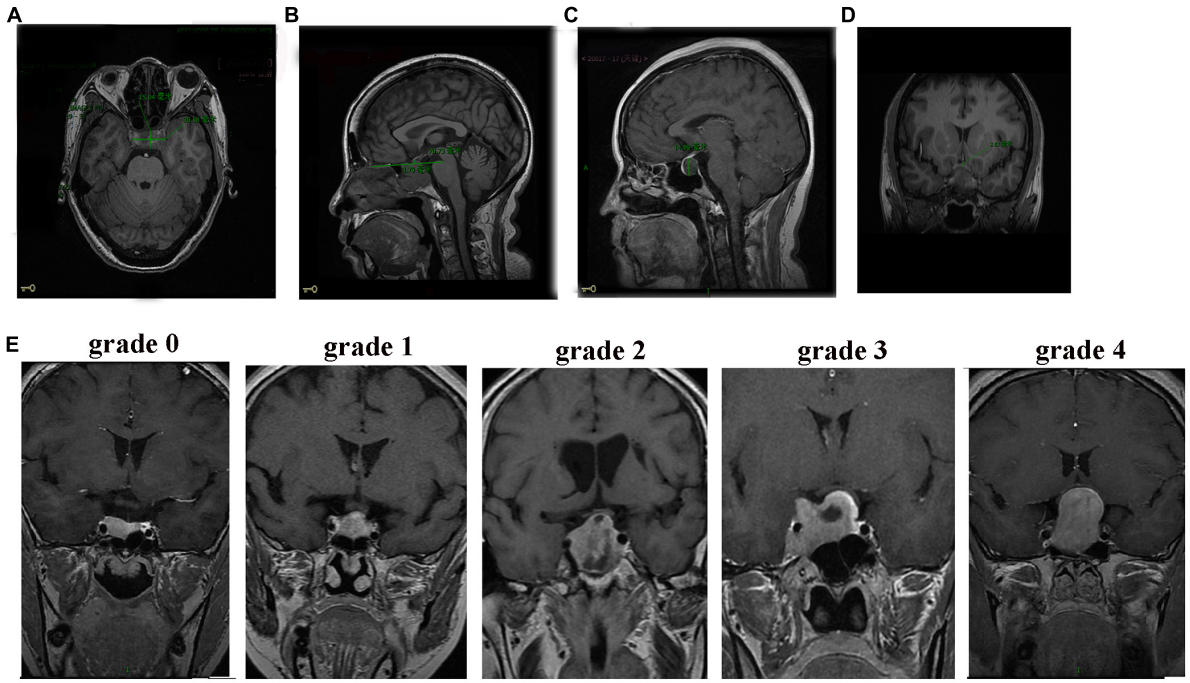


FIGURE 1
MRI images of patients with NFPA. **(A)** Axial position measurement of the horizontal diameter line (30.58 mm) and the anteroposterior diameter line (15.04 mm). **(B)** Sagittal measurements of the epitaxial height of the sella turcica diameter line (3.79 mm). **(C)** Sagittal measurements of the vertical diameter line (14.99 mm). **(D)** Thickness measurement of the optic chiasm. **(E)** Grades (0–4) of tumor compression on the optic chiasm.

10.3389/fnins.2023.1057781

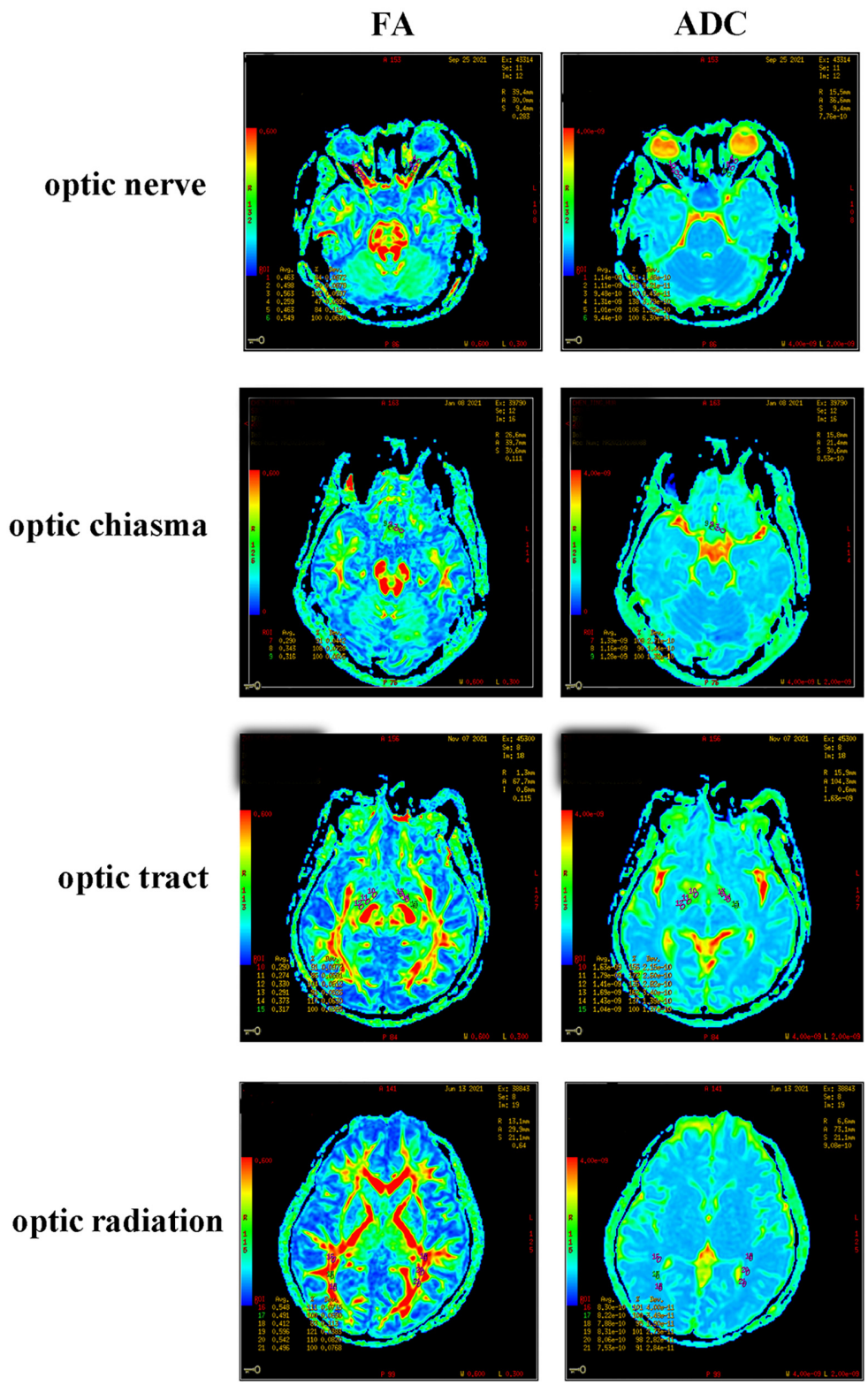


FIGURE 2
FA and ADC diagrams of the optic nerve, optic chiasma, optic tract, and optic radiation.

edge, and the ROI became oval with a size of 8–12 mm^2. To minimize measurement errors, image reconstruction and data measurement were performed by experienced physicians. The FA and ADC values of the optic nerve, optic chiasma, optic tract, and optic radiation were taken as the averages of the three ROIs.

Statistical analysis

IBM SPSS24.0 statistical software was used for the statistical analysis, and the measurement data of normal distribution were expressed as mean ± standard deviation. An independent sample

t-test was used for parameter comparison between groups, and a one-way ANOVA was used for parameter comparison among three groups and multiple groups, followed by a pairwise comparison conducted using the LSD-t test. Spearman correlation was used to evaluate the relationships among the best corrected visual acuity; visual field; RNFL, GCC, GCL, and IPL thicknesses; and the FA and ADC values of the DTI parameters. The area under the receiver operating characteristic (ROC) curve of the FA and ADC values was calculated and compared by conducting the *Z*-test to evaluate the diagnostic ability of the DTI parameters for optic path injury. To reduce the results bias, the statistical analysis of the data was performed by a dedicated researcher who was not selected from a group of radioactive data measurement researchers. A value of $P < 0.05$ was considered statistically significant.

Results

Subject groups and visual field comparison

The 53 patients were classified according to the degree of optic chiasma compression: grade 1 (13 patients), grade 2 (12 patients), grade 3 (11 patients), and grade 4 (17 patients). The average thicknesses of the optic chiasma in these four grades were (2.82 ± 0.98) mm, (2.62 ± 0.85) mm, (1.84 ± 0.73) mm, and (1.39 ± 0.78) mm, respectively, with a significant difference ($F = 9.00$, $P < 0.05$). Among these, there was a statistically significant difference between grades 1 and 2 and grades 3 and 4 but no statistically significant difference between grades 1 and 2 and between grades 3 and 4 ($P = 0.01$ for grade 1 vs. grade 3, $P = 0.00$ for grade 1 vs. grade 4, $P = 0.03$ for grade 2 vs. grade 3, $P = 0.00$ for grade 2 vs. grade 4, $P = 0.55$ for grade 1 vs. grade 2, and $P = 0.18$ for grade 3 vs. grade 4). Therefore, we combined grade 1 and grade 2 to form the mild optic chiasma compression group (case group 1) and grade 3 and grade 4 to form the heavy optic chiasma compression group (case group 2). The tumor body and optic chiasma in the two case groups were compared. The tumor suprasellar epitaxial height, anteroposterior diameter line, horizontal diameter line, and vertical diameter line in case group 2 were greater than those in case group 1, and the optic chiasma in case group 2 was thinner than that in case group 1 (Table 1).

The BCVA values of case group 1, case group 2, and the control group were (0.15 ± 0.19) LogMAR, (0.71 ± 0.61) LogMAR, and (0.03 ± 0.05) LogMAR, respectively, with significant differences ($F = 79.58$, $P = 0.00$). A comparison between the two groups showed $P = 0.00$ for case group 1 vs. case group 2, $P = 0.03$ for case group 1 vs. the control group, and $P = 0.00$ for case group 2 vs. the control group. The MD values of case group 1, case group 2, and the control group were (1.29 ± 3.17) dB, (8.50 ± 7.53) dB, and (0.62 ± 0.92) dB, respectively, with significant differences ($F = 68.80$, $P = 0.00$). A comparison between the two groups showed $P = 0.00$ for case group 1 vs. case group 2, $P = 0.35$ for case group 1 vs. the control group, and $P = 0.00$ for case group 2 vs. the control group.

Comparison of CP-RNFL thickness among case group 1, case group 2, and the control group

The temporal CP-RNFL of case group 1 was significantly thinner than that of the control group ($P = 0.04$), and the CP-RNFL thickness of the other quadrants had no statistical significance between the two groups (*P*-values of average, nasal, supra, and inferior quadrants were 0.65, 0.56, 0.95, and 0.47, respectively) (Table 2). A comparison of CP-RNFL in all quadrants between case group 2 and the control group indicated statistically significant differences in thickness (all four quadrants had a *P*-value of 0.00) (Table 2).

Comparison of the thicknesses of GCC, IPL, and GCL among case group 1, case group 2, and the control group

Compared to the control group, case group 1 had significantly thinner GCC, IPL, and GCL within the macular center of 1 mm, GCC in the inferior aspect of the inner ring, and GCC in the temporal aspect of the outer ring ($P = 0.00$, 0.00, 0.00, 0.00, and 0.01) (Table 3), indicating that mild optic chiasma compression resulted in local thinning of the GCC in the macular area.

Compared to the control group, case group 2 had significantly thinner GCC, IPL, and GCL in the macular center of 1 mm; nasal-, supra-, temporal-, and inferior-macular inner rings; and nasal-, supra-, temporal-, and inferior-macular outer rings (all had *P*-values < 0.05) (Table 3), indicating that the GCC in the macular area was thinner diffusely with increasing compression of the optic chiasma.

TABLE 1 Visual acuity, visual field, and tumor were compared between mild and heavy optic chiasma compression case groups.

Tumor and optic chiasma parameters (mm)	Classification based on optic chiasma compression			*t*	*P*
	Case group 1	Case group 2	Control group		
BCVA (LogMAR)	0.15 ± 0.19	0.71 ± 0.61	0.03 ± 0.05		0.00
MD (dB)	1.29 ± 3.17	8.50 ± 7.53	0.62 ± 0.92		0.00
Epitaxial height of sella turcica diameter line	3.68 ± 2.76	16.69 ± 6.40		−9.39	0.00
Anteroposterior diameter line	19.98 ± 6.43	30.54 ± 7.91		−5.29	0.00
Horizontal diameter line	18.67 ± 6.43	26.57 ± 8.57		−3.78	0.00
Vertical diameter line	19.74 ± 6.10	36.08 ± 7.79		−8.42	0.00
Optic chiasma thickness	2.72 ± 0.91	1.57 ± 0.78	2.95 ± 0.25	4.97	0.00

TABLE 2 Comparison of CP-RNFL thickness between case group and control group.

CP-RNFL thickness (um)	Case group 1 (25 cases, 50 eyes)	Case group 2 (28 cases, 56 eyes)	Control group (53 cases, 106 eyes)	*F*	*P*
Average CP-RNFL	105.50 ± 15.82	81.55 ± 26.55	106.60 ± 11.65	40.27	0.00
Nasal CP-RNFL	73.60 ± 18.34	56.76 ± 26.26	71.72 ± 13.01	14.60	0.00
Supra CP-RNFL	131.68 ± 25.49	100.71 ± 34.56	131.61 ± 20.31	29.22	0.00
Temporal CP-RNFL	78.80 ± 17.30	59.83 ± 24.11	85.02 ± 13.79	36.80	0.00
Inferior CP-RNFL	136.16 ± 29.73	109.94 ± 35.87	139.76 ± 20.74	22.50	0.00

TABLE 3 Comparison of the thickness of GCC, IPL, and GCL between two case groups and control group.

Macular GCC, IPL, and GCL thickness (um)	Case group 1 (50 eyes)	Case group 2 (56 eyes)	Control group (106 eyes)	*F*	*P*
Center (1 mm diameter)	33.46 ± 9.79	34.48 ± 12.88	46.21 ± 10.13	34.05	0.00
	13.50 ± 5.16	15.04 ± 4.70	18.81 ± 4.39	27.88	0.00
	8.84 ± 4.73	8.94 ± 6.02	15.23 ± 4.74	44.53	0.00
Nasal-macular inner ring	105.79 ± 14.19	88.27 ± 22.54	107.93 ± 13.69	26.57	0.00
	39.26 ± 6.36	34.04 ± 9.43	39.81 ± 4.97	13.59	0.00
	46.50 ± 9.14	37.16 ± 13.10	46.74 ± 7.67	20.64	0.00
Supra-macular inner ring	111.89 ± 15.70	95.87 ± 23.57	113.04 ± 10.15	22.06	0.00
	38.91 ± 5.42	35.17 ± 8.70	39.33 ± 4.77	8.35	0.00
	50.17 ± 9.13	40.66 ± 13.25	49.90 ± 5.77	22.93	0.00
Temporal-macular inner ring	97.24 ± 14.52	89.37 ± 19.10	100.85 ± 9.08	12.22	0.00
	37.51 ± 6.20	34.04 ± 7.02	38.50 ± 4.55	10.55	0.00
	42.82 ± 8.90	37.16 ± 10.21	44.13 ± 5.59	15.29	0.00
Inferior-macular inner ring	104.67 ± 16.72	95.50 ± 20.76	111.55 ± 11.78	18.09	0.00
	37.82 ± 5.69	33.60 ± 7.98	38.43 ± 4.27	12.56	0.00
	45.75 ± 9.96	40.35 ± 10.17	48.05 ± 6.34	15.81	0.00
Nasal-macular outer ring	113.08 ± 15.55	99.09 ± 19.26	116.90 ± 10.75	26.45	0.00
	31.10 ± 4.32	26.52 ± 4.96	31.81 ± 3.50	28.66	0.00
	38.86 ± 8.10	32.91 ± 7.54	39.96 ± 5.78	20.49	0.00
Supra-macular outer ring	98.15 ± 14.82	90.55 ± 17.07	102.37 ± 11.02	12.78	0.00
	28.69 ± 3.63	26.69 ± 5.40	29.63 ± 3.32	9.26	0.00
	34.36 ± 7.48	31.66 ± 7.23	35.60 ± 5.20	7.23	0.00
Temporal-macular outer ring	84.08 ± 13.50	83.11 ± 12.48	89.19 ± 8.45	6.77	0.00
	31.76 ± 3.80	30.95 ± 5.37	32.79 ± 3.70	3.55	0.03
	35.37 ± 6.67	33.39 ± 7.59	36.18 ± 5.18	3.78	0.02
Inferior-macular outer ring	98.55 ± 14.28	92.16 ± 15.32	99.45 ± 12.87	5.10	0.01
	27.96 ± 4.08	26.13 ± 5.07	28.41 ± 3.79	4.99	0.00
	34.34 ± 5.93	31.10 ± 6.57	34.28 ± 8.26	4.02	0.02

Comparison of the FA and ADC values among case group 1, case group 2, and the control group

A comparison of case group 1 with the control group showed that the FA values of the optic nerve, optic chiasma, and optic tract in case group 1 decreased significantly (P = 0.00, 0.01, and 0.00), while the ADC values of the optic nerve and optic chiasma increased significantly (P = 0.01 and 0.04) (Table 4). A comparison of case group 2 with the control group showed that the FA values of the optic nerve, optic chiasma, optic tract, and optic radiation in case group 2 decreased significantly (P = 0.00, 0.00, 0.00, and 0.03), while the ADC values of the optic nerve, optic chiasma, and optic radiation increased significantly (P = 0.00, 0.00, and 0.00) (Table 4). There was no statistical difference in the ADC values of the optic tracts in group 1 and group 2 compared to the control group (P = 0.55 and 0.24) (Table 4).

TABLE 4 Comparison of the FA value and ADC value between the two case groups and the control group.

FA value and ADC value ($\times10^{-9}mm^2/s$)	Case group 1 (50 eyes)	Case group 2 (56 eyes)	Control groups (106 eyes)	*F*	*P*
Optic nerve	0.38 ± 0.09	0.35 ± 0.11	0.43 ± 0.10	13.83	0.00
	1.53 ± 0.31	1.64 ± 0.39	1.38 ± 0.27	12.53	0.00
Optic chiasma	0.28 ± 0.08	0.20 ± 0.07	0.32 ± 0.05	30.37	0.00
	1.86 ± 0.51	2.02 ± 0.37	1.66 ± 0.39	5.04	0.01
Optic tract	0.42 ± 0.10	0.34 ± 0.10	0.48 ± 0.09	39.16	0.00
	1.22 ± 0.35	1.25 ± 0.33	1.18 ± 0.31	0.72	0.49
Optic radiation	0.53 ± 0.05	0.49 ± 0.07	0.52 ± 0.05	4.90	0.01
	0.86 ± 0.06	0.88 ± 0.07	0.85 ± 0.06	4.25	0.00

Analysis of the correlation of BCVA and MD values with DTI and OCT parameters

We performed a correlation analysis between the BCVA and MD values and all DTI and OCT parameters and displayed the parameters with the strongest correlation with BCVA and MD values. The BCVAlogMAR and MD values showed the strongest negative correlation with the FA value of the optic chiasma ($r = -0.51$, -0.54, $P = 0.00$), and BCVAlogMAR showed the strongest negative correlation with the superior CP-RNFL thickness ($r = -0.53$, $P = 0.00$) (Table 5). The MD value of the visual field showed the strongest negative correlation with CP-RNFL thickness from the supra and average aspects, respectively (both $r = -0.62$, $P = 0.00$) (Table 5). BCVAlogMAR had the strongest positive correlation with the tumor's suprasellar epitaxial height ($r = 0.44$, $P = 0.00$). The MD value had the strongest positive correlation with the vertical diameter of the tumor ($r = 0.60$, $P = 0.00$) (Table 5). These data suggest that the degree of tumor compression on the vertical diameter line on the optic chiasma indicates the degree of visual field impairment.

Correlation between DTI and OCT parameters and the tumor diameter line

In all patients with NFPA, the correlation of FA and ADC values of the optic nerve, optic chiasma, optic tract, and optic radiation with RNFL thickness, macular IPL, macular GCL, and macular GCC was analyzed, and the strongest correlation was displayed. As shown in Table 6, the strongest positive correlations were observed between the FA value of the optic nerve and average CP-RNFL thickness ($r = 0.58$, $P = 0.00$), between the FA value of the optic chiasma and thickness of IPL and GCC in the inferior-macular inner ring ($r = 0.52$, $P = 0.00$), between the FA value of the optic tract and inferior CP-RNFL thickness ($r = 0.47$, $P = 0.00$), and between the FA value of optic radiation and superior CP-RNFL thickness ($r = 0.46$, $P = 0.00$) (Table 6). These data suggest that the average CP-RNFL thickness, and thickness of inferior-macular inner-ring IPL, GCC, inferior CP-RNFL, and superior CP-RNFL are the best indicators of the impairment of the optic nerve, optic chiasma, optic tract, and optic radiation, respectively.

The strongest negative correlations were observed between the FA value of the optic nerve and optic tract and the tumor vertical diameter line ($r = -0.38$, $P = 0.01$; $r = -0.39$, $P = 0.01$), between the FA value of the optic chiasma and epitaxial height of the tumor suprasellar ($r = -0.62$, $P = 0.00$), and between the FA value of the optic radiation and the tumor horizontal diameter line ($r = -0.43$, $P = 0.00$) (Table 6). These data suggest that the degree of tumor compression on the vertical diameter line to the optic chiasma indicates impaired nerve fiber conduction.

Comparison of the test effectiveness of the FA and ADC values in visual pathway damage

To evaluate the diagnostic ability of the FA and ADC values for damage to the optic nerve, optic chiasma, optic tract, and optic radiation, we calculated the area under the curve (AUC) on the ROC curve of these values. The Z value of the AUC on the ROC curve of the FA and ADC values of the structures of the entire visual pathway showed that the Z optic nerve = 0.71 ($P = 0.48$), Z chiasma = 3.77 ($P = 0.00$), Z optic tract = 4.95 ($P = 0.00$), and Z visual radiation = 1.18 ($P = 0.23$) (Figure 3), indicating that the FA values of the visual chiasma and optic tract were more effective than the ADC values.

Discussion

Retinal RNFL can be evaluated from the changes in two substructures, nasal crossed fibers and temporal uncrossed fibers, which are bounded by the vertical line of the fovea. In our study, patients with NFPA showed early thinning of the temporal CP-RNFL quadrant, the quadrant in which the papillary macular fibrosis was located. According to the literature, the temporal CP-RNFL thickness is correlated with visual function. For example, Kawaguchi et al. (2019) found that the thickness of the temporal CP-RNFL was significantly correlated with visual function recovery, and the thinning of the temporal nerve fiber layer had the greatest impact on vision because the papillomacular fibers converged from the temporal side of the optic disk. Glebauskiene et al. (2018) found that the thickness of the temporal CP-RNFL was positively correlated with the distance between the optic chiasma and the pituitary tumor. However, none of these studies revealed a relationship between changes in CP-RNFL thickness and the degree of chiasmatic compression. In our study, mild optic chiasma compression showed

TABLE 5 Correlation analysis between BCVAlogMAR and MD and DTI parameters and OCT parameters.

Correlation coefficient	BCVAlogMAR	MD
Optic nerve FA		
r	−0.44	−0.50
P	0.00	0.00
Optic chiasma FA		
r	−0.51	−0.54
P	0.00	0.00
Optic tract FA		
r	−0.41	−0.44
P	0.00	0.00
Optic radiation FA		
r	−0.36	−0.41
P	0.01	0.00
CP-RNFL	Supra CP-RNFL	Supra or Average CP-RNFL
r	−0.53	−0.62
P	0.00	0.00
Macular IPL	Inferior outer ring	Supra inner ring
r	−0.45	−0.43
P	0.00	0.00
Macular GCL	Supra inner ring	Supra inner ring
r	−0.37	−0.53
P	0.01	0.00
Macular GCC	Nasal inner ring	Supra inner ring
r	−0.38	−0.51
P	0.01	0.00
PA diameter	Epitaxial height of suprasellar extension	Vertical diameter
r	0.44	0.60
P	0.00	0.00
Optic chiasma thickness		
r	−0.47	−0.46
P	0.00	0.00

thinning of the temporal CP-RNFL, while heavy optic chiasma compression showed thinning of the CP-RNFL in all quadrants, indicating the progression of the CP-RNFL from temporal lesions to diffuse atrophy in patients with NFPA.

Agarwal et al. (2021) suggested that GCL-IPL had a stronger diagnostic ability for optic nerve injury than CP-RNFL. In addition, Zhang et al. (2019) found that the GCL was thinned in patients with early PA without visual field defects. Moon analyzed the OCT multilayer structure of the retinas of patients with PA and found that the thicknesses of both GCL and IPL decreased as compared with the control (Moon and Shin, 2020). Sun et al. (2017) indicated that the GCC thickness in patients with PA presented quadrant characteristics, mainly nasal-macular thinning. Consistent with these studies, we found that the GCC, GCL, and IPL in the macular area of NFPA patients with mild optic chiasma compression were damaged locally. Furthermore, heavy optic chiasma compression caused diffuse thinning of the GCC, IPL, and GCL, indicating that the tumor compresses both crossed and non-crossed nerve fibers with the disease progression. Based on this, we infer that similar to the case of CP-RNFL, thinning of the GCC also starts from a single quadrant and eventually extends to the entire macular region.

The feasibility of OCT indicators in predicting the reversibility of visual impairment remains controversial. Iqbal et al. (2020) found that postoperative visual acuity improved in patients with normal preoperative RNFL thickness but did not in those with preoperative RNFL thinning. An et al. indicated that GCC damage occurred before RNFL thinning (Micieli et al., 2019). However, Lukewich and Micieli (2019) argued that RNFL thinning appeared earlier than GCC thinning. One of the reasons for the controversy is that PA tumors affect the entire visual pathway (from the retina to the visual center of the brain), and the OCT indicators of the retina alone are inadequate for understanding the entire visual pathway impairment. Thus, it is necessary to recognize the objective evaluation effect of DTI parameters on visual impairment. Decreased FA values and increased ADC values arise with the loss of nerve fibers and atrophy of nerve tissue (Lee et al., 2005). Anik et al. (2011) found that a lower preoperative optic nerve FA value and higher mean defect MD value were correlated with poor visual effect at 6 months in patients with PA, and determined DTI parameters to be a good predictor for PA surgery (Anik et al., 2018). Our data showed decreased FA value and increased ADC value in parts of the visual pathway of NFPA patients with mild and heavy optic chiasma compression, indicating that even mild compression of the optic chiasma results in impaired signal transduction. The major contribution of this study was that the decrease in the FA value and the increase in the ADC value were synchronized with quadrant thinning of the CP-RNFL, GCC, GCL, and IPL. This optic nerve appearance indicates that the conduction function and structural abnormalities of nerve fibers are present simultaneously in patients with NFPA. In previous studies, in patients with early PA, CP-RNFL thickening and GCC thinning appeared earlier than visual field defects, while nerve fiber functional impairment appeared later than structural abnormalities (Tieger et al., 2017; Blanch et al., 2018; Monteiro, 2018; Micieli et al., 2019). However, some contrasting findings have also been noted. For example, Lukewich and Micieli (2019) found visual field abnormalities to appear earlier than OCT GCC and CP-RNFL thinning. However, the combination of DTI and OCT results in the present study showed that the functional and structural abnormalities of optic nerve fibers in patients with NFPA appeared simultaneously.

Our exploration of the posterior optic pathway indicated that, as compared to the control group, ADC values did not change in the mild chiasma compression group but increased significantly in the heavy chiasma compression group. However, the ADC values of the optic tract did not change significantly. These results are in agreement with those of John et al., who also found no statistical difference in the MD values of the optic tract in the control group (Rutland et al., 2019). In a previous study, increased MD or ADC values reflected reduced limits on water diffusion due to myelin or axon membrane damage (Lutz et al., 2008). Is there a different mechanism through which the ADC values of the optic tract change? At present, there are a few similar studies, and the specific mechanism needs to be further explored.

We conducted an AUC calculation of the FA and ADC values on the ROC curve of visual pathway injury and found that the area under the ROC curve of the FA value of the optic chiasma

TABLE 6 The strongest correlation between DTI parameters and OCT parameters.

Visual pathway DTI parameters	CP-RNFL thickness	Macular GCC thickness	Macular IPL thickness	Macular GCL thickness	Diameter line of tumor
Optic nerve FA	Average	Center (1 mm diameter)	Center (1 mm diameter)	Center (1 mm diameter)	Vertical diameter line
r	0.58	0.43	0.51	0.43	−0.38
P	0.00*	0.00*	0.00*	0.00*	0.01*
Optic chiasma FA	Average	Inferior-macular inner ring	Inferior-macular inner ring	Inferior-macular inner ring	Epitaxial height of sella turcica diameter line
r	0.44	0.52	0.52	0.48	−0.62
P	0.00*	0.00*	0.00*	0.00*	0.00*
Optic tract FA	Inferior	Nasal-macular outer ring	Nasal-macular outer ring	Center (1 mm diameter)	Vertical diameter line
r	0.47	0.40	0.41	0.40	−0.39
P	0.00*	0.00*	0.00*	0.00*	0.01*
Optic radiation FA	Supra	Temporal-macular outer ring	Nasal-macular outer ring	Temporal-macular outer ring	Horizontal diameter line
r	0.46	0.38	0.45	0.41	−0.43
P	0.00*	0.00*	0.00*	0.00*	0.00*

*$P < 0.001$.

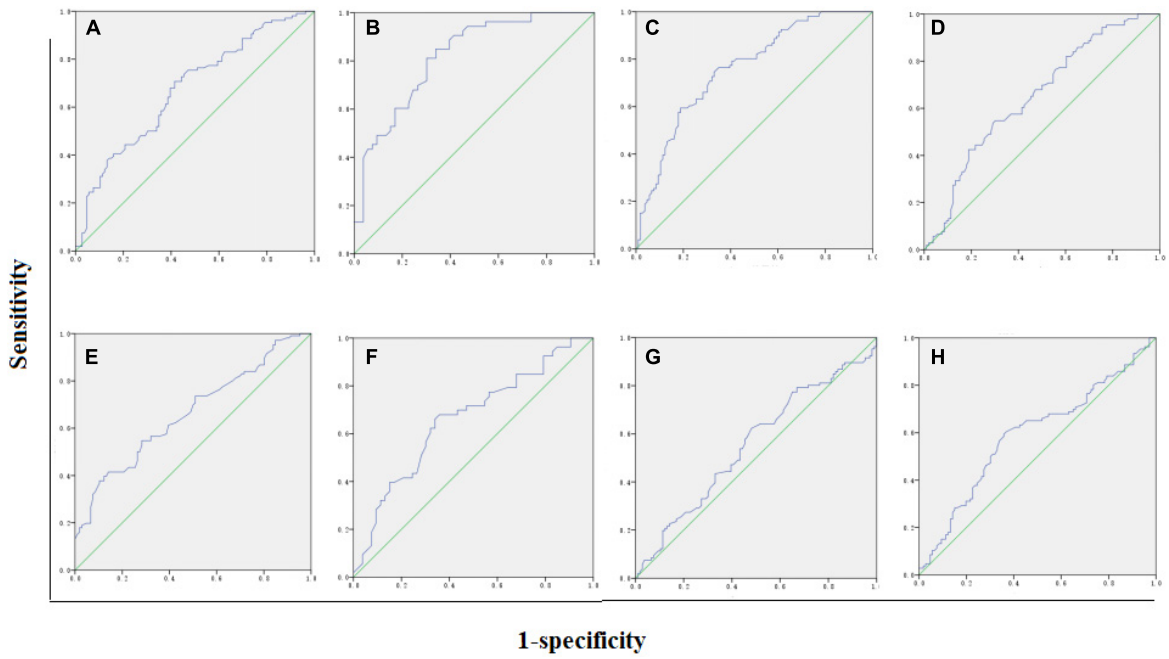


FIGURE 3
AUC of FA and ADC values. **(A)** Optic nerve FA = 0.68. **(B)** Optic chiasma FA = 0.82. **(C)** Optic tract FA = 0.76. **(D)** Optic radiation ADC = 0.65. **(E)** Optic nerve ADC = 0.66. **(F)** Optic chiasma ADC = 0.66. **(G)** Optic tract ADC = 0.55. **(H)** Optic radiation ADC = 0.60.

and the optic tract was larger than that of the ADC value, and the difference was statistically significant. In particular, the area under the ROC curve of the FA value of the optic chiasma was the largest, indicating the best performance for the diagnosis of visual pathway impairment. Although no study has been conducted on the diagnostic ability of DTI parameters for visual pathway injury in patients with NFPA, according to a previous comparison of FA and MD values for visual pathway injury in glaucoma patients, both FA values of the optic nerve and those of optic radiation showed high sensitivity and specificity (Sidek et al., 2014). The reason why the FA value is better than the ADC value may be that the ADC value is affected by complex factors.

This study also has some limitations. First, as the imaging principle of DTI is an echo sequence that is more sensitive to motion, motion artifacts are inevitable. Second, artificial measurement errors are inevitable due to several factors (i.e., the complex distribution of white matter fibers, long visual pathways, and multiple cross fibers, especially in the optic chiasma, and the complex position relationship between PA and optic chiasma). Third, DTI detection is susceptible to interference factors (e.g., skull and gas, hemorrhage artifacts, old brain injury, and primary encephalopathy). All these factors need to be addressed in the future.

In conclusion, the DTI parameter FA and OCT parameters RNFL and macular GCC, GCL, and IPL can be used to evaluate

visual pathway impairment in patients with NFPA. In particular, the FA value of the optic chiasma has a high diagnostic ability for visual pathway impairment. The coupled DTI and OCT can be used to comprehensively understand the microscopic changes in the structure and function of the visual pathway and to more objectively evaluate the visual pathway damage induced by PA. We further explored the predictive ability of this preoperative assessment for postoperative visual recovery.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Affiliated Hospital of Guangdong Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YHP, ZT, and JXZ contributed to the conception and design of the work. XXC, ZHL, XY, and QZ were responsible for the acquisition and analysis of data. BQH, WM, and QSZ interpreted the data. YHP drafted the work. WM substantively revised it. All authors reviewed the manuscript and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

Agarwal, R., Jain, V. K., Singh, S., Charlotte, A., Kanaujia, V., Mishra, P., et al. (2021). Segmented retinal analysis in pituitary adenoma with chiasmal compression: A prospective comparative study. *Indian J. Ophthalmol.* 69, 2378–2384. doi: 10.4103/ijo.IJO_2086_20

Altintas, O., Gumustas, S., Cinik, R., Anik, Y., Ozkan, B., and Karabas, L. (2017). Correlation of the measurements of optical coherence tomography and diffuse tension imaging of optic pathways in amblyopia. *Int. Ophthalmol.* 37, 85–93. doi: 10.1007/s10792-016-0229-0

Alves, C., Batista, S., d'Almeida, O. C., Sousa, L., Cunha, L., Bernardes, R., et al. (2018). The retinal ganglion cell layer predicts normal-appearing white matter tract integrity in multiple sclerosis: A combined diffusion tensor imaging and optical coherence tomography approach. *Hum. Brain Mapp.* 39, 1712–1720. doi: 10.1002/hbm.23946

Anik, I., Anik, Y., Cabuk, B., Caklili, M., Pirhan, D., Ozturk, O., et al. (2018). Visual outcome of an endoscopic endonasal transsphenoidal approach in pituitary macroadenomas: Quantitative assessment with diffusion tensor imaging early and long-term results. *World Neurosurg.* 112, e691–e701. doi: 10.1016/j.wneu.2018.01.134

Anik, I., Anik, Y., Koc, K., Ceylan, S., Genc, H., Altintas, O., et al. (2011). E of early visual recovery in pituitary macroadenomas after endoscopic endonasal transphenoidal surgery: Quantitative assessment with diffusion tensor imaging (DTI). *Acta Neurochir.* 153, 831–842. doi: 10.1007/s00701-011-0942-4

Blanch, R. J., Micieli, J. A., Oyesiku, N. M., Newman, N. J., and Biousse, V. (2018). Optical coherence tomography retinal ganglion cell complex analysis for the detection of early chiasmal compression. *Pituitary* 21, 515–523. doi: 10.1007/s11102-018-0906-2

Chung, Y. S., Na, M., Yoo, J., Kim., W., Jung, I. H., Moon, J. H., et al. (2020). Optical coherent tomography predicts long-term visual outcome of pituitary adenoma surgery: New perspectives from a 5-year follow-up study. *Neurosurgery* 88, 106–112. doi: 10.1093/neuros/nyaa318

Glebauskiene, B., Liutkeviciene, R., Zlatkute, E., Kriauciuniene, L., and Zaliuniene, D. (2018). Association of retinal nerve fibre layer thickness with quantitative magnetic resonance imaging data of the optic chiasm in pituitary adenoma patients. *J. Clin. Neurosci.* 50, 1–6. doi: 10.1016/j.jocn.2018.01.005

Hubers, A., Muller, H. P., Dreyhaupt, J., Bohm, K., Lauda, F., Tumani, H., et al. (2016). Retinal involvement in amyotrophic lateral sclerosis: A study with optical coherence tomography and diffusion tensor imaging. *J. Neural Transm.* 123, 281–287. doi: 10.1007/s00702-015-1483-4

Iqbal, M., Irfan, S., Goyal, J. L., Singh, D., Singh, H., and Dutta, G. (2020). An analysis of retinal nerve fiber layer thickness before and after pituitary adenoma surgery and its correlation with visual acuity. *Neurol. India* 68, 346–351. doi: 10.4103/0028-3886.280634

Kawaguchi, T., Ogawa, Y., and Tominaga, T. (2019). Retinal nerve fiber layer thickness measurement for predicting visual outcome after transsphenoidal surgery: Optic disc atrophy is not the deciding indicator. *World Neurosurg.* 127, e427–e435. doi: 10.1016/j.wneu.2019.03.143

Lee, S. K., Kim, D. I., Kim, J., Kim, D. J., Kim, H. D., Kim, D. S., et al. (2005). Diffusion-tensor MR imaging and fiber tractography: A new method of describing aberrant fiber connections in developmental CNS anomalies. *Radiographics* 25, 53–65; discussion66–8. doi: 10.1148/rg.251045085

Li, M., Ke, M., Song, Y., Mu, K., Zhang, H., and Chen, Z. (2019). Diagnostic utility of central damage determination in glaucoma by magnetic resonance imaging: An observational study. *Exp. Ther. Med.* 17, 1891–1895. doi: 10.3892/etm.2018.7134

Lukewich, M. K., and Micieli, J. A. (2019). Chronic chiasmal compression and persistent visual field defect without detectable changes in optical coherence tomography of the macular ganglion cell complex. *Am. J. Ophthalmol. Case Rep.* 16:100533. doi: 10.1016/j.ajoc.2019.100533

Lutz, J., Jager, L., de Quervain, D., Krauseneck, T., Padberg, F., Wichnalek, M., et al. (2008). White and gray matter abnormalities in the brain of patients with fibromyalgia: A diffusion-tensor and volumetric imaging study. *Arthritis Rheum.* 58, 3960–3969. doi: 10.1002/art.24070

Melmed, S., Kaiser, U. B., Lopes, M. B., Bertherat, J., Syro, L. V., Raverot, G., et al. (2022). Clinical biology of the pituitary adenoma. *Endocr. Rev.* 43, 1003–1037. doi: 10.1210/endrev/bnac010

Micieli, J. A., Newman, N. J., and Biousse, V. (2019). The role of optical coherence tomography in the evaluation of compressive optic neuropathies. *Curr. Opin. Neurol.* 32, 115–123.

Monteiro, M. L. R. (2018). Macular ganglion cell complex reduction preceding visual field loss in a patient with chiasmal compression with a 21-month follow-Up. *J. Neuroophthalmol.* 38, 124–127. doi: 10.1097/WNO.0000000000000625

Moon, J. S., and Shin, S. Y. (2020). Segmented retinal layer analysis of chiasmal compressive optic neuropathy in pituitary adenoma patients. *Graefes. Arch. Clin. Exp. Ophthalmol.* 258, 419–425. doi: 10.1007/s00417-019-04560-3

Pang, Y., Tan, Z., Mo, W., Chen, X., Wei, J., Guo, Q., et al. (2022). A pilot study of combined optical coherence tomography and diffusion tensor imaging method for evaluating microstructural change in the visual pathway of pituitary adenoma patients. *BMC Ophthalmol.* 22:115. doi: 10.1186/s12886-022-02320-2

Rutland, J. W., Padormo, F., Yim, C. K., Yao, A., Arrighi-Allisan, A., Huang, K. H., et al. (2019). Quantitative assessment of secondary white matter injury in the visual pathway

by pituitary adenomas: A multimodal study at 7-Tesla MRI. *J. Neurosurg.* 132, 333–342. doi: 10.3171/2018.9.JNS182022

Santorini, M., Ferreira De Moura, T., Barraud, S., Litre, C. F., Brugniart, C., Denoyer, A., et al. (2022). Comparative evaluation of two SD-OCT macular parameters (GCC, GCL) and RNFL in chiasmal compression. *Eye Brain* 14, 35–48. doi: 10.2147/EB.S337333

Sidek, S., Ramli, N., Rahmat, K., Ramli, N. M., Abdulrahman, F., and Tan, L. K. (2014). Glaucoma severity affects diffusion tensor imaging (DTI) parameters of the optic nerve and optic radiation. *Eur. J. Radiol.* 83, 1437–1441. doi: 10.1016/j.ejrad.2014.05.014

Subramanian, V., Lee, R. S. M., Howell, S., Gregson, S., Lahart, I. M., Kaushal, K., et al. (2021). Non-functioning pituitary macroadenomas: Factors affecting postoperative recurrence, and pre- and post-surgical endocrine and visual function. *Endocrine* 73, 407–415. doi: 10.1007/s12020-021-02713-1

Sun, M., Zhang, Z., Ma, C., Chen, S., and Chen, X. (2017). Quantitative analysis of retinal layers on three-dimensional spectral-domain optical coherence tomography for pituitary adenoma. *PLoS One* 12:e0179532. doi: 10.1371/journal.pone.0179532

Tieger, M. G., Hedges, T. R. III, Ho, J., Erlich-Malona, N. K., Vuong, L. N., Athappilly, G. K., et al. (2017). Ganglion cell complex loss in chiasmal compression by brain tumors. *J. Neuroophthalmol.* 37, 7–12. doi: 10.1097/WNO.0000000000000424

Yamada, H., Yamamoto, A., Okada, T., Kanagaki, M., Fushimi, Y., Porter, D. A., et al. (2016). Diffusion tensor imaging of the optic chiasm in patients with intra- or parasellar tumor using readout-segmented echo-planar. *Magn. Reson. Imaging* 34, 654–661. doi: 10.1016/j.mri.2016.01.002

Zhang, X., Ma, J., Wang, Y. H., Gan, L. Y., Li, L., Wang, X. Q., et al. (2019). [The correlation of ganglion cell layer thickness with visual field defect in non-functional pituitary adenoma with chiasm compression]. *Zhonghua Yan Ke Za Zhi* 55, 186–194. doi: 10.3760/cma.j.issn.0412-4081.2019.03.007

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 20 February 2023
DOI 10.3389/fnins.2023.1144421

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China
Qi-Chen Yang,
Sichuan University, China
Kangcheng Liu,
Affiliated Eye Hospital of Nanchang University, China
Jing Zou,
Xiangya Hospital, Central South University, China

*CORRESPONDENCE
Xiao-Dong Zhou
✉ xdzhou2013@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 14 January 2023
ACCEPTED 02 February 2023
PUBLISHED 20 February 2023

CITATION
Shi W-Q, Wan T, Li B, Li T and Zhou X-D (2023) EFEMP1 is a potential biomarker of choroid thickness change in myopia. *Front. Neurosci.* 17:1144421. doi: 10.3389/fnins.2023.1144421

EFEMP1 is a potential biomarker of choroid thickness change in myopia

Wen-Qing Shi[1†], Ting Wan[1†], Bing Li[2], Tao Li[1] and Xiao-Dong Zhou[1*]

[1]Department of Ophthalmology, Jinshan Hospital of Fudan University, Shanghai, China, [2]Central Laboratory, Jinshan Hospital of Fudan University, Shanghai, China

Purpose: To explore the possible molecular mechanism by which epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) regulates choroid thickness (CT) in the development of myopia.

Methods: In total, 131 subjects were divided into the emmetropia (EM) group, non-high myopia (non-HM) group and high myopia (HM) group. Their age, refraction, intraocular pressure, and other ocular biometric parameters were collected. A 6 × 6 mm area centered on the optic disc was scanned by coherent optical tomography angiography (OCTA) to measure CT, and the tear concentrations of EFEMP1 were quantified using enzyme-linked immunosorbent assay (ELISA) analysis. Twenty-two guinea pigs were divided into the control group and the form-deprivation myopia (FDM) group. The right eye of the guinea pig in the FDM group was covered for 4 weeks, and the diopter and axial length of the right eye of the guinea pig were measured before and after the treatment. After the measurement, the guinea pig was euthanized, and the eyeball was removed. Quantitative reverse transcription polymerase chain reaction, western blotting assays and immunohistochemistry were used to assess the expression of EFEMP1 in the choroid.

Results: There were significant differences in CT among the three groups ($p < 0.001$). CT was positively correlated with age in HM ($r = -0.3613$, $p = 0.0021$), but no significant correlation with SE ($p > 0.05$) was observed. Furthermore, there were increased levels of EFEMP1 in the tears of myopic patients. After 4 weeks of covering the right eye of the FDM guinea pigs, there was a significant increase in axial length and a decrease in diopter ($p < 0.05$). The mRNA and protein expression of EFEMP1 was significantly increased in the choroid.

Conclusion: Choroidal thickness was significantly thinner in myopic patients, and the expression level of EFEMP1 in the choroid increased during the development of FDM. Therefore, EFEMP1 may be involved in the regulation of choroidal thickness in myopia patients.

KEYWORDS

myopia, choroid thickness, EFEMP1, OCT, FDM

Introduction

Myopia has become a growing global threat to public health in recent years (Baird et al., 2020). In Asia, the prevalence of myopia is over 80%. Myopia is a major risk factor for ocular diseases, such as cataracts, glaucoma, and choroidal neovascularization, and it is also associated with similar levels of risk for hypertension, coronary heart disease, and stroke (Baumgarten et al., 2018). According to a study by Naidoo et al. (2019), the potential global economic loss associated with myopia in 2015 was nearly $250 billion. Hence, myopia is a medical problem as well as a heavy social problem that has received increasing attention.

Studies have shown that the choroid plays an important role in regulating ocular development. Blood vessels are a major component of the choroid, and studies have shown that choroidal vessel density decreases with age in healthy people over 30 years old (Fujiwara et al., 2016), while changes in choroidal vessel lumen area can directly affect choroidal thickness (Li et al., 2018). Changes in choroidal thickness (CT) affect the diffusion dynamics between the retina and the sclera, mediating the effect of retinal signals on the sclera and thus regulating scleral growth and refractive state of the eye (Summers, 2013). Numerous previous studies have found that the choroid thickness in myopia differs significantly from emmetropia (Tian et al., 2021; Xiuyan et al., 2021). Nishida et al. (2012) found a negative correlation between CT and age and diopter in patients with high myopia, which is an important predictor of visual acuity. Similarly, consistent findings were obtained in animal models of myopia, where CT was found to be significantly thinner in both chicks and guinea pigs after myopia modeling (Fitzgerald et al., 2002; Lu et al., 2009; Zhang et al., 2019). However, the mechanism leading to CT changes is not clear.

Epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) is a 55 kDa disulfide-bonded secreted extracellular matrix glycoprotein widely expressed in epithelial and endothelial cells (Cheng et al., 2020). Numerous studies have found that EFEMP1 is highly expressed in elastin-rich tissues of human and mouse eyes, especially in the corneal and choroidal retinal pigment epithelium (Livingstone et al., 2020). EFEMP1 is typically characterized by a structural domain rich in epidermal growth factor (EGF), while EGF has been proven to be highly associated with myopia (Jonas et al., 2021). Meanwhile, Daniel et al. (2020) observed structural deformities in the cornea by altering the expression of EFEMP1. The cornea is part of the refractive system. These studies suggest that EFEMP1 may play a role in the development of myopia. In addition, Cheng et al. (2020) found that EFEMP1-overexpressing HUVECs showed a significant increase in tube formation and proliferation. Since CT depends on its perfusion rate, we speculate that EFEMP1 may act as a signaling molecule to regulate the alteration of CT in the development of myopia.

To date, whether the protein levels of EFEMP1 in the choroid change during the development of form-deprivation myopia (FDM) has not been reported. Therefore, we scanned the choroid of myopic subjects using OCTA to analyze the changes in thickness, and EFEMP1 levels in the participants' tears were measured and subsequently validated in the choroidal tissue of FDM guinea pigs to investigate the possible molecular mechanisms involved in the regulation of CT changes by EFEMP1 in myopic patients.

Materials and methods

Study subjects

One hundred thirty-one subjects (131 eyes) were recruited at Jinshan Hospital of Fudan University from October 2019 to October 2021. The inclusion criteria of all subjects were: (1) age range from 18 to 70 years old, (2) spherical equivalence (SE) $\leq$ + 0.5 D, (3) intraocular pressure $\leq$21 mmHg, (4) no ocular lesions, and (5) right-handedness. The exclusion criteria were: (1) smokers and alcoholics, (2) eye diseases, (3) systemic diseases, such as diabetes and hypertension, (4) pregnant or breastfeeding patients, (5) long-term chronic treatment, and (6) history of ophthalmic surgery. All subjects were divided into three groups: (1) Emmetropia (EM) group: SE ranged from −0.5 to +0.5 (D); (2) Non-high myopia (Non-HM) group: SE $>$ −6.0 D; (3) High myopia (HM) group: SE $\leq$ −6.0 D. All examinations were performed with the patient's informed consent.

Clinical examinations

All adult subjects underwent ocular examination, including diopter (CT-5000; Topcan Corporation, Tokyo, Japan), best corrected visual acuity (BCVA) in angle of minimum resolution (LogMAR), intraocular pressure measurement (CT- 80; Canon Inc., Tokyo, Japan), slit lamp examination (SL-F7; Topcan Corporation, Tokyo, Japan), and fundus photography (VISUCAM 200; Carl Zeiss, Jena, Germany). Eyeball biological parameters were measured in 131 adult subjects using A-ultrasound (Aviso; Quantel, France) and IOL-Master 500 (Carl Zeiss, Jena, Germany), respectively.

Image acquisition and analysis

Fundus scans of the optic disc area were performed on all subjects using a 6 mm × 6 mm 3D Disc mode in Triton OCTA (Topcan, Tokyo, Japan) (Liu et al., 2020). The choroid was automatically stratified using OCTA software, and the measurement partitions were divided into four quadrants (S, Superior; I, Inperior; N, Nasal; T, Temporal) according to the diagonal of the two quadrants by the Early Treatment Diabetic Retinopathy Study (ETDRS) (**Figure 1**). The exclusion criteria for OCTA examination were signal intensity index <40 and images with severe artifacts due to poor eye fixation.

Tear collection and preservation

A 0.5 mm diameter polyethylene capillary was placed in the conjunctival sac of the lower dome, and 5–20 μL of tears were collected from both eyes by siphoning into Eppendorf tubes and stored at −80°C for examination.

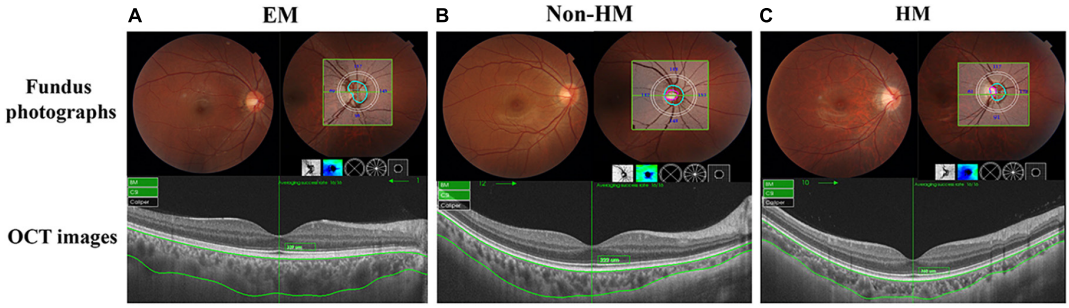


FIGURE 1
Representative fundus photographs, optical coherence tomography (OCT) images, and thickness graph of EM **(A)**, non-HM **(B)**, and HM **(C)** groups. EM, emmetropia; HM, high myopia.

Enzyme-linked immunosorbent assay

A human EFEMP1 ELISA kit (ZN2177; Xi'an Baiolaib Biotechnology Co., Ltd) was used for the determination of EFEMP1 content in tear fluid. One hundred microliters of tear fluid per tube was used for enzyme-linked immunosorbent assay (ELISA) analysis. EFEMP1 content in tear fluid was measured after preparation of 3.125–100 ng/ml standards (Supplementary Figure 1).

Animal experiments

Two-week-old, 100–150 g tricolor male guinea pigs were purchased from Danyang Changyi Laboratory Animal Breeding Co (Jiangsu, China). All animals are kept in natural light conditions to ensure an adequate daily supply of food, water, and fresh vegetables. All animal experiments were approved by Shanghai Public Health Clinical Center Laboratory Animal Welfare & Ethics Committee and in accordance with the ARVO Statement on the Use of Animals in Ophthalmic and Vision Research.

Animal grouping and model preparation

Twenty-two guinea pigs were randomly divided into the FDM group and the control group. For the FDM group, a translucent latex balloon mask was used to cover the right eye, and all other parts of the face were exposed. No treatment was used for the control group. All animals were kept in a natural light environment.

Measurement of ocular biometric features

All guinea pigs were labeled and numbered, and SE was measured in a dark room with streak retinoscopy (YZ24; Six-Six Vision Technology Co., Ltd). Measurements were averaged three times and were accurate to 0.01 D. The conjunctival sac was filled with 1% tropicamide drops three times, each time at 5 min intervals, and the axial length of the eye was measured using A-scan ultrasound (OD1-A, Kaixin Electronic Instrument Co., Ltd., China). Manual measurements were averaged 10 times and were accurate to 0.01 mm. Fundus photography was used to take pictures of the fundus of the right eye of each guinea pig after anesthesia (VISUCAM 200; Carl Zeiss, Jena, Germany).

TABLE 1 The primer sequences.

Primer	Forward 5′-3′	Reverse 5′-3′
EFEMP1	GGACGCACAACTGTAGAGCAGAC	CTTTGGTGGCAATATGGAGGGATGG
β-actin	CTGGGTATGGAATCCTGTGGCATC	CTGTGTTGGCATAGAGGTCCTTACG

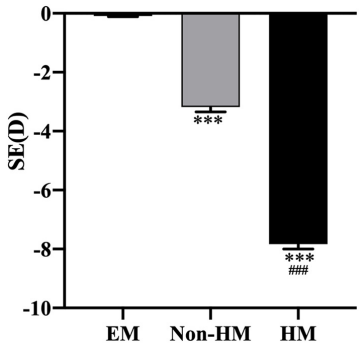


FIGURE 2
Spherical equivalent in the right eyes of the three groups. $^{***}p < 0.001$, compared with EM group; $^{###}p < 0.001$, compared with non-HM group. SE, spherical equivalent; EM, emmetropia; HM, high myopia.

Tissue preparation

The animals were euthanized by overdose with an intraperitoneal injection of sodium pentobarbital. The right eyeball was removed, and a portion of the eyeball was separated from the choroidal tissue, quickly placed in a lyophilization tube and stored at −80°C for western blot experiments. The remaining eyeball was fixed by perfusion and paraffin-embedded for immunohistochemical experiments.

TABLE 2 Demographics and clinical measurements by groups.

Parameter	EM	Non-HM	HM	*p*-value	*p*-value$_1$	*p*-value$_2$	*p*-value$_3$
Sample Size	15	46	70	NA	NA	NA	NA
Age(years)	33.67 ± 12.12	36.39 ± 12.29	33.90 ± 12.09	0.721	1.000	1.000	1.000
SEX, F/M	7/8	16/30	28/42	0.698	–	–	–
SE(Diopter)	−0.15 ± 0.25	−3.31 ± 1.46	−7.70 ± 1.36	< 0.001	< 0.001	< 0.001	< 0.001
BCVA, logMAR	−0.01 ± 0.05	0.00 ± 0.02	0.05 ± 0.12	0.002	0.573	0.002	0.002
IOP(mmHg)	15.40 ± 2.53	14.96 ± 1.71	14.83 ± 2.63	0.737	0.899	0.824	0.985

p-value among the three groups; *p*-value$_1$, *p*-value EM and non-HM; *p*-value$_2$, *p*-value between EM and HM; *p*-value$_3$, *p*-value between non-HM and HM. EM, emmetropia; HM, high myopia; NA, not applicable; SE, spherical equivalent.

Quantitative reverse transcription polymerase chain reaction

Extraction of total RNA from choroidal tissues using Tissue RNA Purification Kit (Yishan Biotechnology, China). The gene transcription was quantified by quantitative RT-PCR with PrimeScript^{TM} RT Master Mix Kit (Takara, Shiga, Japan). The sequence of the primers are shown in Table 1 (Sangon Biotechnology, China).

Western blotting

Frozen choroidal tissues were added proportionally (10 mg) to 100 μL of radioimmunoprecipitation assay buffer (Beyotime, China), and 1 mm phenylmethanesulfonyl fluoride (PMSF, Beyotime, China) was mixed and then homogenized on ice using a tissue homogenizer. Protein concentrations were then determined using a BCA protein assay kit (Beyotime, China). Protein samples (25 μg) were separated by 10% SDS–PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane. After blocking with 5% skim milk for 1 h at room temperature, the membrane was incubated with EFEMP1 antibody (1:1000, ab106429, Abcam, Cambridge, MA, USA) at 4°C overnight and then incubated with species-specific HRP-conjugated secondary antibodies (diluted 1:5000, CoWin Biosciences, Cambridge, MA, USA) for 1 h at room temperature. Then, radioautography-enhanced chemiluminescence (ECL, Thermo Scientific) was performed. β-actin (1:5000, 69,009-1, ProteinTech, Chicago, IL, USA) was used as an internal standard. The grayscale values of each band were calculated and statistically analyzed.

Immunohistochemistry

Paraffin-embedded sections containing eyeball tissue were deparaffinized in a 60°C oven for 50 min and then rehydrated in xylene and graded alcohol solutions. First, tissue antigens were extracted in a rice cooker at 95°C containing citrate antigen retrieval solution, and tissue sections were then incubated in 3% H_2O_2 solution for 15 min at room temperature to inactivate endogenous peroxidase. Nonspecific binding was blocked with goat serum for 10 min at room temperature. Slides were incubated overnight at 4°C with EFEMP1 primary antibody (ab106429, Abcam, Cambridge, MA, USA) diluted to 1:50. After 18 h, the slides were washed with PBS (0.01 M; pH 7.4) for 5 min, which was repeated three times. After incubating with secondary antibody for 30 min, a 3,3′-diaminobenzidine-dine chromogen kit was used for staining, followed by staining with hematoxylin for 2.5 min at room temperature, rinsing the sample with tap water, using dilute hydrochloric acid to acidify the hematoxylin and allowing the slide to dry. Then, they were sealed with neutral gum. PBS was used as a negative control instead of primary antibody.

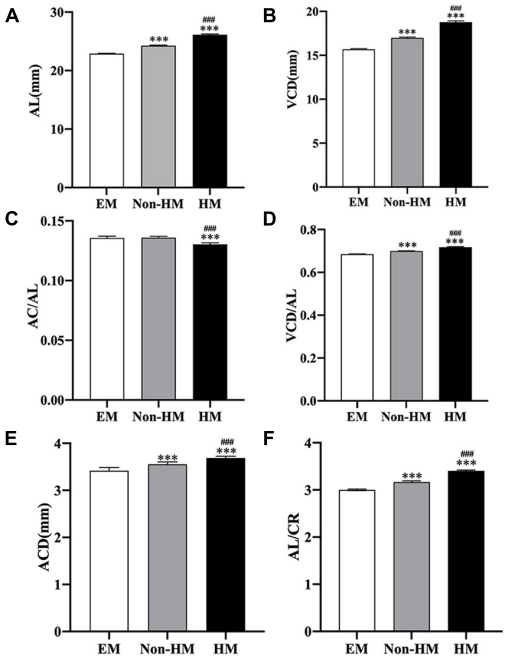


FIGURE 3
Ocular biometric parameters of the right eye measured by A-scan and IOL-Master. **(A)** AL in three groups. **(B)** VCD in three groups. **(C)** AC/AL in three groups. **(D)** VCD/AL in three groups. **(E)** ACD in three groups. **(F)** AL/CR in three groups. $^{***}p < 0.001$, compared with EM group; $^{###}p < 0.001$, compared with non-HM group. EM, emmetropia; HM, high myopia; AL, axial length; VCD, vitreous chamber depth; AC, anterior chamber depth; ACD, anterior chamber depth; CR, radius of corneal curvature.

Statistical analysis

All data were analyzed using the statistical software SPSS (version 25, SPSS Inc., Chicago, IL, USA). Data are presented

TABLE 3 Choroidal parameters in the participant's right eye.

Parameter	EM	Non-HM	HM	*p*-value	*p*-value$_1$	*p*-value$_2$	*p*-value$_3$
Sample size	15	46	70	NA	NA	NA	NA
CT (μm)	270.87 ± 62.01	233.76 ± 80.94	195.91 ± 63.23	< 0.001	0.029	0.001	0.015
CTs (μm)	163.00 ± 58.96	165.33 ± 60.58	143.57 ± 46.52	0.079	1.000	0.606	0.100
CTi (μm)	131.87 ± 47.00	125.43 ± 52.86	99.67 ± 34.91	0.005	1.000	0.03	0.006
CTt (μm)	150.33 ± 63.65	125.35 ± 52.90	104.19 ± 40.33	0.009	0.247	0.003	0.065
CTn (μm)	165.87 ± 62.35	154.50 ± 59.18	146.40 ± 45.19	0.383	0.903	0.607	0.817

p-value among the three groups; *p*-value$_1$, *p*-value EM and non-HM; *p*-value$_2$, *p*-value between EM and HM; *p*-value$_3$, *p*-value between non-HM and HM. EM, emmetropia; HM, high myopia; NA, not applicable; CT, choroid thickness; s, superior; i, inferior; t, temporal; n, nasal.

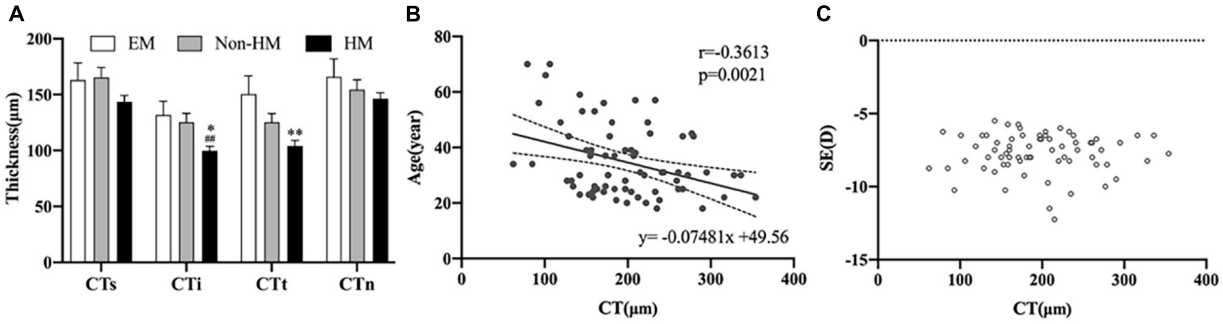


FIGURE 4
Choroidal thickness of right eye. **(A)** Thickness in three groups. **(B)** Scatter plot of correlation between age and CT. **(C)** Scatter plot of correlation between SE and CT. $^{**}p < 0.01$, compared with EM group; $^{\#\#}p < 0.01$, compared with non-HM group. EM, emmetropia; HM, high myopia; CT, choroid thickness; SE, spherical equivalence; s, superior; i, inferior; t, temporal; n, nasal. $^{*}p < 0.05$.

as the means ± SEM. Comparisons between groups were performed using independent *t* test, one-way ANOVA, Welch, and Brown-Forsythe test followed by Bonferroni and Tamhini multiple comparison tests. A *p*-value < 0.05 was considered statistically significant.

Results

General information and ocular parameters

In total, 131 subjects (51 females and 80 males) were collected in this study and divided into EM (−0.5 D ≤ SE ≤ + 0.5 D), non-HM (−6 D < SE < −0.5 D), and HM groups (SE ≤ −6 D); the mean SE of the three groups were −0.15 ± 0.25 D, −3.31 ± 1.46 D, and −7.70 ± 1.36 D, respectively. There was a significant difference in SE among the three groups ($p < 0.001$) (Figure 2 and Table 2). The demographic characteristics and clinical data of the subjects in the three groups are shown in Figures 3 and Supplementary Tables 1, 2.

CT in different choroid regions and correlation between CT, SE, and age

The mean choroidal thicknesses of the EM group, non-HM group, and HM group were 270.87 ± 62.01 μm, 233.76 ± 80.94 μm, and 195.91 ± 63.23 μm, respectively, with significant differences between the three groups ($p < 0.001$). The segmentation of choroidal regions using ETDRS showed that the thickness of CTi and CTt in the EM group was significantly thinner than that in the HM group ($p < 0.05$), and the thickness of CTi in the HM group was significantly thinner than that in the non-HM group ($p < 0.01$) (Table 3). In the high myopia group, there was a significant positive correlation between CT and age ($r = -0.3613$, $p = 0.0021$), while there was no significant correlation with SE ($p > 0.05$) (Figure 4). Signal intensity index of OCTA examination was >40 in all subjects.

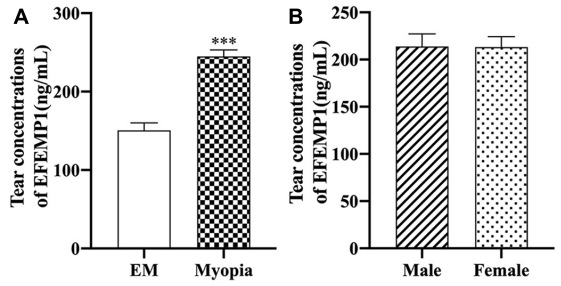


FIGURE 5
EFEMP1 concentrations of tear between different groups. **(A)** Tear EFEMP1 concentration in emmetropia group and myopic group. **(B)** EFEMP1 concentration of tears in male and female groups. $^{***}p < 0.001$, compared with EM group. EM, emmetropia.

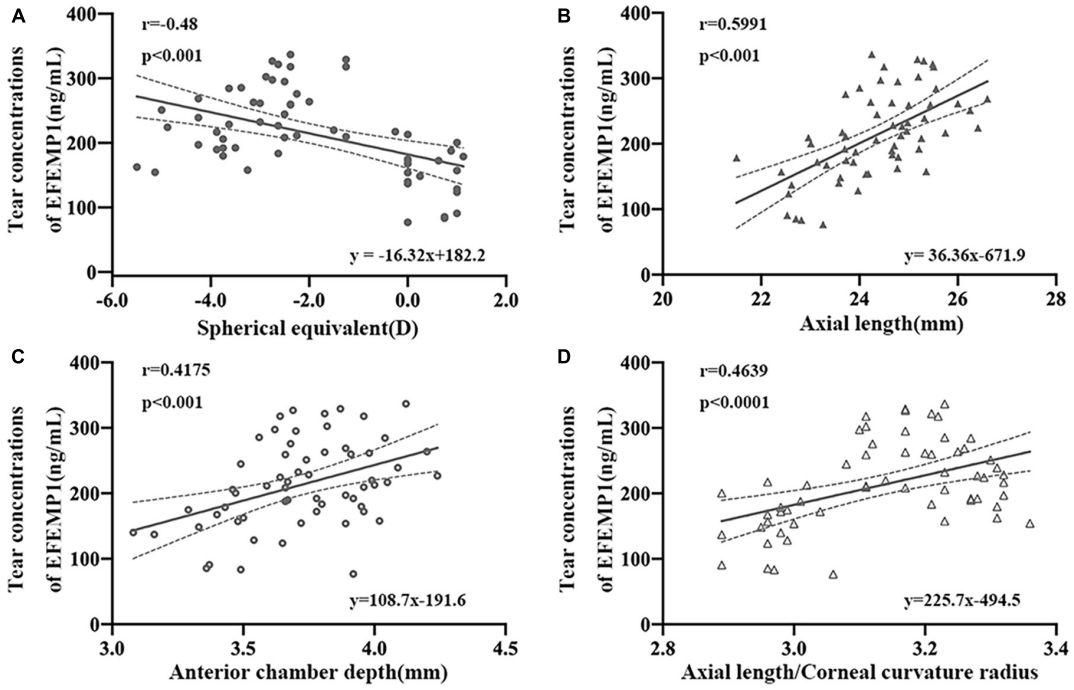


FIGURE 6
Correlation between tear concentrations of EFEMP1 and ocular biometric parameters in two groups. **(A)** Spherical equivalence. **(B)** Axial length. **(C)** Anterior chamber depth. **(D)** Axial length/Corneal curvature radius. ●, ○, △, ▲: scatter plot.

Tear concentrations of EFEMP1 and correlations between EFEMP1 levels and ocular biometric parameters

The average levels of EFEMP1 tears in the myopia group and the emmetropia group were 150.53 ± 42.37 ng/ml and 244.89 ± 51.35 ng/ml, respectively, with a significant difference between the two groups ($p < 0.001$) (Figure 5). Tear concentrations of EFEMP1 in the myopia group were negatively correlated with SE ($r = -0.4800$, $p < 0.001$), and AL ($r = 0.5991$, $p < 0.001$), ACD ($r = 0.4175$, $p < 0.001$), and AL/CR ($r = 0.4639$, $p < 0.001$) were positively correlated (Figure 6).

Diopter and axial length and expression of EFEMP1 mRNA and protein in guinea pigs

After 4 weeks of covering, the FDM group showed a significant decrease in diopter ($p < 0.001$), a gradual increase in axial length (Figure 7), and a significant increase in EFEMP1 mRNA and protein expression in the choroid compared with the control group (all $p < 0.001$) (Figure 8).

EFEMP1 immunohistochemical staining

As shown in Figure 9, EFEMP1 was mainly localized in the ganglion cell layer, inner plexiform layer, inner nuclear layer, RPE layer, scleral extracellular matrix, choroidal vessel wall, and other structures.

Discussion

The choroid is a highly vascularized tissue structure located between the retina and the sclera. In addition to providing oxygen and nutrients to the outer retina, it also mediates visual signals to regulate refractive development and plays an important role in the process of emmetropia or myopia progression (Nickla and Wallman, 2010). Histological studies have suggested that choroidal thickness is closely related to changes in choroidal blood flow (CBF) (Fitzgerald et al., 2002), and numerous studies have been conducted to measure changes in CBF between myopia and emmetropia groups, yet the results have been inconsistent. Some studies have concluded insignificant changes in CBF between myopia and emmetropia (Milani et al., 2018), but others have shown a significant decrease in CBF in myopic patients (Wu et al., 2021). Furthermore, Al-Sheikh et al. (2017) found that blood flow density in the retinal layer was reduced in highly myopic eyes compared to controls, while blood perfusion in the choroidal capillary layer was increased. The inconsistency of the results of the above studies may be related to various factors, such as the observation site and the sensitivity of the instrument. As an objective indicator of choroidal changes, CT provides a clearer indication of the role it may play in myopia development than CBF (Ohno-Matsui and Jonas, 2019). An abundance of studies has found that CT decreases with increasing axial length, especially in patients with high myopia (Duan et al., 2019). In the present study, we found significant differences in CT values among the three groups of subjects, in

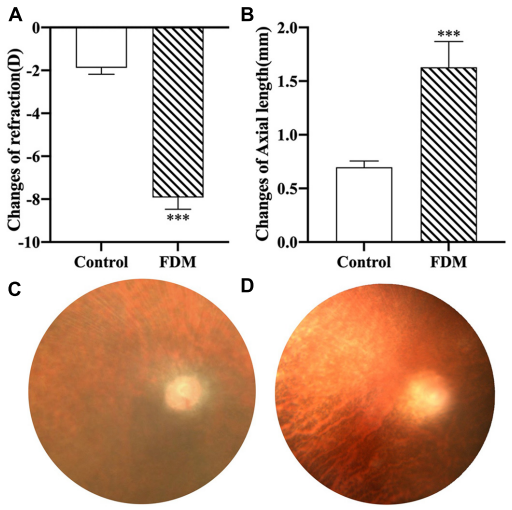


FIGURE 7
Changes in right eye diopter and axial length in two groups. **(A)** Changes of refraction. **(B)** Changes of axial length. **(C)** Fundus of guinea pigs in control group. **(D)** Fundus of guinea pigs in FDM group. ***$p < 0.001$, compared with control group. FDM, form-deprivation myopia.

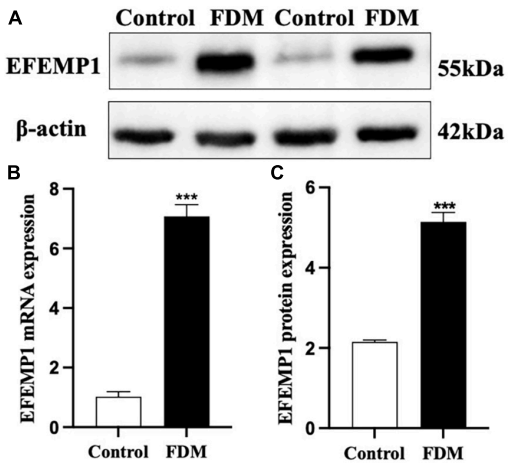


FIGURE 8
Expression of EFEMP1 mRNA and protein. **(A)** Western blot image of the two groups. **(B)** mRNA expression of EFEMP1. **(C)** Protein expression of EFEMP1. ***$p < 0.001$, compared with control group, $n = 5$.

particular, the HM group was significantly different from the other two groups, suggesting a possible association between the myopia degrees and choroidal thicknesses, which is consistent with the results of previous studies (Deng et al., 2018).

In the process of searching for the molecular mechanism of choroid changes in myopia patients, the collection of tears from myopia patients is a feasible operation, which has the advantages of being noninvasive and acceptable. At present, tear detection is mainly used for the analysis of ocular surface diseases, such as dry eye and conjunctivitis (O'Neil et al., 2019). Fewer studies have observed changes in tear composition in myopia studies; however, we consider this to be an exploration of ways to find relevant molecular mechanisms. First, tears are secreted by lacrimal glands and conjunctival goblet cells. The blood flow of the lacrimal gland, conjunctiva, and choroid all belong to the branches of ophthalmic arteries and return to the cavernous sinus through the superior ophthalmic vein (Wang, 2002), with a certain homology between the three, which may indirectly reflect molecular changes in the choroid or in the eye. Second, tears have been used as a biomarker to assess the progression and prognosis of ophthalmic diseases, such as cataracts and age-related macular degeneration (Brown et al., 2018; Engelbrecht et al., 2020). This study is the first to report the relationship between the tear concentration of EFEMP1 and biological parameters of myopic eyes. We found that the EFEMP1 concentration in myopic tears was significantly increased and showed a significant negative correlation with SE and a significant positive correlation with AL, ACD, and AL/CR, suggesting that upregulation of EFEMP1 may be associated with the development of axial myopia.

Combined with the decrease in choroid thickness in myopic patients and the significant increase in EFEMP1 concentrations in the tears of myopic patients in this study, we speculated that EFEMP1 might be involved in the molecular mechanism of CT changes during the development of myopia. Due to the limitations of tear detection, we subsequently verified this hypothesis with FDM guinea pigs and found that EFEMP1 was significantly increased in choroid tissue. However, how EFEMP1 acts on downstream molecules and the possible specific mechanism of EFEMP1 in the FDM process need to be further studied.

Previous studies have found that altered EFEMP1 expression is closely associated with many ocular diseases (Springelkamp et al., 2015), mutations in EFEMP1 expression (R345 W) may lead to Malattia Leventinese (Tsai et al., 2021), and the expression of EFEMP1 is abnormally high in the ciliary body of patients with open-angle glaucoma (Collantes et al., 2022). However, studies on changes in EFEMP1 protein in the choroid in myopia models have not been reported. EFEMP1 contains 493 amino acids and belongs to a family of proteins containing the EGF structural domain. All known sequences of epidermal growth factor receptor (EGFR)-binding proteins have been compared with EFEMP1 sequences, and it was found that EFEMP1 has high homology with EGF and interacts with EGFR (Camaj et al., 2009; Chu and Peters, 2008). Dong et al. (2020) showed that EGFR expression is significantly increased in myopic guinea pigs and that intravitreal injection of EGFR antibody is dose-dependently associated with axial elongation of lens-induced myopia in guinea pigs, suggesting that EFEMP1 may synergize with EGFR in myopia regulation. In addition, (Roybal et al., 2005) found that overexpression of EFEMP1 in RPE cells effectively activated the unfolded protein response (UPR), leading to upregulation of vascular epidermal growth factor (VEGF) expression. UPR is thought to be a protective mechanism for cells against external stimuli, but when endoplasmic reticulum stress is persistent, UPR is unable to correct the imbalance protein homeostasis and initiates apoptosis-related signaling. Studies have shown that as myopia progresses, especially high myopia, RPE cell density decreases (Zhang and Wildsoet, 2015). Meanwhile, anti-VEGF was proven to have an inhibitory effect on myopia development in a chick model of deprivation myopia (Mathis et al., 2014). Scholars have found that vascular endothelial factor A (VEGF-A) is highly expressed in the choroidal vessel wall and may play a large role in the response to choroidal

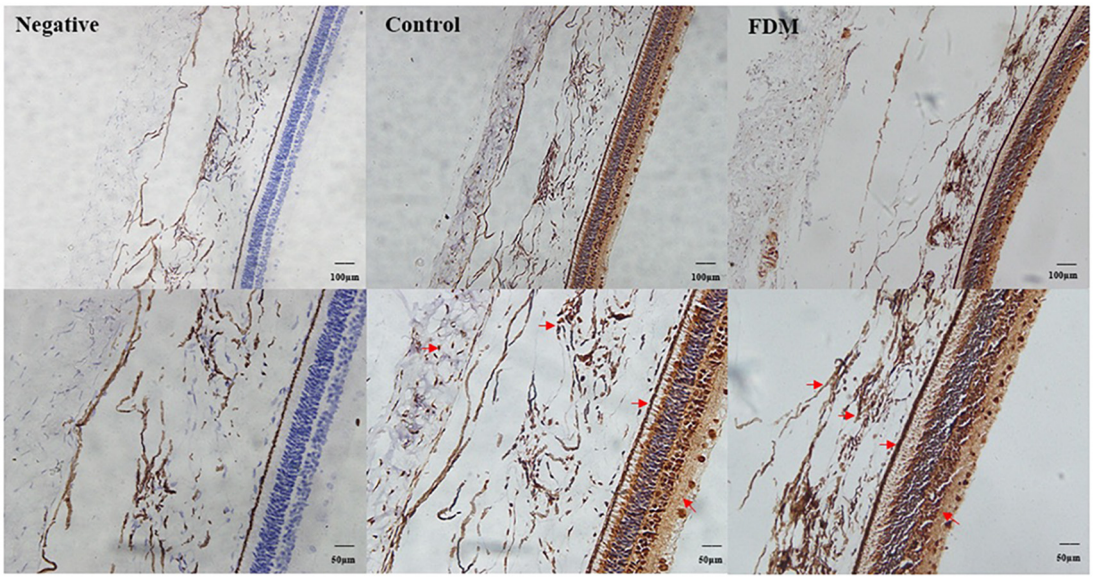


FIGURE 9
Immunohistochemical staining of EFEMP1. A brown stain (red arrow) represents the location of EFEMP1 and a blue stain represents the nucleus.

thickening caused by myopic defocusing (Mathis et al., 2014). Enhanced EFEMP1 expression around the choroidal canal wall was seen in the IHC results of this study. Therefore, we speculate that the increase in EFEMP1 may be associated with the increase in EGFR and VEGF during myopia development. Moreover, EFEMP1 is important for maintaining the integrity of the basement membrane and the binding of other extracellular matrices, such as elastic fibers and basement membranes (Zhang et al., 2020). Some researchers have found that EFEMP1 interacts with elastin as well as collagen 15A1 (a component of collagen fibrils) to produce proteins that affect the expression and function of these proteins (Jorgenson et al., 2015). Peng et al. (2022) found reduced EFEMP1 expression in the abdominal fascia of patients with inguinal hernia. Albig et al. (2006) concluded that EFEMP1-deficient female rats had broken elastic fibers in the vaginal wall. The above studies suggest that EFEMP1 has a specific role in the integrity of elastic fibers. The choroid is rich in elastin, and as the choroidal thickness thins due to the growth of the eye axis in myopic patients, the elastin content is altered, which may stimulate the altered expression of EFEMP1. Targeted regulation of EFEMP1 may be useful for investigating the molecular mechanism of CT changes during myopia development.

In summary, OCTA was used to measure the CT thickness of patients with different degrees of myopia, and significant differences were found between groups. Meanwhile, tears of patients with emmetropia and myopia were collected and analyzed, and EFEMP1 concentrations in tears were significantly increased. Subsequently, FDM guinea pigs were used to validate EFEMP1 for CT thickness changes. The expression of EFEMP1 mRNA and protein was upregulated in the choroid of FDM guinea pigs. We speculate that EFEMP1 may be involved in the development of myopia, especially in choroid thickness. EFEMP1 may be a potential target for the prevention and treatment of choroid thickness changes in myopia. However, the limitations of this study are the small sample size collected and the need to further explore the potential mechanisms of EFEMP1 and CT changes in molecular as well as different experimental models in the future.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Committee of Jinshan Hospital of Fudan University (JIEC 2020-S39-02). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

W-QS, TW, BL, TL, and X-DZ made substantial contributions to this research. W-QS and BL performed the experiments and collected the data. TW and TL designed the current study. W-QS and X-DZ given final approval of the version to be published. W-QS wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by Shanghai Shenkang Hospital Development Center (Grant No. SHDC2020CR1043B-004),

Shanghai Jinshan District Health Commission (Grant Nos. JSZK2019H01 and JSKJ-KTQN-2021-02), and Jinshan Hospital affiliated to Fudan University (Grant No. JYQN-JC-202203).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1144421/full#supplementary-material

SUPPLEMENTARY FIGURE 1
ELISA standard curve for the concentration of EFEMP1 in tear fluid.

SUPPLEMENTARY TABLE 1
Ocular biometric parameters of the right eye measured by A-scan.

SUPPLEMENTARY TABLE 2
Ocular biometric parameters of the right eye measured by IOL-Master.

References

Albig, A. R., Neil, J. R., and Schiemann, W. P. (2006). Fibulins 3 and 5 antagonize tumor angiogenesis in vivo. *Cancer Res.* 66, 2621–2629. doi: 10.1158/0008-5472.CAN-04-4096

Al-Sheikh, M., Phasukkijwatana, N., Dolz-Marco, R., Rahimi, M., Iafe, N. A., Freund, K. B., et al. (2017). Quantitative OCT angiography of the retinal microvasculature and the choriocapillaris in myopic eyes. *Invest. Ophthalmol. Vis. Sci.* 58, 2063–2069. doi: 10.1167/iovs.16-21289

Baird, P. N., Saw, S. M., Lanca, C., Guggenheim, J. A., Smith Iii, E. L., Zhou, X., et al. (2020). Myopia. *Nat. Rev. Dis. Primers* 6:99. doi: 10.1038/s41572-020-00231-4

Baumgarten, S., Schiller, P., Hellmich, M., Walter, P., Agostini, H., Junker, B., et al. (2018). Vitrectomy with and without encircling band for pseudophakic retinal detachment with inferior breaks: VIPER Study Report No. 3. *Graefes. Arch. Clin. Exp. Ophthalmol.* 256, 2069–2073. doi: 10.1007/s00417-018-4106-6

Brown, C. N., Green, B. D., Thompson, R. B., den Hollander, A. I., Lengyel, I., and Eye-Risk consortium. (2018). Metabolomics and age-related macular degeneration. *Metabolites* 9:4. doi: 10.3390/metabo9010004

Camaj, P., Seeliger, H., Ischenko, I., Krebs, S., Blum, H., De Toni, E. N., et al. (2009). EFEMP1 binds the EGF receptor and activates MAPK and Akt pathways in pancreatic carcinoma cells. *Biol. Chem.* 390, 1293–1302. doi: 10.1515/BC.2009.140

Cheng, L., Chen, C., Guo, W., Liu, K., Zhao, Q., Lu, P., et al. (2020). EFEMP1 overexpression contributes to neovascularization in age-related macular degeneration. *Front. Pharmacol.* 11:547436. doi: 10.3389/fphar.2020.547436

Chu, T. J., and Peters, D. G. (2008). Serial analysis of the vascular endothelial transcriptome under static and shear stress conditions. *Physiol. Genomics* 34, 185–192. doi: 10.1152/physiolgenomics.90201.2008

Collantes, E. R. A., Delfin, M. S., Fan, B., Torregosa, J. M. R., Siguan-Bell, C., Vincent de Guzman Florcruz, N., et al. (2022). EFEMP1 rare variants cause familial juvenile-onset open-angle glaucoma. *Hum. Mutat.* 43, 240–252. doi: 10.1002/humu.24320

Daniel, S., Renwick, M., Chau, V. Q., Datta, S., Maddineni, P., Zode, G., et al. (2020). Fibulin-3 knockout mice demonstrate corneal dysfunction but maintain normal retinal integrity. *J. Mol. Med.* 98, 1639–1656. doi: 10.1007/s00109-020-01974-z

Deng, J., Li, X., Jin, J., Zhang, B., Zhu, J., Zou, H., et al. (2018). Distribution pattern of choroidal thickness at the posterior pole in chinese children with myopia. *Invest. Ophthalmol. Vis. Sci.* 59, 1577–1586. doi: 10.1167/iovs.17-22748

Dong, L., Shi, X. H., Li, Y. F., Jiang, X., Wang, Y. X., Lan, Y. J., et al. (2020). Blockade of epidermal growth factor and its receptor and axial elongation in experimental myopia. *FASEB J.* 34, 13654–13670. doi: 10.1096/fj.202001095RR

Duan, F., Yuan, Z., Deng, J., Wong, Y. L., Yeo, A. C., and Chen, X. (2019). Choroidal thickness and associated factors among adult myopia: a baseline report from a medical university student cohort. *Ophthalm. Epidemiol.* 26, 244–250. doi: 10.1080/09286586.2019.1597899

Engelbrecht, C., Sardinha, L. R., and Rizzo, L. V. (2020). Cytokine and chemokine concentration in the tear of patients with age-related cataract. *Curr. Eye Res.* 45, 1101–1106. doi: 10.1080/02713683.2020.1715445

Fitzgerald, M. E., Wildsoet, C. F., and Reiner, A. (2002). Temporal relationship of choroidal blood flow and thickness changes during recovery from form deprivation myopia in chicks. *Exp. Eye Res.* 74, 561–570. doi: 10.1006/exer.2002.1142

Fujiwara, A., Morizane, Y., Hosokawa, M., Kimura, S., Kumase, F., Shiode, Y., et al. (2016). Factors affecting choroidal vascular density in normal eyes: quantification using en face swept-source optical coherence tomography. *Am. J. Ophthalmol.* 170, 1–9. doi: 10.1016/j.ajo.2016.07.006

Jonas, J. B., Dong, L., Da Chen, S., Neumaier, M., Findeisen, P., Panda-Jonas, S., et al. (2021). Intraocular epidermal growth factor concentration, axial length, and high axial myopia. *Graefes. Arch. Clin. Exp. Ophthalmol.* 259, 3229–3234. doi: 10.1007/s00417-021-05200-5

Jorgenson, E., Makki, N., Shen, L., Chen, D. C., Tian, C., Eckalbar, W. L., et al. (2015). A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat. Commun.* 6:10130. doi: 10.1038/ncomms10130

Li, C., Fitzgerald, M. E. C., Del Mar, N., Haughey, C., and Reiner, A. (2018). Defective choroidal blood flow baroregulation and retinal dysfunction and pathology following sympathetic denervation of choroid. *Invest. Ophthalmol. Vis. Sci.* 59, 5032–5044. doi: 10.1167/iovs.18-24954

Liu, K., Xu, H., Jiang, H., Wang, H., Wang, P., Xu, Y., et al. (2020). Macular vessel density and foveal avascular zone parameters in patients after acute primary angle closure determined by OCT angiography. *Sci. Rep.* 10:18717. doi: 10.1038/s41598-020-73223-9

Livingstone, I., Uversky, V. N., Furniss, D., and Wiberg, A. (2020). The pathophysiological significance of fibulin-3. *Biomolecules* 10:1294. doi: 10.3390/biom10091294

Lu, F., Zhou, X., Jiang, L., Fu, Y., Lai, X., Xie, R., et al. (2009). Axial myopia induced by hyperopic defocus in guinea pigs: a detailed assessment on susceptibility and recovery. *Exp. Eye Res.* 89, 101–108. doi: 10.1016/j.exer.2009.02.019

Mathis, U., Ziemssen, F., and Schaeffel, F. (2014). Effects of a human VEGF antibody (Bevacizumab) on deprivation myopia and choroidal thickness in the chicken. *Exp. Eye Res.* 127, 161–169. doi: 10.1016/j.exer.2014.07.022

Milani, P., Montesano, G., Rossetti, L., Bergamini, F., and Pece, A. (2018). Vessel density, retinal thickness, and choriocapillaris vascular flow in myopic eyes on OCT angiography. *Graefes. Arch. Clin. Exp. Ophthalmol.* 256, 1419–1427. doi: 10.1007/s00417-018-4012-y

Naidoo, K. S., Fricke, T. R., Frick, K. D., Jong, M., Naduvilath, T. J., Resnikoff, S., et al. (2019). Potential lost productivity resulting from the global burden of myopia: systematic review, meta-analysis, and modeling. *Ophthalmology* 126, 338–346. doi: 10.1016/j.ophtha.2018.10.029

Nickla, D. L., and Wallman, J. (2010). The multifunctional choroid. *Prog. Retin. Eye Res.* 29, 144–168. doi: 10.1016/j.preteyeres.2009.12.002

Nishida, Y., Fujiwara, T., Imamura, Y., Lima, L. H., Kurosaka, D., and Spaide, R. F. (2012). Choroidal thickness and visual acuity in highly myopic eyes. *Retina* 32, 1229–1236. doi: 10.1097/IAE.0b013e318242b990

Ohno-Matsui, K., and Jonas, J. B. (2019). Posterior staphyloma in pathologic myopia. *Prog. Retin. Eye Res.* 70, 99–109. doi: 10.1016/j.preteyeres.2018.12.001

O'Neil, E. C., Henderson, M., Massaro-Giordano, M., and Bunya, V. Y. (2019). Advances in dry eye disease treatment. *Curr. Opin. Ophthalmol.* 30, 166–178. doi: 10.1097/ICU.0000000000000569

Peng, X., Guo, Z., Zhang, Y., Sun, B., and Zhang, Q. (2022). EFEMP1 in direct inguinal hernia: correlation with TIMP3 and regulation toward elastin homoeostasis as well as fibroblast mobility. *J. Invest. Surg.* 35, 203–211. doi: 10.1080/08941939.2020.1811812

Roybal, C. N., Marmorstein, L. Y., Vander Jagt, D. L., and Abcouwer, S. F. (2005). Aberrant accumulation of fibulin-3 in the endoplasmic reticulum leads to activation of the unfolded protein response and VEGF expression. *Invest. Ophthalmol. Vis. Sci.* 46, 3973–3979. doi: 10.1167/iovs.05-0070

Springelkamp, H., Mishra, A., Hysi, P. G., Gharahkhani, P., Höhn, R., Khor, C. C., et al. (2015). Meta-analysis of genome-wide association studies identifies novel loci associated with optic disc morphology. *Genet. Epidemiol.* 39, 207–216. doi: 10.1002/gepi.21886

Summers, J. A. (2013). The choroid as a sclera growth regulator. *Exp. Eye Res.* 114, 120–127. doi: 10.1016/j.exer.2013.03.008

Tian, F., Zheng, D., Zhang, J., Liu, L., Duan, J., Guo, Y., et al. (2021). Choroidal and retinal thickness and axial eye elongation in chinese junior students. *Invest. Ophthalmol. Vis. Sci.* 62:26. doi: 10.1167/iovs.62.9.26

Tsai, Y. T., Li, Y., Ryu, J., Su, P. Y., Cheng, C. H., Wu, W. H., et al. (2021). Impaired cholesterol efflux in retinal pigment epithelium of individuals with juvenile macular degeneration. *Am. J. Hum. Genet.* 108, 903–918. doi: 10.1016/j.ajhg.2021.04.006

Wang, J. L. (2002). The arterial supply to the eye of the bactrian camel (*Camelus bactrianus*). *Vet. Res. Commun.* 26, 505–512. doi: 10.1023/a:1020310213203

Wu, H., Xie, Z., Wang, P., Liu, M., Wang, Y., Zhu, J., et al. (2021). Differences in retinal and choroidal vasculature and perfusion related to axial length in pediatric anisomyopes. *Invest. Ophthalmol. Vis. Sci.* 62:40. doi: 10.1167/iovs.62.9.40

Xiuyan, Z., Qingmei, T., Qiuxin, W., Tailiang, L., Jing, X., Guodong, T., et al. (2021). Thickness, vessel density of retina and choroid on OCTA in young adults (18-24 years old). *Microvasc. Res.* 136:104169. doi: 10.1016/j.mvr.2021.104169

Zhang, N., Liao, Z., Wu, P., Fang, H., and Cai, G. (2020). Hypermethylation of EFEMP1 in the hippocampus may be related to the deficit in spatial memory of rat neonates triggered by repeated administration of propofol. *Biomed. Res. Int.* 2020:8851480. doi: 10.1155/2020/8851480

Zhang, S., Zhang, G., Zhou, X., Xu, R., Wang, S., Guan, Z., et al. (2019). Changes in choroidal thickness and choroidal blood perfusion in guinea pig myopia. *Invest. Ophthalmol. Vis. Sci.* 60, 3074–3083. doi: 10.1167/iovs.18-26397

Zhang, Y., and Wildsoet, C. F. (2015). RPE and choroid mechanisms underlying ocular growth and myopia. *Prog. Mol. Biol. Transl. Sci.* 134, 221–240. doi: 10.1016/bs.pmbts.2015.06.014

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 24 March 2023
DOI 10.3389/fnins.2023.1148855

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Hongzhi Kuai,
Maebashi Institute of Technology, Japan
Yaofei Xie,
Xuzhou Medical University, China

*CORRESPONDENCE
Wenfeng Duan
✉ ndyfy02345@ncu.edu.cn

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 20 January 2023
ACCEPTED 06 March 2023
PUBLISHED 24 March 2023

CITATION
Wan Z, Li M, Liu S, Huang J, Tan H and Duan W (2023) EEGformer: A transformer–based brain activity classification method using EEG signal.
Front. Neurosci. 17:1148855.
doi: 10.3389/fnins.2023.1148855

EEGformer: A transformer–based brain activity classification method using EEG signal

Zhijiang Wan[1,2,3], Manyu Li[2], Shichang Liu[4], Jiajin Huang[5], Hai Tan[6] and Wenfeng Duan[1]*

[1]The First Affiliated Hospital of Nanchang University, Nanchang University, Nanchang, Jiangxi, China, [2]School of Information Engineering, Nanchang University, Nanchang, Jiangxi, China, [3]Industrial Institute of Artificial Intelligence, Nanchang University, Nanchang, Jiangxi, China, [4]School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, China, [5]Faculty of Information Technology, Beijing University of Technology, Beijing, China, [6]School of Computer Science, Nanjing Audit University, Nanjing, Jiangsu, China

Background: The effective analysis methods for steady-state visual evoked potential (SSVEP) signals are critical in supporting an early diagnosis of glaucoma. Most efforts focused on adopting existing techniques to the SSVEPs-based brain–computer interface (BCI) task rather than proposing new ones specifically suited to the domain.

Method: Given that electroencephalogram (EEG) signals possess temporal, regional, and synchronous characteristics of brain activity, we proposed a transformer–based EEG analysis model known as EEGformer to capture the EEG characteristics in a unified manner. We adopted a one-dimensional convolution neural network (1DCNN) to automatically extract EEG-channel-wise features. The output was fed into the EEGformer, which is sequentially constructed using three components: regional, synchronous, and temporal transformers. In addition to using a large benchmark database (BETA) toward SSVEP-BCI application to validate model performance, we compared the EEGformer to current state-of-the-art deep learning models using two EEG datasets, which are obtained from our previous study: SJTU emotion EEG dataset (SEED) and a depressive EEG database (DepEEG).

Results: The experimental results show that the EEGformer achieves the best classification performance across the three EEG datasets, indicating that the rationality of our model architecture and learning EEG characteristics in a unified manner can improve model classification performance.

Conclusion: EEGformer generalizes well to different EEG datasets, demonstrating our approach can be potentially suitable for providing accurate brain activity classification and being used in different application scenarios, such as SSVEP-based early glaucoma diagnosis, emotion recognition and depression discrimination.

KEYWORDS

brain activity classification, SSVEPs, EEGformer, EEG characteristics, deep learning

1. Introduction

Glaucoma is known as a "silent thief of sight," meaning that patients do not notice the health condition of their visual function until vision loss and even blindness occur (Abdull et al., 2016). According to the world health organization, the number of people with glaucoma worldwide in 2020 is 76 million, and the patient number would be increased to 95.4 million in 2030. As the population ages, the number with this condition will also increase substantially (Guedes, 2021). Glaucoma causes irreversible optic nerve vision damage. It is crucial to provide accurate early screening to diagnose patients in their early stages so that they can receive appropriate early treatment. Steady-state visual evoked potentials (SSVEPs), which refer to a stimulus-locked oscillatory response to periodic visual stimulation commonly exerted in the visual pathway of humans, can be used to evaluate the functional abnormality of the visual pathway that is essential for the complete transmission of visual information (Zhou et al., 2020). SSVEPs are always measured using electroencephalogram (EEG) measurement and have been widely used in the study of brain–computer interface (BCI). Because peripheral vision loss is a key diagnostic sign of glaucoma, patients cannot be evoked by certain repetitive stimuli with a constant frequency from vision loss regions (Khok et al., 2020). Therefore, stimuli with the corresponding frequency are not detected by the primary visual cortex. Thus, the SSVEPs-based BCI applications can be used in the early diagnosis of visual function detection for patients with glaucoma.

The effective analysis method for SSVEPs is critical in the accurate early diagnosis of glaucoma. SSVEPs are EEG activity with a spatial-spectral-temporal (SST) pattern. It is easy to understand that SSVEP signals, such as the EEG signal measured over time, could be analyzed using time series analysis methods. Brain functional connectivity (BFC) can be used to capture spatial patterns from multiple brain regions by analyzing the correlations between brain activities detected from different regions. The spectral pattern extraction method is the most popular method for analyzing the frequency characteristics of EEG signals. For instance, power spectra density–based analysis (PSDA) is a commonly used frequency detection method that can classify various harmonic frequencies from EEG signals (Zhang et al., 2020). In addition, canonical correlation analysis (CCA) (Zhuang et al., 2020) and other similar algorithms, such as multivariate synchronization index (MSI) (Qin et al., 2021) and correlated component analysis (COCA) (Zhang et al., 2019), are effective frequency detection algorithms based on the multivariate statistical analysis method. Although SST pattern extraction algorithms have demonstrated satisfactory results, most patterns or features extracted from raw EEG data require a manually predefined algorithm based on expert knowledge. The procedure of learning handcrafted features for SSVEP signals is not flexible and might limit the performance of these systems in brain activity analysis tasks.

In recent years, deep learning (DL) methods have achieved excellent performance in processing EEG-based brain activity analysis tasks (Li Z. et al., 2022; Schielke and Krekelberg, 2022). Currently, the mainstream technologies of using DL to process SSVEP signal could be divided into two aspects: convolutional neural network (CNN) based methods and transformer-based methods. For the CNN-based methods, Li et al. (2020) propose a CNN-based nonlinear model, i.e. convolutional correlation analysis (Conv-CA), to transform multiple channel EEGs into a single EEG signal and use a correlation layer to calculate correlation coefficients between the transformed single EEG signal and reference signals. Guney et al. (2021) propose a deep neural network architecture for identifying the target frequency of harmonics. Waytowich et al. (2018) design a compact convolutional neural network (Compact-CNN) for high-accuracy decoding of SSVEPs signal. For the transformer-based methods, Du et al. (2022) propose a transformer-based approach for the EEG person identification task that extracts features in the temporal and spatial domains using a self-attention mechanism. Chen et al. (2022) propose SSVEPformer, which is the first application of the transformer to the classification of SSVEP. Li X. et al. (2022) propose a temporal-frequency fusion transformer (TFF-Former) for zero-training decoding across two BCI tasks. The aforementioned studies demonstrate the competitive model performance of DL methods in performing SSVEPs-based BCI tasks. However, most existing DL efforts focused on applying existing techniques to the SSVEPs-based BCI task rather than proposing new ones specifically suited to the domain. Standard well-known network architectures are designed for data collected in natural scenes and do not consider the peculiarities of the SSVEP signals. Therefore, further research is required to understand how these architectures can be optimized for EEG-based brain activity data.

The main question is what is the specificity of the SSVEP signal analysis domain and how to use machine learning methods (particularly DL methods) to deal with the signal characteristics. Because the SSVEP signal is EEG-based brain activity, we can answer the question by analyzing the EEG characteristics in the brain activity analysis domain. Specifically, EEG characteristics are reflected in three aspects: temporal, regional, and synchronous characteristics. The temporal characteristics (e.g., mean duration, coverage, and frequency of occurrence) are easily traceable in standard EEG data and provide numerous sampling points in a short time (Zhang et al., 2021), thereby providing an efficient way to investigate trial-by-trial fluctuations of functional spontaneous activity. The regional characteristics refer to different brain regions that are linked to distinct EEG bands (Nentwich et al., 2020). The synchronous characteristics refer to the synchronous brain activity pattern over a functional network including several brain regions with similar spatial orientations (Raut et al., 2021). Traditionally, brain response to a flickering visual stimulation has been considered steady-state, in which the elicited effect is believed to be unchanging in time. In fact, the SSVEPs belongs to a signal with non-stationary nature, which indicates dynamical patterns and complex synchronization between EEG channels can be used to further understand brain mechanisms in cognitive and clinical neuroscience. For instance, Ibáñez-Soria et al. explored the dynamical character of the SSVEP response by proposing a novel non-stationary methodology for SSVEP detection, and found dynamical detection methodologies significantly improves classification in some stimulation frequencies (Ibáñez-Soria et al., 2019). Tsoneva et al. (2021) studied the mechanisms behind SSVEPs generation and propagation in time and space. They concluded that the SSVEP spatial properties appear sensitive to input frequency with higher stimulation frequencies showing a faster propagation speed. Thus, we hypothesize that a machine learning method that can capture the EEG characteristics in a unified manner can suit the SSVEPs-based BCI domain and improve the model performance in EEG-based brain activity analysis tasks.

In this study, we propose a transformer–based EEG analysis model known as the EEGformer (Vaswani et al., 2017) to capture the EEG characteristics in the SSVEPs-based BCI task. The EEGformer is an end-to-end DL model, processing SSVEP signals from the EEG to the prediction of the target frequency. The component modules of the EEG former are depicted as follows:

(1) Depth-wise convolution-based one-dimensional convolution neural network (1DCNN). The depth-wise convolution-based 1DCNN is first used to process the raw EEG input. Assuming the raw data is collected from C EEG channels, there are M depth-wise convolutional filters for generating M feature maps. Each convolutional filter is responsible for shifting across the raw data in an EEG-channel-wise manner and extracting convolutional features from the raw data of each EEG channel to form a feature map. Unlike other techniques that manually extract temporal or spectrum features based on the time course of the EEG signal, we use the depth-wise convolutional filter to extract the EEG features in a completely data–driven manner. Because the feature map is generated by the same depth-wise convolutional filter, each row of the feature map shares the same convolutional property. Follow-up convolutional layers are allocated with several depth-wise convolutional filters to enrich the convolutional features and deepen the 1DCNN network. A three-dimensional (3D) feature matrix is used to represent the output of the 1DCNN network. The x, y, and z dimensions of the 3D feature matrix represent temporal, spatial, and convolutional features, respectively.

(2) EEGformer encoder. This component module consists of three sub-modules: temporal, synchronous, and regional transformers, which are used in learning the temporal, synchronous, and regional characteristics, respectively. The core strategy of learning EEG characteristics by our model mainly include two steps: input tokens that serve as the basic elements of learning the temporal, synchronous, and regional characteristics are sliced from the 3D feature matrix along the temporal, convolutional, and spatial dimension, respectively. And then, self-attention mechanism is employed to measure the relationships between pairs of input tokens and give tokens more contextual information, yielding more powerful features for representing the EEG characteristics. The three components could be performed in a sequential computing order, allowing the encoder to learn the EEG characteristics in a unified manner.

(3) EEGformer decoder. This module contains three convolutional layers and one fully connected (FC) layer. The output of the last FC layer is fed to a softmax function which produces a distribution over several category labels. The categorical cross entropy combined with regularization was used as the loss function for training the entire EEGformer pipeline. The EEGformer decoder is used to deal with specific tasks, such as target frequency identification, emotion recognition, and depression discrimination. In addition to using a large benchmark database (BETA) (Liu et al., 2020) to validate the performance of the SSVEP-BCI application, we validate the model performance on two additional EEG datasets, one for emotion analysis using EEG signals [SJTU emotion EEG dataset (SEED)] (Duan et al., 2013; Zheng and Lu, 2015) and the other for a depressive EEG database (DepEEG) (Wan et al., 2020) obtained from our previous study, to support our hypothesis that highlights the significance of learning EEG characteristics in a unified manner for EEG-related data analysis tasks.

The main contributions of this study are as follows: (1) current mainstream DL models have superior ability in processing data collected in natural scenes and might not adept at dealing with SSVEP signals. To achieve a DL model that can be applied to the specificity of the SSVEP signal analysis domain and obtain better model performance in SSVEPs-based frequency recognition task, we propose a transformer–based EEG analysis model known as the EEGformer to capture the EEG characteristics in a unified manner. (2) To obtain a flexible method for addressing the SSVEPs-based frequency recognition and avoid the model performance limited by manual feature extraction, we adopt 1DCNN to automatically extract EEG-channel-wise features and fed them into the EEGformer. This operation transforms our method into a complete data–driven manner for mapping raw EEG signals into task decisions. (3) To ensure the effectiveness and generalization ability of the proposed model, we validate the performance of the EEGformer on three datasets for three different EEG-based data analysis tasks: target frequency identification, emotion recognition, and depression discrimination.

2. Materials and methods

2.1. Dataset preparation

Table 1 shows some detailed information about the three datasets (BETA, SEED, and DepEEG) that we used as benchmarks to validate the effectiveness of this study. The participants' column in the table describes how many subjects joined in the corresponding data collection. The experiment per participant (EPP) column shows how many experiments were performed by each participant. The trails per experiment (TPE) column shows how many trails are executed in one experiment. The channel number (CHN) column shows the CHN of the EEG dataset. The sampling rate (SR) column shows the down-sampling rate of the EEG signal. The time length per trail (TLPT) column shows the time length of a single trail in seconds. The labels column shows the categorical emotion labels for the classification task and emotional intensity for the regression task. Specifically, for the target frequency identification task, we classified 40 categories of harmonic frequencies and the frequency range is 8–15.8 HZ with 0.2 HZ intervals. For the emotion recognition task, we used arousal, valence, and dominance rating scores as the dataset labels. For the depression discrimination task, we classified EEG samples from depressive or normal control.

2.2. Pipeline of EEGformer–based brain activity analysis

Figure 1 shows the pipeline of EEGformer–based brain activity analysis. The core modules of the pipeline include 1DCNN, EEGformer encoder, and decoder. The input of the 1DCNN is an

TABLE 1 Detail information on the three datasets.

Dataset	Participants	EPP	TPE	CHN	SR (HZ)	Labels	TLPT
BETA	70 healthy subjects	4	40	64	250	40 harmonics, e.g., $f_j \in \{8,8.2,\ldots,15.8\}$	2/3 s
SEED	15 healthy subjects	3	15	62	200	Positive, neutral, negative	305 s
DepEEG	12 healthy subjects and 23 depressives	1	1	6	500	Depressive, normal control	≥480 s

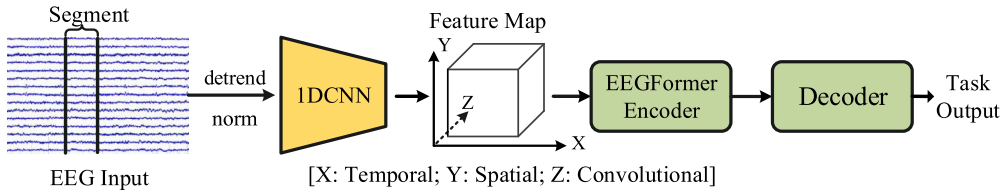


FIGURE 1
Pipeline of EEGformer for different tasks of brain activity analysis.

EEG segment represented using a two-dimensional (2D) matrix of size $S \times L$, where S represents the number of EEG channels, and L represents the segment length. The EEG segment is de-trend and normalized before being fed into the 1DCNN module, and the normalized EEG segment is represented by $x \in R^{S \times L}$. The 1DCNN adopts multiple depth-wise convolutions to extract EEG-channel-wise features and generate 3D feature maps. It shifts across the data along the EEG channel dimension for each depth-wise convolution and generates a 2D feature matrix of size $S \times L_f$, where L_f is the length of the extracted feature vector. The output of the 1DCNN module is a 3D feature matrix of size $S \times C \times L_e$, where C is the number of depth-wise convolutional kernels used in the last layer of the 1DCNN module, L_e is the features length outputted by the last layer of the 1DCNN module. More specifically, the 1DCNN is comprised of three depth-wise convolutional layers. Hence, we have the processing $x \rightarrow z_1 \rightarrow z_2 \rightarrow z_3$, where z_1, z_2, and z_3 denote the outputs of the three layers. The size of the depth-wise convolutional filters used in the three layers is 1×10, valid padding mode is applied in the three layers and the stride of the filters is set to 1. The number of the depth-wise convolutional filter used in the three layers is set to 120, ensuring sufficient frequency features for learning the regional and synchronous characteristics. We used a 3D coordinate system to depict the axis meaning of the 3D feature matrix. The X, Y, and Z axes represent the temporal, spatial, and convolutional feature information contained in the 3D feature matrix, respectively. The output of the 1DCNN module is fed into the EEGformer encoder for encoding the EEG characteristics (regional, temporal, and synchronous characteristics) in a unified manner. The decoder is responsible for decoding the EEG characteristics and inferencing the results according to the specific task.

2.3. EEGformer encoder

The EEGformer encoder is used to provide a uniform feature refinement for the regional, temporal, and synchronous characteristics contained in the output of the 1DCNN module. Figure 2 illustrates the EEGformer architecture and shows that the EEGformer encoder uses a serial structure to sequentially refine the EEG characteristics. The temporal, regional, and synchronous characteristics are refined using temporal, regional, and synchronous transformers, respectively. The outputs of the 1DCNN are defined as $z_3 \in R^{S \times C \times L_e}$ and are represented using black circles in the green rectangle box.

The specific computing procedures of each transformer module are depicted as follows:

2.3.1. Regional transformer module

The input of the regional transformer module is represented by $z_3 \in R^{C \times L_e \times S}$. The 3D matrix z_3 is first segmented into S 2D submatrices along the spatial dimension. Each submatrix is represented by $X_i^{reg} \in R^{C \times L_e}$ (i = 1,2,3,...,S). The input of the regional transformer module is represented by S black circles in the green rectangle box and each circle represents a submatrix. The vector $X_{(i,c)}^{reg} \in R^{L_e}$ is sequentially taken out from the X_i^{reg} along the convolutional feature dimension and fed into the linear mapping module. According to the terminology used in the vision of transformer (ViT) studies, we defined the vector $X_{(i,c)}^{reg}$ as a patch of the regional transformer module. Each $X_{(i,c)}^{reg}$ is represented by a tiny yellow block in the Figure 2. The $X_{(i,c)}^{reg}$ is linearly mapped into a latent vector $z_{(i,c)}^{(reg,0)} \in R^D$ using a learnable matrix $M \in R^{D \times L_e}$:

$$z_{(i,c)}^{(reg,0)} = MX_{(i,c)}^{reg} + e_{(i,c)}^{pos}, \quad (1)$$

where $e_{(i,c)}^{pos} \in R^D$ denotes a positional embedding added to encode the position for each convolutional feature changing over time. The regional transformer module also consists of $K \geq 1$ encoding blocks, each block contains two layers: a multi-head self-attention layer and a position-wise fully connected feed-forward network. The resulting $z_{(i,c)}^{(reg,0)}$ is defined as a token representing the inputs of each block, and the $z_{(0,0)}^{(reg,0)}$ indicates the classification token. The l-th block produces an encoded representation $z_{(i,c)}^{(reg,l)}$ for each token in the input sequence by incorporating the attention scores. Specifically, at each block l, three core vectors, including $q_{(i,c)}^{(l,a)}$, $k_{(i,c)}^{(l,a)}$, and $v_{(i,c)}^{(l,a)}$ are computed from the representation $z_{(i,c)}^{(reg,l-1)}$ encoded by the preceding layer:

$$q_{(i,c)}^{(l,a)} = W_Q^{(l,a)} LN(z_{(i,c)}^{(reg,l-1)}) \in R^{D_h}, \quad (2)$$

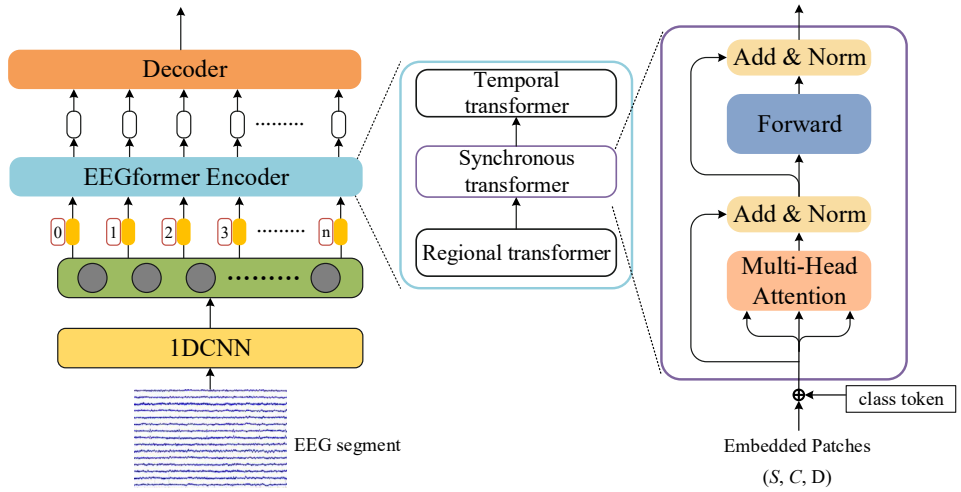


FIGURE 2
Illustration of the EEGformer architecture.

$$k_{(i,c)}^{(l,a)} = W_K^{(l,a)} LN(z_{(i,c)}^{(reg,l-1)}) \in R^{D_h}, \quad (3)$$

$$v_{(i,c)}^{(l,a)} = W_V^{(l,a)} LN(z_{(i,c)}^{(reg,l-1)}) \in R^{D_h}, \quad (4)$$

where $W_Q^{(l,a)}$, $W_K^{(l,a)}$, and $W_V^{(l,a)}$ are the matrixes of query, key, and value in the regional transformer module, respectively. LN() denotes the LayerNorm operation, and a $\in$ {1, 2, 3, . . ., A} is an index over the multi-head self-attention units. A is the number of units in a block. D_h is the quotient computed by D/A and denotes the dimension number of three vectors. The regional self-attention (RSA) scores for $z_{(i,c)}^{(reg,l-1)}$ in the a-th multi-head self-attention unit is given as follows:

$$\alpha_{(i,c)}^{(l,a)reg} = \sigma\left(\frac{q_{(i,c)}^{(l,a)}}{\sqrt{D_h}} \cdot \left[k_{(0,0)}^{(l,a)} \left\{k_{(i,c)}^{(l,a)}\right\}_{c=1,\ldots,C}\right]\right) \in R^C, \quad (5)$$

where σ denotes the softmax activation function, and the symbol · denotes the dot product for computing the similarity between the query and key vectors. $k_{(i,c)}^{(l,a)}$ and $q_{(i,c)}^{(l,a)}$ represent the corresponding key and query vectors, respectively. The equation shows that the RSA scores are merely computed over convolutional features of single brain region. That is, the RSA can calculate the contribution of a changing mono-electrode convolutional feature to the final model decision at a specific EEG channel. An intermediate vector $s_{(i,c)}^{(l,a)}$ for encoding $z_{(i,c)}^{(reg,l-1)}$ is given as follows:

$$s_{(i,c)}^{(l,a)} = \alpha_{(i,0)}^{(l,a)} v_{(i,0)}^{(l,a)} + \sum_{j=1}^{C} \alpha_{(i,j)}^{(l,a)} v_{(i,j)}^{(l,a)} \in R^{D_h}. \quad (6)$$

The encoded feature $z_{(i)}^{(reg,l)} \in R^{C \times D}$ by the l-th block is computed by first concatenating the intermediate vectors from all heads, and the vector concatenation is projected by matrix $W_O \in R^{D \times L}$, where L is equal to A $= D_h$. $z_{(i)}^{'(reg,l)}$ is the residual connection result of the projection of the intermediate vectors and the $z_{(i)}^{(reg,l-1)}$ encoded by the preceding block. Finally, the $z_{(i)}^{'(reg,l)}$ normalized by LN() is passed through a multilayer perceptron (MLP) using the residual connection. The output of the regional transformer is represented by $z_4 \in R^{S \times C \times D}$.

2.3.2. Synchronous transformer module

The input of the synchronous transformer module is represented by $z_4 \in R^{S \times L_e \times C}$. The 3D matrix z_4 is first segmented into C 2D submatrices along the convolutional feature dimension. Each submatrix is represented by $X_i^{syn} \in R^{S \times D}$ (i = 1,2,3,. . .,C). The vector $X_{(i,s)}^{syn} \in R^D$ is sequentially taken out from the X_i^{syn} along the spatial dimension and fed into the linear mapping module. The $X_{(i,s)}^{syn}$ is defined as a patch and is linearly mapped into a latent vector $z_{(i,s)}^{(syn,0)} \in R^D$ using a learnable matrix M $\in R^{D \times D}$:

$$z_{(i,s)}^{(syn,0)} = MX_{(i,s)}^{syn} + e_{(i,s)}^{pos}, \quad (7)$$

where $e_{(i,s)}^{pos} \in R^D$ denotes a positional embedding added to encode the spatial position for each EEG channel changing over time. The synchronous transformer also consists of $K \geq 1$ encoding blocks, and each block contains two layers: a multi-head self-attention layer and a position-wise fully connected feed-forward network. The resulting $z_{(i,s)}^{(syn,0)}$ is defined as a token representing the inputs of each block, and the $z_{(0,0)}^{(syn,0)}$ indicates the classification token. The l-th block produces an encoded representation $z_{(i,s)}^{(syn,l)}$ for each token in the input sequence by incorporating the attention scores. Specifically, at each block l, three core vectors, including $q_{(i,s)}^{(l,a)}$, $k_{(i,s)}^{(l,a)}$, and $v_{(i,s)}^{(l,a)}$ are computed from the representation $z_{(i,s)}^{(syn,l-1)}$ encoded by the preceding layer:

$$q_{(i,s)}^{(l,a)} = W_Q^{'(l,a)} LN(z_{(i,s)}^{(syn,l-1)}) \in R^{D_h}, \quad (8)$$

$$k_{(i,s)}^{(l,a)} = W_K^{'(l,a)} LN(z_{(i,s)}^{(syn,l-1)}) \in R^{D_h}, \quad (9)$$

$$v_{(i,s)}^{(l,a)} = W_V^{'(l,a)} LN(z_{(i,s)}^{(syn,l-1)}) \in R^{D_h}, \quad (10)$$

where $W_Q^{'(l,a)}$, $W_K^{'(l,a)}$, and $W_V^{'(l,a)}$ are the matrixes of query, key, and value in the synchronous transformer module, respectively.

Synchronous e self-attention (SSA) scores for $z_{(i,s)}^{(syn,l-1)}$ in the a-th multi-head self-attention unit are given as follows:

$$\alpha_{(i,s)}^{(l,a)syn} = \sigma\left(\frac{q_{(i,s)}^{(l,a)}}{\sqrt{D_h}} \cdot \left[k_{(0,0)}^{(l,a)}\left\{k_{(i,s)}^{(l,a)}\right\}_{s=1,\ldots,S}\right]\right) \in R^S, \quad (11)$$

where $k_{(i,s)}^{(l,a)}$ and $q_{(i,s)}^{(l,a)}$ denote the corresponding key and query vectors, respectively. The equation shows that the SSA scores are merely computed over the feature map extracted by the same depth-wise convolution. The SSA can calculate the contribution of convolution features changing over time to the final model decision at a specific EEG channel. An intermediate vector $s_{(i,s)}^{(l,a)}$ for encoding $z_{(i,s)}^{(syn,l-1)}$ is given as follows:

$$s_{(i,s)}^{(l,a)} = \alpha_{(i,0)}^{(l,a)} v_{(i,0)}^{(l,a)} + \sum_{j=1}^{C} \alpha_{(i,j)}^{(l,a)} v_{(i,j)}^{(l,a)} \in R^{D_h}. \quad (12)$$

The encoded feature $z_{(i)}^{(syn,l)} \in R^{S \times D}$ by the l-th block is computed by first concatenating the intermediate vectors from all heads, and the vector concatenation is projected by matrix $W_O \in R^{D \times L}$. $z_{(i)}^{'(syn,l)}$ is the residual connection result of the projection of the intermediate vectors and the $z_{(i)}^{(syn,l-1)}$ encoded by the preceding block. Finally, the $z_{(i)}^{'(syn,l)}$ normalized by LN() is passed through a multilayer perceptron (MLP) using the residual connection. The output of the synchronous transformer is represented by $z_5 \in R^{C \times S \times D}$.

2.3.3. Temporal transformer module

The input of the temporal transformer module is $z_5 \in R^{C \times S \times D}$. To avoid huge computational complexity, we compress the original temporal dimensionality D of z_5 into dimensionality M. That is, the 3D matrix z_5 is first segmented and then averaged into M 2D submatrices along the temporal dimension. Each submatrix is represented by $X_i^{temp} \in R^{S \times C}$ (i = 1,2,3,...,M) and the M submatrices are concatenated to form $X^{temp} \in R^{M \times S \times C}$. Each submatrix X_i^{temp} is flattened into a vector $X_i^{'temp} \in R^{L1}$, where $L1$ is equal to $S \times C$. The $X_i^{'temp}$ is defined as a patch and is linearly mapped into a latent vector $z_{(i)}^{(temp,0)} \in R^D$ using a learnable matrix $M \in R^{D \times L}$:

$$z_{(i)}^{(temp,0)} = MX_{(i)}^{'temp} + e_{(i)}^{pos}, \quad (13)$$

where $e_{(i)}^{pos} \in R^D$ denotes a positional embedding added to encode the temporal position for each EEG channel changing over the features extracted by different depth-wise convolutional kernels. The module consists of $K \geq 1$ encoding blocks, each block contains two layers: a multi-head self-attention layer and a position-wise fully connected feed-forward network. The resulting $z_{(i)}^{(temp,0)}$ is defined as a token representing the inputs of each block, and the $z_{(0)}^{(temp,0)}$ indicates the classification token. The l-th block produces an encoded representation $z_{(i)}^{(temp,l)}$ for each token in the input sequence by incorporating the attention scores. Specifically, at each block l, three core vectors, including $q_{(i)}^{(l,a)}$, $k_{(i)}^{(l,a)}$, and $v_{(i)}^{(l,a)}$ are computed from the representation $z_{(i)}^{(temp,l-1)}$ encoded by the preceding layer:

$$q_{(i)}^{(l,a)} = W_Q^{''(l,a)} LN(z_{(i)}^{(temp,l-1)}) \in R^{D_h}, \quad (14)$$

$$k_{(i)}^{(l,a)} = W_K^{''(l,a)} LN(z_{(i)}^{(temp,l-1)}) \in R^{D_h}, \quad (15)$$

$$v_{(i)}^{(l,a)} = W_V^{''(l,a)} LN(z_{(i)}^{(temp,l-1)}) \in R^{D_h}, \quad (16)$$

where $W_Q^{''(l,a)}$, $W_K^{''(l,a)}$, and $W_V^{''(l,a)}$ are the matrixes of query, key, and value in the temporal transformer, respectively. The temporal self-attention (TSA) score for $z_{(i,s)}^{(T,l-1)}$ in the a-th multi-head self-attention unit is given as follows:

$$\alpha_{(i)}^{(l,a)temp} = \sigma\left(\frac{q_{(i)}^{(l,a)}}{\sqrt{D_h}} \cdot \left[k_{(0)}^{(l,a)}\left\{k_{(i)}^{(l,a)}\right\}_{i=1,\ldots,M}\right]\right) \in R^M. \quad (17)$$

The equation shows that the TSA scores are merely computed over the temporal dimension. The TSA can calculate the contribution of multiple electrode features changing over different convolutional features to the final model decision at a specific time. An intermediate vector $s_{(i)}^{(l,a)}$ for encoding $z^{(temp,l-1)}$ is given as follows:

$$s_{(i)}^{(l,a)} = \alpha_{(i,0)}^{(l,a)} v_{(i,0)}^{(l,a)} + \sum_{j=1}^{M} \alpha_{(i,j)}^{(l,a)} v_{(i,j)}^{(l,a)} \in R^{D_h}. \quad (18)$$

The encoded feature $z^{(temp,l)} \in R^{M \times L1}$ by the l-th block is computed by first concatenating the intermediate vectors from all heads, and the vector concatenation is projected by matrix $W_O \in R^{L1 \times L}$. $z^{'(temp,l)}$ is the residual connection result of the projection of the intermediate vectors and the $z^{(temp,l-1)}$ encoded by the preceding block. Finally, the $z^{'(temp,l)}$ normalized by LN() is passed through a multilayer perceptron (MLP) using the residual connection. The output of the temporal transformer is represented by $O \in R^{M \times L1}$.

2.4. EEGformer decoder

The EEGformer is used to extract the temporal, regional, and synchronous characteristics in a unified manner, as well as to deal with various EEG-based brain activity analysis tasks. Unlike the original transformer decoder, which uses a multi-head self-attention mechanism to decode the feature output of the corresponding encoder, we designed a convolution neural network (CNN) to perform the corresponding task. The CNN contains three convolutional layers and one fully connected layer. Specifically, the output $O \in R^{M \times L1}$ of the EEGformer encoder is reshaped to $X \in R^{S \times C \times M}$, where M is the dimensional length of the encoded temporal feature. The first layer of our EEGformer decoder (with the weights $w_1 \in R^{C \times 1}$) linearly combined different convolutional features for normalization across the convolutional dimension. Thus, the output data shape of the first layer is $X_1 \in R^{S \times M}$. The motivation to convolve C feature maps along the convolutional dimension of X into one is to allow the network to make data–driven decisions about the contribution of different convolutional features to the final model decision. The second layer of our CNN was responsible for combining information across spatial dimensions of X and extracting the entire information while discarding redundancy or noninformative variations. To this end, our CNN convolved X along the spatial dimension using the weights $w_2 \in R^{S \times N}$ and returns the plane $X_2 \in R^{M \times N}$,

where N denotes the number of convolutional filters used in the second layer. The third layer halved the dimension and reduced the parameter complexity using the weights $w_3 \in R^{(M/2) \times N}$ to produce the output plane $X_3 \in R^{(M/2) \times N}$. The fourth layer of our CNN is a fully connected layer that produced classification results for the brain activity analysis task. The corresponding equation of the loss function is given as follows:

$$Loss = \frac{1}{D_n} \sum_{i=1}^{D_n} -\log\left(p_i\left(y_i\right)\right) + \lambda |w| \tag{19}$$

where D_n is the number of data samples in the training dataset, p_i and y_i are the prediction results produced by the model and the corresponding ground truth label for the i-th data sample, respectively, and λ is the constant of the L1 regularization.

3. Experiment results

3.1. Experimental setup

For generating the input of the EEGformer and other comparison models, we first extract the raw EEG data of each trial of the three datasets to form data samples and assign the corresponding label to each data sample. Further, we apply a sliding window with the step of $ratio \times SR$ (i.e., SR) on each data sample and generate the final input samples in a non-overlapping manner. The data shape of each input sample is $ratio \times SR \times N_c$, and the N_c denotes the number of EEG channels (i.e., 64). The equation for representing the relationship between segment length T and the total number of input samples N is given as follows:

$$\mathrm{N} = \frac{N_{sub} \times EPP \times TPE \times TLPT}{ratio}, \tag{20}$$

where N_{sub} denotes how many subjects joined in the corresponding data collection experiment. Taking the data splitting method for BETA dataset as an example, we remove the EEG data collected during the gaze shifting of 0.5 s guided by a visual cue and an offset of 0.5 s followed by the visual stimulation. The final BETA dataset consists of 11,200 trials and 40 categories. For the first 15 participants and the remaining 55 participants in the BETA dataset, the time length of the flickering visual stimulation in each trial is 2 and 3 s, respectively. When the number of data points of each input sample is 100, meaning the *ratio* is set to 0.4 and the *SR* is equal to 250 Hz, and the time length of each input sample is 0.4 s, the total number of input samples of the BETA dataset for training and testing models is 78,000. Under the same setting, the total number of input samples of the SEED and DepEEG dataset for training and testing models is 514,687 and 42,000.

The state-of-the-art DL models, which have performed well in previous studies, were tested on the three datasets to compare their model performance with ours. In our comparison, we followed the same test procedures for all these methods. The EEGformer and other comparison baselines were trained with a batch size of 64 and Adam optimizer with a learning rate of 0.001. In each transformer module, the number of encoding blocks is equal to three. The models were trained using an early-stop training strategy. Note that all training hypermeters were optimized using the testing data. Pytorch was used to implement these models, which were trained on an NVIDIA Tesla A100 GPU. As mentioned above, we tested our model on three datasets (BETA, SEED, and DepEEG). Fivefold cross-validation was applied to separate the dataset, and the average classification accuracy (ACC) rate, sensitivity (SEN), and specificity (SPE) and the corresponding standard deviation (SD) of them were used as model performance metrics. For multi-category classification, the accuracy rate, which means how many data samples are corrected and labeled out of all the data samples, is calculated as the sum of true positive and true negative divided by the sum of true positive, false positive, false negative, and true negative. The above metrics are calculated using the following formula:

$$\mathrm{ACC} = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}) \tag{21}$$

$$\mathrm{SEN} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}), \tag{22}$$

$$\mathrm{SPE} = \mathrm{TN}/(\mathrm{TN}+\mathrm{FP}), \tag{23}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

3.2. Comparison baselines

To show the effective model performance of EEGformer, we compared several commonly used DL methods in other studies of EEG-based data analysis tasks, which were target frequency identification, emotion recognition, and depression discrimination. The comparison models are described as follows:

(1) EEGNet (Lawhern et al., 2018). It is a Compact-CNN for EEG-based BCIs. The network starts with a temporal convolution operation to learn frequency filters. The operation is made up of F_1 convolutional filters, and each size equals $1 \times N$, where N represents the length of the convolutional filter. It used $D \times F_1$ depth-wise convolutional filters to learn frequency-specific spatial filters and the size of each filter is $C \times 1$. The separable convolution followed by point-wise convolution was used to learn the summary for each feature map and optimally combine them. The network architecture shows that EEGNet considers temporal and spatial information of EEG signals.

(2) Conv-CCA (Waytowich et al., 2018). It is designed for SSVEPs-based target frequency identification and can be used in other EEG-based classification tasks. Unlike pure DL models, the Conv-CCA uses a signal-CNN with three-layers to transform multiple channel EEGs ($N_s \times N_c \times 1$) into a single $\bar{x}$ with a shape of $N_s \times 1 \times 1$, where N_s and N_c are the numbers of sampling points and channels, respectively. Another reference CNN with two-layers was used to transform the reference signal ($N_s \times N_f \times N_c$) into a 2D signal $\bar{Y}$ with a shape of $N_s \times N_f$, where N_f is the number of target frequencies. Correlation analysis was used to calculate the coefficients of $\bar{x}$ and each $\bar{Y}^n$ for all $n \in [1, N_f]$. A dense layer with N_f units and a softmax activation function was used as the final layer for classification.

(3) 4DCRNN (Shen et al., 2020). It is a DL model known as a four-dimensional (4D) convolutional recurrent neural network that extracts and fuses frequency, spatial and temporal information from raw EEG data to improve model performance of emotion recognition. It is not an end-to-end model for BCI tasks because it requires the Butterworth filter to decompose frequency bands and manually extract differential entropy features from each frequency band. The model input is represented as a 4D structure $X \in R^{h \times w \times d \times 2T}$, where h and w are the height and width of the 2D brain topographical map, respectively, d denotes the number of frequency bands and T denotes the length of the signal segment. CNN was used to extract the frequency and spatial information from each temporal segment of an EEG sample, and long short-term memory (LSTM) was adopted to extract temporal information from CNN outputs.

(4) EmotionNet (Wang et al., 2018). Instead of using 2D convolution filters to extract features from input data, EmotionNet used a 3D convolution filter to learn spatial and temporal features from raw EEG data. The first two layers and the third layer of the model used a 3D convolution filter to learn spatiotemporal and fuse spatial features, respectively. The fourth and fifth layers of the model performed temporal feature extraction using a 2D convolutional filter. The sixth layer of the model is a fully connected layer for dense predictions.

(5) PCRNN (Yang et al., 2018). The model is an end-to-end DL model known as a parallel convolutional recurrent neural network for EEG-based emotion classification tasks. It also takes 3D shape ($X \in R^{h \times w \times T}$) of raw EEG data as model input. CNN model was first used to learn spatial feature maps from each 2D map, and the LSTM was used to extract temporal features from the CNN outputs. Note that the CNN and LSTM were organized by a parallel structure to extract the spatial and temporal features from the model input. The outputs of the parallel structure were integrated to classify emotions.

3.3. Ablation studies

3.3.1. Effect of the EEGformer decoder constructed by different transformer combination

We conducted an ablation study to show the effectiveness of the EEGformer by constructing the encoder with different combinations of temporal, synchronous, and regional transformers. The classification results (ACC, SPE, SEN, and their corresponding SDs) on the three EEG datasets using different transformer module combinations to construct the EEGformer encoder are shown in Table 2. The table shows that the EEGformer encoder constructed by the combinations of the three transformers achieves the best classification results. For BETA dataset, the average sensitivity, specificity, and accuracy are 69.86, 75.86, and 70.15%, respectively. For SEED dataset, the average sensitivity, specificity, and accuracy are 89.14, 92.75, and 91.58%, respectively. For DepEEG dataset, the average sensitivity, specificity, and accuracy are 77.83, 70.95, and 72.19%, respectively. The result supports our hypothesis that a machine learning method can capture EEG characteristics in a unified manner that can suit the EEG-based brain activity analysis tasks.

The table also demonstrates that the EEGformer that contains a synchronous transformer achieves better model performance than the EEGformer without a synchronous transformer. For instance, the EEGformer constructed using a single synchronous transformer outperforms the EEGformer constructed using the other two types of single transformers, with better accuracy of 57.29, 80.12, and 60.12% on BETA, SEED, and DepEEG, respectively. The EEGformer constructed using a transformer pair consisting of a synchronous transformer outperforms the EEGformer constructed using the transformer pair without a synchronous transformer, with better accuracy on the BETA, SEED, and DepEEG datasets. The results indicate the significance of learning spatial distribution characteristics of EEG activity generated by multiple brain regions for the task of SSVEPs-based frequency discrimination. In addition, the EEGformer constructed using synchronous transformer and regional transformer outperforms the EEGformer constructed using other transformer pairs, with better classification results on SEED and DepEEG dataset. On the one hand, the result demonstrates that the convolutional features could represent regional and spatial characteristics of EEH signal well. On the other hand, the result indicates that the integration of the synchronous and regional EEG characteristics improves discrimination ability of our model.

3.3.2. Effect of using 1DCNN or not to construct the EEGformer pipeline

The model performance affected by using 1DCNN or not is validated to show the rationality of using a 1D depth-wise convolutional filter to learn regional characteristics in a completely data–driven manner. Figure 3 compares the results of using 1DCNN or not constructing the EEGformer pipeline. The figure shows that using a 1D depth-wise convolutional filter to learn regional characteristics is beneficial for improving model performance in EEG-based classification tasks.

3.3.3. Effect of EEG channel number on the model performance

Table 3 reports the classification results (ACC, SPE, SEN, and their corresponding SDs) of our model with varying number of EEG channel. The EEG CHN and the corresponding name of brain regions are illustrated as follows: 3 (O1, Oz, and O2), 6 (O1, Oz, O2, POz, PO3, and PO4), 9 (O1, Oz, O2, Pz, PO3, PO5, PO4, PO6, and POz), 32 channels (all channels from occipital, parietal, central-parietal regions and C3, C1, Cz, C2, C4, and FCz) as well as all 64 channels. From the table, we can know that as the EEG CHN increases, the classification results of the EEGformer show an upward trend. This result indicates that although the EEG channels that are placed over the occipital and parietal regions provide perhaps the most informative SSVEP signals, other channels are informative as well. The result also illustrates the data mining ability of our model, which can learn representational features from complex data structure.

3.4. Comparison studies

Leave-one-subject-out (LOSO) cross-validation method is utilized to compare the model performance between EEGformer

TABLE 2 Classification results (ACC, SPE, SEN, and their corresponding SDs) on the three EEG datasets by using different transformer module combinations to construct EEGformer encoders.

Combinations	BETA			SEED			DepEEG		
	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)
Reg	41.63 ± 5.91	46.59 ± 3.58	35.67 ± 3.26	76.53 ± 1.68	77.26 ± 2.41	73.58 ± 1.94	58.78 ± 5.21	60.51 ± 2.58	57.25 ± 3.42
Syn	57.29 ± 6.50	62.86 ± 5.89	55.28 ± 4.69	80.12 ± 5.12	82.83 ± 4.65	78.86 ± 2.71	60.12 ± 4.86	65.94 ± 3.59	55.26 ± 4.27
Temp	45.36 ± 7.18	53.38 ± 6.38	43.86 ± 5.68	77.28 ± 4.12	78.29 ± 3.83	76.69 ± 3.82	61.73 ± 4.12	65.82 ± 4.78	60.83 ± 2.65
Temp + Syn	66.52 ± 3.82	70.25 ± 2.97	62.23 ± 4.32	85.36 ± 3.61	88.36 ± 4.75	83.45 ± 2.86	70.15 ± 3.18	68.97 ± 3.56	75.65 ± 4.81
Temp + Reg	59.29 ± 3.27	65.93 ± 2.65	58.79 ± 3.54	80.12 ± 3.19	82.33 ± 2.08	79.16 ± 3.19	65.21 ± 2.89	62.14 ± 4.72	72.31 ± 3.75
Syn + Reg	65.72 ± 2.91	70.85 ± 2.58	61.23 ± 5.12	86.73 ± 2.95	88.04 ± 2.36	83.77 ± 3.76	71.46 ± 2.85	61.96 ± 2.36	75.64 ± 3.19
Temp + Syn + Reg	70.15 ± 2.18	75.86 ± 2.04	69.86 ± 3.29	91.58 ± 2.77	92.75 ± 3.72	89.14 ± 2.98	72.19 ± 2.67	70.95 ± 2.38	77.83 ± 2.15

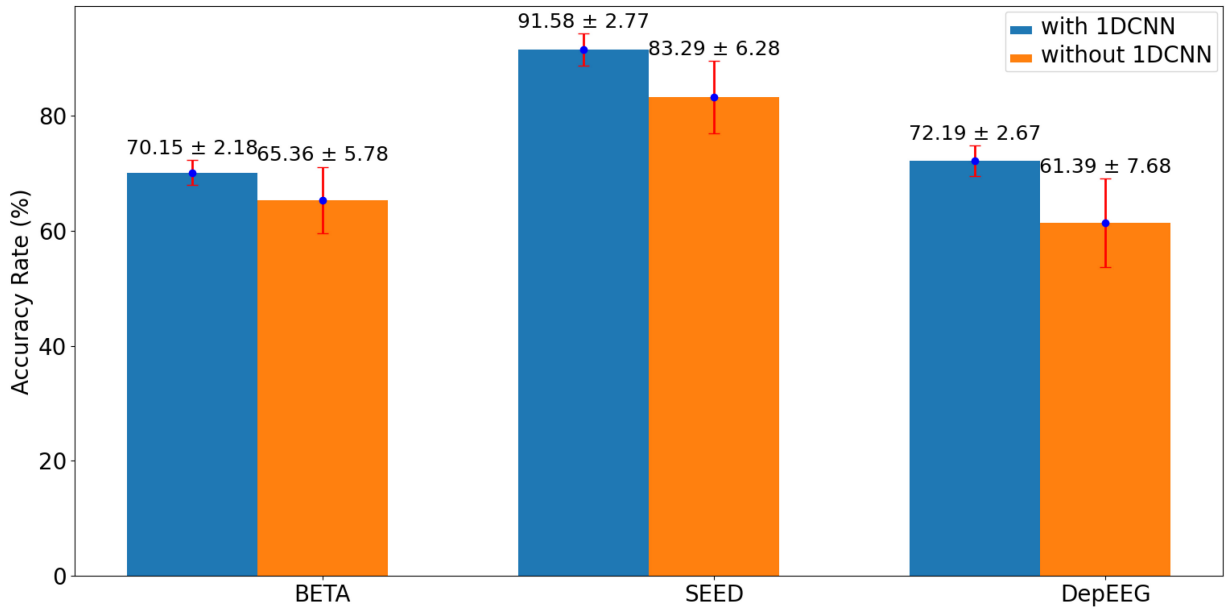


FIGURE 3
Comparison results of using 1DCNN or not to construct the EEGformer pipeline.

TABLE 3 Classification results (ACC, SPE, SEN, and their corresponding SDs) of our model is reported versus varying number of channels and 1.0 s of stimulation.

Channel number	BETA			SEED			DepEEG		
	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)
3	42.73 ± 3.60	50.73 ± 5.17	36.83 ± 4.39	69.54 ± 3.86	70.49 ± 2.96	66.76 ± 4.85	51.29 ± 2.99	50.86 ± 3.75	55.71 ± 4.51
6	50.86 ± 4.49	63.69 ± 2.38	55.17 ± 6.73	73.21 ± 2.83	74.62 ± 3.79	73.61 ± 2.73	56.74 ± 3.85	54.14 ± 2.64	60.26 ± 3.29
9	56.52 ± 2.17	70.46 ± 3.96	65.89 ± 5.26	76.37 ± 3.72	77.24 ± 4.21	78.18 ± 3.82	61.21 ± 4.74	59.75 ± 3.82	65.78 ± 2.79
32	65.21 ± 3.05	72.17 ± 2.57	65.36 ± 4.74	85.98 ± 3.16	86.91 ± 2.64	86.27 ± 4.54	68.56 ± 2.38	65.37 ± 3.57	70.39 ± 4.26
64	70.15 ± 2.18	75.86 ± 2.04	69.86 ± 3.29	91.58 ± 2.77	92.75 ± 3.72	89.14 ± 2.98	72.19 ± 2.67	70.95 ± 2.38	77.83 ± 2.15

and other five comparison methods. As shown in Figure 4, the upper figure shows accuracy comparison results between EEGformer and Conv-CCA across using BETA dataset, and the lower figure shows standard deviation comparison between EEGformer and other five comparison methods across subjects using BETA dataset. The reason of only choosing Conv-CCA to compare with EEGformer is both of them achieve high accuracy on the BETA dataset. From the Figure 5, we can find that EEGformer achieves the lowest standard deviation among other comparison methods, indicating the proposed method generalizes well on unseen data and potentially requires little to model training and calibration for new users, suitable for SSVEP classification tasks.

1. Accuracy comparison between EEGformer and Conv-CCA across subjects using BETA dataset.

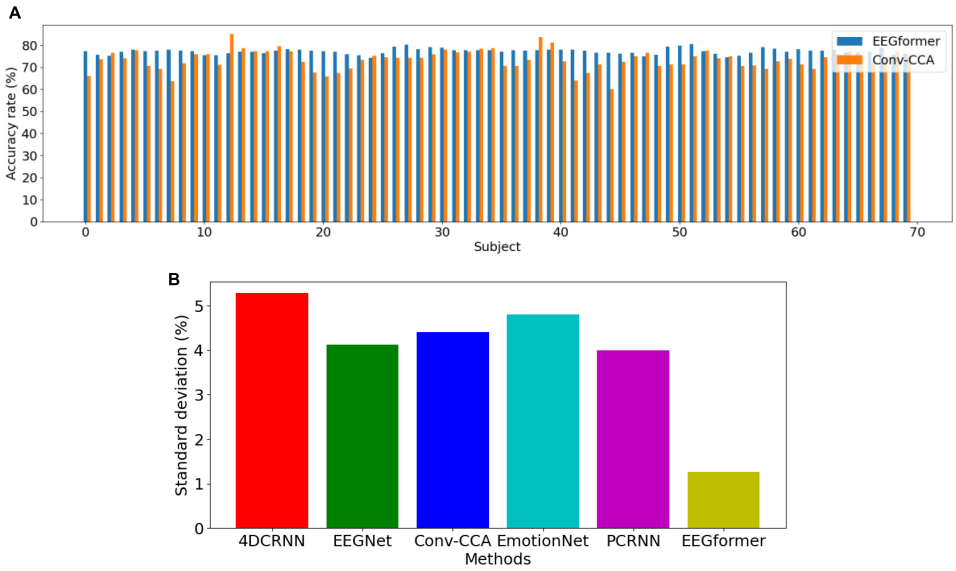


FIGURE 4
Performance comparison between EEGformer and other five comparison methods using leave-one-subject-out cross-validation method based on BETA dataset. **(A)** Accuracy comparison between EEGformer and Conv-CCA across subjects using BETA dataset. **(B)** Standard deviation comparison between EEGformer and other five comparison methods across subjects using BETA dataset.

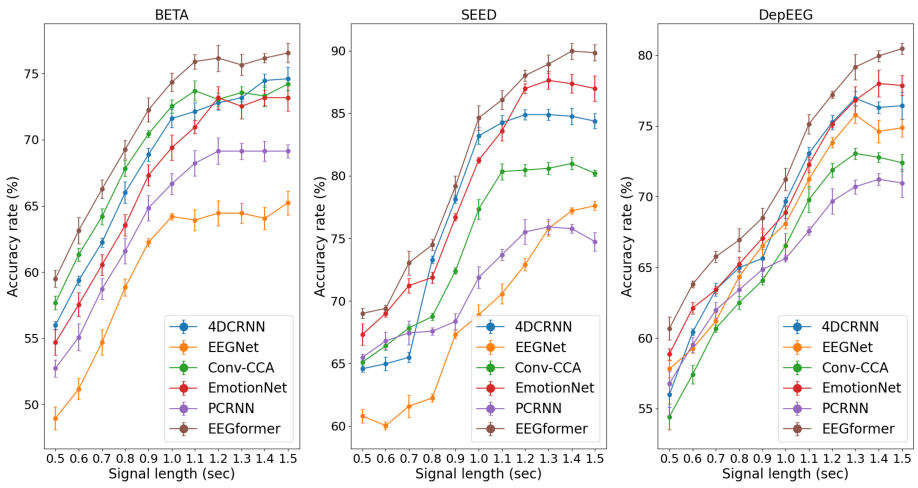


FIGURE 5
Performance (average ACC ± SD %) of segment length *T* using the EEGformer and other comparable models on the three EEG datasets.

2. Standard deviation comparison between EEGformer and other five comparison methods across subjects using BETA dataset.

Furthermore, according to the SSVEP studies, they pursue a higher information transfer rate by not using long EEG segments to execute the target frequency identification task. The model performance can be improved by increasing the segment length *T* because longer EEG segments contain more information about brain activity. Therefore, we investigated the impact of segment length *T* ranges [0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5] on model performance. The performance (average ACC and SD) of segment length *T* using the EEGformer and other comparable models on the three EEG datasets are shown in Figure 5. The figure shows that our model achieves the best accuracy rate across the three datasets. For other comparison baseline models, the model performance reduces in some cases if the segment length *T* exceeds 1.2 s. The model performance of the EEGformer on the three datasets showed an increasing trend as the segment length *T* increases, indicating that our method can extract inherent temporal information from EEG and is unaffected by segment length. In addition, the model performance of 4DRCNN and EmotionNet outperforms the performance of other comparison baselines. Because 4DRCNN and the EmotionNet are models that learn spatiotemporal features simultaneously, this operation may facilitate the DL model to learn better feature representation of EEG regional and synchronous characteristics.

4. Discussion

The abovementioned ablation and comparison studies show the rationality of our EEGformer architecture and demonstrate that our model performs outperforms other comparison baselines. This section covers several noteworthy points and future works:

(1) The unified manner, sequentially maps an input sequence into an abstract continuous representation that holds temporal, convolutional, and spatial information of that input outperforms the 2D and 3D structures that integrate frequency, spatial and temporal information of EEG. The EEGformer achieved the highest accuracy rate compared with other comparison baselines, which could be due to the unified EEG characteristics learning manner. Compared with 4DRCNN, which requires the user to manually extract frequency information from raw EEG data and use it as model input, our model is an end-to-end deep method because it uses depth-wise 1DCNN to learn the feature in an EEG-channel-wise manner. In the EEGformer encoder, we sequentially encode the convolutional results generated by the 1DCNN from temporal, convolutional, and spatial dimensions. The temporal, regional, and synchronous transformers were responsible for learning the temporal, regional, and synchronous characteristics of EEG signals. This type of feature learning strategy contains more cues of EEG characteristics than other model structures and performs better than them.

(2) EEG signals are well-known to exhibit data statistics that can drastically change from one subject to another in various aspects (e.g., regional characteristics), but also share similarities in certain other aspects (e.g., synchronous characteristics). To exploit the commonalities while tackling variations, we require a large data sample to train the model and improve its generalization ability. However, the performance of a DL model is always affected by the dataset size. Compare with the dataset size in the computer vision studies, researchers find it difficult to collect a dataset with a similar size in EEG-based clinical studies. Therefore, increasing the number of EEG datasets used for training DL models is crucial to reduce the influence of small dataset size on model performance. To this end, many studies separate the EEG signal collected in a trial into several segments and label them with the same label. Those segments were then used in cross-subject and within-subject classifications, which are two commonly used experimental designs, to execute model training and validate model performance. Meanwhile, those studies also designed model training strategies to improve the model generalization ability. For instance, Guney et al. (2021) trained their model in two stages: the first stage trains globally with all the available data from all the subjects, and then the second stage fine-tunes the model individually using the data of each subject separately. In the future, we can also design a training strategy to reduce the influence of small dataset size on model performance.

(3) Although the experimental results demonstrated that learning temporal, regional, and spatial characteristics in a unified manner facilitates the EEGformer to achieve promising classification performance across three EEG datasets, this result might be unable to provide strong support for clinical treatment that is associated with EEG biomarkers. Because DL methods are essentially considered black boxes, we require novel methods to open the box and visualize the feature learned by the DL model. To this end, an emerging technique known as explainable artificial intelligence (AI) enables the understanding of how DL methods work and what drives their decision-making. The competitive model performance of DL methods and the explainable AI provided a promising way to support effective EEG-based brain activity analysis. By using the explainable AI method, we could visualize the form of the temporal, regional, and spatial characteristics learned by the EEGformer and use it to connect with BFC, as well as perform brain activity analysis.

5. Conclusion

In this study, we proposed a transformer–based EEG analysis model known as EEGformer to capture EEG characteristics in a unified manner. The EEGformer consists of 1DCNN, an EEGformer encoder (sequentially constructed by three components: regional, synchronous, and temporal transformers), and an EEGformer decoder. We conducted ablation studies to demonstrate the rationality of the EEG former. The results not only supported our hypothesis that a machine learning method capable of capturing the EEG characteristics in a unified manner can be applied to EEG-based brain activity analysis tasks but also demonstrated that convolutional features could accurately represent regional and spatial characteristics of EEG signals. The LOSO cross-validation method is utilized to compare the model performance between EEGformer and other five comparison methods, the result shows the proposed method generalizes well on unseen data and potentially requires little to model training and calibration for new users, suitable for SSVEP classification tasks. We also investigate the impact of segment length T on model performance, and the results show that our method can extract inherent temporal information from EEG and is unaffected by the segment length. The proposed EEGformer outperforms the comparison models, which perform well in other studies on the three EEG datasets.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://bci.med.tsinghua.edu.cn/download.html and https://bcmi.sjtu.edu.cn/home/seed/seed.html.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of Beijing Anding Hospital of Capital Medical University. The patients/participants provided their written informed consent to participate in the

data collection. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable data included in this article.

Author contributions

ZW, ML, and WD contributed to the conception and design of the study. SL and ML performed the data analysis. ZW and WD drafted the manuscript. JH and HT participated in editing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant (62161024), China Postdoctoral Science Foundation under Grant (2021TQ0136 and 2022M711463), and the State Key Laboratory of Computer Architecture (ICT, CAS) Open Project under Grant (CARCHB202019).

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

Abdull, M. M., Chandler, C., and Gilbert, C. J. (2016). Glaucoma,"the silent thief of sight": Patients' perspectives and health seeking behaviour in Bauchi, northern Nigeria. *BMC Ophthalmol.* 16:44. doi: 10.1186/s12886-016-0220-6

Chen, J., Zhang, Y., Pan, Y., Xu, P., and Guan, C. (2022). A Transformer-based deep neural network model for SSVEP classification. *arXiv* [Preprint]. arXiv:2210.04172.

Du, Y., Xu, Y., Wang, X., Liu, L., and Ma, P. (2022). EEG temporal–spatial transformer for person identification. *Sci. Rep.* 12:14378.

Duan, R. N., Zhu, J. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG-based emotion classification," in *Proceedings of the international IEEE/EMBS conference on neural engineering* (San Diego, CA), 81–84. doi: 10.1109/NER.2013.6695876

Guedes, R. A. P. (2021). Glaucoma, collective health and social impact. *Rev. Bras.Oftalmol.* 05–07. doi: 10.5935/0034-7280.20210001

Guney, O. B., Oblokulov, M., and Ozkan, H. J. (2021). A deep neural network for ssvep-based brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 69, 932–944. doi: 10.1109/TBME.2021.3110440

Ibáñez-Soria, D., Soria-Frisch, A., Garcia-Ojalvo, J., and Ruffini, G. (2019). Characterization of the non-stationary nature of steady-state visual evoked potentials using echo state networks. *PLoS One* 14:e0218771. doi: 10.1371/journal.pone.0218771

Khok, H. J., Koh, V. T., and Guan, C. (2020). "Deep multi-task learning for SSVEP detection and visual response mapping," in *Proceedings of the IEEE international conference on systems, man, and cybernetics (SMC)*, (Toronto, ON), 1280–1285. doi: 10.1109/SMC42975.2020.9283310

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

Li, X., Wei, W., Qiu, S., and He, H. (2022). "TFF-Former: Temporal-frequency fusion transformer for zero-training decoding of two BCI tasks," in *Proceedings of the 30th ACM international conference on multimedia*, (Lisboa), 51–59. doi: 10.1145/3503161.3548269

Li, Y., Xiang, J., and Kesavadas, T. J. (2020). Convolutional correlation analysis for enhancing the performance of SSVEP-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2681–2690. doi: 10.1109/TNSRE.2020.3038718

Li, Z., Wang, Q., Zhang, S. F., Huang, Y. F., and Wang, L. Q. (2022). Timing of glaucoma treatment in patients with MICOF: A retrospective clinical study. *Front. Med.* 9:986176. doi: 10.3389/fmed.2022.986176

Liu, B., Huang, X., Wang, Y., Chen, X., and Gao, X. J. (2020). BETA: A large benchmark database toward SSVEP-BCI application. *Front. Neurosci.* 14:627. doi: 10.3389/fnins.2020.00627

Nentwich, M., Ai, L., Madsen, J., Telesford, Q. K., Haufe, S., Milham, M. P., et al. (2020). Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI. *Neuroimage* 218:117001. doi: 10.1016/j.neuroimage.2020.117001

Qin, K., Wang, R., and Zhang, Y. (2021). Filter bank-driven multivariate synchronization index for training-free SSVEP BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 934–943. doi: 10.1109/TNSRE.2021.3073165

Raut, R. V., Snyder, A. Z., Mitra, A., Yellin, D., Fujii, N., Malach, R., et al. (2021). Global waves synchronize the brain's functional systems with fluctuating arousal. *Sci. Adv.* 7:eabf2709. doi: 10.1126/sciadv.abf2709

Schielke, A., and Krekelberg, B. (2022). Steady state visual evoked potentials in schizophrenia: A review. *Front. Neurosci.* 16:988077. doi: 10.3389/fnins.2022.988077

Shen, F., Dai, G., Lin, G., Zhang, J., Kong, W., and Zeng, H. (2020). EEG-based emotion recognition using 4D convolutional recurrent neural network. *Cogn. Neurodyn.* 14, 815–828. doi: 10.1007/s11571-020-09634-1

Tsoneva, T., Garcia-Molina, G., and Desain, P. (2021). SSVEP phase synchronies and propagation during repetitive visual stimulation at high frequencies. *Sci. Rep.* 11:4975. doi: 10.1038/s41598-021-83795-9

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30:15. doi: 10.48550/arXiv.1706.03762

Wan, Z., Huang, J., Zhang, H., Zhou, H., Yang, J., and Zhong, N. (2020). HybridEEGNet: A convolutional neural network for EEG feature learning and depression discrimination. *IEEE Access* 8, 30332–30342. doi: 10.1109/ACCESS.2020.2971656

Wang, Y., Huang, Z., McCane, B., and Neo, P. (2018). "EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro, Brazil), 1–7. doi: 10.1109/IJCNN.2018.8489715

Waytowich, N., Lawhern, V. J., Garcia, J. O., Cummings, J., Faller, J., Sajda, P., et al. (2018). Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *J. Neural Eng.* 15, 066031. doi: 10.1088/1741-2552/aae5d8

Yang, Y., Wu, Q., Qiu, M., Wang, Y., and Chen, X. (2018). "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proceedings of the international joint conference on neural networks*, (Rio de Janeiro: IEEE), 1–7. doi: 10.1109/IJCNN.2018.8489331

Zhang, X., Yao, L., Wang, X., Monaghan, J., McAlpine, D., and Zhang, Y. (2021). A survey on deep learning-based non-invasive brain signals: Recent advances and new frontiers. *J. Neural Eng.* 18:031002. doi: 10.1088/1741-2552/abc902

Zhang, Y., Xie, S. Q., Wang, H., and Zhang, Z. J. (2020). Data analytics in steady-state visual evoked potential-based brain–computer interface: A review. *IEEE Sens. J.* 21, 1124–1138. doi: 10.1109/JSEN.2020.3017491

Zhang, Y., Yin, E., Li, F., Zhang, Y., Guo, D., Yao, D., et al. (2019). Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs. *Neural Netw.* 119, 1–9. doi: 10.1016/j.neunet.2019.07.007

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zhou, Y., He, S., Huang, Q., and Li, Y. J. (2020). A hybrid asynchronous brain-computer interface combining SSVEP and EOG signals. *IEEE Trans. Biomed Eng.* 67, 2881–2892. doi: 10.1109/TBME.2020.2972747

Zhuang, X., Yang, Z., and Cordes, D. J. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Hum. Brain Mapp.* 41, 3807–3833. doi: 10.1002/hbm.25090

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 06 April 2023
DOI 10.3389/fnins.2023.1156990

OPEN ACCESS

EDITED BY
Zhi Wen,
Renmin Hospital of Wuhan University,
China

REVIEWED BY
Chen-Xing Qi,
Eye Center,
Renmin Hospital of Wuhan University,
China
Zhijiang Wan,
Nanchang University,
China
Xiang Liu,
The First Affiliated Hospital of Nanchang University,
China

*CORRESPONDENCE
Yu-Lin Zhong
804722489@qq.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 02 February 2023
ACCEPTED 17 March 2023
PUBLISHED 06 April 2023

CITATION
Duan M-M, Liu H and Zhong Y-L (2023) Effect of shape deprivation on retinal thickness in myopic mice using an OCT method. *Front. Neurosci.* 17:1156990. doi: 10.3389/fnins.2023.1156990

Effect of shape deprivation on retinal thickness in myopic mice using an OCT method

Ming-Ming Duan[1], Hui Liu[2] and Yu-Lin Zhong[2]*

[1]Department of ophthalmology, the First People's Hospital of Jiujiang City, Jiujiang, Jiangxi, China, [2]Department of ophthalmology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, Jiangxi, China

Purpose: The purpose of this study was to study in retina thickness changes in myopic mice using optical coherence tomography (OCT).

Methods: There were 18 mice in the form-deprivation myopia (FDM) group，in which the left eye was not treated as a control;18 untreated mice served as a normal control group. The diopter of all mice was measured 21days after birth (P21), before form deprivation. After 4weeks of form deprivation (P49), the refraction, fundus, and retinal sublayer thickness of all mice were measured.

Results: After 4weeks of form deprivation, the refractive power of the right eye in the FDM group was significantly higher than that in the left eye (p <0.05). There was no significant change in the refractive power of the left eye in the FDM group compared with the normal control group. The retina, nerve fiber layer (NFL), inner nuclear layer (INL), and outer nuclear layer (ONL) in the right eye of the FDM group were significantly thinner than those of both the FDM and control groups (p <0.05). There was no significant change in photoreceptor (PR).

Conclusion: Our study highlights that the myopic mice have decreased R thickness, which might reflect the potential pathological mechanism of myopia.

KEYWORDS

form-deprivation myopia, retina thickness, optical coherence tomography, myopia mice, myopia

1. Introduction

Myopia is one of the most common diseases in ophthalmology. The disease can affect almost all of the population and has reached epidemic levels, especially in East and Southeast Asia (Jung et al., 2012; Saw et al., 2002; He et al., 2004). It has become a major public health problem that needs to be solved urgently. It is estimated that, by 2050, nearly half of the world's people will be suffering from myopia (Holden et al., 2016). With the increase in the incidence of early-onset myopia (Myrowitz, 2012), more and more people will develop high myopia and pathological myopia. Pathological myopia leads to blindness, such as glaucoma, retinal detachment, and macular hole. A previous study suggests that the occurrence of myopia is caused by the interaction of genetic and environmental factors (Baird et al., 2020), and the environmental factors include the close work or study time, outdoor activities time and the excessive sugar intake. Some studies suggest the hyperopic defocusing causes axial length growth, which might lead to the decreased retina thickness.

However, the pathogenesis is still unclear. At present, low concentrations of atropine and orthokeratology have achieved results in controlling the development of myopia (Bullimore & Richdale, 2020). However, the pathological structural changes in intraocular structure of the sclera, choroid, and retina (R) caused by myopia are irreversible (Lee et al., 2020; Jonas & Panda-Jonas, 2019).

Optical coherence tomography (OCT) is a non-invasive imaging method *in vivo*, which provides a high-resolution image of the retina, and can be used for quantitative and qualitative evaluation of different regions and layers of retina and optic nerve (Bhende et al., 2018). It can also be used to evaluate the relationship between refractive state and retinal thickness (Bonnin et al., 2015). OCT suggests that the retina and choroid will become thinner in myopic children (Matalia et al., 2018), and with the increase of axial length, the retinal nerve fiber layer (NFL) becomes thinner (Leung et al., 2006). As in myopia, atrophy around the optic papilla and thinning of the retinal NFL can be observed. However, there are also studies that suggest the opposite, in which the whole and temporal NFL of high myopia is thicker than that of the normal control group (Hsu et al., 2013). Therefore, our study of the changes of retinal structure and thickness might reveal the potential pathological mechanism of myopia, providing a new direction for the prevention and control of myopia.

The emergence of myopic animal models has promoted greatly the research on the occurrence, development, and treatment of myopia. Form deprivation in childhood can induce the change of myopia refractive state and prolongation of eye axis. This phenomenon has been confirmed in monkeys, sloths, marmosets, mice, guinea pigs, and chickens (Wiesel & Raviola, 1977; Schaeffel et al., 1988; Barathi et al., 2008; Shaikh et al., 1999; Graham & Judge, 1999), and it provides a basis for the establishment of a suitable animal model of myopia. Mice have become the first choice of myopic animal model because of the possibility of genetic manipulation, abundant available antibodies, and low-cost breeding. A large number of experiments have studied the changes of intraocular structure in myopic mice. However, there are few studies on the interlamellar thickness of the retina in myopic mice. The hypothesis of the study is that the form-deprivation myopia (FDM) causes the changes of retinal interlamellar thickness. Based on the hypothesis, our study is to establish a mouse model of FDM and use OCT method to study the changes of retinal interlamellar thickness in myopia.

2. Materials and methods

2.1. Animals

The study was approved by the Ethics Committee of Jiangxi Provincial People's Hospital (Nanchang, China). The experimental procedure adhered strictly to the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research.

In our study, 36 male C57BL/6 wild type mice (3 weeks old) were obtained from the Jiangxi Laboratory Animal Center. Mice were examined clinically to confirm no injuries or infections to the eyes and the refractive level remained consistent. After being divided randomly into 2 groups, all animals were kept at a constant temperature of 25 deg. Celsius. A 12-h light–dark cycle provided the light conditions for visual development in mice. Refractive errors in the FDM group (n = 18) were induced by wearing a translucent diffuser on the right eye from postnatal day 21 (P21). The contralateral untreated eyes in the FDM group served as a self-control group. The remaining untreated mice (n = 18) served as the control group.

2.2. Form deprivation

After anesthetizing with an injection of avertin, the model of shape deprivation myopia was constructed by covering the right eye of FDM group (n = 18) with a handmade translucent occluded device for 4 weeks. A collar made of thin plastic was attached to the neck to prevent the mice from removing the translucent balloon. For both the FDM group and control group, weight, refraction, and R were measured before and at the end of the 4-week treatment. After 4 weeks of form deprivation, mice with corneal ulcers and lesions were removed. To minimize measurement errors, all measurements were performed by one researcher.

2.3. Biometric measurements

Unanesthetized mice were placed in a dark room and their refractive status was measured with an infrared eccentric photorefractor at a working distance of 50 cm. At 3 min before the measurement, the mice were given eye drops of tropicamide in both eyes to dilate the pupils and facilitate the measurement of refractive status. After gently limiting the movement of the mouse, the head position of the mouse was adjusted to image the center of the pupil successfully, and the eyeball diopter number was read at this time. Each mouse was measured 5 times alternately in both eyes and averaged.

In the present study, we used OCT (Phoenix, MICRON IV) to measure the thickness of R in the mouse eye *in vivo*. After anesthetizing with an injection of avertin, the test eye was dilated with tropicamide, and ofloxacin ointment was applied to the corneal surface to facilitate clear fundus imaging. To improve image clarity, each OCT image is composed of 50 instantaneous layers. After OCT image acquisition, retinal stratification and thickness analysis were performed using Insight software. The thickness of different parts of the mouse R was different; here we focused on the 200 μm area adjacent to the optic nerve for analysis.

2.4. Statistical analysis

Data analysis was conducted using SPSS version 19.0. The deprived eyes and contralateral eyes of FDM mice were compared by paired t-test. Independent t-test was applied to compare the control and FDM eyes. After confirming the normal distribution of the data (Shapiro–Wilk test), analysis of variance (ANOVA) was used to compare the data of each parameter at different times or the four sets of data. Kruskal-Wallis test was used for data that did not conform to the normal distribution. In our study, data were expressed as mean ± standard deviation (SD) in the table or as mean ± SD in the figure. $P<0.05$ was considered statistically significant.

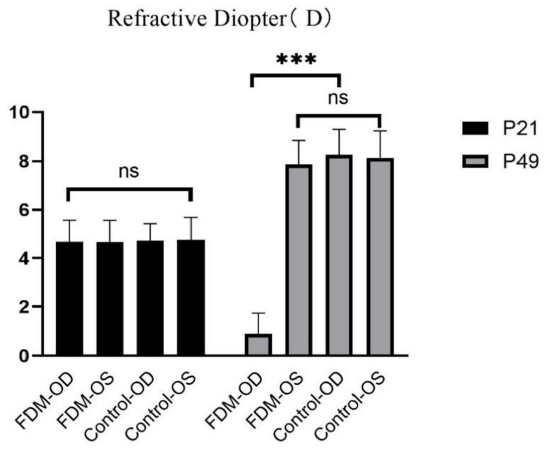


FIGURE 1
Refractions between the FDM and control groups. FDM: form-deprivation myopia; OD: oculus dexter; OS: oculus sinister.

TABLE 1 Refractions between the FDM and control groups.

group	n	Diopter (D)	
		P21	P49
FDM-od	18	4.68 ± 0.89	0.88 ± 0.84
FDM-os	18	4.66 ± 0.90	7.87 ± 0.96*
Control-od	18	4.72 ± 0.70	8.26 ± 1.02*
Control-os	18	4.76 ± 0.93	8.14 ± 1.08*
F		0.048	241.670
P		0.986	0.000

FDM: form-deprivation myopia; OD: oculus dexter; OS: oculus sinister. * indicate P < 0.05

3. Results

As shown in Figure 1 and Table 1, after 4 weeks of continuous form deprivation treatment in FDM group, the refractive power between the left and right eyes was significantly different in the FDM group. There were no statistically significant differences among left eyes in the FDM group and bilateral eyes in the control group. Figure 2 showed the fundus of left and right eyes of mice in FDM group after 4 weeks of form deprivation treatment, without obvious abnormal indications. Figure 3 showed the OCT images of the right and left eyes of mice in the FDM group after 4 weeks of treatment. Combined with Table 2, it could be seen that the retinal thickness (include NFL, INL, and ONL) of the right eye of mice in the FDM group was significantly thinner than that of the left eye and the control group (P<0.001). Figure 4 showed the retinal thickness of each layer in the bar graph between two groups.

4. Discussion

At present, researchers have successfully established common myopia models such as monkey, tree shrew, mouse, guinea pig, and chicken. There are differences in the structure and developmental characteristics of the eyeball in these common myopia models, so their applications in myopia research are also different. Mice are easy to obtain and raise, and their eyeball structure is similar to that of humans. In addition, the mouse is highly fertile and has complete genome information and mature genetic manipulation methods, which make this animal model widely used in myopia research. There are two main types of classical myopia induction methods: form-deprivation myopia (FDM) and lens-induced myopia (LIM). In this study, we found the R, NFL, inner nuclear layer (INL), and outer nuclear layer (ONL) around the optic nerve were thinner in myopic mice with FDM method. Thinning of the R caused by myopia has been reported in many studies, with a large number of studies suggesting that the R thins in myopia (Xie et al., 2009; Zereid & Osuagwu, 2020). However, some studies have reported an increase in retinal thickness in the central macular area of high myopia (Guo et al., 2020; Kim et al., 2019). Further analysis of the thickness of different retinal layers showed that the macular ganglion cell complex (GCC)(Guo et al., 2019), NFL (Ucak et al., 2020), and inner plexiform layer (Lee et al., 2020) were also thinned in myopia. Because previous studies on myopic retinal thickness changes are controversial. To clarify this issue, in this study, we used OCT to analyze the interlayer thickness of the eyes in myopic mice because it is non-invasive and can obtain high-definition images of the intraocular cross-section. Form deprivation caused the thinning of all retinal layers. The number of cells in the nucleated cell layer, photoreceptor cell layer, inner nuclear cell layer and ganglion cell layer decreased, and their arrangement was sparse and disordered. The thinning of the R might be caused by the stretching effect of the increase of the eye axis or by the thinning of different layers of the R.

The thickness of the NFL around the optic nerve in this study was significantly lower than that in the control group. Similar findings have been found in other myopic animals, such as thinning of the NFL in myopic chickens (Swiatczak et al., 2019), and patients with high myopia also have progressive loss of R NFL around the optic nerve papilla (Chopra & Lee, 2019), with some studies suggesting that the thinning of NFL is related to the blood perfusion of the R and the decrease of microvessels (Wang et al., 2016; Ucak et al., 2020; Li et al., 2017), and it is suggested that there is a correlation between retinal ischemia and NFL thinning. Myopia is a risk factor for open-angle glaucoma and studies have shown that the damage of the NFL structure often occurs before the visual-field damage and the thinning of the NFL in myopia might also affect the development of glaucoma in some young patients (Lee et al., 2017). The length of eye axis is correlated negatively with the thickness of NFL and the thinning of the NFL might be a result of the stretching effect of eye axis growth (Ganekal et al., 2021). However, contrary to our results, an earlier study suggests that only the nasal NFL is lower than normal, the thickness of the NFL of the upper and lower nasal side is normal, and even the whole and temporal NFL is thicker than normal (Hsu et al., 2013).

Our study found that the INL and ONL around the optic nerve were also significantly thinner. Similar changes were found in other animal models of form deprivation myopia. For example, INL thinning was found in myopic tree shrew and chickens (Mao et al., 2006; Abbott et al., 2011). Similar to our results, the thickness of the inner and outer layers of the R decreased in high myopia and thinning

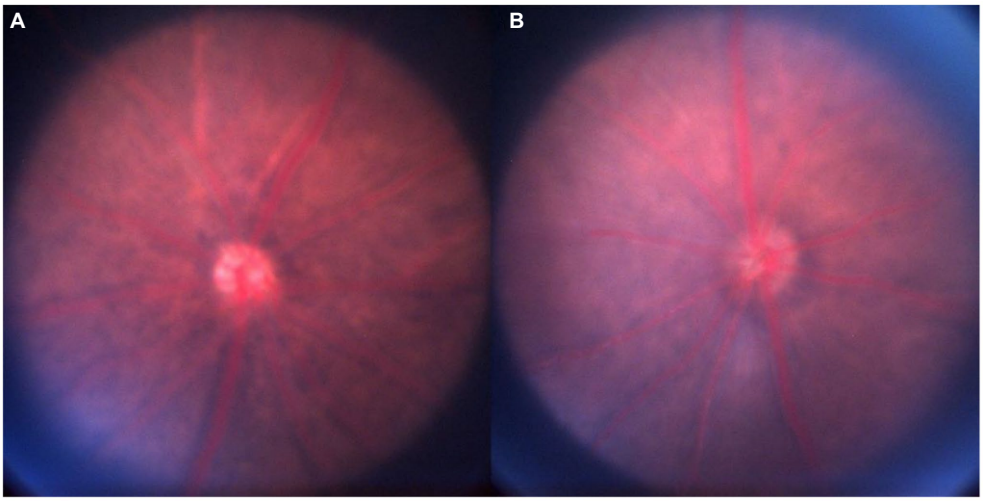


FIGURE 2
Fundus imaging. Notes: **(A)**: FDM-OD; **(B)**: FDM-OS. FDM: form-deprivation myopia; OD: oculus dexter; OS: oculus sinister.

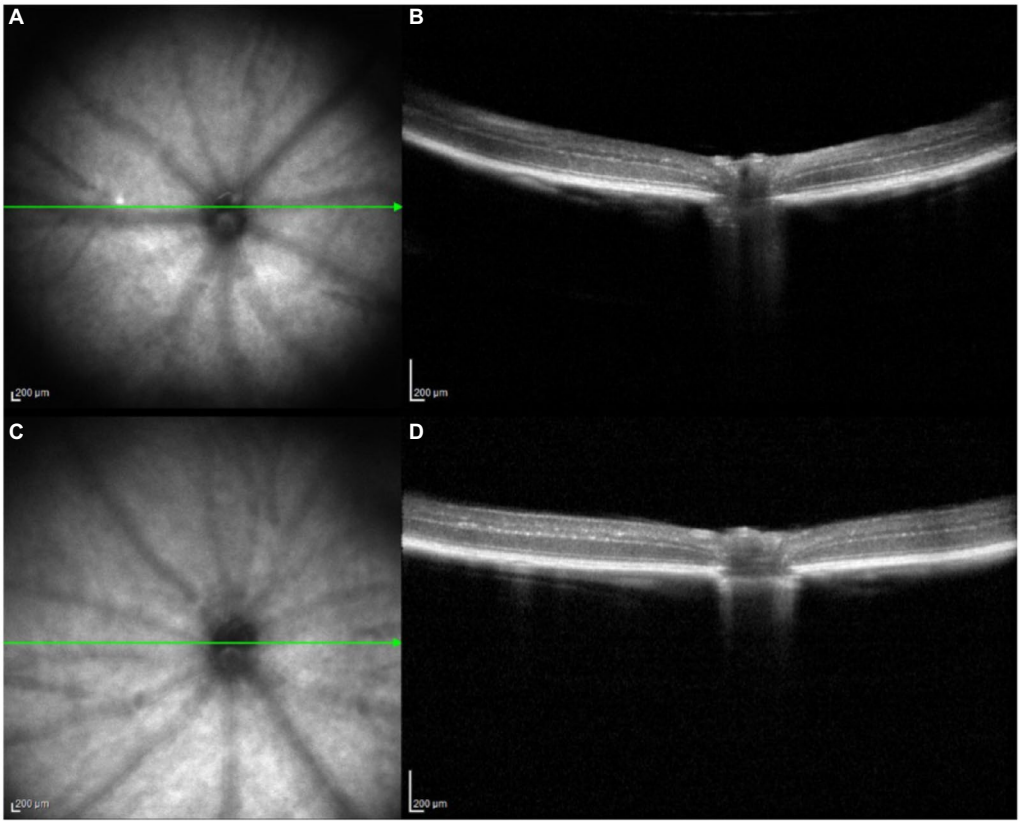


FIGURE 3
OCT images for the FDM-OD and FDM-OS. Notes: **(A, B)**: FDM-OD; **(C, D)**: FDM-OS. FDM: form-deprivation myopia; OD: oculus dexter; OS: oculus sinister; OCT: optical coherence tomography.

of the INL and outer plexiform layer (OPL) was also found in anisometropia myopia (Kim et al., 2020; Kirik et al., 2021). Some studies have suggested that the thinning of the inner R might be caused by the tangential tensile force caused by the elongation of the axial length，and the thinning of the outer R is caused by anterior and posterior traction (Wu et al., 2008). Recent studies have found that increased release of dopamine in myopic chicks leads to cell activation and cell density of bipolar cells and amacrine cells in INL

TABLE 2 Thickness of the retina between the FDM and control groups (mean±s, μm).

Groups	n	R	NFL	INL	ONL	PR
FDM-od	18	201.97 ± 8.59	16.36 ± 1.76	25.45 ± 2.55	47.99 ± 4.49	21.70 ± 2.19
FDM-os	18	220.93 ± 12.83*	23.58 ± 3.61*	29.98 ± 3.17*	57.08 ± 4.09*	21.48 ± 2.17
Control-od	18	224.65 ± 12.38*	23.39 ± 3.75*	29.53 ± 2.73*	55.46 ± 1.68*	22.33 ± 2.50
Control-os	18	221.44 ± 9.56*	25.28 ± 3.67*	28.62 ± 3.20*	56.04 ± 4.53*	23.42 ± 2.29
F		15.862	38.376	22.314	26.967	2.574
P		0.000	0.000	0.000	0.000	0.061

FDM: form-deprivation myopia; R: retina; NFL: nerve fiber layer; INL: inner nuclear layer; ONL: outer nuclear layer; PR: photoreceptor; OD: oculus dexter; OS: oculus sinister.
*indicate $p < 0.05$.

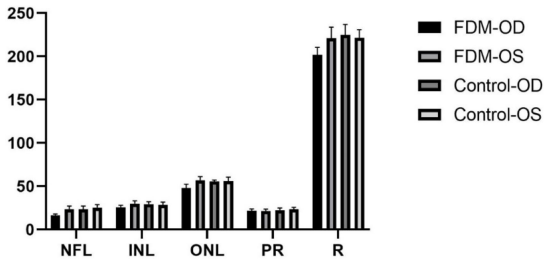


FIGURE 4
Thickness of the retina between the FDM and control groups. FDM: form-deprivation myopia; OD: oculus dexter; OS: oculus sinister; R: retina; NFL: nerve fiber layer; INL: inner nuclear layer; ONL: outer nuclear layer; PR: photoreceptor.

(Mathis et al., 2020). The thinning of the INL might also be the atrophy caused by the decrease of the activation of retinal inner layer cells in myopia.

The ONL is composed mainly of cell bodies of rods and cones and our study suggests that ONL becomes thinner. Similar to our results, studies have confirmed that ONL is thinner in high myopia and the packing density and regularity of cones in the outer layer of the R decreased (Park et al., 2013; Wang et al., 2019). As the axial length lengthens, the elongation and thinning of the R might lead to sparse cones and the choroid of myopia will become thinner and the blood flow will decrease (Zhang et al., 2019; Wong et al., 2017). The oxygen supply of the outer layer of the R is mostly provided by the choroid, therefore, the thinning of the outer layer of the R might also be caused by hypoxia caused by myopia. Some studies have suggested that the change of the thickness of the outer retinal sublayer is related to the microvessel density in the R (Ye et al., 2020). Our study suggests that the changes of photoreceptor are not statistically significant, but, unlike our results, there are other studies that suggest that the photosensitive layer of myopia becomes thinner (Chen et al., 2012) and, in pathological myopia, the myoid and ellipsoid zone also becomes thinner (Ye et al., 2020).

The study still has some shortcomings in the study. First, anterior segment OCT scanning should be performed on myopic mice. This measurement captures a number of indicators of the overall structure of the eye, including axial length, central corneal thickness, anterior chamber depth, retinal thickness, and corneal curvature radius. Secondly, the sample size should be expanded and the success rate of FDM modeling should be improved. Finally, in future experiments, we should not only test the mouse model, but also the guinea pig or macaque model which is more similar to the human visual system, so as to increase the universality and conviction of the theory.

In conclusion, our study highlights the decreased retina thickness in myopia mice, which might reflect the potential pathological mechanism of myopia.

Data availability statement

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

Ethics statement

The animal study was reviewed and approved by the medical ethics committee of the Jiangxi Provincial People's Hospital.

Author contributions

M-MD, HL, and Y-LZ contributed to data collection, statistical analyses, wrote the manuscript, designed the protocol, contributed to the MRI analysis, designed the study, oversaw all clinical aspects of study conduct, and manuscript preparation. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

Abbott, C. J., Grunert, U., Pianta, M. J., and McBrien, N. A. (2011). Retinal thinning in tree shrews with induced high myopia: optical coherence tomography and histological assessment [J]. *Vis. Res.* 51, 376–385. doi: 10.1016/j.visres.2010.12.005

Baird, P. N., Saw, S. M., Lanca, C., Guggenheim, J. A., Smith III, E. L., Zhou, X., et al. (2020). Myopia [J]. *Nat. Rev. Dis. Primers.* 6:99. doi: 10.1038/s41572-020-00231-4

Barathi, V. A., Boopathi, V. G., Yap, E. P., and Beuerman, R. W. (2008). Two models of experimental myopia in the mouse [J]. *Vis. Res.* 48, 904–916. doi: 10.1016/j.visres.2008.01.004

Bhende, M., Shetty, S., Parthasarathy, M. K., and Ramya, S. (2018). Optical coherence tomography: a guide to interpretation of common macular diseases [J]. *Indian J. Ophthalmol.* 66, 20–35. doi: 10.4103/ijo.IJO_902_17

Bonnin, S., Tadayoni, R., Erginay, A., Massin, P., and Dupas, B. (2015). Correlation between ganglion cell layer thinning and poor visual function after resolution of diabetic macular edema [J]. *Invest. Ophthalmol. Vis. Sci.* 56, 978–982. doi: 10.1167/iovs.14-15503

Bullimore, M. A., and Richdale, K. (2020). Myopia control 2020: where are we and where are we heading?[J]. *Ophthalmic Physiol. Opt.* 40, 254–270. doi: 10.1111/opo.12686

Chen, W., Wang, Z., Zhou, X., Li, B., and Zhang, H. (2012). Choroidal and photoreceptor layer thickness in myopic population [J]. *Eur. J. Ophthalmol.* 22, 590–597. doi: 10.5301/ejo.5000092

Chopra, V., and Lee, C. C. (2019). Progressive OCT retinal nerve fiber layer loss in myopia [J]. *Ophthalmology* 126, 529–530. doi: 10.1016/j.ophtha.2018.10.015

Ganekal, S., Sadhwini, M. H., and Kagathur, S. (2021). Effect of myopia and optic disc area on ganglion cell-inner plexiform layer and retinal nerve fiber layer thickness [J]. *Indian J. Ophthalmol.* 69, 1820–1824. doi: 10.4103/ijo.IJO_2818_20

Graham, B., and Judge, S. J. (1999). The effects of spectacle wear in infancy on eye growth and refractive error in the marmoset (Callithrix jacchus)[J]. *Vis. Res.* 39, 189–206. doi: 10.1016/s0042-6989(98)00189-8

Guo, Y., Pang, Y., Kang, Y., Zhang, X., Zhang, H., Zhang, G., et al. (2020). Correlations among peripapillary vasculature, macular superficial capillaries, and eye structure in healthy and myopic eyes of Chinese young adults (STROBE)[J]. *Medicine (Baltimore)* 99:e22171. doi: 10.1097/MD.0000000000022171

Guo, Y., Sung, M. S., and Park, S. W. (2019). Assessment of superficial retinal microvascular density in healthy myopia [J]. *Int. Ophthalmol.* 39, 1861–1870. doi: 10.1007/s10792-018-1014-z

He, M., Zeng, J., Liu, Y., Xu, J., Pokharel, G. P., and Ellwein, L. B. (2004). Refractive error and visual impairment in urban children in southern China [J]. *Invest. Ophthalmol. Vis. Sci.* 45, 793–799. doi: 10.1167/iovs.03-1051

Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K. S., Sankaridurg, P., et al. (2016). Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050[J]. *Ophthalmology* 123, 1036–1042. doi: 10.1016/j.ophtha.2016.01.006

Hsu, S. Y., Chang, M. S., Ko, M. L., and Harnod, T. (2013). Retinal nerve fibre layer thickness and optic nerve head size measured in high myopes by optical coherence tomography [J]. *Clin. Exp. Optom.* 96, 373–378. doi: 10.1111/cxo.12052

Jonas, J. B., and Panda-Jonas, S. (2019). Epidemiology and anatomy of myopia][J. *Ophthalmologe* 116, 499–508. doi: 10.1007/s00347-019-0858-6

Jung, S. K., Lee, J. H., Kakizaki, H., and Jee, D. (2012). Prevalence of myopia and its association with body stature and educational level in 19-year-old male conscripts in Seoul, South Korea [J]. *Invest. Ophthalmol. Vis. Sci.* 53, 5579–5583. doi: 10.1167/iovs.12-10106

Kim, J. H., Lee, S. H., Han, J. Y., Kang, H. G., Byeon, S. H., Kim, S. S., et al. (2019). Comparison of individual retinal layer thicknesses between highly myopic eyes and Normal control eyes using retinal layer segmentation analysis. *Sci. Rep.* 9:14000. doi: 10.1038/s41598-019-50306-w

Kim, T. Y., Lee, M. W., Baek, S. K., and Lee, Y. H. (2020). Comparison of retinal layer thicknesses of highly myopic eyes and Normal eyes [J]. *Korean J. Ophthalmol.* 34, 469–477. doi: 10.3341/kjo.2020.0012

Kirik, F., Ekinci, C., Akbulut, E., Bayraktar, H., and Ozdemir, H. (2021). Regional analysis of segmented-macular structure in patients with myopic anisometropia [J]. *Int. Ophthalmol.* 41, 3713–3726. doi: 10.1007/s10792-021-01934-7

Lee, J. E., Sung, K. R., Park, J. M., Yoon, J. Y., Kang, S. Y., Park, S. B., et al. (2017). Optic disc and peripapillary retinal nerve fiber layer characteristics associated with glaucomatous optic disc in young myopia [J]. *Graefes Arch. Clin. Exp. Ophthalmol.* 255, 591–598. doi: 10.1007/s00417-016-3542-4

Lee, M. W., Nam, K. Y., Park, H. J., Lim, H. B., and Kim, J. Y. (2020). Longitudinal changes in the ganglion cell-inner plexiform layer thickness in high myopia: a prospective observational study [J]. *Br. J. Ophthalmol.* 104, 604–609. doi: 10.1136/bjophthalmol-2019-314537

Lee, M. W., Lee, S. E., Lim, H. B., and Kim, J. Y. (2020). Longitudinal changes in axial length in high myopia: a 4-year prospective study [J]. *Br. J. Ophthalmol.* 104, 600–603. doi: 10.1136/bjophthalmol-2019-314619

Leung, C. K., Mohamed, S., Leung, K. S., Cheung, C. Y., Chan, S. L., Cheng, D. K., et al. (2006). Retinal nerve fiber layer measurements in myopia: an optical coherence tomography study [J]. *Invest. Ophthalmol. Vis. Sci.* 47, 5171–5176. doi: 10.1167/iovs.06-0545

Li, M., Yang, Y., Jiang, H., Gregori, G., Roisman, L., Zheng, F., et al. (2017). Retinal microvascular network and microcirculation assessments in high myopia [J]. *Am J. Ophthalmol.* 174, 56–67. doi: 10.1016/j.ajo.2016.10.018

Mao, J., Liu, S., Wen, D., Tan, X., and Fu, C. (2006). Basic fibroblast growth factor suppresses retinal neuronal apoptosis in form-deprivation myopia in chicks [J]. *Curr. Eye Res.* 31, 983–987. doi: 10.1080/02713680600910510

Matalia, J., Anegondi, N. S., Veeboy, L., and Roy, A. S. (2018). Age and myopia associated optical coherence tomography of retina and choroid in pediatric eyes [J]. *Indian J. Ophthalmol.* 66, 77–82. doi: 10.4103/ijo.IJO_652_17

Mathis, U., Feldkaemper, M., Wang, M., and Schaeffel, F. (2020). Studies on retinal mechanisms possibly related to myopia inhibition by atropine in the chicken [J]. *Graefes Arch. Clin. Exp. Ophthalmol.* 258, 319–333. doi: 10.1007/s00417-019-04573-y

Myrowitz, E. H. (2012). Juvenile myopia progression, risk factors and interventions [J]. *Saudi J Ophthalmol* 26, 293–297. doi: 10.1016/j.sjopt.2011.03.002

Park, S. P., Chung, J. K., Greenstein, V., Tsang, S. H., and Chang, S. (2013). A study of factors affecting the human cone photoreceptor density measured by adaptive optics scanning laser ophthalmoscope [J]. *Exp. Eye Res.* 108, 1–9. doi: 10.1016/j.exer.2012.12.011

Saw, S. M., Carkeet, A., Chia, K. S., Stone, R. A., and Tan, D. T. H. (2002). Component dependent risk factors for ocular parameters in Singapore Chinese children [J]. *Ophthalmology* 109, 2065–2071. doi: 10.1016/s0161-6420(02)01220-4

Schaeffel, F., Glasser, A., and Howland, H. C. (1988). Accommodation, refractive error and eye growth in chickens [J]. *Vis. Res.* 28, 639–657. doi: 10.1016/0042-6989(88)90113-7

Shaikh, A. W., Siegwart, J. J., and Norton, T. T. (1999). Effect of interrupted lens wear on compensation for a minus lens in tree shrews [J]. *Optom. Vis. Sci.* 76, 308–315. doi: 10.1097/00006324-199905000-00019

Swiatczak, B., Feldkaemper, M., Schraermeyer, U., and Schaeffel, F. (2019). Demyelination and shrinkage of axons in the retinal nerve fiber layer in chickens developing deprivation myopia [J]. *Exp. Eye Res.* 188:107783. doi: 10.1016/j.exer.2019.107783

Ucak, T., Icel, E., Yilmaz, H., Karakurt, Y., Tasli, G., Ugurlu, A., et al. (2020). Alterations in optical coherence tomography angiography findings in patients with high myopia [J]. *Eye (Lond.)* 34, 1129–1135. doi: 10.1038/s41433-020-0824-1

Ucak, T., Icel, E., Yilmaz, H., Karakurt, Y., Tasli, G., Ugurlu, A., et al. (2020). Alterations in optical coherence tomography angiography findings in patients with high myopia [J]. *Eye (Lond.)* 34, 1129–1135. doi: 10.1038/s41433-020-0824-1

Wang, X., Kong, X., Jiang, C., Li, M., Yu, J., and Sun, X. (2016). Is the peripapillary retinal perfusion related to myopia in healthy eyes? A prospective comparative study [J]. *BMJ Open* 6:e10791. doi: 10.1136/bmjopen-2015-010791

Wang, Y., Ye, J., Shen, M., Yao, A., Xue, A., Fan, Y., et al. (2019). Photoreceptor degeneration is correlated with the deterioration of macular retinal sensitivity in high myopia [J]. *Invest. Ophthalmol. Vis. Sci.* 60, 2800–2810. doi: 10.1167/iovs.18-26085

Wiesel, T. N., and Raviola, E. (1977). Myopia and eye enlargement after neonatal lid fusion in monkeys [J]. *Nature* 266, 66–68. doi: 10.1038/266066a0

Wong, C. W., Phua, V., Lee, S. Y., Wong, T. Y., and Cheung, C. M. G. (2017). Is choroidal or scleral thickness related to myopic macular degeneration?[J]. *Invest. Ophthalmol. Vis. Sci.* 58, 907–913. doi: 10.1167/iovs.16-20742

Wu, P. C., Chen, Y. J., Chen, C. H., Chen, Y. H., Shin, S. J., Yang, H. J., et al. (2008). Assessment of macular retinal thickness and volume in normal eyes and highly myopic eyes with third-generation optical coherence tomography [J]. *Eye (Lond.)* 22, 551–555. doi: 10.1038/sj.eye.6702789

Xie, R., Zhou, X. T., Lu, F., Chen, M., Xue, A., Chen, S., et al. (2009). Correlation between myopia and major biometric parameters of the eye: a retrospective clinical study [J]. *Optom. Vis. Sci.* 86, E503–E508. doi: 10.1097/OPX.0b013e31819f9bc5

Ye, J., Wang, M., Shen, M., Huang, S., Xue, A., Lin, J., et al. (2020). Deep retinal capillary plexus decreasing correlated with the outer retinal layer alteration and visual acuity impairment in pathological myopia [J]. *Invest. Ophthalmol. Vis. Sci.* 61:45. doi: 10.1167/iovs.61.4.45

Zereid, F. M., and Osuagwu, U. L. (2020). Myopia and regional variations in retinal thickness in healthy eyes [J]. *J. Ophthalmic Vis. Res.* 15, 178–186. doi: 10.18502/jovr.v15i2.6735

Zhang, S., Zhang, G., Zhou, X., Xu, R., Wang, S., Guan, Z., et al. (2019). Changes in choroidal thickness and choroidal blood perfusion in Guinea pig myopia [J]. *Invest. Ophthalmol. Vis. Sci.* 60, 3074–3083. doi: 10.1167/iovs.18-26397

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 13 April 2023
DOI 10.3389/fnins.2023.1160040

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Ruiqing Wu,
University of Electronic Science and Technology of China, China
Xiaohui Tao,
University of Southern Queensland, Australia
Ying Fang,
Shangqiu Normal University, China

*CORRESPONDENCE
Renping Zhu
✉ xgczrp@ncu.edu.cn
Wenfeng Duan
✉ 328301281@qq.com

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 06 February 2023
ACCEPTED 27 March 2023
PUBLISHED 13 April 2023

CITATION
Wan Z, Cheng W, Li M, Zhu R and Duan W (2023) GDNet-EEG: An attention-aware deep neural network based on group depth-wise convolution for SSVEP stimulation frequency recognition.
Front. Neurosci. 17:1160040.
doi: 10.3389/fnins.2023.1160040

GDNet-EEG: An attention-aware deep neural network based on group depth-wise convolution for SSVEP stimulation frequency recognition

Zhijiang Wan[1,2,3], Wangxinjun Cheng[4], Manyu Li[2], Renping Zhu[2,3,5]* and Wenfeng Duan[1]*

[1]The First Affiliated Hospital of Nanchang University, Nanchang University, Nanchang, Jiangxi, China, [2]School of Information Engineering, Nanchang University, Nanchang, Jiangxi, China, [3]Industrial Institute of Artificial Intelligence, Nanchang University, Nanchang, Jiangxi, China, [4]Queen Mary College of Nanchang University, Nanchang University, Nanchang, Jiangxi, China, [5]School of Information Management, Wuhan University, Wuhan, China

Background: Steady state visually evoked potentials (SSVEPs) based early glaucoma diagnosis requires effective data processing (e.g., deep learning) to provide accurate stimulation frequency recognition. Thus, we propose a group depth-wise convolutional neural network (GDNet-EEG), a novel electroencephalography (EEG)-oriented deep learning model tailored to learn regional characteristics and network characteristics of EEG-based brain activity to perform SSVEPs-based stimulation frequency recognition.

Method: Group depth-wise convolution is proposed to extract temporal and spectral features from the EEG signal of each brain region and represent regional characteristics as diverse as possible. Furthermore, EEG attention consisting of EEG channel-wise attention and specialized network-wise attention is designed to identify essential brain regions and form significant feature maps as specialized brain functional networks. Two publicly SSVEPs datasets (large-scale benchmark and BETA dataset) and their combined dataset are utilized to validate the classification performance of our model.

Results: Based on the input sample with a signal length of 1 s, the GDNet-EEG model achieves the average classification accuracies of 84.11, 85.93, and 93.35% on the benchmark, BETA, and combination datasets, respectively. Compared with the average classification accuracies achieved by comparison baselines, the average classification accuracies of the GDNet-EEG trained on a combination dataset increased from 1.96 to 18.2%.

Conclusion: Our approach can be potentially suitable for providing accurate SSVEP stimulation frequency recognition and being used in early glaucoma diagnosis.

KEYWORDS

group depth-wise convolution, EEG attention, SSVEPs, stimulation frequency recognition, EEG signal

1. Introduction

Glaucoma is one of the leading causes of blindness in the world. The damage to visual function caused by glaucoma is irreversible, and it can be difficult for the patients to realize this disease until their vision is damaged. According to the World Health Organization (WHO), the number of people living with glaucoma worldwide reached 76 million in 2020 and will rise to 95.4 million by 2030 (Guedes, 2021). China is one of the countries with the largest number of glaucoma patients. In 2020, the number of glaucoma patients in China reached 21 million, of which 5.67 million were blind (Soh et al., 2021). Glaucoma is generally not preventable, but most patients can maintain adequate vision in later life if detected early and appropriately treated. Therefore, early detection and diagnosis are significant for glaucoma blindness prevention. Traditional methods for assessing functional loss in glaucoma always adopt standard automated perimetry (SAP), which requires considerable subjective response from patients. The subjective assessment is limited by large test-retest variability, and may result in late diagnosis or delayed detection of progressive degeneration of retinal ganglion cells (RGCs).

Steady-state visual evoked potentials (SSVEPs) are typically recorded by electroencephalography (EEG) and reliably applied to brain-computer interface systems (BCIs). When exposed to a fixed frequency of visual stimuli, the brain's visual cortex produces a continuous frequency-dependent response (Nuzzi et al., 2018). This response known as SSVEPs can be used to assess functional abnormalities in visual pathways (Geethalakshmi et al., 2022). For glaucoma patients, due to the loss of peripheral vision, some constant frequency of repeated stimuli can no longer be received, so the corresponding stimulation frequency cannot be detected from the EEG brain signal (Lin et al., 2015; Chen et al., 2021, 2022a,b). Therefore, SSVEP can be considered as an objective assessment of visual field damage caused by glaucoma. For example, Lin et al., 2015 hypothesized that a brain region corresponding to a visual field deficit would be less perceivable and thereby would result in weaker SSVEP amplitude. Their study demonstrated that the SSVEP dynamics in terms of amplitude is capable of serving as objective biomarkers to assess visual field loss in glaucoma. Medeiros et al., 2016 produced nGoggle, a portable brain-based device, to assess the visual function deficits in glaucoma. Moreover, Nakanishi et al., 2017 investigated the ability of nGoggle equipment to discriminate glaucomatous from healthy subjects in a clinic-based setting. The aforementioned studies demonstrate the feasibility of using SSVEP signal to provide objective assessment of visual field damage.

The SSVEPs-based early detection for glaucoma requires effective analysis methods for recognizing stimulation frequencies. Traditional analysis methods for SSVEP signal can be mainly divided into two categories: spatial-spectral-temporal (SST) based method (Mora-Cortes et al., 2018; Salelkar and Ray, 2020; Zhang et al., 2020) and canonical correlation analysis (CCA) based method (Liu Q. et al., 2020; Cherloo et al., 2022; Ma et al., 2022). The former tries to extract SST features from the EEG signal and use them to execute classification tasks. Based on statistical analysis, the latter attempts to identify and measure the associations between the SSVEP signal and reference signal (e.g., sinusoidal signal). For example, Chen et al. (2015) construct the filter bank CCA (FBCCA) which decompose SSVEPs into multiple sub-band components under multiple pre-processing filters, then fuse the classifications from all sub-band. Although both achieve satisfactory results in SSVEPs-based applications, they require manually predefined algorithms based on expert knowledge to extract handcrafted features. This procedure is not flexible and may limit the usage of the method in SSVEPs-based applications. In recent years, convolution neural network (CNN) based deep learning (DL) methods have been widely used in processing SSVEPs-based frequency recognition tasks and achieved excellent performance (Khok et al., 2020). Combined with existing methods (e.g., SST analysis, CCA), CNN models use multiple layers to progressively extract higher-level features from model input and perform automatic feature extraction. Many advanced CNN-based technologies have been proposed in the recent years. For example, Li et al., 2022 proposed DSCNN, a dilated shuff CNN model for actualizing EEG-based SSVEP signal classification. Yao et al., 2022 constructed FB-EEGNet by fusing features of multiple neural networks for SSVEP target detection. To achieve reasonable model architecture with superior model performance, many studies designed the deep learning models specifically suited to the domain of EEG-based SSVEP signal classification. For example, Waytowich et al., 2018 proposed a Compact-CNN for classifying asynchronous SSVEPs. The Compact-CNN's architecture is similar to EEGNet (Lawhern et al., 2018), which performs two convolutional steps (temporal convolution and depth-wise convolution) sequentially to learn frequency and frequency-specific spatial filters, respectively. Guney et al., 2021 designed a novel deep neural network (DNN) to process the multi-channel SSVEP with convolutions across sub-bands of harmonics, channels, and time and classify them at a fully connected layer. Li et al., 2020 implemented a CNN-based non-linear model, i.e., convolutional correlation analysis (Conv-CA), which first uses CNNs at the top of a self-defined correlation layer.

Further, it utilizes the correlation layer to calculate the correlation coefficients between EEG and reference signals.

Previous studies of CNN-based SSVEP stimulation frequency recognition (Waytowich et al., 2018; Li et al., 2020; Guney et al., 2021) have usually adopted one-dimensional (1D) temporal convolution to mimic a bandpass frequency filter for filtering the signal of each EEG channel, followed by depth-wise spatial convolutions to combine the channels to obtain a better frequency pattern. Because the same 1D convolutional filter filters the data of each EEG channel, different rows in the same feature map contain the same EEG frequency components. The following depth-wise spatial convolution is used to learn spatial filters for each temporal filter, enabling the efficient extraction of frequency-specific spatial filters. However, the brain signal generated from different regions presents different harmonics in the same period (Atasoy et al., 2016; Retter et al., 2021), the frequency-specific spatial characteristics might be insufficient to reflect the diversity of brain signals in different brain regions. In addition, regional neural complexity and network functional connectivity may relate to the brain's information processing (McDonough and Nashiro, 2014). The regional neural complexity reflects the richness or diversity of brain signals in different brain regions, the more complex the regional neural activity, the higher functional connectivity this region has with other brain regions. Thus, it is reasonable to believe that diverse frequency combinations across different EEG channels may play an essential role in EEG-based brain activity classification. To simulate the regional characteristics of the EEG signal and reflect the diversity, we are interested in creating the different rows in the single feature map containing different frequency components. This motivates us to use different convolutional filters to process the EEG signal of different EEG channels.

Our brain is a coherent information processing system integrated by distributed and specialized networks (Ferraro et al., 2018). The current theory of brain functional networks suggests that the integration of specialized networks in the brain is facilitated by a set of essential nodes (Shine et al., 2016; Ferraro et al., 2018). The theory highlighted the significance of specialized networks and the relation between different specialized networks in evaluating brain function. Instead of using the connectivity of all brain regions, the connectivity features of partial brain regions might be more effective in representing different brain activities accurately. However, most existing combination studies of the DL and brain functional connectivity (BFC) focus on automatically learning the global connectivity feature of all brain regions (Babaeeghazvini et al., 2021; Avberšek and Repovš, 2022; Lin et al., 2022). Few concentrate on automatically learning the local connectivity features of specialized networks and the relations between different specialized networks. Considering different brain states involve different functional connectivity networks, we have reasons to believe the EEG characteristics over the local BFC network may contain useful classification information for discriminating different brain activities. The critical step of learning specialized network characteristics by the CNN model is identifying essential nodes. The attention mechanism (Vaswani et al., 2017; Lv et al., 2021) provides an automatic solution to identify essential nodes from whole brain regions since it can assign high attention weights for important regions. According to the definition in the field of computer vision (Chen et al., 2020), temporal-wise attention can assign weights to different EEG temporal segments collected in one experiment trail. Channel-wise attention can assign a higher weight to a more important feature map and refine feature maps. Spatial-wise attention can identify important feature regions in a single feature map. For example, Woo et al., 2018 propose convolutional block attention module (CBAM), sequentially infers attention maps using channel-wise attention and spatial-wise attention, then the attention maps are multiplied to the input feature map for adaptive feature refinement. To differentiate the three attention methods mentioned above, we use the terminology of EEG channel-wise to describe the attention operation for identifying important EEG channels (i.e., essential nodes) from a single feature map. The weight vector learned by the EEG channel-wise attention helps us to identify the EEG channels which are not important for the specialized network and emphasize the EEG channels which are essential to the specialized network. In addition, we re-term channel-wise attention as specialized network-wise attention to make our study easier to comprehend.

This study addresses the SSVEPs-based frequency recognition task as a multi-category classification problem. It proposes a novel CNN model named group depth-wise convolutional neural network (GDNet-EEG) to execute the task. To overcome the problem of the frequency-specific spatial characteristics might be insufficient to reflect the diversity of brain signals in different brain regions, we construct group depth-wise convolutional filter, which comprises C 1D depth-wise convolutional filter, to extract as diverse regional characteristics as possible from raw EEG data. Furthermore, to automatically learn the local connectivity features of specialized networks and the relations between different specialized networks, we propose EEG attention to sequentially infer attention maps along two dimensions (EEG channel and feature map): the former identifies essential brain regions to form a specialized network in a single feature map, and the latter infers important specialized networks across multiple feature maps. More specifically, the GDNet-EEG model is comprised of several group depth-wise convolutional layer. Each layer consists of multiple group depth-wise convolutional filter that employs C different 1D depth-wise convolutional filters to process the data outputted by the previous layer. Each depth-wise convolutional filter is separately utilized to process the signal of a single EEG channel and learn regional characteristics originating from different brain regions. C denotes the number of EEG channels, i.e., the row number of the feature map in every convolution layer is the same as the EEG channel number. We set K group depth-wise convolutional filters to generate K feature maps and adopt the same operation in the following convolution layers. Further, the EEG attention is embedded into the GDNet-EEG for learning essential nodes (i.e., significant EEG channel) and meaningful specialized networks (i.e., important feature map). For a feature map generated by a group depth-wise convolution layer, EEG attention first infers attention maps along the EEG channel dimension. Then the attention maps are multiplied by the feature maps for adaptive feature refinement. The refined feature map concerns important brain regions essential to a specialized network. After that, specialized network-wise attention is utilized to give further feature refinement to the different feature maps, highlighting the significance of different specialized networks. The main contributions of this study are depicted as follows:

(1) Unlike the previous studies adopted 1D temporal convolution followed by depth-wise spatial convolutions to extract frequency-specific spatial characteristics, we propose a deep neural network named GDNet-EEG, utilizing group depth-wise convolutional filter to extract regional characteristics from raw EEG data, for SSVEP stimulation frequency recognition. The advantage of using group depth-wise convolutional filter is that it can learn the regional characteristics of the EEG signal and reflect the diversity. The diverse frequency combinations across different EEG channels may be beneficial for EEG-based brain activity classification.
(2) Instead of using DL models to automatically learning the global connectivity feature of all brain regions from BFC matrix, we introduce attention mechanism to identify essential nodes and form specialized connectivity feature of the nodes to improve the performance of SSVEP stimulation frequency recognition. The EEG attention, containing EEG channel-wise attention and specialized network-wise attention, is proposed to identify important EEG channels from a single feature map and recognize important feature map as meaningful specialized networks.
(3) We have used two publicly available SSVEP datasets and their combination dataset consisting of the EEG data of 105 subjects with 40 target characters to validate the model performance of the GDNet-EEG. The related results have been presented to support the correctness of our study.

2. Materials and methods

2.1. Data description

Two SSVEP datasets (a benchmark dataset for SSVEPs-based BCI (Wang et al., 2016) (benchmark for short) and a large-scale benchmark database toward SSVEP-BCI application (BETA for short) (Liu B. et al., 2020)) and their combination dataset are used to validate the classification performance of the GDNet-EEG model. Each experiment of the benchmark dataset contains six sessions, and each session is comprised of 40 trials. The time length of each trial is 6 s which consists of three parts: gaze shifting of 0.5 s guided by a visual cue, visual stimulation of 5 s, and an offset of 0.5 s followed by the visual stimulation. A target character flickers at a specific frequency on screen in each trial, and the subject is asked to gaze at the flickering character for visual stimulation. The 40 stimulation frequencies are 8–15 Hz with 0.2 Hz strides, and there is a 0.5πphase difference between adjacent frequencies. The EEG data collected in each trial is down-sampled to 250 Hz.

The BETA dataset is similar to the benchmark dataset, and the main difference between them is illustrated as follows. The character matrix layout resembling the traditional QWERTY keyboard is used for the stimulus presentation in the experiment of BETA collection. In contrast, the corresponding layout in the experiment of the benchmark dataset is arranged in a square. The BETA dataset is collected from 70 healthy subjects. Each subject is asked to participate in 4 sessions of the experiment, and each session also consists of 40 trials. The time length of each trial is also comprised of three parts: gaze shifting of 0.5 s guided by a visual cue, visual stimulation of 2 or 3 s, and a rest time of 0.5 s followed by the visual stimulation. Visual stimulation of 2 s and 3 s are given to the first 15 subjects and the remaining 55 subjects, respectively. The EEG data collected in each trial is also down-sampled to 250 Hz.

2.2. Data preprocessing

A Chebyshev TypeIfilter filters the EEG signal collected in each trial with cutoff frequencies from 6 to 90 Hz and stopband corner frequencies from 4 to 100 Hz. The multi-channel EEG data collected in one trial is a 2D time series which can be represented by a data matrix X of size C^*Len, where C denotes the number of EEG channels, and Len means the signal length of visual stimulation in one-trial EEG record. The record is split into t segments $\{X_1, X_2,..., X_t\}$. The size of each segment X_i is C^*l, where l is the ratio of Len and t. Each segment X_t has a corresponding classification label L_t, and segments collected from the same trial have the same label. The L_t means the target frequency of the visual stimulus given to the subject in the corresponding trial.

2.3. GDNet-EEG construction

Figure 1 shows the architecture of the GDNet-EEG model, which contains a regular convolution layer, four group depth-wise convolution layers, a depth-wise separable convolution layer, and a dense layer. Note that the regular convolution layer and the depth-wise separable convolution layer are inherited from the EEGNet model to support the feature learning. Considering the pooling operation in the convolution results may cause the loss of meaningful features, we did not add a pooling layer to the GDNet-EEG model. Table 1 shows the specific parameters setting of the GDNet-EEG model. The specific operations of the GDNet-EEG are illustrated as follows:

2.3.1. Regular convolution layer

This layer aims at generating multiple frequency-specific feature maps which will be fed into the group depth-wise convolution layer for further feature learning. The input of the regular convolution layer is represented by $X_i \in R^{C*Ns}$ (i.e., a volume of 64 × 50 in the case of C = 64, Ns = 50 = T × = f_s with T = 0.2 s and f_s = 250 Hz). As shown in Table 1, 64 convolutional filters are utilized to process the input data, and the size of each filter is set to 1 × 17. Every filter sweeps the temporal and EEG channel dimensions in one stride. This layer is followed by batch normalization and linear activation layer. It utilizes the "SAME" padding mode to pad the input of the convolutional layer if the filter does not fit the input. The output of the layer is represented by $z_1 \in R^{C*Ns*64}$.

2.3.2. Group depth-wise convolution layer

The motivation for using this layer is to learn diverse regional EEG characteristics and deepen the neural network for achieving more abstract EEG features. This layer contains three subparts: group depth-wise convolutional layer, a batch normalization layer,

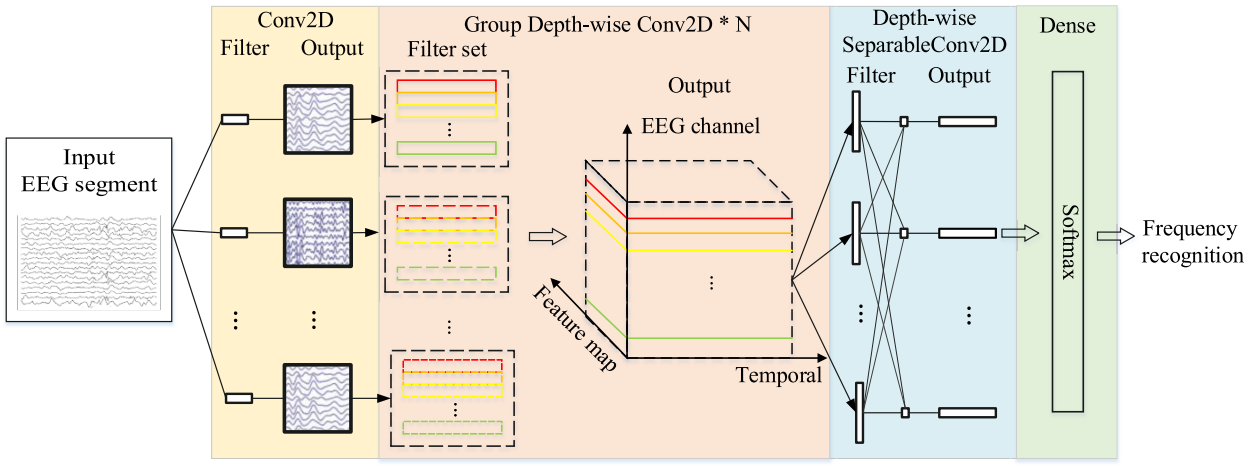


FIGURE 1
The architecture of the GDNet-EEG model for SSVEP stimulation frequency recognition.

TABLE 1 Specific parameters setting in the GDNet-EEG model, where *C* means the number of EEG channels, *T* denotes the number of time points, and *N* indicates the number of SSVEP stimulation frequencies.

Layer	Layer type	Output size	Hyperparameters
1	Input	(*C*, *Ns*)	
2	Conv2D	(*C*, *Ns*, 64)	$\left\{\begin{array}{l}1 \times 17,\ 64,\ stride\ 1\\ BatchNorm\\ Linear\ Activation\end{array}\right\} \times 1$, mode = same
3	Group depth-wise Conv2D	(*C*, *Ns* /16, 64)	$\left\{\begin{array}{l}1 \times 17,\ 64,\ stride\ 2\\ BatchNorm\\ Linear\ Activation\end{array}\right\} \times 4$, mode = same
4	Dropout	(*C*, *Ns* /16, 64)	rate = 0.5
5	Depth-wise Conv2D	(1, *Ns* /16, 64)	$\left\{\begin{array}{l}C \times 1,\ 64,\ stride\ 1\\ ELU\ Activation\end{array}\right\} \times 1$, mode = valid
6	Point-wise Conv2D	(1, *Ns* /16, 16)	$\left\{\begin{array}{l}1 \times 1,\ 16,\ stride\ 1\\ BatchNorm\\ ELU\ Activation\end{array}\right\} \times 1$, mode = same
7	Dropout	(1, *Ns* /16, 16)	rate = 0.5
8	Dense	*Nclass*	

and a linear activation layer. Unlike the traditional depth-wise separable convolutional operation, which utilizes a single depth-wise convolution to convolve the data of each feature map, the group depth-wise convolution employs C 1D depth-wise convolutional filters to convolve the EEG data of C channels simultaneously. More specifically, we can consider the C 1D depth-wise convolutional filters as a filter set that can produce a 2D feature map, and K (i.e., K = 64) filter sets produce K 2D feature maps. The Figure 1 has K dashed line frames in black, and each contains a filter set. The long frames with different colors (e.g., red, yellow, blue, or green) represent different depth-wise convolutional filters. The output of the group depth-wise convolution layer is represented by a three-dimensional (3D) feature cube comprised of a feature map, temporal, and EEG channel dimensions. If l = 0, layer l is the input layer, with the input being EEG fragment $X_m \in \mathrm{R}^{C*Ns*64}$. Let l ($1 \le l \le N$) be a group depth-wise convolution block. Then, the input of block l comprises m^{l-1} feature maps from the previous block. The output of block l consists of m^l feature maps. $Y_i^{c,l}$ denotes the row of the i^{th} feature map in block l where $c \in [1, C]$. The $Y_I^{c,l}$ is computed as follows:

$$Y_I^{c,l} = f\left(B_i^{c,l} + \sum_{j=1}^{m^{l-1}} K_{i,j}^{c,l} * Y_j^{c,l-1}\right) (l > \ = 1), \qquad (1)$$

where $B_i^{c,l}$ is bias matric, and $K_{i,j}^{c,l}$ is the convolution filter connecting the j^{th} feature map in block l-1 with the i^{th} feature map in block l. After the convolution operation, the leaky rectified linear unit (LeakyReLU) is used as the activation function f(·). The i^{th} feature map is obtained by stacking $Y_i^{c,l}s$ together. Every convolution filter shifts along the temporal dimension by stride

$s1$ (i.e., $s1$ = 2). The block l is followed by the dropout layer with a dropout rate of 0.5 and adopts the "SAME" padding mode considering the original elements in the layer input. From Table 1, we can see that the filter size (i.e., 1 × 17) equals the size used in the 2D convolutional filter. There are 4 group depth-wise convolution block in the layer, and the final output of the layer is represented by $z_2 \in R^{C*(Ns/16)*64}$. Compared with the depth-wise convolution layer in the Compact-CNN to classify 12 categories of SSVEP stimulus frequency, the group depth-wise convolution layer in our model covers the receptive field of the same size. It has a deeper model architecture with fewer parameters which is beneficial for avoiding over-fitting.

2.3.3. Depth-wise separable convolution layer

The motivation for using this layer is to (1) reduce the number of parameters to fit and (2) explicitly decouple the relationship within and across feature maps by first learning a kernel summarizing each feature map individually, then optimally merging the outputs afterward. More specifically, it firstly uses depth-wise spatial convolution in which the kernel shape is C*1 to convolve each 2D feature map into a 1D vector along the temporal dimension of each feature map. Then it utilizes point-wise convolution to combine information across feature map dimensions. The depth-wise spatial convolution layer employs exponential linear unit (ELU)'s nonlinearity and "VALID" padding mode. The filter number of the depth-wise spatial convolution layer is set to 64, and the output of the layer is represented by $z_3 \in R^{(Ns/16)*64}$. It is noteworthy that the depth-wise spatial convolution filter sweeps the data along temporal and EEG channel dimension in one stride and C stride, respectively. The point-wise layer is followed by batch normalization and dropout layer. The ELU activation and "SAME" padding mode are adopted in the point-wise convolutional layer. The point-wise convolutional layer employs the convolution filter with size of 1 × 1 to process the data, and the filter number of the point-wise convolution is set to 16 to reduce the number of parameters to fit. The output of the point-wise convolutional layer is denoted by $z_4 \in R^{(Ns/16)*16}$.

2.3.4. Dense layer and the corresponding loss function

The feature maps outputted by the depth-wise separable convolution layer are flattened and concatenated into one vector, fed into the dense layer. It is noteworthy that the GDNet-EEG model only contains one dense layer for avoiding high computation complexity. Let l be a dense layer, the identity activation function is utilized as activation function g(·), and the output of the i^{th} unit in layer l is computed as follows:

$$Z_i^l = g\left(\sum_{j=1}^{Ns} w_{i,j}^l Z_j^{l-1}\right), \tag{2}$$

where $w_{i,j}^l$, and Z_j^{l-1} denote the weights of the i^{th} unit in layer l and the outputs of layer (l-1), respectively. The outputs of the dense layer are passed into a softmax function for yielding stimulation frequency recognition results. Thus, the very first input X_i is predicted as $\hat{y}$ argmax s(Z_i^l), where s∈[0,1]^{Nclass} (i.e., $Nclass$ = 40) is the softmax output of the dense layer.

2.4. EEG attention module

Figure 2 shows the overall process of the EEG attention module. In the GDNet-EEG, the group depth-wise convolution block output is defined as feature map $F \in R^{C \times M \times Len}$, in which C represents the number of EEG channels, M means the number of feature maps, and Len indicates the length of convolution feature. F is fed into the EEG attention module as input. The EEG attention module sequentially infers a 2D EEG channel-wise attention map $M_{EC} \in R^{C \times M \times 1}$ and a 1D specialized network-wise attention vector $M_{SN} \in R^{M \times 1 \times 1}$. The process of the EEG attention module could be illustrated as:

$$F' = M_{EC}(F) \times F, \tag{3}$$

$$F'' = M_{SN}\left(F'\right) \times F', \tag{4}$$

where F' is the EEG channel-wise refined feature, calculated by multiplying EEG channel-wise attention map M_{EC} and the input feature F. The final output F," the feature for refining the specialized network, is calculated by multiplying specialized network attention M_{SN} and the EEG channel refined feature F'. The final output F" is fed into the next group depth-wise convolution block.

Figure 3 shows the overall process of the EEG attention module. The module includes two sequential parts: EEG channel-wise attention sub-module and specialized network-wise attention sub-module. The EEG channel-wise attention sub-module chooses essential brain regions from each feature map, regarded as a specialized network. The specialized network-wise attention sub-module acts on the feature map refined by the EEG channel-wise attention and generates an attention vector to represent the importance of different specialized networks. As the top part of Figure 3 shows, we have generated the EEG channel-wise attention map along the feature map dimension. Every feature map generated by the previous convolution layer is downsampled along the convolution feature dimension using both average and maximum pooling. Every feature map is down-sampled into a 1D vector whose length is the same as the EEG channel number. The data representation of the average-pooled feature $F_{avg}^{EC} \in R^{C \times M \times 1}$ and max-pooled feature $F_{max}^{EC} \in R^{C \times M \times 1}$ are 2D matrix, in which the row represents the EEG channel, and the column means feature map. We stack the F_{avg}^{EC} and F_{max}^{EC} together as the input of a separable convolution layer, which uses M 1*1 convolution filters to separately convolve the pooled feature stack along the EEG channel axis and generate M vectors. Every vector is passed into a sigmoid function to assign attention weight for EEG channels in every feature map. M attention weight vectors constitute the 2D EEG channel-wise attention map M_{EC}. The EEG channel-wise attention map is computed as follows:

$$M_{EC}(F) = \sigma\left(f^{M;1*1}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) = \sigma\left(f^{M;1*1}\left(\left[F_{avg}^{EC}, F_{max}^{EC}\right]\right)\right), \tag{5}$$

where σ means the sigmoid function and $f^{M;1*1}$ denotes a separable convolution network.

As the bottom part of Figure 2 illustrates, the input of the specialized network-wise attention is the feature maps

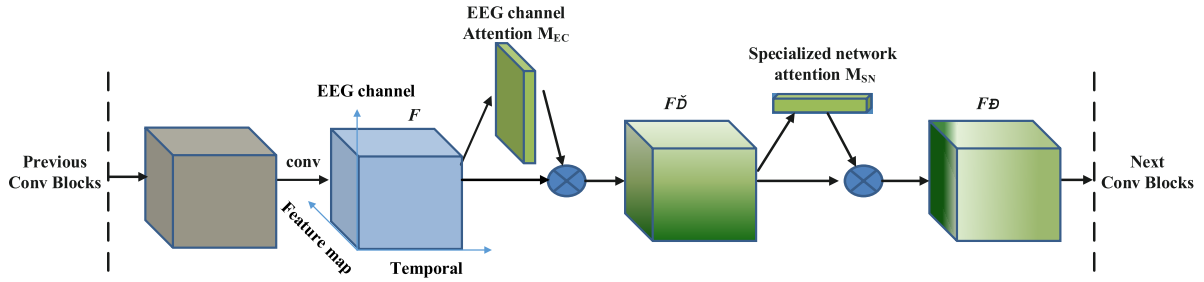


FIGURE 2
EEG attention integrated with a convolution block in GDNet-EEG.

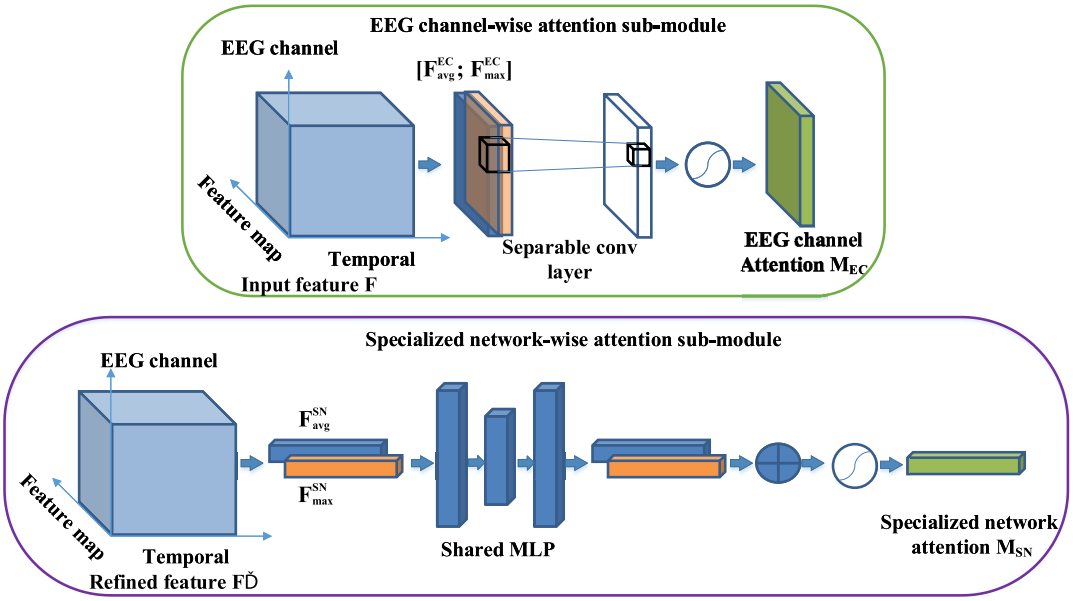


FIGURE 3
The overall process of the EEG attention module. The module includes two sequential parts: EEG channel-wise attention sub-module and specialized network-wise attention sub-module.

refined by the EEG channel-wise attention sub-module. These are the dot multiplication results of the 2D EEG channel-wise attention map M_{EC} and the original feature map F. The feature maps refined by the EEG channel-wise attention sub-module are pooled by using two pooling operations: average-pooled feature $F^{\mathrm{SN}}_{avg} \in \mathrm{R}^{M \times 1 \times 1}$ and max-pooled feature $F^{\mathrm{SN}}_{\mathrm{max}} \in \mathrm{R}^{M \times 1 \times 1}$. The two vectors are forwarded separately to a shared network composed of a multi-layer perceptron (MLP) with one hidden layer to produce two refined pooled vectors. After the shared network is applied to each descriptor, we merge the output feature vectors using element-wise summation. The specialized network-wise attention is computed as follows:

$$\begin{aligned} M_{SN}(F) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}\left(F''\right)\right) + MLP\left(MaxPool\left(F''\right)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F^{SN}_{avg}\right)\right) + W_1\left(W_0\left(F^{SN}_{max}\right)\right)\right), \end{aligned} \quad (6)$$

where σ denotes the sigmoid function, W_0 and W_1 are the MLP weights shared for average-pooled vector F^{SN}_{avg} and max-pooled vector F^{SN}_{max}.

3. Results

3.1. Experimental setup

The EEG data collected during the visual stimulation period is kept. To split the raw EEG data collected in each session into EEG segments, we remove the EEG data collected during the gaze shifting of 0.5 s guided by a visual cue and an offset of 0.5 s followed by the visual stimulation. The benchmark dataset contains 8,400 trials and 40 categories, and the time length of the flickering visual stimulation in each trial is 5 s. The BETA dataset consists of 11,200 trials and 40 categories. For the first 15 participants and the remaining 55 participants in the BETA dataset, the time length of the flickering visual stimulation in each trial is 2 and 3 s, respectively. For generating the input of the GDNet-EEG and

other comparison models, we first extract the raw EEG data of each trial of the two datasets to form data samples and assign the corresponding flickering character as the label to each data sample. Further, we apply a sliding window with the step of *ratio* × 250 on each data sample and generate the final input samples in a non-overlapping manner. For example, assuming the *ratio* equals 0.4, the data shape of each input sample is 100uN_c, and the N_c denotes the number of EEG channels (i.e., 64).

Because longer EEG segments contain more information about brain activity, the model performance for target frequency identification can be improved by increasing the segment length T. Considering this fact, we investigated the impact of segment length T ranges [0.2, 0.4, 0.6, 0.8, and 1.0] on model performance. More specifically, when the number of data points of each input sample is 50, meaning the *ratio* is set to 0.2, and segment length T representing the time length of each input sample is 0.2 s, the total number of input samples of the combination dataset for training and testing models is 366,000. The models are trained with a batch size of 64, and mini-batch gradient descent and Adam optimizer with a learning rate of 0.001 are used to optimize the model parameters. An early-stop training strategy is adopted to train the models. Ten-fold cross-validation is applied to divide the dataset into training data and testing data, and the average classification accuracy (ACC) rate, sensitivity (SEN), and specificity (SPE) and the corresponding standard deviation (STD) of them are employed as model performance metrics. The above metrics are calculated using the following formulas:

$$\mathrm{ACC} = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}), \tag{7}$$

$$\mathrm{SEN} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}), \tag{8}$$

$$\mathrm{SPE} = \mathrm{TN}/(\mathrm{TN}+\mathrm{FP}), \tag{9}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

3.2. Model training and further details

The GDNet-EEG and other comparable models are implemented by Pytorch and trained with a Tesla A100 GPU. The GDNet-EEG model is initialized by sampling the network weights from Gaussian distribution with 0 mean and 0.01 variance. Categorical cross-entropy is used as the loss function to train the model by comparing the probability distribution with true distribution. More specifically, the EEG data collected in one trial is represented by (X, Y), where $X \in \mathrm{R}^{C*Len}$ and $Y \in \mathrm{R}^{Nclass}$. As mentioned above, X is split into t segments $\{X_1, X_2, ..., X_t\}$ and segments collected from the same trial have the same label Y. To train the GDNet-EEG, we select the EEG signal of D_b trials as a batch of data to train the model in each iteration. The loss function of the categorical cross-entropy is computed as follows:

$$\frac{-1}{t*D_b}\sum_{i=1}^{t*D_b}\sum_{j=1}^{Nclass} y_{ij}\, log\left(s_{ij}\right) + \lambda|w|^2, \tag{10}$$

where λ (i.e., λ = 0.001) denotes the constant of the L2 regularization. $s_{ij} \in [0,1]^{Nclass}$ and y_i represent softmax output for the input segment X_i and the corresponding frequency label of the input segment X_i, respectively. w means the weights of the GDNet-EEG model. The GDNet-EEG model is trained by two stages: the first stage is trained by the benchmark dataset and the second stage is trained by the BETA dataset. Note that the second stage re-initializes the network with the weights trained by the first stage and fine-tunes the weights to fit the data distribution of the BETA dataset. The model training strategy originates from the consideration of inter-dataset statistical variations.

3.3. Comparison baselines

Five kinds of CNN models are reproduced as baseline approaches for result comparison. To perform the SSVEPs-based stimulation frequency recognition task, we reconstruct the output layer of these models to distinguish 40 target stimulation frequencies. The simplified description of the baseline approaches is depicted as follows:

EEGNet (Lawhern et al., 2018): The network starts with a temporal convolution to learn frequency filters and then uses depth-wise convolution to learn frequency-specific spatial filters. The depth-wise convolution combines all EEG channels to obtain a better frequency pattern.

Compact-CNN (Waytowich et al., 2018): The network is a variant of the EEGNet for classifying the SSVEP signals. Unlike the EEGNet, the dense layer of the Compact-CNN does not adopt the max-norm constraint function to the kernel weights matrix.

DeepConvNet (Schirrmeister et al., 2017): The model is a deep convolution network for end-to-end EEG analysis. It is comprised of four convolution-max-pooling blocks and a dense softmax classification layer. The first convolutional block is split into a first convolution across time and a second convolution across space (electrodes). The following blocks utilize standard convolution operation with a large filter whose width is equivalent to the number of feature maps.

Shallow ConvNet (Schirrmeister et al., 2017): The network is a shallow version of the DeepConvNet and contains one convolution-max-pooling block and a dense softmax classification layer. Compared with the deep ConvNet, the temporal convolution of the shallow ConvNet adopts a larger kernel size. After the two convolutions of the shallow ConvNet, a squaring nonlinearity, a mean pooling layer, and a logarithmic activation function followed.

Convolutional correlation analysis (Li et al., 2020): The network consists of a signal-CNN branch and a reference-CNN branch. The former is comprised of three convolutional layers, and the latter contains two convolutional layers. The output of the two branches is fed into the dropout layer for regularization. A correlation layer is followed by the dropout layer for calculating the correlation coefficients of the output of the two branches. A dense layer and softmax activation function is applied as the final classification layer.

FB-SSVEPformer (Chen et al., 2022c). This is the first Transformer-based deep learning model for SSVEP classification. The frequency spectrum of the SSVEP signals is extracted by filter bank technology and fed into SSVEPformer, which further learns spectral and spatial characteristics by self-attention mechanism for final frequency classification.

Filter bank CCA (Chen et al., 2015). This method tries to make use of harmonic SSVEP components to enhance the CCA-based frequency detection. By incorporating the fundamental and harmonic SSVEP components in target identification, the method significantly improves the performance of the SSVEP-based BCI.

3.4. Ablation studies

On the one hand, we design a comparison experiment to compare the classification performance of the GDNet-EEG model and its variations. The motivation of designing this comparison experiment is to validate the main innovations of our model, such as group depth-wise convolution and EEG attention module. On the other hand, the effect of EEG channel number on the model performance is also validated for demonstrating whether our model can recognize more informative SSVEP features from the signal of multiple EEG channels or not.

3.4.1. Comparison results between the GDNet-EEG model and its variations

The main innovation of our model mainly includes two aspects: (1) GDNet-EEG is a deep convolution architecture using a group depth-wise convolutional filter to extract as diverse regional characteristics as possible from raw EEG data. (2) EEG attention consisting of EEG channel and specialized network-wise attention is proposed to refine EEG feature of single EEG channel and recognize specialized networks to improve the model performance of SSVEPs-based target stimulation frequency recognition. To validate the model performance of the GDNet-EEG affected by the above two aspects, we design the following models: (1) we adopt a regular convolutional filter to substitute the group depth-wise convolutional filter in the GDNet-EEG; (2) we implement a shallow version of the GDNet-EEG, comprised of two group depth-wise convolutional layers; (3) we remove the EEG attention module of the GDNet-EEG; (4) the EEG channel-wise attention is removed from the GDNet-EEG; (5) the specialized network-wise attention is removed from the GDNet-EEG; (6) Instead of using EEG attention module, we embedded CBAM block into the GDNet-EEG model for refining the feature maps learned by the group depth-wise convolution layer We use model 1 model 6 to denote the five models for simplification.

The model performance affected by the signal length of the input sample is investigated. Figure 4 gives average classification accuracies obtained by the GDNet-EEG and model 1 model 6 over 10-fold cross-validation, and error bars indicate standard errors. The figure shows that the GDNet-EEG outperforms other models in classification accuracy across the three datasets in various signal lengths. As the signal length increases, the classification accuracy of different models shows an upward trend. This result shows that the EEG signal with a longer time length contains a more apparent characteristic pattern, which facilitates the deep learning models to generate more accurate decisions. Especially in the signal length of 1 s, the GDNet-EEG model achieves the highest classification accuracy of 84.11, 85.93, and 93.35% on the benchmark, BETA, and combination datasets, respectively. The models trained on the combination dataset obtained better model performance than the models trained on the benchmark dataset and BETA dataset, which may be attributed to the impact of dataset size on the deep learning model. Compared with the model 1 which is implemented by a regular convolutional filter, the GDNet-EEG obtains better classification accuracy, indicating the superiority and rationality of the group depth-wise convolution layer. The shallow GDNet-EEG (model 2) achieves the lowest accuracy, indicating the deep layer structure might provide an accuracy increment for the GDNet-EEG. The superiority of the EEG attention is also validated by comparing model 3 model 5 with the classification accuracy of the GDNet-EEG. More specifically, the classification rate of the model 3 is lower than the classification rate of our model, as well as the classification performance of model 4 or model 5 is also worse than the classification performance of the GDNet-EEG, demonstrating the EEG attention module can improve the classification performance of the GDNet-EEG. The comparison results between classification rate of model 4 and model 5 indicate the specialized network-wise attention seems to be capable of better boosting the classification performance of our model. By comparing the classification performance of model 6 with the classification performance of the GDNet-EEG, we can know the EEG attention module might be more suitable for refining representational EEG feature and improve the model performance for target frequency identification.

3.4.2. Effect of EEG channel number on the model performance

Note that the EEG channel location is arranged by international 10-10 EEG system. Although previous studies demonstrated the EEG channels that are placed over the occipital and parietal regions provide perhaps the most informative SSVEP signals, we want to validate the effectiveness of our approach on using the data of varying number of EEG channel. Table 2 gives the classification results (ACC, SPE, SEN, and their corresponding STDs) of our model is reported versus varying number of channels and 1.0 s of stimulation. We conducted five experiments to validate the effect of varying number of EEG channel on the model performance, the channel number and the corresponding channel name are given as follows:

- three EEG channels (labeled by O1, Oz, and O2) that are placed over the occipital (O) regions;
- six EEG channels (labeled by O1, Oz, O2, POz, PO3, and PO4) that are placed over the occipital and parietal- occipital (PO) regions, it is noteworthy that PO denotes the EEG channel placed between occipital and parietal regions;
- on the basis of the six EEG channels, we add another three EEG channels that are placed over PO regions, the nine EEG channels are labeled by O1, Oz, O2, Pz, PO3, PO5, PO4, PO6, and POz;
- thirty-two EEG channels that are placed over occipital, parietal, central, and central-parietal regions.
- Sixty-four EEG channels are placed over all brain regions.

The results demonstrate that there is an increasing tendency of the classification metrics of our approach as the EEG channel number increases, indicating the data collected from all EEG channels can help to improve the model performance. In addition, it is noteworthy that based on the combination

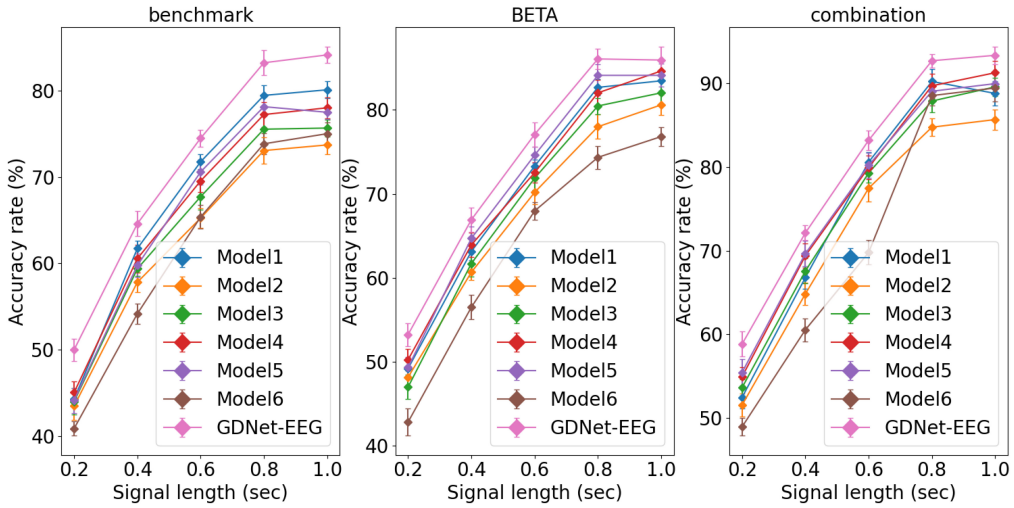


FIGURE 4
Average classification accuracies obtained by GDNet-EEG and model 1~model 6 over 10-fold cross-validation. Error bars indicate standard errors.

TABLE 2 Classification results (ACC, SPE, SEN, and their corresponding STDs) of our model is reported versus varying number of channels and 1.0 s of stimulation.

Channel number	Benchmark			BETA			Combination		
	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)	ACC (%)	SPE (%)	SEN (%)
3	65.32 ± 1.96	68.95 ± 2.12	63.58 ±	70.52 ± 1.74	69.72 ± 3.26	72.56 ± 2.51	86.16 ± 2.07	83.47 ± 1.85	88.73 ± 2.36
6	68.89 ± 2.52	70.32 ± 1.73	65.49 ±	72.46 ± 1.38	69.89 ± 2.79	73.85 ± 1.86	86.73 ± 1.96	84.59 ± 2.20	89.39 ± 1.82
9	75.28 ± 1.15	78.64 ± 1.58	73.24 ±	76.57 ± 2.21	74.87 ± 2.58	77.31 ± 2.70	91.27 ± 1.47	89.76 ± 1.63	92.26 ± 2.18
32	80.19 ± 1.09	81.79 ± 1.17	79.37 ±	82.91 ± 1.93	79.41 ± 2.90	83.46 ± 1.93	91.52 ± 2.15	89.50 ± 2.37	91.87 ± 2.60
64	84.11 ± 1.28	85.27 ± 0.93	83.81 ± 1.70	85.93 ± 1.36	83.26 ± 2.14	86.97 ± 2.36	93.35 ± 1.59	91.24 ± 1.54	94.12 ± 1.67

dataset, the classification metrics of 9 EEG channels are close to the classification metrics of 32 EEG channels while lower than the classification metrics of 64 EEG channels. This result indicates the EEG channels that are placed over the occipital and parietal regions might provide the most informative SSVEP signals while other channels might be informative as well.

3.5. Comparison studies

The ablation study shows that the GDNet-EEG model achieves the best classification accuracies based on the three datasets with the input sample length of 0.8 and 1 s. To further validate the model performance of the GDNet-EEG, we present average classification accuracies obtained by GDNet-EEG and five other models over 10-fold cross-validation using the signal length of 0.8 and 1 s. Figure 5 shows that the average classification accuracies of the other five model baselines trained on a combination dataset decreased from 1.96 to 18.2% compared to the GDNet-EEG. It indicates that the GDNet-EEG can produce more robust features than existing EEG-oriented deep learning methods and improve the discriminability between different stimulation frequencies. Compare with FB-SSVEPformer, our model achieves better classification rate based on the combination dataset, indicating the superiority of the GDNet-EEG based on the dataset with larger scale. In addition, the average classification accuracies of the FBCCA are lower than the classification accuracies of the GDNet-EEG model across the three EEG datasets, while the Conv-CA trained on the benchmark and BETA datasets outperformed the GDNet-EEG in average classification accuracies. Since the technical route of the Conv-CA and the GDNet-EEG is different, it gives us a cue for adapting the model architecture of the GDNet-EEG by integrating the CCA method to discriminate stimulation frequencies.

4. Discussion

Glaucoma is a common eye condition caused by a damaged optic nerve and can lead to vision loss if not diagnosed and treated early. The SSVEPs-based BCI application can generate brain signals when human looks at something flickering. If a patient has a blind area in a region, the signals extracted from these stimuli are weak, and it is reflected on the visual response map. That is, the patient cannot accept the stimulation from the flickering object at the field of vision loss occurred. Thus, the SSVEPs-based BCI application, e.g., visual speller, can diagnose glaucoma (Lin et al., 2015; Nakanishi et al.,

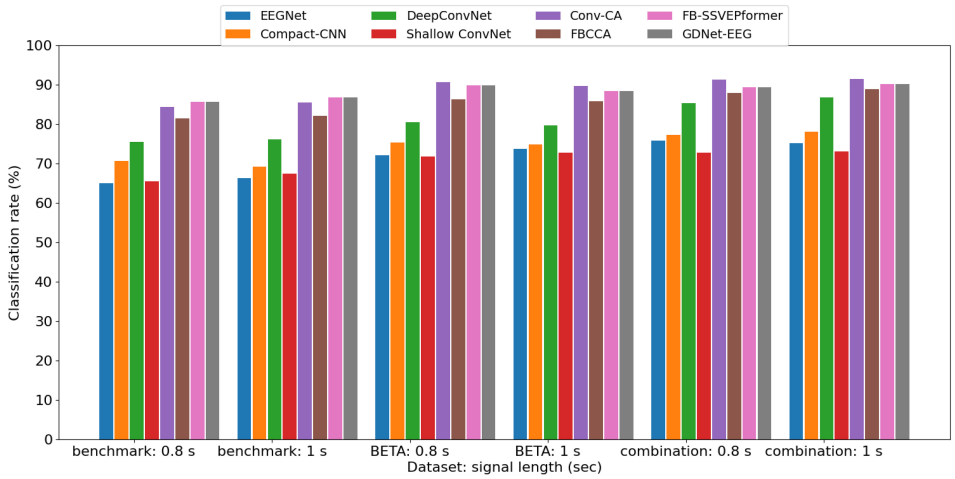


FIGURE 5
The average classification accuracies obtained by GDNet-EEG and five other models over 10-fold cross-validation using a signal length of 0.8 and 1 s.

2017; Khok et al., 2020). Based on SSVEPs-based BCI application, accurate glaucoma diagnosing requires effective EEG analysis methods to discriminate stimulation frequencies. Machine learning methods, especially deep learning, can achieve high accuracy in EEG-based classification tasks. However, most EEG-oriented deep learning methods focused on applying existing techniques to the EEG-based brain activity analysis task rather than proposing new ones specifically suited to the domain (Rasheed and Extraction, 2021). The standard well-known network architectures were designed for the data collected in natural scenes (e.g., natural images) and did not consider the EEG-based brain activity's peculiarities. Therefore, research must understand how these architectures can be optimized for SSVEPs-based classification tasks.

The peculiarities of EEG-based brain activity at least include the following two aspects: regional characteristics and network characteristics. The former can be represented by the temporal and spectral features of the signal generated from a single brain region. The BFC can represent the latter *via* learning all brain regions' global and local connectivity features. Although many existing studies extract temporal, spectral, and spatial features to represent the regional and network characteristics and feed them into deep learning models for generating decision results (Rocca et al., 2014; Amin et al., 2019; Su et al., 2020), they are not end-to-end deep learning frameworks. Convolution operation using the 1D convolutional filter is the priority choice for building the end-to-end deep learning framework for SSVEPs-based BCI applications (Waytowich et al., 2018). Unlike the previous studies using the regular 1D convolutional filter to learn EEG features, we utilize group depth-wise convolution operations containing a set of 1D convolutional filters and use each filter to convolve the data of the corresponding single brain region. An attention mechanism is adopted to identify important EEG channels from a single feature map and recognize significant feature maps as specialized brain networks.

An ablation study and comparison study are implemented to validate the performance of our proposed method in discriminating stimulation frequencies. From the experiment results described in **Figures 4**, **5** we can conclude that the average classification accuracies achieved by the models trained on the combination dataset are better than the average classification accuracies of the models trained on the benchmark and BETA datasets. The average classification accuracies obtained *via* the models trained on the BETA dataset are better than the models trained on the benchmark dataset. The reason can be explained from the aspect of deep learning model performance affected by the dataset size. As we know, insufficient training data can lead to poor performance of deep learning models. Small training and testing datasets will result in underfitting the deep learning model, generating an optimistic and high variance estimation of model performance. By observing the experiment results of the ablation study, we can see an upward trend of average classification accuracies along with the signal length increasing. This result coincides with the experiment result of other studies (Li et al., 2020; Guney et al., 2021), which indicates better classification accuracy can be obtained by lengthening the stimulation duration (i.e., signal length of input sample). In addition, the comparison results between the average classification accuracies obtained by the GDNet-EEG using a regular 1D convolutional filter. Additionally, our method demonstrates the superiority of the group depth-wise convolution operation. Compared with EEGNet and Compact-CNN, our model's group depth-wise convolution layer covers the receptive field of the same size and has a deeper model architecture with fewer parameters. The higher classification accuracies achieved by our model indicate that the architecture of our model can capture more robust EEG features to discriminate stimulation frequencies. The ablation study also validates that using an attention mechanism can improve the classification accuracies of models in discriminating different stimulation frequencies.

Our proposed GDNet-EEG has three potential improvement directions: (1) This study is a pilot study for glaucoma diagnosing by implementing an effective deep learning method for SSVEPs-based stimulation frequency discrimination. The datasets used in this study are collected from healthy participants. Collecting an SSVEP dataset from glaucoma patients is a feasible route for making our method more available in SSVEPs-based BCI application of early glaucoma diagnosis. (2) Inspired by the method of using CCA to discriminate stimulation frequencies, we plan to use a self-attention mechanism (e.g., Transformer model) (Vaswani et al., 2017) to calculate how similar between stimulation signals and reference signals and utilize the similarity to generate more robust EEG feature for discriminating stimulation frequencies. (3) Although the experimental results have demonstrated that group depth-wise convolution and EEG attention facilitates the GDNet-EEG to achieve promising classification performance in discriminating SSVEPS-based stimulation frequencies, this result may be unable to provide strong support for clinical treatment that is associated with EEG biomarkers. Because DL methods are essentially black boxes, we require novel methods to open the box and visualize the feature learned by the DL model. To this end, an emerging technique known as explainable artificial intelligence (AI) (Gunning et al., 2019) enables the understanding of how DL methods work and what drives their decision-making. We plan to use the explainable AI method to visualize the critical brain regions and significant specialized networks and further validate our method's performance.

5. Conclusion

In this study, we propose a novel deep learning model named the GDNet-EEG, which is tailored to learn regional characteristics and network characteristics of EEG-based brain activity to perform the SSVEPs-based stimulation frequency recognition task. The group depth-wise convolution is proposed to extract temporal and spectral features from the EEG signal of each brain region and represent regional characteristics as diverse as possible. Based on the output of the group depth-wise convolutional layer, EEG attention consisting of EEG channel-wise attention and specialized network-wise attention is designed to identify essential brain regions and form significant feature maps as the specialized brain functional networks. The experiment results demonstrate that our method outperforms the existing deep learning models tailored to process EEG data on two publicly SSVEPs datasets (large-scale benchmark and BETA dataset) and their combined dataset. Our approach could be potentially suitable for providing accurate stimulation frequency discrimination and being used in the early glaucoma diagnosis using SSVEP signals.

Data availability statement

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

Author contributions

ZW and WD contributed to the conception and design of the study and drafted the manuscript. WC and ML performed the data analysis. RZ provided technique and writing guidance. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

Amin, S., Alsulaiman, M., Muhammad, G., Bencherif, M., and Hossain, M. (2019). Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. *IEEE Access* 7, 18940–18950. doi: 10.1109/ACCESS.2019.2895688

Atasoy, S., Donnelly, I., and Pearson, J. (2016). Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* 7:10340. doi: 10.1038/ncomms10340

Avberšek, L., and Repovš, G. (2022). Deep learning in neuroimaging data analysis: applications, challenges, and solutions. *Front. Neuroimaging* 1:981642. doi: 10.3389/fnimg.2022.981642

Babaeeghazvini, P., Rueda-Delgado, L., Gooijers, J., Swinnen, S., and Daffertshofer, A. (2021). Brain structural and functional connectivity: a review of combined works of diffusion magnetic resonance imaging and electro-encephalography. *Front. Hum. Neurosci.* 15:721206. doi: 10.3389/fnhum.2021.721206

Chen, W., Zhang, D., Li, M., and Lee, D. (2020). Stcam: spatial-temporal and channel attention module for dynamic facial expression recognition. *IEEE Trans. Affect. Comput.* 14, 800–810. doi: 10.1109/TAFFC.2020.3027340

Chen, X., Cheng, G., Wang, F. L., Tao, X., Xie, H., and Xu, L. (2022a). Machine and cognitive intelligence for human health: systematic review. *Brain Informat.* 9:5. doi: 10.1186/s40708-022-00153-9

Chen, X., Tao, X., Wang, F. L., and Xie, H. (2022b). Global research on artificial intelligence-enhanced human electroencephalogram analysis. *Neural Comput. Appl.* 34, 11295–11333. doi: 10.1007/s00521-020-05588-x

Chen, J., Zhang, Y., Pan, Y., Xu, P., and Guan, C. (2022c). A transformer-based deep neural network model for SSVEP classification. *arXiv [preprint]*. doi: 10.48550/arXiv.2210.04172

Chen, X., Wang, Y., Gao, S., Jung, T. P., and Gao, X. (2015). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J. Neural Eng.* 12:046008. doi: 10.1088/1741-2560/12/4/046008

Chen, X., Zhang, X., Xie, H., Tao, X., Wang, F. L., Xie, N., et al. (2021). A bibliometric and visual analysis of artificial intelligence technologies-enhanced brain MRI research. *Multimedia Tools Appl.* 80, 17335–17363. doi: 10.1007/s11042-020-09062-7

Cherloo, M., Amiri, H., and Daliri, M. (2022). Spatio-spectral CCA (SS-CCA): a novel approach for frequency recognition in SSVEP-based BCI. *J. Neurosci. Methods* 371:109499. doi: 10.1016/j.jneumeth.2022.109499

Ferraro, G., Moreno, A., Min, B., Morone, F., Pérez-Ramírez, Ú, Pérez-Cervera, L., et al. (2018). Finding influential nodes for integration in brain networks using optimal percolation theory. *Nat. Commun.* 9:2274. doi: 10.1038/s41467-018-04718-3

Geethalakshmi, R., Vani, R., and Cruz, M. (2022). "A study of glaucoma diagnosis using brain-computer interface technology," in *Proceedings of ICCIML Computational Intelligence in Machine Learning*, (Berlin: Springer), 271–279. doi: 10.1007/978-981-16-8484-5_25

Guedes, R. (2021). Glaucoma, collective health and social impact. *Rev. Bras. Oftalmol.* 80, 05–07. doi: 10.5935/0034-7280.20210001

Guney, O., Oblokulov, M., and Ozkan, H. (2021). A deep neural network for ssvep-based brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 69, 932–944. doi: 10.1109/TBME.2021.3110440

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., and Yang, G. (2019). XAI—Explainable artificial intelligence. *Sci. Robot.* 4:eaay7120. doi: 10.1126/scirobotics.aay7120

Khok, H., Koh, V., and Guan, C. (2020). "Deep multi-task learning for SSVEP detection and visual response mapping," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (Toronto, ON), 1280–1285. doi: 10.1109/SMC42975.2020.9283310

Lawhern, V., Solon, A., Waytowich, N., Gordon, S., Hung, C., and Lance, B. (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, M., Ma, C., Dang, W., Wang, R., Liu, Y., and Gao, Z. (2022). DSCNN: dilated shuffle CNN model for SSVEP signal classification. *IEEE Sens. J.* 22, 12036–12043. doi: 10.1109/JSEN.2022.3173433

Li, Y., Xiang, J., and Kesavadas, T. (2020). Convolutional correlation analysis for enhancing the performance of SSVEP-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2681–2690. doi: 10.1109/TNSRE.2020.3038718

Lin, K., Jie, B., Dong, P., Ding, X., Bian, W., and Liu, M. (2022). Convolutional recurrent neural network for dynamic functional MRI analysis and brain disease identification. *Front. Neurosci.* 16:933660. doi: 10.3389/fnins.2022.933660

Lin, Y., Wang, Y., Jung, T., Gracitelli, C., Abe, R., Baig, S., et al. (2015). Using multifocal steady-state visual evoked potentials for objective assessment of visual field loss in glaucoma. *Investig. Ophthalmol. Vis. Sci.* 56:486.

Liu, B., Huang, X., Wang, Y., Chen, X., and Gao, X. (2020). BETA: a large benchmark database toward SSVEP-BCI application. *Front. Neurosci.* 14:627. doi: 10.3389/fnins.2020.00627

Liu, Q., Jiao, Y., Miao, Y., Zuo, C., Wang, X., Cichocki, A., et al. (2020). Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA. *Neurocomputing* 378, 36–44. doi: 10.1016/j.neucom.2019.10.049

Lv, Z., Qiao, L., Singh, A., and Wang, Q. (2021). Fine-grained visual computing based on deep learning. *ACM Trans. Multimidia Comput. Commun. Appl.* 17, 1–19. doi: 10.1145/3418215

Ma, P., Dong, C., Lin, R., Ma, S., Jia, T., Chen, X., et al. (2022). A classification algorithm of an SSVEP brain-Computer interface based on CCA fusion wavelet coefficients. *J. Neurosci. Methods* 371:109502. doi: 10.1016/j.jneumeth.2022.109502

McDonough, I., and Nashiro, K. (2014). Network complexity as a measure of information processing across resting-state networks: evidence from the Human Connectome Project. *Front. Hum. Neurosci.* 8:409. doi: 10.3389/fnhum.2014.00409

Medeiros, F., Zao, J., Wang, Y., Nakanishi, M., Lin, Y., Jung, T. P., et al. (2016). The nGoggle: a portable brain-based method for assessment of visual function deficits in glaucoma. *Investig. Ophthalmol. Vis. Sci.* 57:3940.

Mora-Cortes, A., Ridderinkhof, K., and Cohen, M. (2018). Evaluating the feasibility of the steady-state visual evoked potential (SSVEP) to study temporal attention. *Psychophysiology* 55:e13029. doi: 10.1111/psyp.13029

Nakanishi, M., Wang, Y. T., Jung, T. P., Zao, J. K., Chien, Y. Y., Diniz-Filho, A., et al. (2017). Detecting glaucoma with a portable brain-computer interface for objective assessment of visual function loss. *JAMA Ophthalmol.* 135, 550–557.

Nuzzi, R., Dallorto, L., and Rolle, T. (2018). Changes of visual pathway and brain connectivity in glaucoma: a systematic review. *Front. Neurosci.* 12:363. doi: 10.3389/fnins.2018.00363

Rasheed, S., and Extraction, K. (2021). A review of the role of machine learning techniques towards brain–computer interface applications. *Mach. Learn. Knowl. Extract.* 3, 835–862. doi: 10.3390/make3040042

Retter, T., Rossion, B., and Schiltz, C. (2021). Harmonic amplitude summation for frequency-tagging analysis. *J. Cognit. Neurosci.* 33, 2372–2393. doi: 10.1162/jocn_a_01763

Rocca, D., Campisi, P., Vegso, B., Cserti, P., Kozmann, G., Babiloni, F., et al. (2014). Human brain distinctiveness based on EEG spectral coherence connectivity. *IEEE Trans. Biomed. Eng.* 61, 2406–2412. doi: 10.1109/TBME.2014.2317881

Salelkar, S., and Ray, S. (2020). Interaction between steady-state visually evoked potentials at nearby flicker frequencies. *Sci. Rep.* 10, 1–16. doi: 10.1038/s41598-020-62180-y

Schirrmeister, R., Springenberg, J., Fiederer, L., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain. Mapp.* 38, 5391–5420. doi: 10.1002/hbm.23730

Shine, J., Bissett, P., Bell, P., Koyejo, O., Balsters, J., Gorgolewski, K., et al. (2016). The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron* 92, 544–554. doi: 10.1016/j.neuron.2016.09.018

Soh, Z., Yu, M., Betzler, B., Majithia, S., Thakur, S., Tham, Y., et al. (2021). The global extent of undetected glaucoma in adults: a systematic review and meta-analysis. *Ophthalmology* 128, 1393–1404. doi: 10.1016/j.ophtha.2021.04.009

Su, C., Xu, Z., Pathak, J., and Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. *Transl. Psychiatry* 10:116. doi: 10.1038/s41398-020-0780-3

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. Advances in neural information processing systems. *arXiv [preprint]*. Available online at: https://arxiv.org/pdf/1706.03762.pdf (accessed December 6, 2017).

Wang, Y., Chen, X., Gao, X., and Gao, S. (2016). A benchmark dataset for SSVEP-based brain–computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1746–1752. doi: 10.1109/TNSRE.2016.2627556

Waytowich, N., Lawhern, V., Garcia, J., Cummings, J., Faller, J., Sajda, P., et al. (2018). Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *J. Neural Eng.* 15:066031. doi: 10.1088/1741-2552/aae5d8

Woo, S., Park, J., Lee, J., and Kweon, I. S. (2018). "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, (Munich), 3–19.

Yao, H., Liu, K., Deng, X., Tang, X., and Yu, H. (2022). FB-EEGNet: a fusion neural network across multi-stimulus for SSVEP target detection. *J. Neurosci. Methods* 379:109674. doi: 10.1016/j.jneumeth.2022.109674

Zhang, Y., Xie, S., Wang, H., and Zhang, Z. (2020). Data analytics in steady-state visual evoked potential-based brain–computer interface: a review. *IEEE Sens. J.* 21, 1124–1138. doi: 10.1109/JSEN.2020.3017491

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 20 April 2023
DOI 10.3389/fnins.2023.1182388

OPEN ACCESS

EDITED BY
Zhi Wen,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Bing Lin Huang,
Jiangxi University of Traditional Chinese Medicine, China
Zhijiang Wan,
Nanchang University, China
Yu Ji,
The First Affiliated Hospital of Nanchang University, China
Tianming Huo,
Wuhan University, China

*CORRESPONDENCE
Shui-Lian Zhou
✉ zhoushuilian_12358@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Visual Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 08 March 2023
ACCEPTED 27 March 2023
PUBLISHED 20 April 2023

CITATION
Gan F, Liu H, Qin W-G and Zhou S-L (2023) Application of artificial intelligence for automatic cataract staging based on anterior segment images: comparing automatic segmentation approaches to manual segmentation.
Front. Neurosci. 17:1182388.
doi: 10.3389/fnins.2023.1182388

Application of artificial intelligence for automatic cataract staging based on anterior segment images: comparing automatic segmentation approaches to manual segmentation

Fan Gan[1,2†], Hui Liu[2†], Wei-Guo Qin[3] and Shui-Lian Zhou[2*]

[1]Medical College of Nanchang University, Nanchang, China, [2]Department of Ophthalmology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, China, [3]Department of Cardiothoracic Surgery, The 908th Hospital of Chinese People's Liberation Army Joint Logistic Support Force, Nanchang, China

Purpose: Cataract is one of the leading causes of blindness worldwide, accounting for >50% of cases of blindness in low- and middle-income countries. In this study, two artificial intelligence (AI) diagnosis platforms are proposed for cortical cataract staging to achieve a precise diagnosis.

Methods: A total of 647 high quality anterior segment images, which included the four stages of cataracts, were collected into the dataset. They were divided randomly into a training set and a test set using a stratified random-allocation technique at a ratio of 8:2. Then, after automatic or manual segmentation of the lens area of the cataract, the deep transform-learning (DTL) features extraction, PCA dimensionality reduction, multi-features fusion, fusion features selection, and classification models establishment, the automatic and manual segmentation DTL platforms were developed. Finally, the accuracy, confusion matrix, and area under the receiver operating characteristic (ROC) curve (AUC) were used to evaluate the performance of the two platforms.

Results: In the automatic segmentation DTL platform, the accuracy of the model in the training and test sets was 94.59 and 84.50%, respectively. In the manual segmentation DTL platform, the accuracy of the model in the training and test sets was 97.48 and 90.00%, respectively. In the test set, the micro and macro average AUCs of the two platforms reached >95% and the AUC for each classification was >90%. The results of a confusion matrix showed that all stages, except for mature, had a high recognition rate.

Conclusion: Two AI diagnosis platforms were proposed for cortical cataract staging. The resulting automatic segmentation platform can stage cataracts more quickly, whereas the resulting manual segmentation platform can stage cataracts more accurately.

KEYWORDS

anterior segment images, artificial intelligence, cortical cataract, multi-feature fusion, automatic segmentation

1. Introduction

Cataract is one of the leading causes of blindness worldwide, accounting for over 50% of cases of blindness in low- and middle-income countries (Wu et al., 2019). It is a visual impairment characterized by cloudiness or opacification of the crystalline lens, and most cataracts are age-related, although they can also be attributed to disease, trauma, or congenital factors (Do et al., 2013; Gao et al., 2015; Satyam et al., 2015). The pathogenesis of cataract is quite complex and results from the long-term comprehensive effect of various internal and external factors on the lens. Surgical removal of the lens and implantation of intraocular lens are the only effective treatments of a visually significant cataract (Son et al., 2022).

Cortical cataract is the most common type of the senile (age-related) cataract. Depending on its severity, cortical cataract is divided into four stages: (1) incipient stage, in which the lens is partially opaque, with spokes and vacuoles, and wedge-shaped opacity; (2) intumescent stage (immature stage), during which lens thickness is increased and the depth of the anterior chamber becomes shallow; (3) mature stage, in which the lens is completely opaque; and (4) the hypermature stage, which has a shrunken and wrinkled anterior capsule owing to water leakage out of the lens and might also have calcium deposits. In the incipient stage, because the lesion rarely involves the pupil area, vision is affected rarely. Some measures can be taken to slow cataract progression, such as by wearing anti-glare sunglasses (Gao et al., 2015). In the intumescent stage, for patients with anatomic factors of angle-closure glaucoma, an acute glaucoma attack can be induced by anterior chamber shallowing. By the mature stage, the patient will have severe vision loss and will require surgical treatment. In the hypermature stage, patients will have serious complications, such as phacolytic glaucoma and phacoanaphylactic uveitis. Therefore, for timely cataract treatment, to prevent complications, and to improve quality of life, accurate staging is important.

Currently, the diagnosis of cataract relies on the rich experience of the ophthalmologist and slit-lamp biomicroscopy examination. However, the distribution of medical resources is far from satisfactory for cataract diagnosis and management (Wu et al., 2019). The COVID-19 pandemic has also led to a shift from on-site medical needs to telemedicine. In the previous research, Xie et al. (2020) applied a semiautomated telemedicine platform combining a deep learning system with human assessment to achieve the best economic return for diabetic retinopathy (DR) screening in Singapore, resulting in potential savings of approximately 20% of the annual cost. Therefore, it is particularly important to develop an artificial intelligence (AI) diagnosis platform for cataracts to achieve high-precision automated diagnosis and lay the foundation for the combination of AI and telemedicine in the future.

Recently, artificial intelligence (AI) has made remarkable progress in medicine (Amjad et al., 2021). An increasing number of AI diagnostic models for ophthalmologic diseases have been proposed. Lin et al. (2020) used the random forest (RF) and adaptive boosting (Ada) algorithms for the identification of congenital cataracts. Gao et al. (2015) used deep-learning algorithms to grade nuclear cataracts. Hasan et al. (2021) used a transfer-learning algorithm to detect cataracts. All these models exhibit excellent performance. However, to the best of our knowledge, there has not been research applying AI for automatic cortical cataract staging. Most previous studies used the traditional machine learning or deep learning based on original slit-lamp images. Compared with traditional methods, transfer-learning represents an important way of solving the fundamental problem of insufficient training data in deep learning (He et al., 2020). In addition, for similar experimental conditions, a pre-trained network can be adjusted quickly through transfer-learning, which can reduce the training time greatly (Lin et al., 2021). It has also been suggested that the image features derived from segmented images yield increased accuracy than those from non-segmented images (Zhang et al., 2020). Automatic segmentation can be faster and more reproducible compared with manual delineation but might not have the same accuracy as manual segmentation (Huang et al., 2019; Tsuji et al., 2020).

Therefore, unlike previous studies, we combined segmentation with a deep transfer-learning algorithm and multi-feature fusion to create two AI platforms for automatic cortical cataract staging. One is based on an automatic segmentation method, whereas the other is based on a manual segmentation method; the flowchart of the detailed processes within this study is shown in **Figure 1**.

2. Materials and methods

2.1. Data collection

We collected the anterior segment images of cataract-affected eyes from the Department of Ophthalmology, Jiangxi Provincial People's Hospital. All images were diffuse-illuminated photographs that were collected from the same slit-lamp digital microscopy and taken by experienced ophthalmic technologists using standardized techniques. All images were screened, the images that clearly demonstrate the characteristics and reflect different stages of cataracts were retained, and blurry images

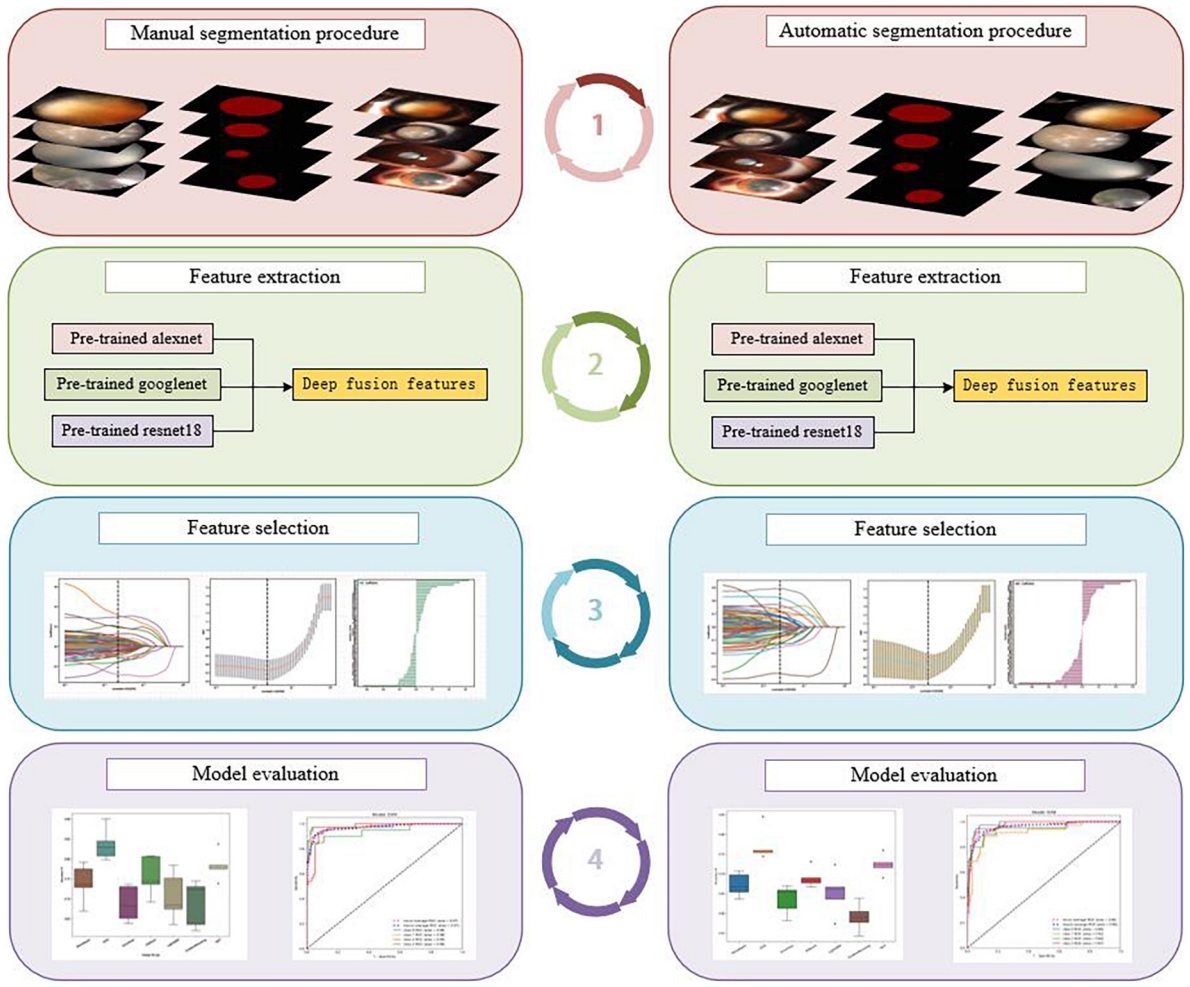


FIGURE 1
The flowchart of the detailed processes of the study: manual segmentation procedure **(left)**, automatic segmentation procedure **(right)**.

and images of corneal disease that affected lens observation significantly were excluded. The screened high-quality images were then divided randomly into a training and test set using a stratified random-allocation technique at a ratio of 8:2, in which stratification was by staging of cortical cataracts. This means the distribution of data in each stage was random in both the training and testing sets.

2.2. The region of interest delineating and cataract labeling

An experienced ophthalmologist used the LabelMe software to delineate the region of interest (ROI), which was the lens regions of the images. Based on the diagnostic reports obtained from the electronic medical record system and combined with the opacity of the lens in the images, the ROIs were labeled as "label 0", indicating the incipient stage, "label 1", indicating the intumescent stage (immature stage), "label 2", indicating the mature stage, and "label 3" indicating the hypermature stage.

2.3. Establishment of the automatic segmentation DTL platform

2.3.1. Establishment of the automatic segmentation model

First, we trained the automatic segmentation model with the FCNResnet50 architecture. The FCN model 'learns' a pixel's class by finding optimal values for the model parameters through minimizing the prediction error against the target data set (Larsen et al., 2021). The images that the experienced ophthalmologist had delineated the lens regions of were then used as the gold standard. The model was run for 30 'epochs', each time training on 80% of the dataset and evaluating model performance on a 20% hold-out set. Finally, the trained model was applied to the whole dataset and the segmentations of lens regions were obtained.

2.3.2. Deep fusion features extraction

First, the alexNet, googleNet, and resnet18 models were pretrained on the natural image dataset ImageNet dataset[1],

1 www.image-net.org/

respectively. Then the pretraining parameters obtained from the ImageNet dataset were used to initialize our models. The resulting pretrained alexNet, googleNet, and resnet18 models were then utilized to extract DTL features from the output of the avgPool layer, respectively. To reduce dimensionality, we employed principal components analysis (PCA). Subsequently, we utilized channel concat to combine the output features after PCA dimension reduction, and this resulted in the deep fusion features.

2.3.3. Feature selection

The final deep fusion features used to construct the model were selected in the training set. The least absolute shrinkage and selection operator (LASSO) algorithm was used to construct the feature selection model. First, all the deep fusion features were standardized to a mean of 0 and a variance of 1 by the regularization method. The formula used is shown here:

$$column = \frac{\text{column} - \text{mean}}{\text{std}}$$

Then, the LASSO model selected features using a tuning parameter (λ). The optimal λ was chosen based on a ten-fold cross-validation. Depending on the regulation weight λ, LASSO shrinks all regression coefficients toward zero and sets the coefficients of many irrelevant features exactly to zero (Lao et al., 2017). The features with non-zero coefficients were retained.

2.3.4. Establishment of the classification model

After features selection, the selected features were used to establish the classification models. Seven machine-learning algorithms were imported from the scikit-learn python library to establish seven classification models, respectively, including naive bayes (NB), support vector machines (SVM), extremely randomized trees (Extra Trees, ET), extreme gradient boosting (XGBoost, XGB), light gradient boosting machine (LightGBM), gradient boosting (GB), and multilayer perceptron (MLP) models. To prevent overfitting, five-fold cross-validation was used to fit each classification model.

2.4. Establishment of the manual segmentation DTL platform

Manual segmentation of the DTL platform included manual segmentation, deep fusion features extraction, feature selection, and the classification model establishment.

The rest was the same as the automatic segmentation DTL platform, except that the segmentation was different. First, based on the lens regions delineated by an experienced ophthalmologist, a ROI was segmented manually from each image. Then, based on the ROI, as with automatic segmentation DTL platform, pretrained alexNet, pretrained googleNet, and pretrained resNet18 models were used to extract DTL features, respectively. Next, PCA was used for dimensionality reduction. The reduced DTL features were fused. The LASSO model was used to select features. Finally, seven different classification models, NB, SVM, ET, XGB, LightGBM, GB, and MLP were established.

2.5. Model validation and performance evaluation

The trained models were applied to the test set for independent testing. Different quantitative metrics, such as pixel accuracy (PA), intersection over union (IoU), and Dice coefficient (Dice), were adopted to evaluate the performance of the automatic segmentation model and the classification model. PA is the simplest indicator of image segmentation, which is the percentage of correctly classified pixels out of the total pixels in each image (Larsen et al., 2021). IoU is a concept used in object detection, which measures the overlap between two boundaries: the predicted boundary and the truth boundary (Kim and Hidaka, 2021). The higher the IoU, the more accurate is the position of the prediction boundary. The Dice coefficient is a score that indicates the similarity between two samples (Takahashi et al., 2021). It used to measure the amount of overlap of regions.

To the classification models, the classification accuracy, confusion matrix, and the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were also introduced to evaluate the performance. The classification accuracy is computed as the ratio of the correctly classified number of samples and the total number of samples (Masood and Farooq, 2019). The confusion matrix is a visualization tool used typically in multiclass supervised learning and contains information about the actual classifications and the classifications predicted by a classification model (Bang et al., 2021). ROC curve and AUC was another class of indicators to evaluate the classification accuracy. The closer the ROC curve is to the upper left corner, the larger the AUC value, and the better the classification effect.

2.6. Statistical analysis

ROI was delineated, segmented, and labeled using an open-source annotation tool LabelMe. All statistical calculations and the drawing of statistical graphs were performed in Python (version 3.9.7).

3. Results

3.1. Imaging dataset

A total of 647 high quality anterior segment images were included into the dataset. One hundred ninety one incipient stage images, 171 intumescent stage images, 100 mature stage images and 183 hypermature stage images. Through stratified random division, with 80% of the images used for training and 20% for testing. 517 images were included in training set, of these, included 153 incipient stage images, 136 intumescent stage images, 80 mature stage images and 147 hypermature stage images. One hundred thirty images were included in test set, of these, included 38 incipient stage images, 35 intumescent stage images, 20 mature stage images and 36 hypermature stage images.

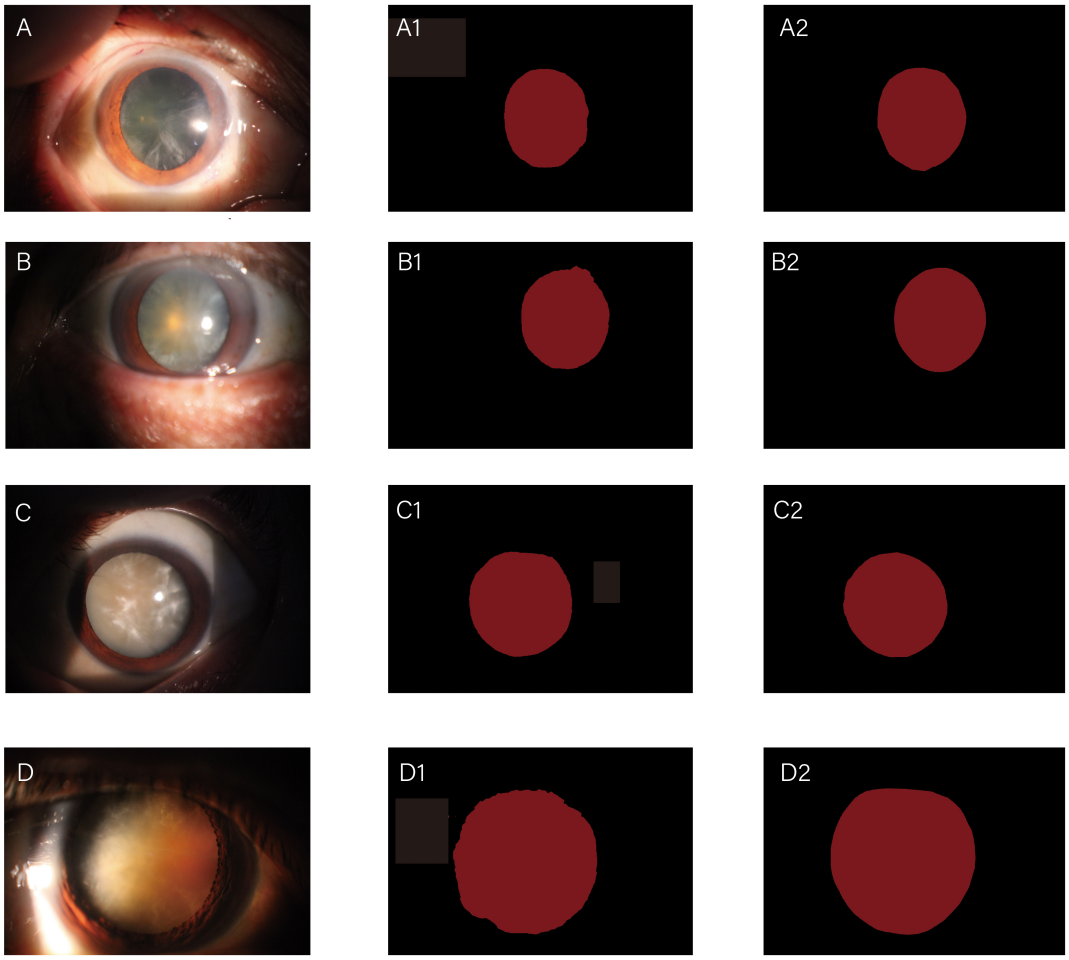


FIGURE 2
Diagrams of comparisons between automatic and manual segmentation: original images of cataracts at different stage **(A–D)**; the corresponding automatic segmentation mask **(A1–D1)**; the corresponding manual segmentation mask **(A2–D2)**.

3.2. Segmentation performance of the automatic segmentation model

The whole automatic segmentation process took 2 min and 43 s. While the manual segmentation process from the experienced ophthalmologist required approximately a week. The segmentation results graph of the automated and manual segmentations as shown in Figure 2. The visualization of the FCNResnet50 model training process was shown in Figure 3. The loss value decreases gradually with epoch and stabilizes at 5 epochs, and the accuracy reaches 95% in the test set. It can be seen that the contours obtained manually often fit better with the true contour of the lens compared with the automatic segmentation. And the results of automatic segmentation showed that the PA was 98.9, the mean IoU was 93.3, and the mean Dice score was 96.4%.

3.3. Results of the feature extraction and fusion

In the automatic segmentation platform, the automatic segmentation lens images were input to the three pretrained models, the extracted features were output from the last fully connected layer. 512, 1024, 9216 DTL features of each image were extracted from pre-trained alexNet, pre-trained googleNet and pre-trained resNet 18, respectively. After PCA dimension reduction, 31 features of each image from each model were obtained. And then after features fusion, the feature subset included 93 features of each image were obtained. In the manual segmentation platform, the manual segmentation lens images were input to the three pretrained models. After features extraction, PCA dimension reduction and features fusion, the 93 features of each image were also obtained.

3.4. Results of feature selection

The optimal λ (λ = 0.025595) was chosen based on a ten-fold cross-validation. Depending on the optimal λ, 49 features were retained in the automatic segmentation platform, including 21 features of alexNet model, 12 features of resNet model and 16 features of googleNet model. 51 features were retained in the manual segmentation platform, including 20 features of alexnet model, 17 features of resNet model and 14 features of googleNet model. The selection process was shown in Figure 4.

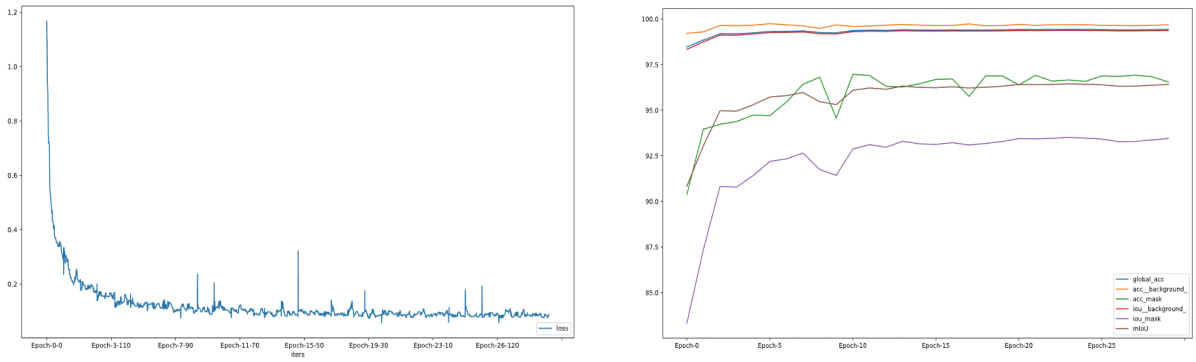


FIGURE 3
The loss convergence and accuracy curves of FCNResNet50 model in the test. The loss convergence curve **(left)**, the pixel accuracy (PA) and Intersection over Union (IoU) curve **(right)**.

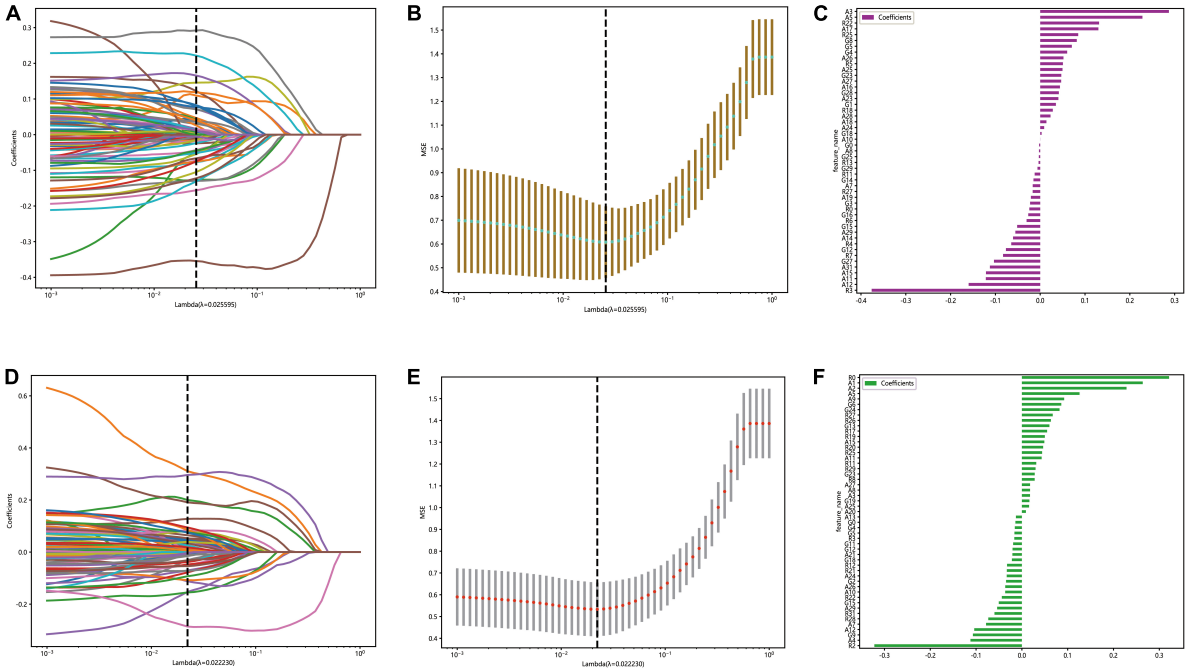


FIGURE 4
The figure of LASSO coefficient distribution (left): the colored curve shows the path of the coefficients for each input feature as lambda varies; the figure of partial likelihood deviation of the LASSO coefficient distribution (middle): the vertical dashed line represents the optimal value of the regularization parameter determined by cross-validation; feature weight coefficient graph (right). Automatic segmentation platform **(A–C)**; manual segmentaion platform **(D–F)**.

3.5. The classification performance of the automatic segmentation DTL platform and the manual segmentation DTL platform

By five-fold cross-validation, the result of classification accuracy revealed that the accuracy of SVM model was the best in both the automatic segmentation DTL platform or the manual segmentation DTL platform. In the automatic segmentation DTL platform, the accuracy of the model in the training set and the test set were 94.59 or 84.50%, respectively. In the manual segmentation DTL platform, the accuracy of the model in the training set and the test set were 97.4 or 90.00%, as shown in Table 1. The range of classification accuracy rates were shown in Figure 5.

The result of AUC also revealed that the performance of SVM model was best in both two platforms. In the automatic segmentation DTL platform, the micro and macro average AUC of SVM model both were 96% in the test set. In the manual segmentation DTL platform, the micro and macro average AUC of SVM model both were 97% in the test set. And in both two platforms, the AUC for each classification was all more that 90% in the test set. As shown in Figure 6.

The results of 4 × 4 matrix show the number of correct and incorrect classifications by the SVM model in each stage of cataract. In the automatic segmentation DTL platform, the recognition rates

TABLE 1 The accuracy of classification models in the automatic and manual segmentation DTL platforms.

Group	Model name	Accuracy	Train/Test
Manual	NaiveBayes	81.82%	Train
	NaiveBayes	79.23%	Test
	SVM	97.48%	Train
	SVM	90.00%	Test
	ExtraTrees	100%	Train
	ExtraTrees	80.77%	Test
	XGBoost	96.71%	Train
	XGBoost	78.46%	Test
	LightGBM	96.71%	Train
	LightGBM	78.46%	Test
	GradientBoosting	89.17%	Train
	GradientBoosting	72.31%	Test
	MLP	95.94%	Train
	MLP	78.46%	Test
Automatic	NaiveBayes	73.55%	Train
	NaiveBayes	69.77%	Test
	SVM	94.59%	Train
	SVM	84.50%	Test
	ExtraTrees	100%	Train
	ExtraTrees	58.14%	Test
	XGBoost	100%	Train
	XGBoost	66.67%	Test
	LightGBM	94.79%	Train
	LightGBM	63.57%	Test
	GradientBoosting	84.94%	Train
	GradientBoosting	54.26%	Test
	MLP	92.66%	Train
	MLP	75.97%	Test

of incipient stage, intumescent stage, hypermature stage were all high. The hypermature stage had the highest recognition rate. Of the 36 images, 34 of them were correctly recognized (94.5%) and only 2 were incorrectly recognized (5.5%). While, the mature stage had the lowest recognition rate. Of the 20 images, 14 of them were correctly recognized (70%) and 6 were incorrectly recognized (6%). In the manual segmentation DTL platform, the results also show that all stages except for mature had high recognition rate. Of the 20 images in the mature stage, 15 of them were correctly recognized (75%) and 5 were incorrectly recognized (5%). The recognition rate of the intumescent stage was highest, of the 35 images, 34 of them were correctly recognized (97.1%) and only 1 were incorrectly recognized (2.9%) as shown in **Figure 7**.

In addition to the SVM model, we also drew the ROC curves and confusion matrix of the other models, which included NB, ET, XGB, LightGBM, GB, and MLP models of the manual segmentation platform and the automatic segmentation platform. The ROC curves of other models were shown in the **Supplementary Figures 1**, **2**. The confusion matrix of other models were shown in the **Supplementary Figures 3**, **4**.

4. Discussion

Precision medicine is an emerging medical model that has great promise for the prevention, diagnosis, and treatment of many diseases (McGonigle, 2016). Accurate staging of cataracts is a precise classification of the different states and processes of a disease, which is the embodiment of precision medicine strategies. It is also essential to planning of appropriate treatment, assessing outcome, and future prognosis. The establishment of an automated cataract diagnosis platform not only makes medical services more convenient and efficient, but also contributes to epidemic prevention and control. In this study, we developed two AI platforms based on using a deep transfer-learning algorithm and a multi-feature fusion method. The results of our study indicated that both platforms can stage cataract well. In the automatic segmentation DTL platform, the segmentation process completed in just 2 min and 43 s, with training and test set accuracies of 94.59 and 84.50% respectively. On the other hand, the manual segmentation DTL platform required approximately a week for an experienced ophthalmologist to manually segment. However, the model achieved higher accuracies in the training and test sets, at 97.48 and 90.00%, respectively. On the whole, the manual segmentation DTL platform was more precise, whereas the automatic segmentation DTL platform was more rapid.

The grading of cataracts is based on the opacity of the lens, and good segmentation performance is the basis of classification. The difference in the tissue outside the lens might affect the classification results. In the automatic segmentation DTL platform, based on the FCNResnet50 model, we have presented a method for the automatic segmentation of the lens from cataract images. The segmentation results showed that the proposed model was able to segment the lens accurately. Compared with previous research, the PA, IoU, and Dice improved by 8.4, 14.9, and 9.5%, respectively (Cai et al., 2021).

In previous studies, Gao et al. (2015) used a deep learning method to grade nuclear cataracts, but the accuracy only reached 70%. Lin used a convolutional recursive neural network to develop an AI platform for diagnosing childhood cataracts and the accuracy was 87.4%, whereas the accuracy of our study reached 90.00%. In addition to the task itself, the reason is possibly caused by the algorithmic upgrading. In this study, we adopt three pretrained models trained on the ImageNet and then fine-tuned into our dataset, which makes up for the insufficient datasets and leads to a reduction in the learning time. We also adopt the early fusion approaches for the classification task. Early fusion is also called feature level fusion, which emphasizes data combination before the classification (Zhang et al., 2017), which reduced the influence of single feature inherent defects and realized feature complementarity. Multiple studies have also confirmed that the combination of different features presents better classification results than individual features (Fang et al., 2019; Wan and Tan, 2019; Nemoto et al., 2020).

Comparing the results of the ROC curves of the two platforms, the macro average calculates the indicators of

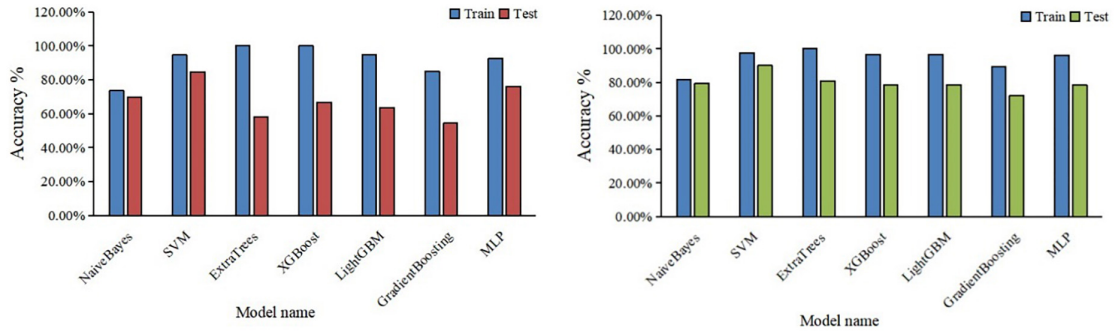


FIGURE 5
Boxplots for the range of accuracy of each classification model, the automatic segmentation platform **(left)**; the manual segmentation platform **(right)**.

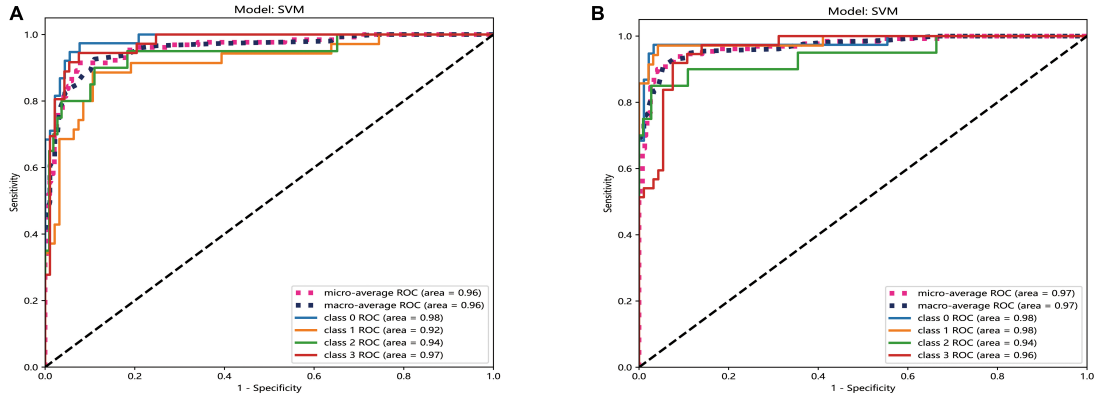


FIGURE 6
The ROC curves of SVM model of two different platforms in the test set. The automatic segmentation platform **(A)**; the manual segmentation platform **(B)**, "Class 0" indicated incipient stage, "class 1" indicated intumescent stage, "class 2" indicated mature stage, and "class 3" indicated hypermature stage.

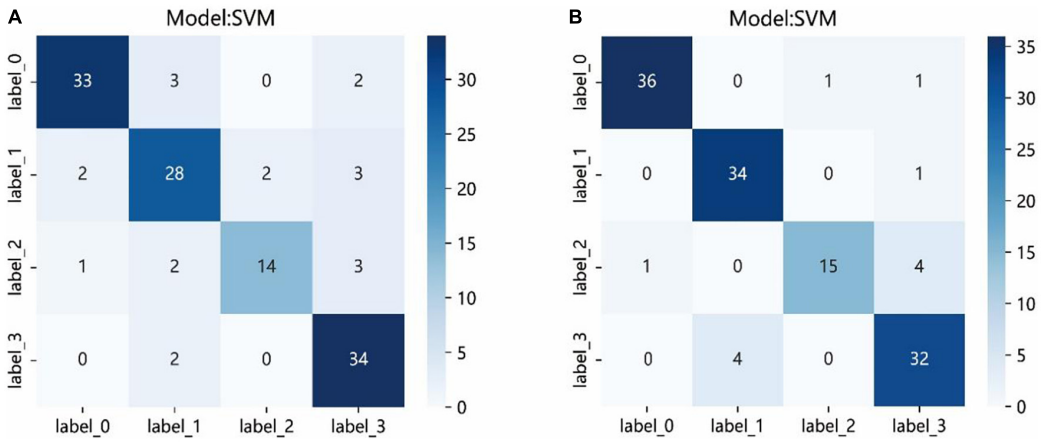


FIGURE 7
The confusion matrix of the SVM model of two different platforms in the test set. Each column represents the predicted class and each row of the matrix represents the actual class. The automatic segmentation platform **(A)**; the manual segmentation platform **(B)**. "Label 0" indicated incipient stage, "label 1" indicated intumescent stage, "label 2" indicated mature stage, and "label 3" indicated hypermature stage.

each class independently and then takes the mean value to treat all classes equally; the micro average aggregates the contributions of all classes to calculate the average indicator (Huang et al., 2022). The results of the two platforms can reach >95%, indicating that both show good performance. The AUC for each classification was >90% in the

test set, indicating that both platforms have excellent classification accuracy.

The confusion matrices showed the prediction results of each sample in the test set. Although the results showed that all stages, except for mature, had high recognition rate, the probabilities of correct identification (PCIs) of mature stage in the two platforms achieved >70%. The images of the mature stage were misassigned to the hypermature stage easily. The major reason for this result might be that the staging of the cataracts is determined by the opacity of the lens, also, sometimes it is hard to define clear boundaries of adjacent stage, and a large sample size might be required. Compared with other stages, the sample size of the mature stage was the smallest.

This study had some other limitations. All data are only based on the diffuse-illuminated photographs, it is important to note that slit-lamp photography, fundus photography, and clinical data can also provide valuable insights into the disease. And this study was only based on clinical diagnosis of the disease; however, individualized treatment is an integral and mandatory part of precision medicine. Therefore, in future studies, we will increase the sample size and combine multiple modal data to combine diagnosis and treatment, to build a more perfect and convenient AI platform for clinical diagnosis and treatment.

5. Conclusion

In this study, two AI diagnosis platforms have been proposed for cortical cataract staging. Through the multi-feature transfer-learning method combined with an automatic or manual segmentation algorithm, the resulting automatic segmentation platform can stage cataracts more quickly, whereas the resulting manual segmentation platform can stage cataracts more accurately.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Declaration of Helsinki and was approved by the Medical Ethics Committee of the Jiangxi Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

FG, HL, W-GQ, and S-LZ contributed to data collection, statistical analyses, and wrote the manuscript. All authors read and approved the final manuscript, contributed to the manuscript and approved the submitted version.

Acknowledgments

Some of our experiments were carried out on OnekeyAI platform. Thanks OnekeyAI and it's developers' help in this scientific research work.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1182388/full#supplementary-material

References

Amjad, A., Khan, L., and Chang, H. (2021). Effect on speech emotion classification of a feature selection approach using a convolutional neural network. *PeerJ Comput. Sci.* 7:e766. doi: 10.7717/peerj-cs.766

Bang, C., Yoon, J., Ryu, J., Chun, J., Han, J., Lee, Y., et al. (2021). Automated severity scoring of atopic dermatitis patients by a deep neural network. *Sci. Rep.* 11:6049. doi: 10.1038/s41598-021-85489-8

Cai, W., Xu, J., Wang, K., Liu, X., Xu, W., Cai, H., et al. (2021). EyeHealer: A large-scale anterior eye segment dataset with eye structure and lesion annotations. *Precis. Clin. Med.* 4, 85–92. doi: 10.1093/pcmedi/pbab009

Do, D., Gichuhi, S., Vedula, S., and Hawkins, B. (2013). Surgery for post-vitrectomy cataract. *Cochrane Database Syst. Rev.* 12:CD006366. doi: 10.1002/14651858.CD006366.pub3

Fang, T., Zhang, Z., Sun, R., Zhu, L., He, J., Huang, B., et al. (2019). RNAm5CPred: Prediction of RNA 5-methylcytosine sites based on three different kinds of nucleotide composition. *Mol. Ther. Nucleic Acids* 18, 739–747. doi: 10.1016/j.omtn.2019.10.008

Gao, X., Lin, S., and Wong, T. (2015). Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans. Biomed. Eng.* 62, 2693–2701. doi: 10.1109/TBME.2015.2444389

Hasan, M., Tanha, T., Amin, M., Faruk, O., Khan, M., Aljahdali, S., et al. (2021). Cataract disease detection by using transfer learning-based intelligent methods. *Comput. Math. Methods Med.* 2021:7666365. doi: 10.1155/2021/7666365

He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J., et al. (2020). A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci. Rep.* 10:15072. doi: 10.1038/s41598-020-71914-x

Huang, C., Tian, J., Yuan, C., Zeng, P., He, X., Chen, H., et al. (2019). Fully automated segmentation of lower extremity deep vein thrombosis using convolutional neural network. *Biomed. Res. Int.* 2019:3401683. doi: 10.1155/2019/3401683

Huang, T., Yang, R., Shen, L., Feng, A., Li, L., He, N., et al. (2022). Deep transfer learning to quantify pleural effusion severity in chest X-rays. *BMC Med. Imaging* 22:100. doi: 10.1186/s12880-022-00827-0

Kim, S., and Hidaka, Y. (2021). Breathing pattern analysis in cattle using infrared thermography and computer vision. *Animals (Basel)* 11:207. doi: 10.3390/ani11010207

Lao, J., Chen, Y., Li, Z., Li, Q., Zhang, J., Liu, J., et al. (2017). A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* 7:10353. doi: 10.1038/s41598-017-10649-8

Larsen, A., Hanigan, I., Reich, B., Qin, Y., Cope, M., Morgan, G., et al. (2021). A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. *J. Expo. Sci. Environ. Epidemiol.* 31, 170–176. doi: 10.1038/s41370-020-0246-y

Lin, D., Chen, J., Lin, Z., Li, X., Zhang, K., Wu, X., et al. (2020). A practical model for the identification of congenital cataracts using machine learning. *EBioMedicine* 51:102621. doi: 10.1016/j.ebiom.2019.102621

Lin, H., Lee, H., Tague, N., Lugagne, J., Zong, C., Deng, F., et al. (2021). Microsecond fingerprint stimulated Raman spectroscopic imaging by ultrafast tuning and spatial-spectral learning. *Nat. Commun.* 12:3052. doi: 10.1038/s41467-021-23202-z

Masood, N., and Farooq, H. (2019). Investigating EEG patterns for dual-stimuli induced human fear emotional state. *Sensors (Basel)* 19:522. doi: 10.3390/s19030522

McGonigle, I. (2016). The collective nature of personalized medicine. *Genet. Res. (Camb)* 98:e3. doi: 10.1017/S0016672315000270

Nemoto, K., Shimokawa, T., Fukunaga, M., Yamashita, F., Tamura, M., Yamamori, H., et al. (2020). Differentiation of schizophrenia using structural MRI with consideration of scanner differences: A real-world multisite study. *Psychiatry Clin. Neurosci.* 74, 56–63. doi: 10.1111/pcn.12934

Satyam, S., Bairy, L., Pirasanthan, R., and Vaishnav, R. (2015). Grape seed extract and Zinc containing nutritional food supplement delays onset and progression of Streptozocin-induced diabetic cataract in Wistar rats. *J. Food Sci. Technol.* 52, 2824–2832. doi: 10.1007/s13197-014-1305-y

Son, K., Ko, J., Kim, E., Lee, S., Kim, M., Han, J., et al. (2022). Deep learning-based cataract detection and grading from slit-lamp and retro-illumination photographs: Model development and validation study. *Ophthalmol. Sci.* 2:100147. doi: 10.1016/j.xops.2022.100147

Takahashi, S., Takahashi, M., Kinoshita, M., Miyake, M., Kawaguchi, R., Shinojima, N., et al. (2021). Fine-tuning approach for segmentation of gliomas in brain magnetic resonance images with a machine learning method to normalize image differences among facilities. *Cancers (Basel)* 13:1415. doi: 10.3390/cancers13061415

Tsuji, S., Sekiryu, T., Sugano, Y., Ojima, A., Kasai, A., Okamoto, M., et al. (2020). Semantic segmentation of the choroid in swept source optical coherence tomography images for volumetrics. *Sci. Rep.* 10:1088. doi: 10.1038/s41598-020-57788-z

Wan, X., and Tan, X. (2019). A study on separation of the protein structural types in amino acid sequence feature spaces. *PLoS One* 14:e0226768. doi: 10.1371/journal.pone.0226768

Wu, X., Huang, Y., Liu, Z., Lai, W., Long, E., Zhang, K., et al. (2019). Universal artificial intelligence platform for collaborative management of cataracts. *Br. J. Ophthalmol.* 103, 1553–1560. doi: 10.1136/bjophthalmol-2019-314729

Xie, Y., Nguyen, Q., Hamzah, H., Lim, G., Bellemo, V., Gunasekeran, D., et al. (2020). Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: An economic analysis modelling study. *Lancet Digit. Health* 2, e240–e249. doi: 10.1016/S2589-7500(20)30060-1

Zhang, L., Xu, Z., Xu, D., Ma, J., Chen, Y., and Fu, Z. (2020). Growth monitoring of greenhouse lettuce based on a convolutional neural network. *Hortic. Res.* 7:124. doi: 10.1038/s41438-020-00345-6

Zhang, P., Wang, X., Chen, J., and You, W. (2017). Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload. *Sensors (Basel)* 17:2315. doi: 10.3390/s17102315

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 27 April 2023
DOI 10.3389/fnins.2023.1174937

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Tingwei Quan,
Huazhong University of Science and Technology, China
Lin-Lin Zhu,
Peking University Sixth Hospital, China
Wu Jian,
Nanchang Hangkong University, China

*CORRESPONDENCE
Hai Tan
✉ 270602@nau.edu.cn
Zhijiang Wan
✉ wandndn@gmail.com

RECEIVED 27 February 2023
ACCEPTED 05 April 2023
PUBLISHED 27 April 2023

CITATION
Chen L, Zhou Y, Gao S, Li M, Tan H and Wan Z (2023) ARA-net: an attention-aware retinal atrophy segmentation network coping with fundus images.
Front. Neurosci. 17:1174937.
doi: 10.3389/fnins.2023.1174937

ARA-net: an attention-aware retinal atrophy segmentation network coping with fundus images

Lei Chen[1], Yuying Zhou[1], Songyang Gao[2], Manyu Li[2], Hai Tan[3]* and Zhijiang Wan[1,2,4]*

[1]Second College of Cinical Medicine, Nanchang University, Nanchang, Jiangxi, China, [2]Information Engineering College, Nanchang University, Nanchang, Jiangxi, China, [3]School of Computer Science, Nanjing Audit University, Nanjing, Jiangsu, China, [4]Industrial Institute of Artificial Intelligence, Nanchang University, Nanchang, Jiangxi, China

Background: Accurately detecting and segmenting areas of retinal atrophy are paramount for early medical intervention in pathological myopia (PM). However, segmenting retinal atrophic areas based on a two-dimensional (2D) fundus image poses several challenges, such as blurred boundaries, irregular shapes, and size variation. To overcome these challenges, we have proposed an attention-aware retinal atrophy segmentation network (ARA-Net) to segment retinal atrophy areas from the 2D fundus image.

Methods: In particular, the ARA-Net adopts a similar strategy as UNet to perform the area segmentation. Skip self-attention connection (SSA) block, comprising a shortcut and a parallel polarized self-attention (PPSA) block, has been proposed to deal with the challenges of blurred boundaries and irregular shapes of the retinal atrophic region. Further, we have proposed a multi-scale feature flow (MSFF) to challenge the size variation. We have added the flow between the SSA connection blocks, allowing for capturing considerable semantic information to detect retinal atrophy in various area sizes.

Results: The proposed method has been validated on the Pathological Myopia (PALM) dataset. Experimental results demonstrate that our method yields a high dice coefficient (DICE) of 84.26%, Jaccard index (JAC) of 72.80%, and F1-score of 84.57%, which outperforms other methods significantly.

Conclusion: Our results have demonstrated that ARA-Net is an effective and efficient approach for retinal atrophic area segmentation in PM.

KEYWORDS

retinal atrophy, segmentation, self-attention, multi-scale, 2D fundus images

1. Introduction

The eyes are one of the essential sensory organs in humans; many people worldwide have myopia, which causes many inconveniences in their lives. Holden et al. (2016) performed a meta-analysis of myopia prevalence. They predicted that by 2050, 49.8 and 9.8% of the world's population would suffer from myopia and high myopia, respectively. High myopia has the risk of deteriorating into pathological myopia. Retinal changes caused by myopia include fundus

tessellation, parapapillary atrophy, optic disc tilting, myopic maculopathy, and retinal detachment. Retinal atrophy is a condition that leads to the loss of retinal layers, affecting vision quality. It is associated with choroidal retinal thinning and attenuation of the parapapillary retinal pigment epithelium (RPE) adjacent to the optic nerve head (ONH). Myopia, glaucoma, and age-related macular degeneration (AMD) are among the diseases that can cause retinal atrophy (Jonas et al., 1988; Manjunath et al., 2010; Srinivas et al., 2018). Accurate segmentation of retinal atrophic regions from OCT or fundus images is essential for eye condition diagnosis, monitoring, and treatment. It enables personalized interventions and plays a crucial role in improving the overall management of these ocular conditions.

The degree of retinal atrophy is a valuable medical assessment indicator as it is correlated closely with the severity of ophthalmic diseases and conditions, including glaucomatous optic nerve damage, visual field defects, and myopia (Park et al., 1996; Uchida et al., 1998; Dai et al., 2013). Consequently, the segmentation of retinal atrophic regions has become a significant part of diagnosing ophthalmic diseases. Although experienced ophthalmologists can give accurate results, manual segmentation is laborious and time-consuming, and different ophthalmologists might make different treatments. The development of an automatic segmentation model to accurately segment the retinal atrophy regions is thus vital, as it offers a reliable, efficient, and arguably more consistent diagnosis for ophthalmic diseases. The automatic segmentation models always adopt fundus images to perform the segmentation task of retinal atrophy areas. Compare with three-dimensional (3D) fundus image, two-dimensional (2D) fundus image are more widely available and easier to be acquired. The 3D fundus images require special equipment and technology that may not be accessible or affordable for many clinics or researchers. In addition, the 2D fundus images can provide sufficient information for segmenting retinal atrophy areas, which are mainly located on the surface of the retina.

In previous studies, most segmentation models are based on traditional image segmentation algorithms with manually designed features. Lu et al. (2010) segmented and quantified the optic disc and parapapillary area automatically using a combination of techniques, such as scanning filter, thresholding, region growing, and a modified Chan-Vese model (Chan and Vese, 2001) with a shape constraint. Li et al. (2018) proposed a novel parapapillary atrophy segmentation algorithm that utilizes evenly-oriented radial line segments and ellipse fitting. Although the traditional methods utilize machine learning to implement the image segmentation algorithm, most of them require manual feature selection and are not end-to-end solutions. Recent strides in deep learning have enabled the utilization of deep learning-based techniques in the medical domain, surpassing traditional methods in image segmentation with a higher degree of accuracy. Current mainstream deep learning methods for object segmentation can be divided into convolution-based and transformer-based methods. Long et al. (2014) proposed full convolutional networks (FCN), a model foundation of many segmentation networks for pixel-wise semantic segmentation tasks. Transformer-based (Vaswani et al., 2017) image segmentation models have emerged because they learn a global understanding of images which facilitates image segmentation models to achieve accurate segmentation results.

Deep learning models, especially UNet, have been widely adopted in various studies for segmenting areas of retinal atrophy based on fundus images. The UNet is a convolutional neural network designed for biomedical applications (Ronneberger et al., 2015). The core module, the FCN, utilizes the skip connections between the encoder and decoder to improve model performance. Due to its low demand on dataset size and the U-shaped structure containing contextual information, UNet has become a prevalent choice for medical segmentation and yields promising results. Furthermore, variations of UNet have been proposed to enhance the model performance. Zhou et al. (2018) proposed UNet++ and re-designed the skip pathways to reduce the semantic gap between the feature maps of the encoder and decoder networks. Guo et al. (2020) proposed a novel Lesion-aware segmentation network inspired by the UNet encoder-decoder structure and contained a binary classifier. The feature flows were integrated into the decoder to absorb various scales of feature maps. Ruben et al. (2020) evaluated the detection of pathological myopia (PM) using deep learning and the semantic segmentation of myopia-related lesions from fundus images. They used UNet++ as their network and used ResNet-18 as encoders. Chai et al. (2020) proposed a novel multi-task fully convolutional network (MFCN) model for peripapillary atrophy area segmentation from retinal images by transforming the atrophic area into two regions with relatively regular and uniform shapes. Wan et al. (2021) proposed OT-Unet, combining parallel partial decoder, edge attention, and reverse attention modules to enhance the segmentation accuracy.

Although the existing UNet-based retinal area segmentation algorithms achieved good results, the performance of the segmentation model is challenged by the following characteristics of retinal atrophic regions in the 2D fundus image, such as blurred boundaries, irregular shape, and size variation. These characteristics can make the segmentation models challenging to segment the areas accurately. As depicted in Figure 1, the top and bottom parts show a 2D fundus image and the corresponding retinal atrophy areas (i.e., areas in white), respectively. The bottom parts of subfigures (a) and (b) exhibit larger areas of retinal atrophy, while the bottom parts of subfigures (c), (d), and (e) show significantly reduced areas of atrophy. From the figure, we know that retinal atrophic areas can vary significantly in size from patient to patient, and the areas of each patient are randomly distributed in the fundus image. In addition, the model's generalization ability is restricted by the limited availability of annotated fundus image datasets.

To overcome these challenges, we proposed an attention-aware retinal atrophy segmentation network based on UNet structure, containing an encoder, a decoder, skip self-attention (SSA) connection blocks and multi-scale feature flow (MSFF), to segment retinal atrophic areas from 2D fundus image. The SSA connection block contains two distinct components: a shortcut and a parallel polarized self-attention (PPSA) block. The shortcut helps to preserve the original features, while the PPSA block can enhance feature learning capabilities. The PPSA block can capture contextual information over long distances and map semantic information from the encoder in both channel and spatial dimensions. Thus, it enables the model to learn features more robustly for alleviating the challenges, such as blurred boundaries and irregular shapes of the retinal atrophy region. Furthermore, a MSFF is added between the SSA connection blocks to address the challenge of size diversity in the retinal atrophy region. In addition, the transfer learning strategy and data augmentation are introduced to improve the performance and generalization ability

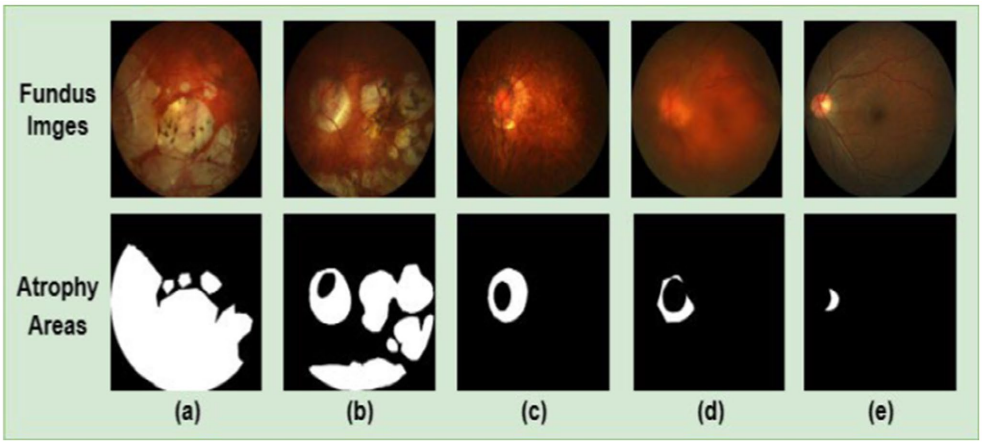


FIGURE 1
2D fundus images (top part) and the corresponding retinal atrophic areas (bottom part). Subfigures **(A–D)** are fundus images of patients with PM. Subfigure **(E)** is a fundus image of patients with non-pathological myopia. PM images have more significant retinal atrophic regions than non-pathological myopia images.

of the model. Due to the significant time and computational resources required for developing effective deep learning models, transfer learning has become a widely used strategy. Transfer learning uses the knowledge gained from pre-trained models to improve the performance of new tasks. Pre-trained models can train models efficiently for new segmentation tasks, reducing time and computational costs. And the transfer learning can help combat overfitting by providing a starting point for the model and reducing the reliance on training data. To improve the sensitivity rate of the retinal atrophy segmentation, a customized hybrid loss was employed to assign a higher weight to false negative detections. It enabled the algorithm to be more sensitive to false negative detections, thus leading to more precise segmentation results. The main contributions of our work can be enumerated as follows:

1. We proposed a novel skip connection block named the SSA connection block, which can be easily integrated into existing UNet-based architectures. The SSA connection block can better capture the global structure of the retinal atrophy, allowing the model to learn features more robustly. It is capable of dealing with blurred boundaries and irregular shapes of the retinal atrophic region. Additionally, it only requires a minimal increase in computational overheads.
2. We proposed an MSFF between the SSA connection blocks, allowing the network to capture multi-scale semantic information and significantly enhancing the self-attention mechanism's ability to capture multi-scale spatial and channel features. Thus, it improves the segmentation performance and more accurate detection of retinal atrophy and effectively addresses the challenge of size diversity in the retinal atrophic region.
3. We introduced a learning strategy to improve the performance and generalization of the model. By employing pre-trained models with large datasets to initialize the model weights, adaptation to new datasets with reduced training data is expedited. This improves segmentation accuracy, shortens training time, and reduces computing resources.

2. Materials and methods

2.1. Data preparation

Retinal images were obtained from the "Detection of Pathological Myopia from Retinal Images" challenge (iChallenge-PALM) held at the IEEE International Symposium on Biomedical Imaging, organized in 2019 (Fu et al., 2019). The training and validation datasets contain 311 fundus images and 271 fundus images, respectively. The Zeiss VISUCAM device took these fundus images at an angle of 45° with a resolution of 2,124 × 2,056, or 30 ° angle with a resolution of 1,444 × 1,444. To improve the computational efficiency and conserve computing resources, all fundus images were resized to 512 × 512 and normalized to facilitate faster and more stable processing by the neural network. Finally, a logical AND operation of the network's predicted mask and the original image was performed to generate the resulting color output.

2.2. Model architecture

Figure 2 illustrates the proposed deep learning network for segmenting retinal atrophic areas from 2D fundus images. The proposed network consists of an encoder, a decoder, and the PPSA block. The encoder is responsible for extracting the features of an input image, while the decoder is responsible for recovering the image details and capturing the boundaries of the retinal atrophy region. The PPSA blocks act as a bridge between the encoder and decoder, providing a source of feature information to the decoder. This connection is essential in allowing the network to reconstruct high-frequency details.

2.2.1. UNet backbone

UNet is a fully convolutional network architecture for medical research applications such as segmentation and classification. It consists of an encoder and a decoder based on convolutional neural networks (CNNs). The encoder utilizes 3 × 3 convolutional layers,

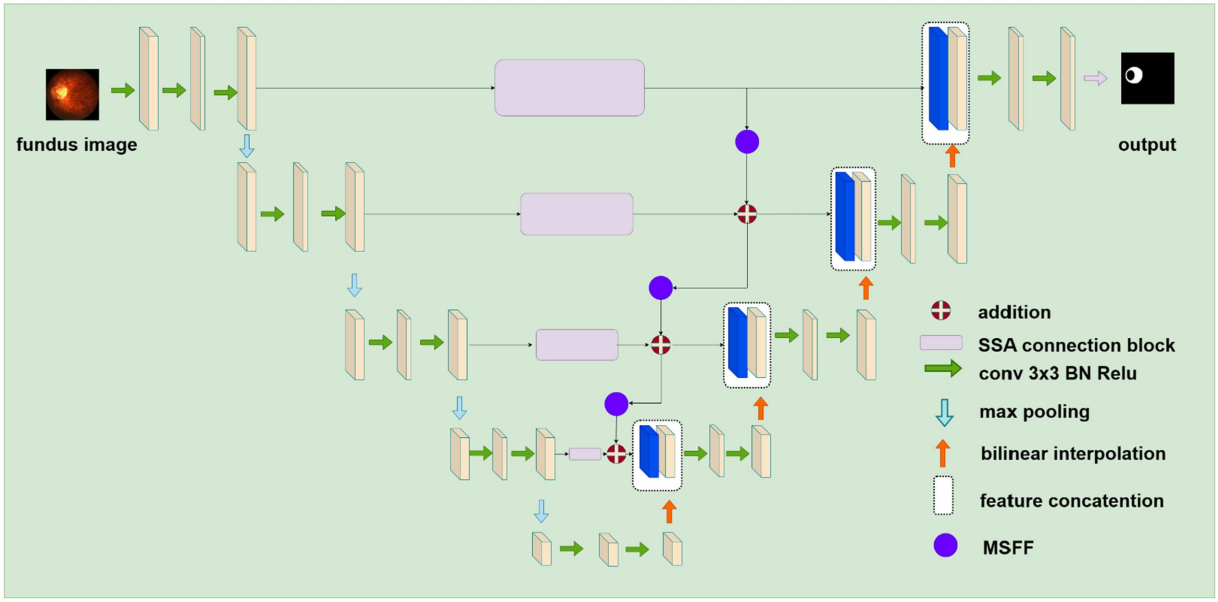


FIGURE 2
The structure of the proposed deep learning network for segmenting retinal atrophic areas from 2D fundus images. The encoder on the left extracts features, and the decoder on the right recovers image details. The SSA connection block represented by purple rectangular flows the feature from the encoder to the decoder.

batch normalization, 2×2 max-pooling layers, and ReLu activation functions to extract features from the input image. In contrast, the decoder contains transposed convolutional layers (bilinear interpolation), 3×3 convolutional layers, batch normalization, 2×2 max-pooling layers, and ReLu activation functions to reduce the number of channels and to segment pixels into different regions. The skip connections between the encoder and decoder networks further facilitate the flow of low-level features from the encoder to the decoder and improve segmentation performance. UNet can learn complex feature representations and provide high-quality performance in biomedical image segmentation tasks.

2.2.2. SSA connection block

In the original UNet architecture, features from the encoder flow directly to the decoder *via* a skip connection. However, the features received by the decoder are mostly background information, do not provide meaningful semantic information, and are not on the same semantic level as the encoder. To overcome this problem and make the network more attentive to atrophic region edges and shape information, we propose the SSA Connection block, which can better capture long-range dependencies in the feature maps. In addition, the parallel polarization design of the self-attention mechanism allows the block to learn feature maps in both the spatial and channel dimensions, allowing it to capture features of the retinal atrophy region effectively.

Our skip self attention blocks not only retain the original Unet skip connection which allow direct connections between the encoder and decoder layers, preserving low-level features that can then be combined with high-level features, but also include a PPSA branch in which contains two key modules: polarized self-attention and mapping enhancement. Figure 3 shows the proposed SSA block, subfigure (a) gives a shortcut path and PPSA block paths, and subfigure (b) introduces the detailed structure of the PPSA block. As shown in subfigure (b), the PPSA block has two convolution layers, followed by polarized self-attention, which contains two branches: Spatial-only self-attention and Channel-only self-attention. Polarized filtering is a design technique in deep learning that involves maintaining the internal resolution of both the channel and spatial attention computations, while reducing the dimensionality of the input data. This helps to filter out irrelevant data and preserve important details, allowing the model to focus on the most important features. Mapping enhancement is a design strategy that involves mapping the output of the model to a distribution that more closely resembles a typical fine-grained regression. In retinal atrophy segmentation, the output can be mapped to a 2D Binomial distribution that represents the probability of each pixel belonging to the segmented object. This design helps the model to better fit the output to the desired distribution, resulting in more accurate predictions. The computation method of the spatial-only self-attention is given as follows:

$$Out_{sp}(X) = Sigmoid \left(\left(Softmax\left(GP\left(Conv_{1\times1}(X) \right) \right)^{R} \times \left(Conv_{1\times1}(X) \right)^{R} \right)^{R} \right), \quad (1)$$

where $Conv_{1\times1}$ is standard convolutional layer using 1×1 convolution, Softmax and Sigmoid are activation functions, R is tensor reshape operation, and GP is global average operation. The computation method of the channel-only self-attention is described as follows:

$$Out_{ch}(X) = Sigmoid \left(LN\left(Conv_{1*1}\left(Softmax\left(Conv_{1*1}(X) \right)^{R} \times Conv_{1*1}(X) \right)^{R} \right) \right) \quad (2)$$

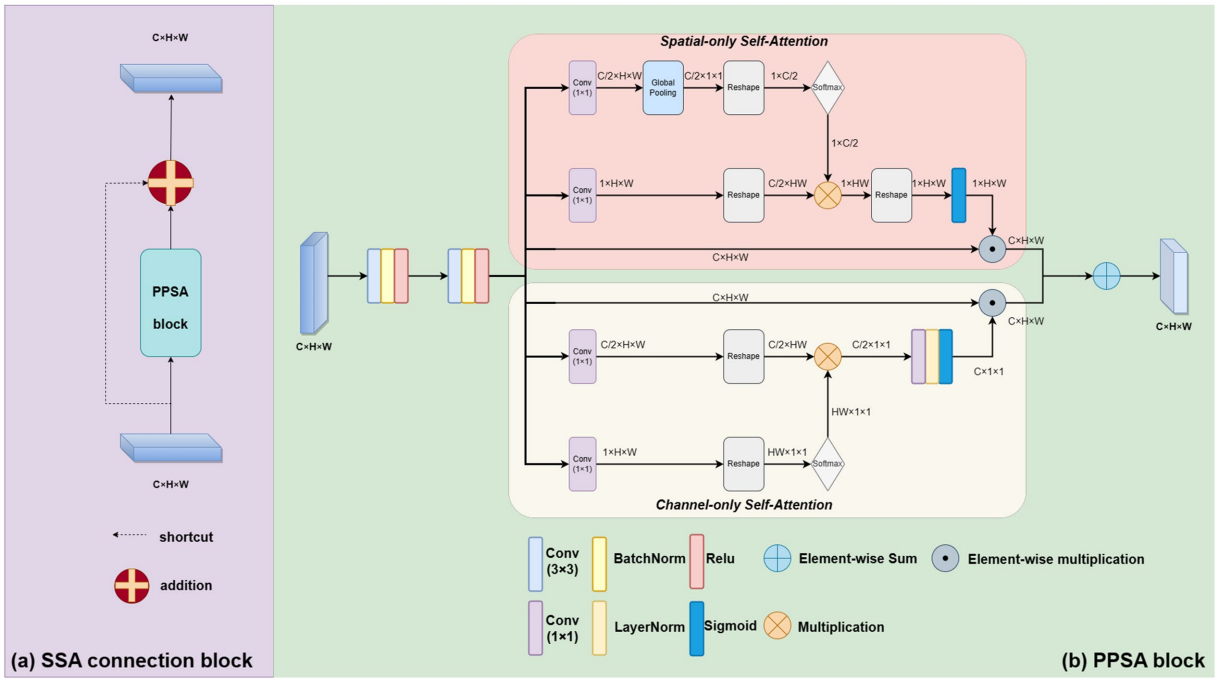


FIGURE 3
The proposed novel SSA connection block. As shown in subfigure **(A)**, it contains a shortcut path and PPSA block paths. The PPSA block is shown in subfigure **(B)**.

where LN means layer normalization. The final output of the PPSA block is listed as follows:

$$PPSA(X) = CBR\left(CBR\left(Out_{sp}(X) \underset{sp}{\odot} X + Out_{ch}(X) \underset{ch}{\odot} X\right)\right), \quad (3)$$

where CBR is a combination of the convolution layer, BatchNorm, and ReLu activation function, $\underset{sp}{\odot}$ and $\underset{ch}{\odot}$ are multiplication operators in spital and channel dimensions, respectively. The output of the skip connection block is given as follows:

$$Block_{skip}(X) = X + PPSA(X). \quad (4)$$

2.2.3. Multi-scale feature flow

The MSFF is a critical component of our proposed model, as it integrates features from multiple resolutions to understand the input data comprehensively. It enhances the ability of the self-attention mechanism to capture spatial and channel features at different resolutions. This is especially important in retinal atrophy segmentation, where the size and shape of the atrophic lesions can vary greatly. MSFF is implemented using a 1×1 convolutional layer followed by a 2×2 average pooling layer that downsamples the high-resolution feature map to a low-resolution feature map. This affects the resolution of the feature maps, but by inserting MSFF blocks between SSA blocks to enhance the model's ability to perceive multi-scale semantic information. In this case, the MSFF blocks reduce the resolution of the feature maps without affecting the overall perceptual capability of the network, as the feature maps from different scales are already fused together. This ultimately improves the model's segmentation accuracy, thus mitigating information loss and distortion. The output of the skip connection block is given as follows:

$$MSFF(X) = \mathrm{AP}_{2*2}\left(Conv_{1*1}(X)\right). \quad (5)$$

2.3. Loss function

The image segmentation task can be viewed as a pixel-level classification problem, which refers to the process of categorizing every pixel in an image into its corresponding semantic class, such as atrophic regions or healthy tissue. In the context of retinal atrophy segmentation, pixel-level prediction enables the identification and localization of atrophic regions with high precision. This is achieved by assigning every pixel in the image a label based on its semantic class (i.e., atrophic or healthy tissue), allowing for the creation of a segmented image highlighting the area of atrophic regions. Therefore, we used binary cross-entropy loss (BCE), shown in Eq. 6, as a part of the loss function. However, since the pixel number of the region of interest (ROI) generally accounts for less than 20% of the total pixel number in an image, only using the BCE loss function alone may not be sufficient for accurate ROI classification as it prioritizes overall accuracy over foreground or background classification. In addition, the foreground pixels in the most fundus images, which includes the retina and blood vessels, makes up about 90% of the image. The

pixels of atrophic area, makes up less than 10% of the image. The problem with this sample imbalance is that if a model is trained using a simple cross-entropy loss function, which is a common loss function used in image classification tasks, the model may learn to predict the outcome tendency as the foreground. In other words, the model may learn to ignore the atrophic area and only focus on the larger, more prominent foreground. This might be a problem in medical image analysis, because it may lead to false negatives or missed diagnosis. The Tversky Loss introduces the Tversky Index, with two adjustable parameters, α and β, that balance false positives and false negatives while prioritizing false negatives for small ROIs. Incorporating the Tversky Loss into the loss function can improve the accuracy of ROI classification, leading to better overall performance in the task. Therefore, we introduce the Tversky Loss (L_{TL}) (Abraham and Khan, 2018), shown in Eq. 7, in the loss function to improve sensitivity for small ROIs. The Tversky Loss introduces the Tversky Index, with two adjustable parameters, α and β, that balance false positives and false negatives while prioritizing false negatives for small ROIs. The final combination loss function is shown in Eq. 8,

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[\left(1-y_i\right)\cdot\log\left(1-\left(p\left(y_i\right)\right)\right)+y_i\cdot\log\left(p\left(y_i\right)\right)\right], \quad (6)$$

$$L_{TL} = 1-\frac{\sum_{i=1}^{N} pa(y_i)ga(y_i)+\xi}{\sum_{i=1}^{N} pa(y_i)ga(y_i)+\alpha\sum_{i=1}^{N} pb(y_i)ga(y_i)+\beta\sum_{i=1}^{N} pa(y_i)gb(y_i)+\xi}, \quad (7)$$

$$L_f = \mu\times L_{BCE}+\left(1-\mu\right)\times L_{TL}, \quad (8)$$

where the total number of pixels is represented by N, and the pixels y_i are labeled by 1 to represent the pixels located in the atrophic area. The pixels in the image background are labeled by 0. $p_a(y_i)$ is the probability of the pixel belonging to the atrophic area, and $p_b(y_i)$ is used to denote the probability of the pixel in the image background. $g_a(y_i)$ is the probability of the pixel in ground truth belonging to the atrophied area, and $g_b(y_i)$ is the probability of the pixel in ground truth belonging to the background. We set $\propto$ to 0.75, α to 0.6 and β to 0.4.

3. Experiments and results

3.1. Evaluation metric

To evaluate the performance of our model and compare it with other models, we use six evaluation metrics: Dice coefficient (DICE) (Dice, 1945), Jaccard index (JAC, IoU) (Jaccard, 1912), Precision (PRE), Sensitivity (SEN, also known as recall), Accuracy (ACC) and F1-score. The six metrics are defined as follows:

$$DICE = \frac{2\cdot|G\cap P|}{|G| + |P|} = \frac{2\cdot TP}{FP+FN+2\cdot TP}, \quad (9)$$

$$JAC = \frac{|G\cap P|}{|G\cup P|} = \frac{TP}{FP+FN+TP}, \quad (10)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}, \quad (11)$$

$$PRE = \frac{TP}{TP+FP}, \quad (12)$$

$$SEN = \frac{TP}{TP+FN}, \quad (13)$$

$$F1-score = \frac{2\cdot PRE\cdot SEN}{PRE+SEN}, \quad (14)$$

where G and P refer to ground truth and predicted mask, respectively, TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. The DICE and JAC metrics are used to measure how accurate the predicted segmentation produced by a model is when compared to the ground truth. The scores range from 0 to 1, with higher scores indicating a higher level of accuracy. These metrics are useful for evaluating model performance and comparing different models against each other. The F1-score is a combination of PRE and SEN. PRE measures how often the model accurately identifies positives, while SEN measures how often the model identifies true positives and is a measure of sensitivity. The F1-score ranges from 0 to 1, with higher scores indicating a better balance between precision and sensitivity. This metric is important because it provides an overall assessment of how well the model is able to identify true positives and true negatives while minimizing false positives and false negatives. Overall, these metrics are important for evaluating a model's ability to accurately identify the target object without including false positives or false negatives.

3.2. Implementation details

The implementation of the proposed network is based on the Pytorch computing library and was performed on a system equipped with RTX A5000, offering 24 GB of memory, and Tesla P100 with 16 GB of memory. To improve the generalization performance and prevent overfitting of our model, we employed various data augmentation techniques, including random flipping and image cropping. By setting a horizontal and vertical flip probability of 0.5, we randomly flip each image, while cropping the training images to 320×320 allows us to extract multiple sub-images from a single image, increasing the quantity and variability of our training data. To optimize the network, the initial learning rate of the AdamW optimizer was set to 0.008, the betas to (0.9, 0.999) and the weights decayed to 1e-4. Mini-batch is a popular optimization algorithm used for the training of deep neural networks. It works by dividing the training data into smaller batches and updating the model parameters

based on the gradients computed from each batch. The batch size in our experiment was set to 32, ensuring comparable training and consistency from all experiments.

3.3. Ablation study

3.3.1. Effectiveness of blocks

We conducted an ablation study to investigate the contribution of each block to the model performance. We began with the native UNet and then sequentially added novel SSA connection blocks and MSFF to the architecture to show the effectiveness of each of our proposed blocks. For the experiments conducted, four groups of models were developed, each consisting of an ARA-Net model combined with a different backbone ResNet18 (ARA-Net(ResNet18)), MobileNet-v3 (ARA-Net(MobileNet-v3)), and EfficientNet-b3 (ARA-Net(EfficientNet-b3)). ResNet18, with its 18 layers and residual connections, allows the network to learn deeper features and prevent vanishing gradients. MobileNet-v3, on the other hand, uses depth wise separable convolutions to reduce the number of parameters and computation needed while maintaining good accuracy. EfficientNet-b3 uses a compound scaling method to optimize the network architecture for different resource constraints, making it a relatively lightweight choice for real-time applications. Three ablation experiments were conducted for each group, namely type a (as the blank control group), type b (with novel SSA connection blocks added), and type c (with multi-scale feature flows added on top of type b). The results are presented in Table 1. The computation is calculated regarding the number of floating point operators (FLOPs). In all these experiments, we used the loss function defined in Eq. 8.

As illustrated in Table 1, among the four experimental groups, the type c model has higher performance metrics in DICE, JAC, and F1-score, yielding a uniform 0.5% improvement compared to type b and a significant improvement compared to the control group, type a. The superior performance achieved by the proposed method with minimal computational overhead increase demonstrates that the novel SSA connection block and MSFF can provide a better solution to the difficulties of segmenting retinal atrophy.

3.3.2. Effectiveness of transfer learning

To measure the effectiveness of using transfer learning, we also set up three groups of experiments, and the results are shown in Table 2. The results show that using transfer learning achieves the best results. Type 'a' represents the control groups without transfer learning, and type 'b' represents groups that use transfer learning. Note that, in all these experiments, we use the loss function defined in Eq. 8. The application of transfer learning to segmentation methods has been proven to produce significant performance gains.

Specifically, when this strategy was implemented in segmentation tasks, the DICE, JAC, and F1-score realized an average improvement of 4, 6, and 4%, respectively. This demonstrates the potential of transfer learning to develop more accurate segmentation models. Moreover, the boost in performance can result in higher quality predictions of segmentation masks and increased usability in medical applications.

3.3.3. Effectiveness of loss functions

To show the model performance affected by the loss function, we utilize three backbones (ResNet18, MobileNet-v3, and EfficientNet-b3) to build and compare their model performance across three loss functions (BCE loss, Tversky loss, and their combination). The results obtained are shown in Table 3. There is about a 2% improvement of DICE and F1-score when combining the two loss functions compared to using any one of the two loss functions alone.

3.4. Comparison study

We compare our proposed method with other UNet-based methods, including UNet++ (Zhou et al., 2018), AttentionUNet (Oktay et al., 2018), R2UNet (Alom et al., 2018), and UNeXt (Valanarasu and Patel, 2022). Table 4 shows the comparison results, demonstrating that our model performs better than the comparison methods. Results from Table 4 indicate that ARA-Net (EfficientNet-b3) performed best among all architectures, achieving a DICE of 84.57%, a JAC of 72.80%, an ACC of 96.95%, a PRE of 89.09%, an SEN of 80.49%, and an F1-score of 84.57%. Further, this was followed by ARA-Net (MobileNet-v3) and UNeXt. However, their performance is

TABLE 1 Performance comparison, number of parameters, and FLOPs of different models for ablation experiments.

Methods	Type	DICE (%)	JAC (%)	ACC (%)	PRE (%)	SEN (%)	F1-score (%)	Param (M)	FLOPs (G)
ARA-Net	a	10.31	5.43	88.67	40.52	45.00	42.64	16.47	40.29
	b	71.29	55.38	95.13	75.35	68.01	71.49	2.78	30.77
	c	71.63	55.80	95.00	76.25	68.26	72.03	2.78	30.77
ARA-Net (ResNet18)	a	70.40	54.32	94.50	73.60	68.87	71.16	81.53	15.71
	b	73.33	57.90	95.16	73.40	74.32	73.86	108.79	21.07
	c	74.27	59.06	95.17	77.93	72.39	75.06	108.79	21.07
ARA-Net (MobileNet-v3)	a	76.37	61.77	95.51	75.36	78.24	76.77	13.18	3.24
	b	76.15	61.49	95.12	77.66	77.48	77.57	14.35	4.16
	c	77.27	62.96	95.54	83.09	73.94	78.25	14.35	4.16
ARA-Net (EfficientNet-b3)	a	76.64	62.13	95.49	81.50	74.19	77.67	44.51	5.51
	b	77.03	62.65	95.65	82.62	73.18	77.62	57.68	7.17
	c	78.60	64.74	95.82	81.53	77.08	79.24	57.68	7.17

TABLE 2 Performance comparison of the proposed method with and without using transfer learning.

Methods	Type	DICE (%)	JAC (%)	ACC (%)	PRE (%)	SEN (%)	F1-score (%)
ARA-Net (ResNet18)	a	74.27	59.06	95.17	77.93	72.39	75.06
	b	76.47	61.90	95.75	79.29	75.66	77.43
ARA-Net (MobileNet-v3)	a	77.27	62.96	95.54	83.09	73.94	78.25
	b	82.30	69.91	96.37	86.96	79.17	82.88
ARA-Net (EfficientNet-b3)	a	78.60	64.74	95.82	81.53	77.08	79.24
	b	84.57	72.80	96.95	89.09	80.49	84.57

TABLE 3 Performance comparison of the methods across different loss functions.

Methods	Loss function	DICE (%)	JAC (%)	ACC (%)	PRE (%)	SEN (%)	F1-score (%)
ARA-Net (ResNet18)	BCE	68.74	52.36	94.61	68.97	68.30	68.63
	Tversky Loss	74.91	59.89	94.83	77.28	73.92	75.56
	BCE + Tversky Loss	76.47	61.90	95.75	79.29	75.66	77.43
ARA-Net (MobileNet-v3)	BCE	76.56	62.02	96.18	84.82	69.94	76.66
	Tversky Loss	78.03	63.97	94.81	75.42	84.07	79.51
	BCE + Tversky Loss	82.30	69.91	96.37	86.96	79.17	82.88
ARA-Net (EfficientNet-b3)	BCE	82.39	70.05	96.66	89.43	76.38	82.39
	Tversky Loss	82.99	70.93	96.57	90.37	77.23	83.29
	BCE + Tversky Loss	84.57	72.80	96.95	89.09	80.49	84.57

TABLE 4 Comparison of segmentation performance, number of parameters, and FLOPs of different methods.

Methods	DICE (%)	JAC (%)	ACC (%)	PRE (%)	SEN (%)	F1-score (%)	Param (M)	FLOPs (G)
UNet++(ResNet-18)	76.84	62.39	96.12	86.67	69.66	77.24	60.92	64.05
AttentionUNet	71.00	55.04	94.50	67.79	78.97	72.95	54.56	67.21
R2UNet	68.13	51.67	94.10	78.29	66.06	71.65	388.88	262.76
UNeXt	77.01	62.62	95.58	80.00	74.71	77.26	5.61	0.57
ARA-Net (ResNet18)	76.47	61.90	95.75	79.29	75.66	77.43	108.79	21.07
ARA-Net (MobileNet-v3)	82.30	69.91	96.37	86.96	79.17	82.88	14.35	4.16
ARA-Net (EfficientNet-b3)	84.57	72.80	96.95	89.09	80.49	84.57	57.68	7.17

commendable due to fewer parameters and computation resources compared to EfficientNet-b3.

Figure 4 compares DICE scores for various segmentation methods, where the parameters and GLOPs utilized in each method are varied. It can be seen from the results that the ARA-Net (EfficientNet-b3) performs most efficiently in terms of segmentation performance. This is due to the comparatively low GLOPs requirement and the number of parameters needed for this model compared to other methods. UNeXt and ARA-Net (MobileNet-v3) perform significantly better than any other networks in terms of computational complexity and parameter count, which are crucial considerations for practical applications. Figure 5 depicts the change in loss function values during training and DICE values on the validation set. As illustrated in Figure 5, the training process of the networks proceeded gradually, except for R2Unet, which had a more stable loss. On the validation set, the Dice scores for ARA-Net (EfficientNet-b3) and ARA-Net (MobileNet-v3) were both favorable, with ARA-Net (EfficientNet-b3) displaying a more consistent result. Sample qualitative results from various methods are shown in Figure 6. ARA-Net (EfficientNet-b3) and ARA-Net (MobileNet-v3) effectively segment both large and tiny atrophic regions of the retina, as shown in Figure 6. On the other hand, the other networks either produced excessive segmentation of smaller areas or failed to detect more expansive areas of atrophy effectively. Compared to other techniques, ARA-Net (EfficientNet-b3) generates high-quality segmentation predictions, making it a viable option for atrophic area segmentation.

4. Discussion

The automated segmentation of the retinal atrophy in fundus images is a valuable tool for ophthalmologists in detecting and diagnosing myopia, as the size of the retinal atrophy is positively correlated with the severity of the condition. Automated

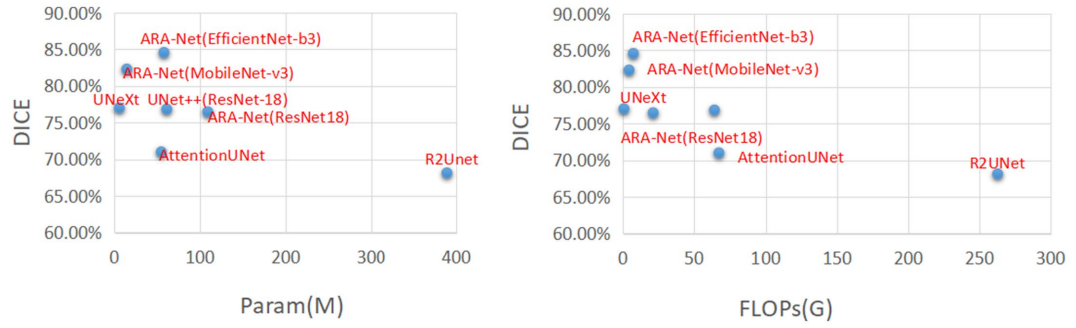


FIGURE 4
Comparison charts. *X*-axis corresponds to FLOPs(G) and the number of parameters (lower the better). *Y*-axis corresponds to DICE (higher the better).

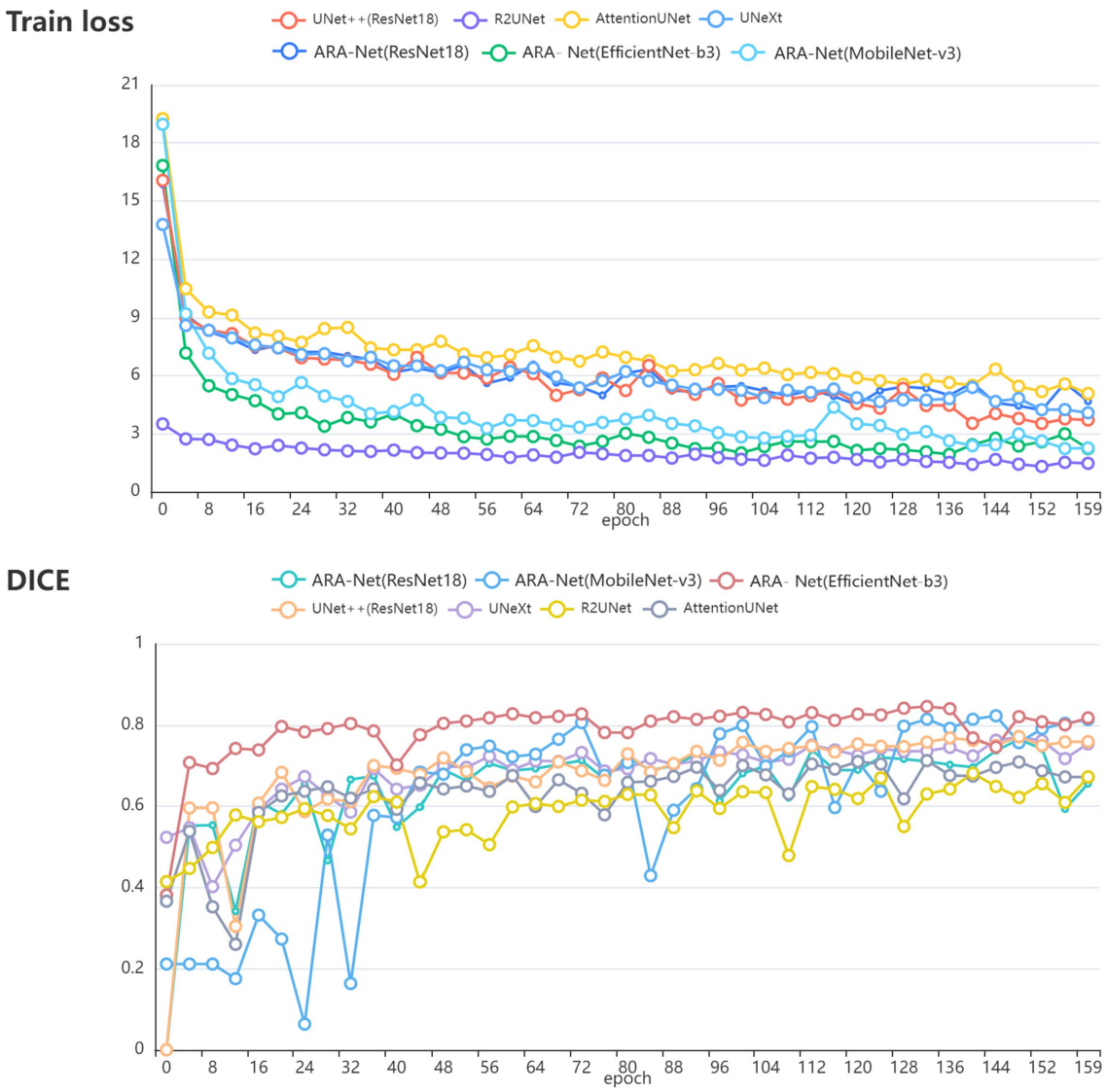


FIGURE 5
(A) The value of the loss function during training. **(B)** The DICE value during validation.

segmentation techniques allow for efficient extraction and representation of image features, reducing the need for human intervention.

The advance of deep learning has the potential to further enhance segmentation accuracy through the development of advanced segmentation network models and improved hardware performance.

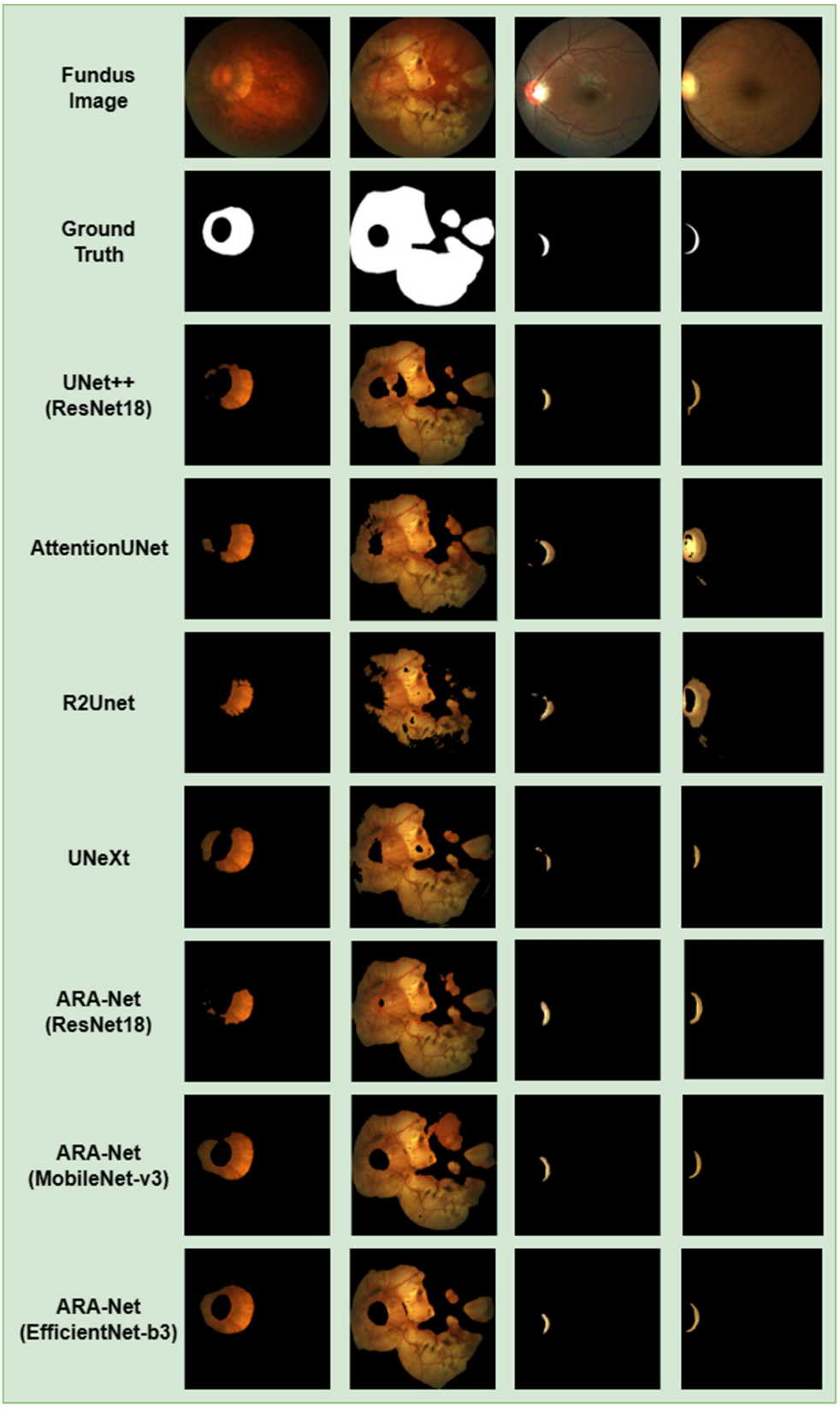


FIGURE 6
Visual segmentation results for myopic parapapillary atrophy. It shows the color fundus input images, the ground truth masks, and the segmentation results for UNet++ (ResNet18), AttentionUNet, R2UNet, UNeXt, ARA-Net (ResNet18), ARA-Net (MobileNet-v3) and ARA- Net (EfficientNet-b3).

However, the field of automated retinal atrophy segmentation is currently under-researched, with the majority of studies utilizing optical coherence tomography (OCT) images (Fang et al., 2017; He et al., 2021; Szeskin et al., 2021) due to their high image resolution and cross-sectional tissue structure. The limited availability of appropriate datasets, the significant variations in retinal atrophy shape and size among patients, and the interference of blood vessels present significant challenges for applying deep learning approaches to retinal atrophy segmentation. Additionally, the proximity of the retinal atrophy to the optic disc, which can be a source of brightness that can mask the retinal atrophy, further complicates the task. These challenges must be addressed to facilitate progress in retinal atrophy segmentation using deep learning.

The experimental results indicate that other Unet-based models struggle to accurately segment retinal atrophic areas of varying sizes, and the segmentation boundaries are not distinct. To address this challenge, novel SSA connection blocks and multi-scale feature flows were incorporated into the model to extract boundary information and shape features of the atrophic regions. The SSA block includes the initial skip connection as well as the PPSA branch. The polarized filtration and enhancement are two crucial modules. By reducing the dimensions of the inputting data while maintaining the accuracy of the channel and spatial attention, our model can focus on the most critical features and retaining critical details, leading to better feature extraction and more accurate predictions. In addition, the enhancement strategy maps the model's results to a desired distribution, resulting in more accurate predictions.

The proposed model demonstrated superior segmentation results on atrophic regions of different sizes, as shown in Figure 6. Additionally, during the training phase, the ARA-Net model utilized three pre-training models (ResNet, MobileNet-v3, and EfficientNet-b3) to extract richer feature information, resulting in improved performance. This approach leverages pre-trained models on large datasets to initialize the model weights, allowing the model to adapt to new datasets with limited training data quickly. Furthermore, using a combination of BCE loss and TVERSKY loss during training enabled the model to focus more on small ROIs, reducing the likelihood of predicting small-size retinal atrophy areas as background.

Due to the scarcity of data, the proposed method can achieve better results to some extent depending on the pre-trained model of ImageNet. In the training phase, we use pre-trained model of ImageNet for training the ARA-Net. As we known, the pre-trained model of ImageNet is trained on the images collected from natural environment, and the model is not specific to the medical domain. Thus, we might use some pre-trained model based on medical images to substitute the pre-trained model of ImageNet, and investigate the effect of different image domain based pre-trained model on the segmentation performance for retinal atrophy areas. Meanwhile, future study might consider utilizing other data improvement methods, such as the use of generative models for data augmentation or the integration of multimodal imaging. Additionally, further improvements can consider few-shot learning methods and design segmentation networks for different myopic stages of retinal atrophy. As retinal atrophy can vary in size and shape across different stages, it is crucial to develop segmentation techniques tailored to each stage. These enhancements hold the potential to improve the accuracy of retinal atrophy segmentation, leading to more precise diagnosis and treatment.

5. Conclusion

In this work, we proposed an ARA-Net model for segmenting the retinal atrophic area from 2D fundus images. In particular, our proposed novel skip-connect blocks named PPSA effectively fuse the feature maps between the encoder and decoder, enabling the network to learn representational feature maps from both channel and spatial dimensions. The MSFF utilized by our model also help to enhance the semantic information of the images, and the combined Tversky and BCE loss functions further improve the model performance. Comprehensive experimental results demonstrate that our proposed ARA-Net model achieves awesome performance on retinal atrophy segmentation, especially in challenging scenarios with blurred boundaries and irregular shapes. Our work has made significant contributions in solving the retinal atrophy segmentation challenges and introducing the PPSA blocks as a new technique for feature fusion in medical image segmentation. Comprehensive experimental results demonstrate the effectiveness of ARA-Net and its potential clinical value in retinal atrophic area segmentation applications.

Data availability statement

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

Author contributions

LC and ML contributed to the conception of the study. LC performed the experiments and wrote the first draft of the manuscript. YZ, SG, and ML performed the data analysis and revised the manuscript. ZW and HT revised the manuscript and supervised the entire study. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

References

Abraham, N., and Khan, N. M. (2018). "A novel focal tversky loss function with improved attention u-net for lesion segmentation," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy*. 683–687. doi: 10.48550/arXiv.1810.07842

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv Comput. Vision Pattern Recogn.* doi: 10.48550/arXiv.1802.06955

Chai, Y., Liu, H., and Xu, J. (2020). A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images. *Appl. Soft Comput.* 86:105890. doi: 10.1016/j.asoc.2019.105890

Chan, T. F., and Vese, L. A. (2001). Active contours without edges. *IEEE Trans. Image Process.* 10, 266–277. doi: 10.1109/83.902291

Dai, Y., Jonas, J. B., Huang, H., Wang, M., and Sun, X. (2013). Microstructure of peripapillary atrophy: beta zone and gamma zone. *Invest. Ophthalmol. Vis. Sci.* 54, 2013–1545. doi: 10.1167/iovs.12-11255

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409.97.2:S594-S594

Fang, L., Cunefare, D., Wang, C., Guymer, R. H., Li, S., and Farsiu, S. (2017). Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative AMD patients using deep learning and graph search. *Biomed. Opt. Express*. 8, 2732–2744. doi: 10.1364/boe.8

Fu, H., Li, F., Orlando, J. I., Bogunović, H., Sun, X., Liao, J., et al. (2019). [Dataset]. iChallenge-PALM: Pathologic Myopia Challenge. IEEE Dataport.

Guo, Y., Wang, R., Zhou, X., Liu, Y., Wang, L., Lv, C., et al. (2020). Lesion-aware segmentation network for atrophy and detachment of pathological myopia on fundus images. *Int. Sympo. Biomed. Imag.*, 1242–1245. doi: 10.1109/isbi45749.2020.9098669

He, Y., Carass, A., Liu, Y., Jedynak, B., Solomon, S. D., Saidha, S., et al. (2021). Structured layer surface segmentation for retina oct using fully convolutional regression networks. *Med. Image Anal.* 68:101856. doi: 10.1016/j.media.2020.101856

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv Comput. Vision Pattern Recogn.*, 770–778. doi: 10.48550/arXiv.1512.03385

Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K., Sankaridurg, P., et al. (2016). Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* 123, 1036–1042. doi: 10.1016/j.ophtha.2016.01.006

Howard, A., Pang, R., Adam, H., Le, Q. V., Sandler, M., Chen, B., et al. (2019). Searching for mobilenetv3. *Int. Conf. Comput. Vision* 2019, 1314–1324. doi: 10.1109/iccv.2019.00140

Jaccard, P. (1912). The distribution of the flora in the alpine zone.1. *New Phytol.* 11, 37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x

Jonas, J. B., Gusek, G. C., and Naumann, G. O. H. (1988). Optic disk morphometry in high myopia. *Graefes Arch. Clin. Exp. Ophthalmol.* 226, 587–590. doi: 10.1007/bf02169209

Li, H., Kang, J., Feng, Y., Xu, J., and Li, H. (2018). Automatic segmentation of ppa in retinal images. *Conf. Indust. Electron. Appl.*, 1791–1796. doi: 10.1109/iciea.2018.8397999

Liu, H., Liu, F., Fan, X., and Huang, D. (2021). Polarized self-attention: Towards high-quality pixel-wise regression[J]. *arXiv preprint*. doi: 10.48550/arXiv.2107.00782

Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi: 10.1109/tpami.2016.2572683

Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv*. doi: 10.48550/arXiv.1711.05101

Lu, C., Tang, T., Alan, F. M., Laude, A., and Dhillon, B. (2010). Automatic parapapillary atrophy shape detection and quantification in colour fundus images. *Biomed. Circuits Syst. Conf.*, 86–89. doi: 10.1109/biocas.2010.5709577

Manjunath, V., Shah, H., Fujimoto, J. G., and Duker, J. S. (2010). Analysis of peripapillary atrophy using spectral domain optical coherence tomography. *Ophthalmology* 118, 531–536. doi: 10.1016/j.ophtha.2010.07.013

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. C. H., Heinrich, M. P., Misawa, K., et al. (2018). Attention u-net: learning where to look for the pancreas. *arXiv Comput. Vision Pattern Recogn.* doi: 10.48550/arXiv.1804.03999

Park, K. H., Tomita, G., Liou, S. Y., and Kitazawa, Y. (1996). Correlation between peripapillary atrophy and optic nerve damage in normal-tension glaucoma. *Ophthalmology* 103, 1899–1906. doi: 10.1016/s0161-6420(96)30409-0

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assis. Intervent.* 9351, 234–241. doi: 10.1007/978-3-319-24574-428

Ruben, H., Bart, E., Matthew, B. B., Julie, J., Ingeborg, S., and Patrick, D. B. (2020). Pathological myopia classification with simultaneous lesion segmentation using deep learning. *Comput. Methods Prog. Biomed.* 199:105920. doi: 10.1016/j.cmpb.2020.105920

Srinivas, R. S., Robyn, G., Frank, G. H., Steffen, S., Christine, A. C., et al. (2018). Consensus definition for atrophy associated with age-related macular degeneration on OCT: classification of atrophy report 3. *Ophthalmology* 125, 537–548. doi: 10.1016/j.ophtha.2017.09.028

Szeskin, A., Yehuda, R., Shmueli, O., Levy, J., and Joskowicz, L. (2021). A column-based deep learning method for the detection and quantification of atrophy associated with amd in oct scans. *Med. Image Anal.* 72:102130. doi: 10.1016/j.media.2021.102130

Tan, M., and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *Int. Conf. Machine Learn.* 11946, 6105–6114. doi: 10.48550/arXiv.1905.11946

Uchida, H., Ugurlu, S., and Caprioli, J. (1998). Increasing peripapillary atrophy is associated with progressive glaucoma. *Ophthalmology* 105, 1541–1545. doi: 10.1016/s0161-6420(98)98044-7

Valanarasu, J. M. J., and Patel, V. M. (2022). Unext: Mlp-based rapid medical image segmentation network. *Med. Image Comput. Comput. Assis. Intervent.* 13435, 23–33. doi: 10.1007/978-3-031-16443-9_3

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Neural Inform. Proces. Syst.* 30, 5998–6008.

Wan, C., Wu, J., Li, H., Yan, Z., Wang, C., Jiang, Q., et al. (2021). Optimized-unet: novel algorithm for parapapillary atrophy segmentation. *Front. Neurosci.* 15:758887. doi: 10.3389/fnins.2021.758887

Zhou, Z., Siddiquee, M. R., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Deep Learn. Med. Image Anal. Multimodal. Learn. Clin. Decis. Support* 11045, 3–11. doi: 10.1007/978-3-030-00889-5_1

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 16 May 2023
DOI 10.3389/fnins.2023.1168962

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Jianqiang Yang,
Emory University, United States
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China

*CORRESPONDENCE
Guiqin Liu
✉ 1665361754@qq.com
Yunping Fan
✉ zhfanyp@163.com

†These authors have contributed equally to this work

RECEIVED 18 February 2023
ACCEPTED 12 April 2023
PUBLISHED 16 May 2023

CITATION
Tu X, Xiong C, Qi H, Ou YM, Rao J, Sun YQ, Fan YP and Liu GQ (2023) Diagnosis and treatment of transnasal endoscopic optic canal decompression for traumatic optic neuropathy. *Front. Neurosci.* 17:1168962. doi: 10.3389/fnins.2023.1168962

Diagnosis and treatment of transnasal endoscopic optic canal decompression for traumatic optic neuropathy

Xiang Tu[1†], Cheng Xiong[2†], Hui Qi[1†], Yangming Ou[2], Jing Rao[2], Yueqi Sun[1], Yunping Fan[1*] and Guiqin Liu[2*]

[1]Department of Otolaryngology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, [2]The Department of Orbital Disease and Ophthalmic Oncology, Shenzhen Eye Hospital, Shenzhen Eye Institute, Jinan University, Shenzhen, China

Objective: To investigate the clinical efficacy and prognostic factors of transnasal endoscopic optic decompression in the treatment of traumatic optic neuropathy (TON).

Methods: A retrospective analysis was performed on 13 TON patients in The Seventh Affiliated Hospital of Sun Yat-sen University and Shenzhen Eye Hospital in Shenzhen City (China) from June 2020 to April 2022. These patients had received transnasal endoscopic optic decompression, and hormonal and neurotrophic drugs were given after surgery. Visual acuity (VA) improvement was used as the criterion to judge clinical efficacy.

Results: In a total of 13 patients, 13 injured eyes (12 men and 1 woman; mean age, 28.0±11.8years) received transnasal endoscopic optic decompression. After surgery, nine patients had improved VA, whereas four patients failed to show any improvement, resulting in a total effective rate of 69.2%. Of the six patients with no light perception preoperatively, three had effective results after the operation, giving an effective rate of 50.0%. Of the seven patients with residual light sensation preoperatively, six had effective results after the operation, giving an effective rate of 85.7%. Of the 10 patients operated on within 7days after injury, seven had effective results, giving an effective rate of 70%. Of the three patients injured and operated on after 7days, two had effective results, giving an effective rate of 66.7%.

Conclusion: Transnasal endoscopic optic nerve decompression is an effective treatment method for TON. The presence of residual light perception and the timing of surgery within 7days are crucial to the prognosis.

KEYWORDS

nasal endoscopy, optic nerve decompression, traumatic optic neuropathy, trauma, glucocorticoid

Introduction

Traumatic optic neuropathy (TON) is a type of injury to the optic nerve resulting from external force impacting the optic nerve canal in the skull or face. This condition is characterized by a sudden or gradual loss of visual acuity (VA) and/or visual field defects following trauma (Yu-Wai-Man, 2015). TON is caused mainly by an external force acting on the optic nerve that

causes optic nerve damage. The basic injury mechanism of this disease can be divided into direct and indirect injury. Direct injury is caused mostly by open craniomaxillofacial trauma, which causes anatomical fracture and avulsion of the optic nerve directly (Yu-Wai-Man and Griffiths, 2013). Direct injury can lead to severe irreversible vision loss or even complete loss, and the treatment response and prognosis are poor. TON caused by indirect injury, which is transmitted to the optic nerve by the energy generated though open or closed craniomaxillofacial trauma, causes the fracture of the osseous optic canal or the compression of the surrounding structures, such as the reactive edema of the optic nerve sheath, indirectly. The vascular supply and neurotrophic supply of retinal ganglion cells (RGCs) might be compromised to trigger apoptosis. The primary and secondary injuries caused by RGC apoptosis also lead to axon transport disorders, inflammatory reactions, and electrochemical disorders, which eventually lead to optic nerve injury (Giacci et al., 2014). However, owing to the complex pathophysiological mechanism of TON, various related molecular pathogenesises still need to be further explored. Treatment options for TON are complex, with some studies suggesting that glucocorticoid pulse or transnasal endoscopic optic nerve decompression might be crucial in preserving surviving RGCs (Yu-Wai-Man and Griffiths, 2013; Sefi-Yurdakul and Koç, 2018; Yu et al., 2020), Despite this, some studies advocate for a conservative approach of observation, because spontaneous visual improvement can occur in some TON patients (Sosin et al., 2016). There is still much debate among medical professionals about the most effective treatment for TON owing to a lack of large-scale clinical evidence. In recent years, advancements in clinical research and technology have led to improved diagnosis and treatment strategies for TON, with transnasal endoscopic decompression surgery gaining widespread recognition (Abhinav et al., 2015). However, various factors can impact the outcome of TON treatment, such as the timing of surgery, preoperative vision, and optic nerve tube fractures (Ma et al., 2018). This study analyzed the clinical data of 13 TON patients retrospectively, and evaluated the factors influencing surgical outcomes to determine surgical efficacy on TON.

Patients and methods

We collected clinical data from patients with TON who were admitted to the Department of Otorhinolaryngology of the Seventh Affiliated Hospital of Sun Yat-sen University and the Department of Orbitopathy and Ophthalmic Oncology of Shenzhen Eye Hospital (China) between June 2020 and April 2022. The Ethics Committees of the two hospitals did not require the application of ethical approval for this type of study, and all aspects of the study were conducted in accordance with the Declaration of Helsinki. In the 13 TON patients, four of whom had a surgical interval of >4 days and also accepted hormone treatment before the operation, whereas the other nine patients accepted no treatment before surgery. To assess the effect of optic nerve decompression, we first conducted ophthalmological examinations to assess the visual acuity of patients, which helps to rule out preexisting optic neuropathy and retinopathy. The best corrected VA (BCVA) was tested initially using a Snellen chart, and a VA of <0.01 was documented with counting fingers (*CF*), hand motion (HM), light perception (LP), and no light perception (NLP). Additionally, ophthalmic equipment inspection, such as optical coherence tomography (OCT) (Carl Zeiss Meditec, United States or Heidelberg Engineering, Germany), visual field (Carl Zeiss, Germany Humphrey HFA), visual evoked potential (ROLAND, Germany), anterior segment photography (CSO, Italy), and fundus photography, were used to appraise visual function. Second, image analysis, such as computed tomography (CT)/magnetic resonance imaging (MRI), was applied to observe the optic canal fracture and orbital fracture. A thin-slice CT scan of the orbit was performed, and CT angiography (CTA) was performed in the case of craniocerebral trauma. Third, complete routine preoperative examination for patients was carried out. Finally, all patients underwent ophthalmological examinations after transnasal endoscopic optic nerve decompression and had postoperative drugs, such as hormones, nutritional nerves, and vasodilators. Efficacy determination: this study divided visual acuity into five different grades: I, NLP; II, LP; III, HM; IV, *CF*; and V, logarithm of minimum angle of resolution (logMAR) with a VA of 0.02 and above. Postoperative VA that improved by one level or more than preoperative was defined as valid; no improvement in VA was defined as ineffective; a reduction of one level or more was defined as a failed operation.

Results

Demographic characteristics

There were 13 patients (12 males, 1 female, average age 28.0 ± 11.8 years) with 13 injured eyes (12 right eyes, 1 left eye). The causes of injury were six car accidents, four blunt object injuries, and three fall injuries, and none of them had simultaneous bilateral optic nerve injury. Of these, six patients had no light sensation (I) before surgery, three had LP (II), one had *CF* (IV), and three had preoperative residual VA (V) (Table 1).

Clinical characteristics

Figure 1 demonstrates the fracture line of the optic nerve tube from the horizontal and coronal positions, respectively. The fracture was located in the right sphenoid sinus and was accompanied by a fracture of the nasal bone (Figure 1). The visual field of the patient was improved on the first day after the operation and improved significantly 6 months after the operation (Figure 2). The fovea of the macula and thickness were normal (Figures 2B1–B3), as was the thickness of the retinal nerve fiber layer (Figures 2C1–C3). Prior to surgery, the amplitude of N2, P2, and N3 had decreased apparently in the left eye compared with the right eye (Figure 2D1), whereas, on the first day post surgery, the amplitude of N2, P2, and N3 had not improved obviously compared with the preoperative surgery (Figure 2D2). By 6 months after the surgery, the amplitude of N2, P2, and N3 had not improved obviously compared with the first day after the surgery (Figure 2D3). The fundus blood vessels were normal and the retina was flat, without bulging after 6 months (Figure 2).

Treatment and prognosis

Surgical method

General anesthesia and supine position were used. (a) The deviated nasal septum that impeded the surgical field was corrected

10.3389/fnins.2023.1168962

TABLE 1 Clinical features for 13 patients.

Case	Age at onset(years)	Involved eye	Time to operation (days)	Cause of injury	Fracture of skull base	Preoperative vision	Postoperative vision for 6 months
1/M	12	OD	0.83	BFT	N	NLP	0.6
2/M	40	OS	1	TA	Y	NLP	NLP
3/M	47	OD	1	FI	Y	NLP	NLP
4/M	30	OD	0.71	TA	N	NLP	LP
5/M	18	OS	10	TA	N	LP	*CF*
6/M	23	OD	30	TA	N	0.16	0.25
7/M	26	OD	4	BFT	N	0.1	0.4
8/M	13	OD	14	BFT	N	0.05	0.05
9/M	41	OD	7	FI	N	LP	*CF*
10/M	24	OS	1	TA	N	LP	*CF*
11/F	29	OD	3	BFT	N	*CF*	0.3
12/M	42	OD	1	TA	N	NLP	NLP
13/M	19	OD	0.67	FI	Y	NLP	0.4

BFT, blunt force trauma; CF, counting fingers; D, days; F, female; FI, falling injury; HM, hand motion; LP, Light perception; M, male; N, no; NLP, no light perception; OD, oculus dexter; OS, oculus sinister; TA, traffic accident; Y, yes.

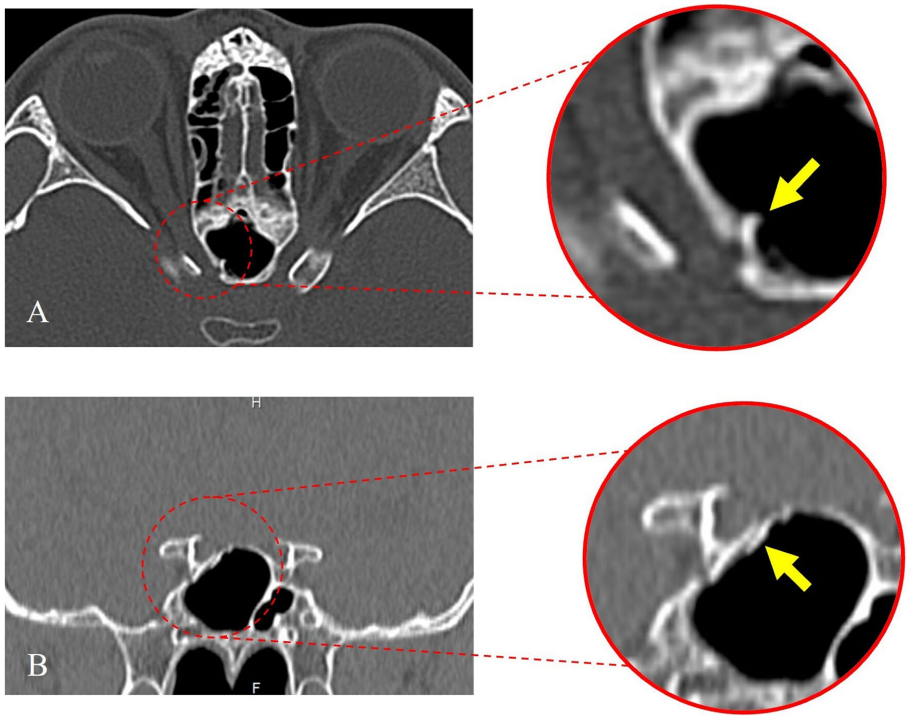


FIGURE 1
Preoperative orbital computed tomography (CT) thin-slice scan. **(A)** The axial position of the optic nerve tube shows the fracture line. **(B)** The coronal position of the optic nerve tube shows the fracture line.

initially. (b) High-definition nasal endoscopy was used to open the maxillary sinus, ethmoid sinus, and sphenoid sinus according to the Messerklinger procedure. (c) The optic tube carina (part of the optic nerve) was confirmed to be free from the sphenoid sinus lumen, and the internal carotid artery carina was exposed, revealing the medial and inferior wall of the orbit and optic nerve tube. (d) Multiple fractures of the neural tube wall and multiple bone defects in the lower orbital wall were observed. (e) Part of the medial orbital wall and the inner as well as lower and upper inner walls of the optic nerve tube were thinned carefully, and a thin bone piece was peeled off carefully with a microstripper. (f) The full length of the optic nerve tube was opened and decompressed by approximately 180°. A gelatin sponge

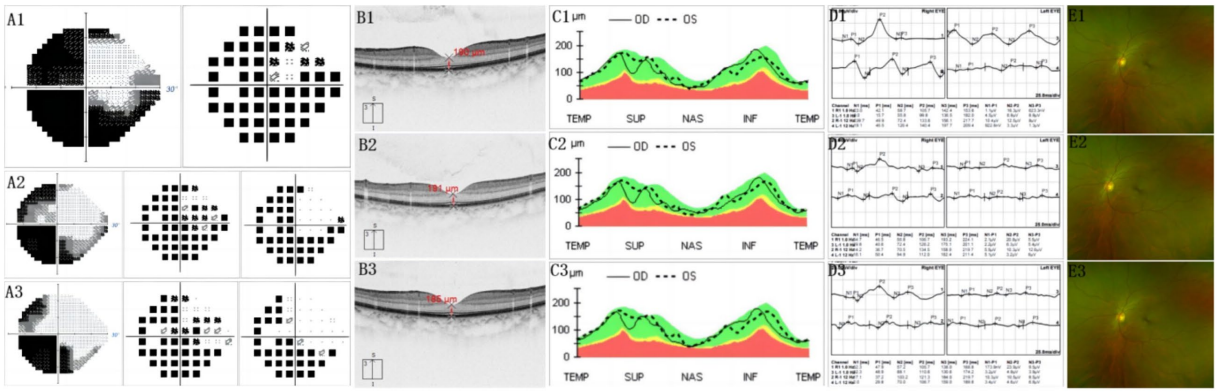


FIGURE 2
(A1) Before surgery, most of the visual field of the injured eye disappeared. Only visible is a small part of the temporal visual field. The pattern deviation chart shows that the threshold of average defect is exceeded. **(A2)** First day after operation, temporal visual field improved significantly. **(A3)** At 6months after surgery, temporal visual field expanded. Also, a small amount of nasal visual field appeared. **(B1.2.3)** Indicates that the fovea of the macula and thickness are normal. **(C1.2.3)** Shows that the thickness of the retinal nerve fiber layer was normal. **(D1)** Before surgery, the amplitude of N2, P2, and N3 descended apparently in the left eye, compared with the right eye. **(D2)** On the first day after surgery, the amplitude of N2, P2, and N3 had not obviously improved, compared with before surgery. **(D3)** At 6months after surgery, the amplitude of N2，P2, and N3 had no obviously improved, compared with the first day after surgery. **(E1.2.3)** Implies that fundus blood vessels were normal and the retina is flat, without bulge. **(B1,C1,E1)** Before surgery. **(B2,C2,E2)** First day after surgery. **(B3,C3,E3)** At 6months after surgery.

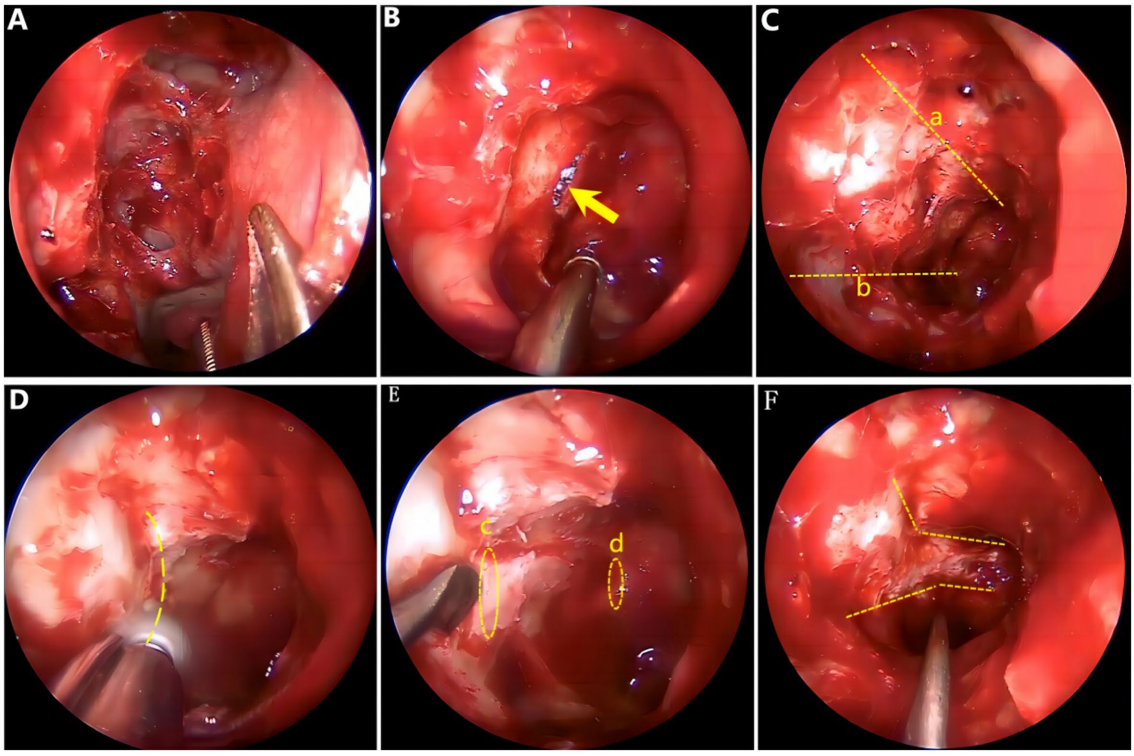


FIGURE 3
Surgical image. **(A)** Open ethmoid sinus and sphenoid sinus. **(B)** Fracture line of the optic nerve bony canal. **(C)** Orbital cranial line (a) and orbital floor line (b). **(D)** Thinning the bone of the optic nerve tube at 180°. **(E)** Starting point for bone removal: orbital ostia (c), end point for bone removal: cranial ostial (d). **(F)** Optic nerve tube decompressed adequately.

soaked with dexamethasone was covered over the optic nerve and an inflated sponge tamponade was used to compress the middle nasal passage to stop bleeding. Intraoperative endoscopic images are shown in Figure 3. After surgery, symptomatic comprehensive treatment, such as antibiotics, glucocorticoids, neurotrophic and vasodilators that can penetrate the blood and brain barrier, was administered. A total of nine patients experienced an improvement in their visual power, whereas four patients failed to show any improvement,

resulting in a postoperative effective rate of 69.2%. The time from trauma to surgery ranged from 0.7 to 30 days (average time 5.7 ± 8.43 days). A total of 10 patients were operated on within 7 days, seven of whom showed improved visual power, resulting in an effective rate of 70%. Another three patients were operated on 7 days later, two of whom showed improved visual power, resulting in an effective rate of 66.7%. Only six patients had no light preoperatively, three of whom had improved visual power after surgery, resulting in an effective rate of 50.0%. A total of seven patients had residual light sensation preoperatively, six of whom had improved visual power after surgery, resulting in an effective rate of 85.7%. The VA of three cases was >0.02, and the VA improved in two cases after surgery (Table 2).

Discussion

TON is a rare but serious complication of closed head injury (Dhaliwal et al., 2016). The risk group is mainly young men aged 30 years range (Lee et al., 2010). Road traffic accidents (especially collisions with motorcycles, electric bicycles, and rickshaws), falls from heights, and other blunt trauma are the main causes of TON (Pirouzmand, 2012).

TON includes both primary and secondary injuries (Huempfner-Hierl et al., 2015). Primary injury is caused by transient external force resulting in optic nerve rupture, optic nerve contusion, or optic nerve sheath hemorrhage, leading to immediate vision loss. Secondary injury after external force can cause optic nerve edema, local vascular compression, vascular occlusion, and other circulatory disorders, which can lead to avascular necrosis of the optic nerve, resulting in delayed vision loss, which usually occurs within hours or days after injury. The optic nerve is divided into intrabulbar, intraorbital, intracanal, and intracranial segments, of which the main rhinological segments related to rhinology are the intraorbital and intracanal segments. Approximately 90% of traumatic optic neuropathy cases occur in the intracanal segment (Steinsapir and Goldberg, 2011). Orbital thin-slice CT scan can determine the location of optic nerve tube injury effectively. A quick eye examination after an injury can assess the patient's vision. Initial fundoscopy can also help rule out pre-existing optic neuropathy and retinopathy, and might identify patients with papillary swelling and adjacent retinal hemorrhage. Visual Evoked Potential (VEP) is an objective indicator of optic nerve function and can be used as a predictor of long-term prognosis, and its reduced amplitude and increased latency are associated closely with visual impairment (Tabatabaei et al., 2011).

TABLE 2 Clinical features for operative efficiency.

Clinical features	Cases (*n*) (Operative efficiency (%))
Total cases	13 (69.2)
Gender	
Male	12 (58.3)
Female	1 (100)
Operation time	
≤ 7 days	10 (70)
>7 days	3 (66.7)
Initial vision	
No light	6 (50.0)
Residual light	7 (85.7)

In recent years, there has still been no unified standard treatment strategy for clinicians to consider for TON. Surgical intervention or glucocorticoid pulse therapy is an effective treatment option in patients with TON (Simsek et al., 2006), although the exact efficacy of optic nerve decompression remains controversial. There are some opinions suggesting that optic nerve decompression improves prognosis, emphasizing the importance of surgery (He et al., 2016). The operation improves vision by opening the full length of the optic nerve bone canal, resecting 1/2 of the circumference of the tube wall, and cutting the optic nerve sheath and the total tendon ring at the front end longitudinally (Sofferman, 1981). The operation improves vision by opening the full length of the optic nerve bone canal, resecting 1/2 of the circumference of the tube wall, and cutting the optic nerve sheath and the total tendon ring at the front end longitudinally (Dhaliwal et al., 2016). By contrast, some studies suggest that optic nerve decompression should not be recommended for patients owing to the fact that surgical decompression of the optic nerve does not result in good clinical benefits, and that surgery itself is a traumatic procedure, coupled with the risks associated with surgery (Sosin et al., 2016). Furthermore, the effect of optic nerve decompression surgery is affected largely by relevant prognostic factors, such as the patient's consciousness after injury and appropriate operation time (Ma et al., 2018).

In this study, 13 patients with TON were assessed, all of whom underwent transnasal endoscopic optic nerve decompression, and the visual improvement of patients before and after surgery was analyzed comparatively; 69.2% of patients had an improvement in VA postoperatively. This suggests that optic nerve decompression has a significant effect on the treatment of TON. To investigate the influence of related prognostic factors, our study divided patients into two groups based on the time of onset to surgery, with one group being within 7 days and the other being >7 days, and analyzed the effective rate of postoperative VA improvement between the two groups. The results showed that the effective rate of visual improvement in patients within 7 days was better than the patients who took more than 7 days to receive treatment. This indicates that the timing of TON surgery is related closely to the pathophysiology of optic nerve injury. Shortening the timing of surgery as much as possible is beneficial in interrupting the progression of pathophysiological processes and minimizing the harm caused by secondary injury to the optic nerve. Although surgery has been reported to improve impaired vision after months of visual impairment (Thakar et al., 2003), most of the center's findings suggest that early surgical intervention is still performed within 7 or even 3 days of visual impairment (Chen et al., 2018).

Studies have demonstrated that the rate of visual improvement after surgery in patients with TON who have residual vision is significantly higher than that of those with non-photosensitive TON (Ma et al., 2018). This study further revealed that the rate of postoperative visual improvement in TON patients without LP before surgery was significantly lower than that in patients with residual vision. This might be attributed to the fact that patients with totally blind TON suffer irreversible optic neuropathological

changes and have fewer retinal ganglion cells (RGCs) than those with residual vision. Nevertheless, preoperative VA and timing are sometimes not isolated and correlated, thus the effects of both should be taken into account when undertaking surgery in patients with TON.

The impact of optic nerve fractures on the final visual outcome of patients with TON remains a matter of debate. Studies have demonstrated that the presence of optic nerve fractures is an independent predictor of poor visual prognosis (Yu et al., 2018). Nevertheless, some reports suggest that the presence or absence of optic nerve fractures does not influence the treatment outcome of patients with TON significantly (Gupta et al., 2007). The use of glucocorticoids in the treatment of TON has also been a subject of debate. There is no uniform standard for the treatment of TON, and its application is based on the treatment of acute spinal cord injury (Bracken et al., 1997). The purpose of glucocorticoid therapy for TON is to reduce optic nerve microcirculation spasm, anti-inflammatory effects, reduce swelling, scavenge free radicals, and protect the optic nerve by antioxidants. Entezari et al. (2007) found no significant difference between the glucocorticoid-based and non-glucocorticoid groups, whereas Kircher et al. (2008) have suggested that glucocorticoids can promote the repair of damaged optic nerves. Taking into account its anti-inflammatory, swelling and other effects, as well as the serious consequences of TON, the leading clinical recommendation is to apply glucocorticoids as soon as possible after the onset, while avoiding adverse reactions, such as gastrointestinal ulcers, osteoporosis, and elevated blood glucose. All patients in this study were treated with glucocorticoids after surgery.

Conclusion

In conclusion, this study found that transnasal endoscopic optic nerve decompression is a safe and effective treatment for TON. Preoperative residual vision and surgery within 7 days are essential for successful outcomes, but should not be strictly adhered to, and the earlier surgery is performed, the better the prognosis, depending on the pathophysiology of the optic nerve injury. Additionally, owing to the involvement of multiple disciplines, such as ophthalmology, otolaryngology, and neurosurgery, there is a disparity in the understanding of the diagnosis and treatment of the disease. Therefore, interdisciplinary collaboration is necessary to develop the most suitable treatment plan for the patient.

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

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

XT and CX designed and conducted the study. XT, CX, HQ, YO, and JR collected, analyzed, managed, and interpreted the data. XT and HQ wrote the manuscript and conducted the statistical analysis. YS, YF, and GL made key revisions to the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Science, Technology and Innovation Commission of Shenzhen Municipality under Grant (number GJHZ20190821113605296).

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

Abhinav, K., Acosta, Y., Wang, W. H., Bonilla, L. R., Koutourousiou, M., Wang, E., et al. (2015). Endoscopic endonasal approach to the optic canal: anatomic considerations and surgical relevance. *Neurosurgery* 11, 431–445. doi: 10.1227/NEU.0000000000000900

Bracken, M. B., Shepard, M. J., Holford, T. R., Leo-Summers, L., Aldrich, E. F., Fazl, M., et al. (1997). Administration of methylprednisolone for 24 or 48 hours of tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury. *JAMA* 277, 1597–1604. doi: 10.1001/jama.1997.03540440031029

Chen, M., Jiang, Y., Zhang, J., and Li, N. (2018). Clinical treatment of traumatic optic neuropathy in children: summary of 29 cases. *Exp. Ther. Med.* 16, 3562–3566. doi: 10.3892/etm.2018.6637

Dhaliwal, S. S., Sowerby, L. J., and Rotenberg, B. W. (2016). Timing of endoscopic surgical decompression in traumatic optic neuropathy: a systematic review of the literature. *Int. Forum Allergy Rhinol.* 6, 661–667. doi: 10.1002/alr.21706

Entezari, M., Rajavi, Z., Sedighi, N., Daftarian, N., and Sanagoo, M. (2007). High-dose intravenous methylprednisolone in recent traumatic optic neuropathy; a randomized double-masked placebo-controlled clinical trial. *Graefes Arch. Clin. Exp. Ophthalmol.* 245, 1267–1271. doi: 10.1007/s00417-006-0441-0

Giacci, M. K., Wheeler, L., Lovett, S., Dishington, E., Majda, B., Bartlett, C. A., et al. (2014). Differential effects of 670 and 830 nm red near infrared irradiation therapy: a comparative study of optic nerve injury, retinal degeneration, traumatic brain and spinal cord injury. *PLoS One* 9:e104565. doi: 10.1371/journal.pone.0104565

Gupta, A. K., Gupta, A. K., Gupta, A., and Malhotra, S. K. (2007). Traumatic optic neuropathy in pediatric population: early intervention or delayed intervention? *Int. J. Pediatr. Otorhinolaryngol.* 71, 559–562. doi: 10.1016/j.ijporl.2006.11.018

He, Z.-H., Lan, Z.-B., Xiong, A., Hou, G.-K., Pan, Y.-W., Li, Q., et al. (2016). Endoscopic decompression of the optic canal for traumatic optic neuropathy. *Chin. J. Traumatol.* 19, 330–332. doi: 10.1016/j.cjtee.2016.03.004

Huempfner-Hierl, H., Bohne, A., Wollny, G., Sterker, I., and Hierl, T. (2015). Blunt forehead trauma and optic canal involvement: finite element analysis of anterior skull base and orbit on causes of vision impairment. *Br. J. Ophthalmol.* 99, 1430–1434. doi: 10.1136/bjophthalmol-2015-306646

Kircher, K., Weigert, G., Resch, H., Garhofer, G., Dorner, G. T., Mayrl, G. F., et al. (2008). Effects of high-dose prednisolone on optic nerve head blood flow in patients with acute optic neuritis. *Graefes Arch. Clin. Exp. Ophthalmol.* 246, 1423–1427. doi: 10.1007/s00417-008-0870-z

Lee, V., Ford, R. L., Xing, W., Bunce, C., and Foot, B. (2010). Surveillance of traumatic optic neuropathy in the UK. *Eye* 24, 240–250. doi: 10.1038/eye.2009.79

Ma, Y. J., Yu, B., Tu, Y. H., Mao, B. X., Yu, X. Y., and Wu, W. C. (2018). Prognostic factors of trans-ethmosphenoid optic canal decompression for indirect traumatic optic neuropathy. *Int. J. Ophthalmol.* 11, 1222–1226. doi: 10.18240/ijo.2018.07.24

Pirouzmand, F. (2012). Epidemiological trends of traumatic optic nerve injuries in the largest Canadian adult trauma center. *J. Craniofac. Surg.* 23, 516–520. doi: 10.1097/SCS.0b013e31824cd4a7

Sefi-Yurdakul, N., and Koç, F. (2018). Risk factors affecting the visual outcome in patients with indirect traumatic optic neuropathy. *Int. Ophthalmol.* 38, 1647–1652. doi: 10.1007/s10792-017-0637-9

Simsek, T., Simsek, E., Ilhan, B., Ozalp, S., Sekercioglu, B., and Zilelioglu, O. (2006). Traumatic optic nerve avulsion. *J. Pediatr. Ophthalmol. Strabismus* 43, 367–369. doi: 10.3928/01913913-20061101-08

Sofferman, R. A. (1981). Sphenoethmoid approach to the optic nerve. *Laryngoscope* 91, 184–196. doi: 10.1288/00005537-198102000-00002

Sosin, M., De La Cruz, C., Mundinger, G. S., Saadat, S. Y., Nam, A. J., Manson, P. N., et al. (2016). Treatment outcomes following traumatic optic neuropathy. *Plast. Reconstr. Surg.* 137, 231–238. doi: 10.1097/PRS.0000000000001907

Steinsapir, K. D., and Goldberg, R. A. (2011). Traumatic optic neuropathy: an evolving understanding. *Am J. Ophthalmol.* 151, 928–933.e2. doi: 10.1016/j.ajo.2011.02.007

Tabatabaei, S. A., Soleimani, M., Alizadeh, M., Movasat, M., Mansoori, M. R., Alami, Z., et al. (2011). Predictive value of visual evoked potentials, relative afferent pupillary defect, and orbital fractures in patients with traumatic optic neuropathy. *Clin. Ophthalmol. (Auckland, NZ)* 5, 1021–1026. doi: 10.2147/OPTH.S21409

Thakar, A., Mahapatra, A. K., and Tandon, D. A. (2003). Delayed optic nerve decompression for indirect optic nerve injury. *Laryngoscope* 113, 112–119. doi: 10.1097/00005537-200301000-00021

Yu, B., Chen, Y., Ma, Y., Tu, Y., and Wu, W. (2018). Outcome of endoscopic trans-ethmosphenoid optic canal decompression for indirect traumatic optic neuropathy in children. *BMC Ophthalmol.* 18:152. doi: 10.1186/s12886-018-0792-4

Yu, B., Ma, Y. J., Tu, Y. H., and Wu, W. C. (2020). Newly onset indirect traumatic optic neuropathy-surgical treatment first versus steroid treatment first. *Int. J. Ophthalmol.* 13, 124–128. doi: 10.18240/ijo.2020.01.18

Yu-Wai-Man, P. (2015). Traumatic optic neuropathy-clinical features and management issues. *Taiwan J. Ophthalmol.* 5, 3–8. doi: 10.1016/j.tjo.2015.01.003

Yu-Wai-Man, P., and Griffiths, P. G. (2013). Steroids for traumatic optic neuropathy. *Cochrane Database Syst. Rev.* 2013:Cd006032. doi: 10.1002/14651858.CD006032.pub4

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fnins.2023.1143422

OPEN ACCESS

EDITED BY
Mohammed Mahfouz Elmogy,
Mansoura University, Egypt

REVIEWED BY
Yuhui Ma,
Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, China
Yanwu Xu,
Baidu, China

*CORRESPONDENCE
Yong Tao
✉ drtaoyong@163.com
Yaping Lu
✉ luyaping@sinopharm.com
Fanxin Zeng
✉ fanxinly@163.com

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

RECEIVED 13 January 2023
ACCEPTED 19 April 2023
PUBLISHED 02 June 2023

CITATION
Huang X, Ai Z, Wang H, She C, Feng J, Wei Q, Hao B, Tao Y, Lu Y and Zeng F (2023) GABNet: global attention block for retinal OCT disease classification. *Front. Neurosci.* 17:1143422. doi: 10.3389/fnins.2023.1143422

GABNet: global attention block for retinal OCT disease classification

Xuan Huang[1,2†], Zhuang Ai[3†], Hui Wang[1†], Chongyang She[1], Jing Feng[1], Qihao Wei[3], Baohai Hao[4], Yong Tao[1*], Yaping Lu[3*] and Fanxin Zeng[5,6*]

[1]Department of Ophthalmology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China, [2]Medical Research Center, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China, [3]Sinopharm Genomics Technology Co., Ltd., Changzhou, Jiangsu, China, [4]AI-Farm (Nanjing) Big Data Services Co., Ltd., Nanjing, China, [5]Department of Clinical Research Center, Dazhou Central Hospital, Dazhou, Sichuan, China, [6]Department of Big Data and Biomedical AI, College of Future Technology, Peking University, Beijing, China

Introduction: The retina represents a critical ocular structure. Of the various ophthalmic afflictions, retinal pathologies have garnered considerable scientific interest, owing to their elevated prevalence and propensity to induce blindness. Among clinical evaluation techniques employed in ophthalmology, optical coherence tomography (OCT) is the most commonly utilized, as it permits non-invasive, rapid acquisition of high-resolution, cross-sectional images of the retina. Timely detection and intervention can significantly abate the risk of blindness and effectively mitigate the national incidence rate of visual impairments.

Methods: This study introduces a novel, efficient global attention block (GAB) for feed forward convolutional neural networks (CNNs). The GAB generates an attention map along three dimensions (height, width, and channel) for any intermediate feature map, which it then uses to compute adaptive feature weights by multiplying it with the input feature map. This GAB is a versatile module that can seamlessly integrate with any CNN, significantly improving its classification performance. Based on the GAB, we propose a lightweight classification network model, GABNet, which we develop on a UCSD general retinal OCT dataset comprising 108,312 OCT images from 4686 patients, including choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, and normal cases.

Results: Notably, our approach improves the classification accuracy by 3.7% over the EfficientNetV2B3 network model. We further employ gradient-weighted class activation mapping (Grad-CAM) to highlight regions of interest on retinal OCT images for each class, enabling doctors to easily interpret model predictions and improve their efficiency in evaluating relevant models.

Discussion: With the increasing use and application of OCT technology in the clinical diagnosis of retinal images, our approach offers an additional diagnostic tool to enhance the diagnostic efficiency of clinical OCT retinal images.

KEYWORDS

retinal OCT, retinal disease classification, attention mechanism, model visualization, GABNet

1. Introduction

The retina, the visual component responsible for sensing light stimuli, is a soft and transparent membrane located adjacent to the inner surface of the choroid, play an essential role in human vision (Zrenner, 2002). In recent years, there has been a significant increase in the number of patients with retinal disease, which can cause irreversible damage to vision and, in severe cases, result in blindness. Age-related macular degeneration (AMD) and diabetic macular edema (DME) are most common retina diseases (Das, 2016). AMD is divided into two types depending on its clinical presentation and pathological changes: dry AMD and wet AMD. In the early stages of dry AMD, yellowish-white, rounded vitreous warts (drusen) of varying sizes can be seen in the posterior pole of the eye. Wet AMD induces the outward growth of new blood vessels in the choroidal capillaries (known as choroidal neovascularization, CNV). All these pathological features could be important diagnostic indicators of disease progression.

Optical coherence tomography (OCT) is currently the most advanced technique for the detection of retinal diseases, with the advantages of being a noncontact, noninvasive and fast imaging method (Chen and Lee, 2007). The use of OCT technology for fundus images has become widespread, leading to an improvement in the clinical diagnosis of retinal diseases. However, the recognition capability of retinal OCT images varies greatly between different regions, particularly in developing countries where there is a significant shortage of expert physicians compared to the number of patients. Therefore, there is a need to develop an automated machine that can perform recognition on retinal OCT images, reducing the workload of specialist physicians.

Due to the outstanding performance of deep learning in image recognition, more and more researchers have begun to explore its applications in the medical field. However, there are still some challenges in applying deep learning to medical image classification, such as the problem of obtaining enough standardized medical images. In view of this, many researchers have proposed different solutions, among which transfer learning is one of the most commonly used methods. Transfer learning is a machine learning method that is used to correlate two different tasks. Specifically, the parameters of a model trained in one task are transferred to the same model in another task. Since case sourcing in the medical field is extremely difficult and medical image recognition based on deep learning requires a large number of cases to achieve a good classification effect. Transferred model utilizing a model trained on a public dataset can effectively alleviate the difficulty of collecting a dataset in the medical field. A large number of researchers have also used transfer learning in the medical field to obtain good classification results (Narayan Das et al., 2022). Chougrad et al. (2018) developed a computer-aided diagnosis system based on deep convolutional neural networks (DCNNs) to help radiologists classify mammographic occupancy lesions. Kassem et al. (2020) proposed a highly accurate model for the classification of skin lesions, utilizing transfer learning and GoogleNet for pre-training. The initial values for the model parameters are used as before, and they are subsequently adjusted during the training process in order to achieve the best capability to classify various types of skin lesions.

Inspired by the human vision system, which can efficiently view focal regions in complex scenes, a variety of plug-and-play attention mechanisms have been investigated in computer vision (Woo et al., 2018; Hu et al., 2020; Hou et al., 2021) and are widely used in multiple computer vision tasks (Aditya et al., 2021; Wang et al., 2022). Liang et al. (2020) proposed a semi-supervised classification approach based on a CNN model and introduced an attention mechanism to balance the sample weights. Additionally, the focal loss was used to alleviate the poor training effect caused by uneven samples. Farag et al. (2022) proposed a new automatic deep learning-based method for the detection of diabetic retinopathy (DR) severity, first using DenseNet169 as a feature extractor and then introducing a convolutional block attention module (CBAM) on top of it to enhance its discriminative power. Finally, the approach was tested on an external real-world dataset, resulting a good classification capability. Deng et al. (2020) produced a benchmark dataset of breast density images divided into four classes: A (fatty), B (fibrous gland), C (uneven dense), and D (dense). The method begins with data enhancement and normalization of breast images, followed by the implementation of a squeeze-and-excitation (SE) attention module. This module helps to recalibrate the image features and improve the classification of breast density.

Many scholars have also conducted researches on the issue of retinal OCT images. Liu et al. (2011) used the global image descriptor and local binary pattern histogram formed by a multiscale spatial pyramid as a feature vector, which could encode the texture and shape information of a retinal OCT image and its edge image, respectively. Sotoudeh-Paima et al. (2022) proposed a multiscale automated method for classifying AMD-related retinal lesions. Their multiscale CNN architecture was designed by feature fusion based on a feature pyramid network architecture, enabling end-to-end training and reducing the computational complexity of the model compared to that of using multiple CNNs in parallel. Fang et al. (2019) proposed a novel iterative fusion CNN method for retinal OCT image classification. In order to exploit the information between different convolutional layers, the proposed method introduces an iterative layer fusion strategy. Specifically, features from the current convolutional layer are iteratively combined with those from all previous layers in the CNN. Experimental results show that iteratively combining feature information from all layers can achieve better classification results.

Therefore, the challenges in constructing a lightweight network model for retinal OCT grading tasks are as follows.

(1) How to create a plug-and-play attention module.
(2) How to efficiently apply an attention module into a classification network model.

2. Related work

2.1. Previous studies on retinal OCT image analysis

In recent years, a large number of researchers have been working on retinal OCT analysis, which can be broadly divided

into two directions: machine learning methods based on semi-automatic feature extraction and deep learning methods based on fully automatic feature extraction.

Machine learning methods based on semiautomatic feature extraction can be divided into two types: feature extraction and classifier design techniques. The main feature extraction methods are local binary patterns (LBPs), histograms of oriented gradients (HOGs) and scale-invariant feature transform (SIFT) (Lee et al., 2015). The main classifiers are random forests, support vector machines, multilayer perceptron, naive bayes, etc. Srinivasan et al. (2014) designed a method to classify OCT retinal lesion images based on the HOG extracted feature to classify healthy retina, AMD, and DME, achieving high classification accuracy. While machine learning approaches have exhibited promising outcomes in classifying retinal OCT images in recent years, they are accompanied by certain limitations. First, machine learning methods based on semi-automatic feature extraction require manual operations, making them time-consuming and unable to guarantee the quality of the results. In addition, the inconsistency in retinal OCT feature extraction among experts in different regions yields incongruity, thus resulting in inaccurate diagnosis, thereby questioning the veracity of classification outcomes generated by the classifier.

A deep learning approach based on fully automated feature extraction for end-to-end retinal OCT image grading was developed (Das et al., 2021). Kayadibi and Güraksin (2023) used a stacked ensemble learning approach based on CNNs to detect DME, vitreous warts and CNV disease in OCT images. First, features were extracted from OCT images using a fine-tuned AlexNet and then applied to classify using homogeneous and heterogeneous stacked ensemble learning methods.

Semi-automatic feature extraction-based retinal OCT classification methods exhibit certain limitations, including intricate feature engineering and inadequate classification accuracy. While deep learning-based automatic feature extraction has several advantages and can achieve end-to-end prediction effects in retinal OCT disease classification, the current retinal OCT-based deep learning network models suffer from several limitations, including a large number of network parameters and slow model training.

2.2. Attention mechanisms

SE (Hu et al., 2020) is a new generic network module architecture unveiled by the autonomous driving company Momenta in 2017 (SE is shown in Figure 1). It models the correlations between feature channels and enhances important features to achieve improved accuracy. The addition of this structure also resulted in an error rate of 2.251% in the top-5 ILSVR competition. SE has been used by a large number of researchers in various industries since its introduction (XinSheng and Yu, 2022). Zhang et al. (2022) proposed a MobileNetV2-SENet-based approach to identify fish foraging behavior. Firstly, the fish images were pre-processed with some operations in order to enhance sample diversity. Then, MobileNetV2 was used to extract fish image features, and an SENet-based feature weighting network was built. Weights were assigned to features with different values. A linear classifier was used to identify the feeding behaviors of the fish. Finally, a method was provided to determine the amount of feeding based on the identification results to reduce feed consumption. Li et al. (2020) fused a DenseNet with SENet modules as the basic classification framework, and conducted extensive experiments on the proposed framework with the public BreakHis dataset, demonstrating the effectiveness of the proposed framework. Yan and Hua (2020) proposed a deep residual SENet (R-SENet) for leaf recognition. The R-SENet employs an SE strategy to learn and obtain the importance level of each channel in each convolutional layer of the residual block to accomplish the recognition task. Subsequently, the weight of each channel is readjusted by the importance level to promote relevant channels and suppress unimportant ones.

CBAM (Woo et al., 2018) is a lightweight convolutional attention module published in the ECCV conference by Woo et al. in 2018 that combines both channel attention module (CAM) and the spatial attention module (SAM) (Figure 1 shows the CBAM). CAM and SAM pay attention to channels and space, respectively. This procedure not only saves parameters and computational power but also ensures that the CAM and SAM can be integrated into existing network architectures as plug-and-play modules. Since its introduction, the CBAM has been used by a large number of researchers (Chen et al., 2020; Li et al., 2021). Luo and Wang (2021) proposed a neural network algorithm incorporating the CBAM in the ResNet architecture by adding the residual blocks of the attention module in the second to fifth layers of the ResNet architecture. The results were finally output by adaptive average pooling and fully connected layers.

Coordinate attention for efficient mobile network design (CoorDatt) is a new attention mechanism designed for lightweight networks that was presented by Hou et al. (2021) at CVPR in 2021 (Figure 1 presents CoorDatt), this mechanism embeds location information into channel attention. Unlike channel attention, which converts the given feature tensor into individual feature vectors via 2-dimensional global pooling, coordinate attention decomposes channel attention into two 1-dimensional feature encoding processes that aggregate features along two spatial directions so that remote dependencies can be captured along one spatial direction while retaining accurate location information. Since the introduction of the CoorDatt module, it has been used by a large number of researchers (Shi et al., 2022; Xiang et al., 2022). Dai et al. (2022) proposed a method for tunnel crack identification based on an improved You Only Look Once version 5 (YOLOv5) architecture. First, tunnel cracks are labeled according to a novel labeling method that uses one labeling box for regular cracks and multiple boxes for irregular cracks. Second, various data augmentations are used to improve the generalization capability of the model. Third, training YOLOv5 in combination with the CoorDatt module can achieve higher tunnel crack recognition rates. Zha et al. (2021) proposed the YOLOv4_MF model. The YOLOv4_MF model utilizes MobileNetV2 as the feature extraction block and replaces the traditional convolution operation with depthwise-separable convolution (DSC) to reduce the number of model parameters. In addition, a coordinate attention mechanism is embedded in MobileNetV2 to enhance feature information. A

symmetric structure consisting of a three-layer spatial pyramid pool is proposed, and an improved feature fusion structure is designed to fuse the target information. For the loss function, the focal loss is used instead of the cross-entropy loss to enhance the network's ability to learn small targets.

SE only considers the internal channel information in the feature map and ignores the importance of location information, while the spatial structures of targets in vision are quite important; the CBAM tries to introduce location information by perform global pooling on the channels, but this approach can only capture local information and not long range-dependent information; the CoorDatt approach encodes spatial information through maximum pooling in the horizontal and vertical directions, then transforms it, and finally fuses the spatial information by weighting it over the channels. This approach is not able to resolve the information differences between features in terms of heights, widths, and channels, without the ability to aggregate the information across dimensions. The GAB encodes the height, width and dimension of each data point on the channel of the input feature map. To reduce the number of attention mechanism parameters, we use DSC instead of normal convolution operations.

As shown in Figure 1, the maximum pooling and average pooling operations are first performed on the height and width, and then the DSC operation is performed on them to extract the spatial information of the given feature map from different angles. The obtained feature map is concerned with the detailed information in the image. In this paper, the channel operations are similar to those of the CBAM, where the importance of each channel is learned through attention mechanisms. Additionally, global average pooling is applied to obtain global feature maps, which focus the significant information of interest in the input image.

This paper presents three innovative research content to tackle the limitations of semi-automatic feature extraction, enhance deep learning-based feature extraction methods, and fully exploit the potential of attention mechanisms. The objective of these research content is to improve the accuracy and efficiency of retinal OCT disease classification.

(1) A new attention mechanism module, a global attention block (GAB), is constructed.
(2) A new lightweight classification network model based on the GAB is constructed.
(3) This network is validated on both internal and external retinal OCT datasets and a fundus photograph dataset with DR.

3. GAB based on retinal OCT disease classification

3.1. System architecture

In this study, a classification model consisting of a GAB for retinal OCT disease classification is proposed. The classification model architecture is divided into two modules: a retinal OCT image preprocessing module, and a model training and prediction module. The system architecture is shown in Figure 2. The image preprocessing module mainly includes image data normalization and scaling operations, which are used to unify retinal OCT images of different sizes and facilitate the training of the network. The model training and prediction module constructs the algorithm model and compares predictions in terms of various evaluation metrics. This module is mainly used to train and predict models from the unified retinal OCT images and thus to test the strengths and weaknesses of various algorithmic models.

3.2. Image preprocessing module

In the dataset preprocessing module, the input images are mainly deflated to widths of 299 and heights of 299 (by bilinear interpolation) and subsequently normalized (formula 1) to facilitate the calculation of the network.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

X_{min} and X_{max} represent the minimum and maximum image pixel values, respectively, and X_{norm} denotes the normalized image pixel values. The normalized image pixel values are restricted to be between 0 and 1, which accelerates the convergence of the neural network and ensures faster convergence when the program is running.

3.3. Model training and prediction module

3.3.1. GAB

Given a feature map T^{H*W*C}, H, W, and C represent the dimensional information of the feature map T (height, width and number of channels, respectively), and m, n, and k represent the size of the feature map (height, width and number of channels, respectively), so the average pooling results in terms of the height and width are obtained via formula 2 and formula 3, respectively.

$$AvgPool\_H = \frac{1}{m}\sum_{i=1}^{m} T_i(W, C) \tag{2}$$

$$AvgPool\_W = \frac{1}{n}\sum_{i=1}^{n} T_i(H, C) \tag{3}$$

The maximum pooling height and width are shown in formula 4 and formula 5, respectively.

$$MaxPool\_H = max(T_0(W, C), T_1(W, C), T_2(W, C), ..., T_m(W, C)) \tag{4}$$

$$MaxPool\_W = max(T_0(H, C), T_1(H, C), T_2(H, C), ..., T_n(H, C)) \tag{5}$$

The global average pooling process conducted on the channels and global maximum pooling are presented in formula 6 and formula 7, respectively.

$$GAP = \frac{1}{m * n}\sum_{i=1}^{m}\sum_{j=1}^{n} T_{i,j}(C) \tag{6}$$

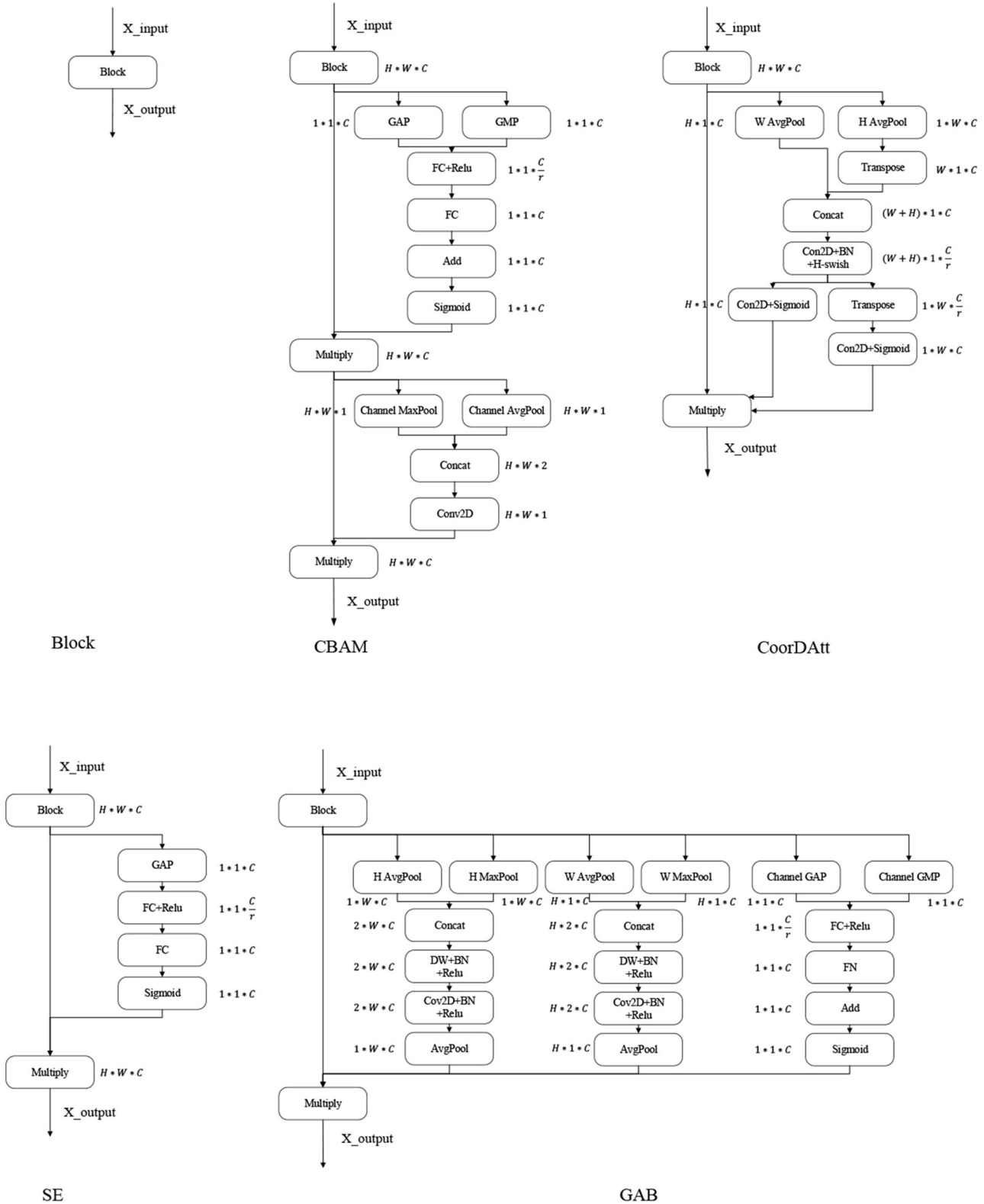


FIGURE 1
Attention mechanism. GAP, global average pooling layer; GMP, global max pooling layer; Channel AvgPool and Channel MaxPool represent the average and maximum pooling operations conducted on the channels, respectively; W AvgPool and W MaxPool represent the maximum and average pooling operations conducted on the width, respectively; H AvgPool and H MaxPool represent the maximum and average pooling operations conducted on the height, respectively; FC, fully connected layer; ReLU, sigmoid and H-swish are the activation functions; BN, batch normalization; Conv2D, convolutional layer; Add, element-by-element addition; Multiply, vector multiplication; Transpose, transposition of high-dimensional vectors; Concat, stitching according to a certain dimension.

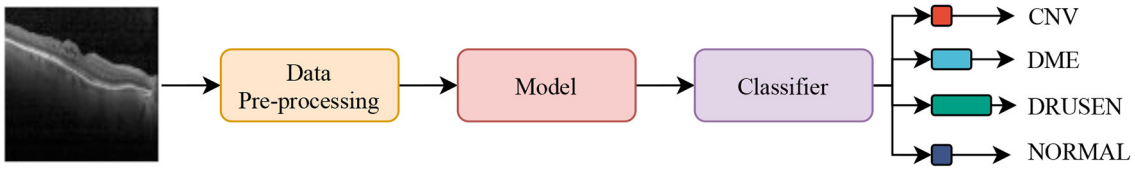


FIGURE 2
System architecture.

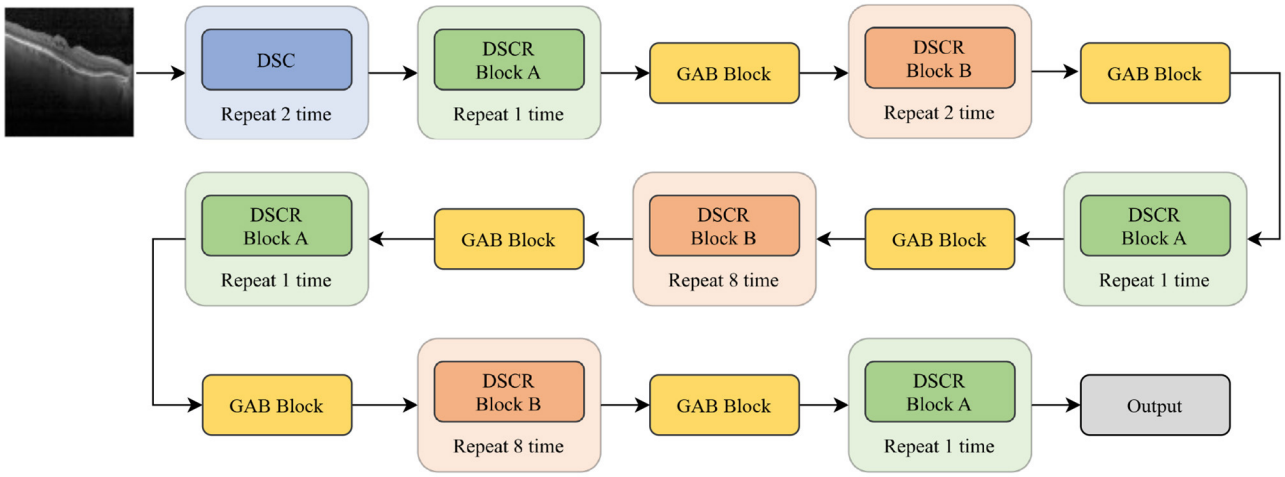


FIGURE 3
GABNet architecture.

$$GMP = max(T_{0,0}(C), T_{0,1}(C), ..., T_{0,n}(C), ..., T_{m,n}(C)) \quad (7)$$

Therefore, the weighted values in the height attention module, width attention module and channel attention module are calculated as in formula 8, formula 9, and formula 10, respectively.

$$H^{1*W*C} = AvgPool\_H(f^{1*1}(d^{1*1}(Concat(AvgPool\_H(T^{H*W*C}), MaxPool\_H(T^{H*W*C}))))) \quad (8)$$

$$W^{H*1*C} = AvgPool\_W(f^{1*1}(d^{1*1}(Concat(AvgPool\_W(T^{H*W*C}), MaxPool\_W(T^{H*W*C}))))) \quad (9)$$

$$C^{1*1*C} = W^1(W^0(GAP(R^{H*W*C}))) + W^1(W^0(GMP(R^{H*W*C}))) \quad (10)$$

where f represents the 1*1 convolution operation, batch normalization (BN) and the fusion operation of the rectified linear unit (ReLU) non-activation function; d represents the 1*1 convolution operation, BN and the fusion operation of the ReLU non-activation function; W_0 is the fusion operation of the fully connected layer and the ReLU non-activation function; W_1 is the fusion operation of the fully connected layer and the sigmoid non-activation function.

For the input feature map R^{H*W*C}, the weighted output feature map is calculated as in formula 11.

$$O^{H*W*C} = R^{H*W*C} * H^{1*W*C} * W^{H*1*C} * C^{1*1*C} \quad (11)$$

3.3.2. GABNet

In this paper, GABNet is used as a feature extraction network after data preprocessing. The overall structure of the network model is shown in Figure 3, and it can be divided into four parts: a DSC module (Figure 4A), a DSC residual (DSCR) Block A (Figure 4B), DSCR Block B (Figure 4C), and a global attention module (GAB; Figure 1 GAB). It has been shown that the DSC operation is effective in reducing the computational complexity of a network while maintaining little variation in its accuracy (Howard et al., 2019). Thus, GABNet makes extensive use of the DSC operation. The use of residual structures can effectively reduce the degradation of the network (He et al., 2016), so two modules, DSCR Block A and DSCR Block B, are constructed throughout the network. The attention mechanism is effective in improving the classification accuracy of the network when the number of parameters does not vary greatly, so the global attention mechanism is followed by an attention mechanism in each residual structure, in order to improve the effect of the network in this case.

4. Experiment

4.1. Experimental conditions

In the experimental environment of this paper, the evaluation is conducted on an Nvidia Tesla V100 with 16 GB of memory. The

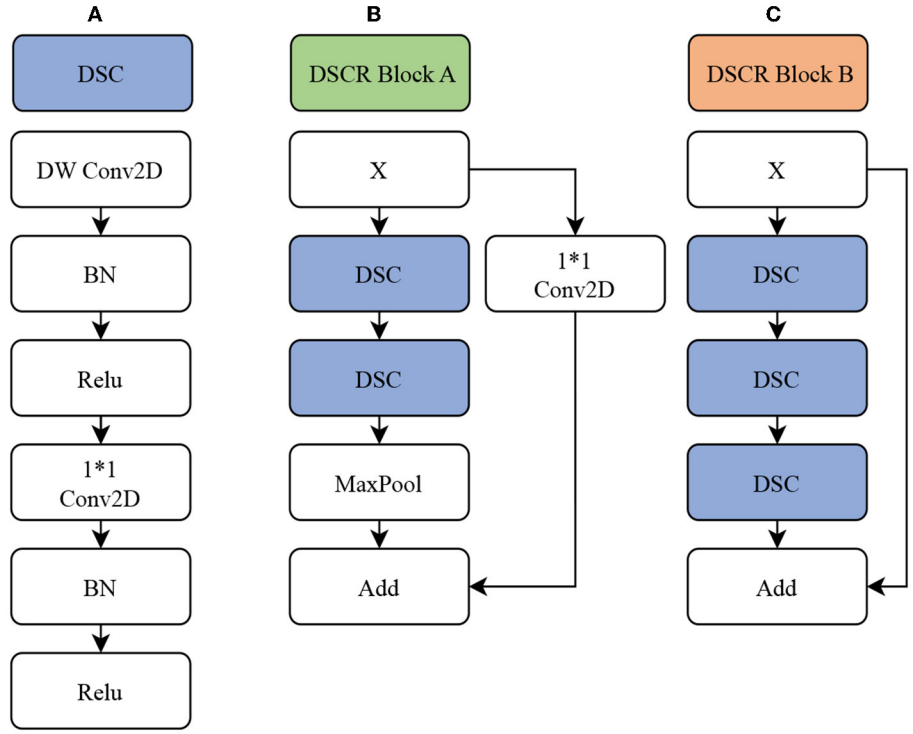


FIGURE 4
GABNet subnetwork architecture. **(A)** DSC module. **(B)** DSC residual (DSCR) Block A. **(C)** DSCR Block B. DW Conv2D, deep separable convolution layer; Conv2D, convolution layer; BN, batch normalization; ReLU, activation function; MaxPool, maximum pooling layer; Add, element-by-element addition.

CUDA, cuDNN, Python, TensorFlow, and Keras versions are 11.6, 8.4.1, 3.8.8, 2.8.0, and 2.4.3, respectively.

4.2. Dataset

The dataset used in this paper is the UCSD retinal OCT dataset, which contains 108,312 OCT images from 4686 patients (Kermany, 2018). These images were acquired using Spectralis OCT from Heidelberg Engineering, Germany. The dataset consists of four categories: CNV, DME, drusen, and normal. The sample sizes of the four categories are 37,206, 11,349, 8,617, and 51,140, respectively. Each category has 250 samples in the test set. The dataset was collected from retrospective cohorts of adult patients by various institutions, including the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and the Beijing Tongren Eye Center, between July 1, 2013 and March 1, 2017. More details about the dataset can be found in the study (Kermany, 2018). Typical images and the sample size distribution of each dataset category are shown in Figure 5.

4.3. Evaluation criteria

To perform a quantitative analysis and obtain objective comparison results, we evaluate the diagnostic performance of the proposed approach according to the Accuracy (Acc), Recall, Precision, Specificity, F1, and area under the curve (AUC) metrics. It should be noted that the aforementioned Acc, Recall, Precision, Specificity, F1, and AUC metrics are all calculated based on weighted averages for each corresponding class, in order to obtain a comprehensive performance evaluation. Table 1 presents the confusion matrix for multi-class classification of Retinal OCT images. For each class i, the true positive (TP_i), false positive (FP_i), false negative (FN_i), and true negative (TN_i) can be easily obtained using the formula 12–15. The number of samples in the class i is represented by $support\_i$. With these four values, we can calculate the Acc, F1 score, precision, specificity and recall. Acc is the proportion of correctly classified samples among all participating samples, as shown in formula 16. Recall is the proportion of correct predictions among all positive samples, as shown in formula 17. Precision is the proportion of correctly classified positive examples among all divided positive examples, as shown in formula 18. Specificity is the proportion of correctly classified negative examples among all divided negative examples, as shown in formula 19. F1 is an index used in statistics to measure the accuracy of binary classification models. It takes both model accuracy and recall into account, as shown in formula 20. The F1 score can be regarded as a harmonic average of the model accuracy and recall, whose maximum value is 1 and minimum value is 0. The AUC is the area under the receiver operating characteristic (ROC) curve enclosed by the coordinate axes, and its value range is between 0.5 and 1.

$$TP_i = X_{ii} \tag{12}$$

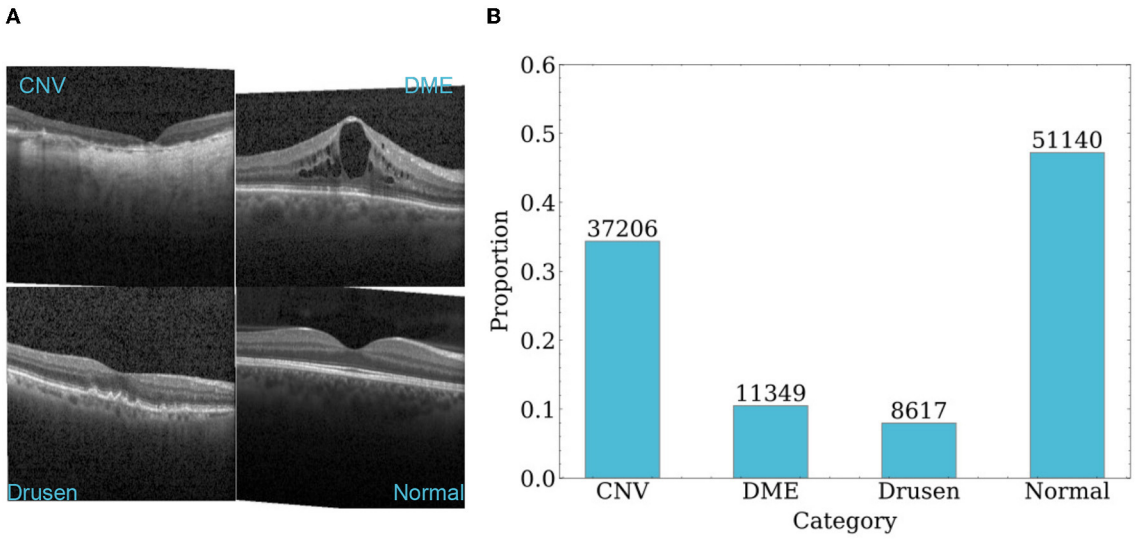


FIGURE 5
An overview of samples in dataset. **(A)** Representative OCT images of each category. **(B)** Sample size distribution of each category.

$$FP_i = \sum_{j=1}^{4} X_{ji} - TP_i \tag{13}$$

$$FN_i = \sum_{j=1}^{4} X_{ij} - TP_i \tag{14}$$

$$TN_i = \sum_{j=1}^{4}\sum_{k=1}^{4} X_{jk} - TP_i - FP_i - FN_i \tag{15}$$

$$Acc = \frac{\sum_{j=1}^{4} X_{jj}}{\sum_{j=1}^{4}\sum_{k=1}^{4} X_{jk}} \tag{16}$$

$$Recall = \frac{\sum_{i=1}^{4} \frac{TP_i * support\_i}{TP_i + FN_i}}{\sum_{i=1}^{4} support\_i} \tag{17}$$

$$Precision = \frac{\sum_{i=1}^{4} \frac{TP_i * support\_i}{TP_i + FP_i}}{\sum_{i=1}^{4} support\_i} \tag{18}$$

$$Specificity = \frac{\sum_{i=1}^{4} \frac{TN_i * support\_i}{TN_i + FP_i}}{\sum_{i=1}^{4} support\_i} \tag{19}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{20}$$

TABLE 1 Confusion matrix for retinal OCT classification.

Confusion matrix		Predicted class			
		CNV	DME	Drusen	Normal
Actual class	CNV	X_{11}	X_{12}	X_{13}	X_{14}
	DME	X_{21}	X_{22}	X_{23}	X_{24}
	Drusen	X_{31}	X_{32}	X_{33}	X_{34}
	Normal	X_{41}	X_{42}	X_{43}	X_{44}

4.4. Experimental results

To ensure objective and fair comparison among the algorithms, we used softmax as the activation function and cross-entropy as the loss function for all compared algorithms. We utilized adam optimizer with an initial learning rate of 0.0001 and a batch size of 10. The model was trained for 300 epochs. Moreover, we employed several optimization strategies, including scaling down the learning rate to 1/10 if the validation accuracy did not improve for 5 consecutive times, and early stopping if the model did not improve validation accuracy for 11 consecutive times.

Regarding the selection of data, we conducted experiments using two distinct approaches. The first approach involved training the models on the complete dataset, which will be subsequently referred to as "complete model" throughout this paper. The second approach consisted of training the models on a balanced subset of the complete dataset, containing corresponding 1,000 images per class, which will be referred to as "limited model" in the following sections.

On the general UCSD retinal OCT dataset, this paper evaluates the model performance in four perspectives: whether to use balanced data, different classification algorithms, whether to use transfer learning, and whether to use attention mechanisms. Our

TABLE 2 Comparative evaluation of various algorithms on different datasets.

Dataset	Classification algorithm	Transfer learning	Attention mechanism	Acc	Recall	Specificity	F1	AUC
Limited model	Xception	No	No	0.94	0.94	0.98	0.9398	0.9937
		Yes	No	0.976	0.976	0.992	0.9761	0.9988
		Yes	SE	0.978	0.978	0.9927	0.9779	0.9988
		Yes	CBAM	0.978	0.978	0.9927	0.978	0.9979
		Yes	CoorDatt	0.985	0.985	0.995	0.985	0.9991
		Yes	GAB	**0.99**	**0.99**	**0.9967**	**0.99**	**0.9994**
	EfficientNetV2	No	No	0.838	0.838	0.946	0.8335	0.9649
		Yes	No	0.971	0.971	0.9903	0.971	0.9985
		Yes	SE	0.971	0.971	0.9903	0.971	0.9984
		Yes	CBAM	0.972	0.972	0.9907	0.972	**0.9986**
		Yes	CoorDatt	0.974	0.974	0.9913	0.9739	0.9975
		Yes	GAB	**0.976**	**0.976**	**0.992**	**0.976**	0.9958
	GABNet	-	No	0.733	0.733	0.9113	0.7208	0.9084
		-	GAB	**0.954**	**0.954**	**0.9846**	**0.954**	**0.998**
Complete model	Xception	No	No	0.95	0.95	**0.9833**	0.9501	0.9961
		Yes	No	0.977	0.977	0.9923	0.977	**0.9992**
		Yes	SE	0.978	0.978	0.9927	0.978	0.9985
		Yes	CBAM	0.961	0.961	0.987	0.9612	0.9985
		Yes	CoorDatt	0.971	0.971	0.9903	0.9709	0.9985
		Yes	GAB	**0.98**	**0.98**	**0.9933**	**0.98**	**0.9992**
	EfficientNetV2	No	No	0.928	0.928	0.976	0.9273	0.9929
		Yes	No	0.968	0.968	0.9893	0.9681	0.9977
		Yes	SE	0.971	0.971	0.9903	0.9711	**0.9986**
		Yes	CBAM	0.973	0.973	0.991	0.973	0.9976
		Yes	CoorDatt	0.966	0.966	0.9887	0.966	0.9979
		Yes	GAB	**0.978**	**0.978**	**0.9927**	**0.9781**	0.9983
	GABNet	-	No	0.965	0.965	0.9883	0.9648	0.991
		-	GAB	**0.965**	**0.965**	**0.9883**	**0.965**	**0.9971**

Value in bold means the best of the same class.
"-" Means that the algorithm does not have this feature or property.

proposed GAB attention mechanism is tested against different attention mechanisms, such as SE, the CBAM and CoorDatt. The proposed GABNet is compared with the Xception and EfficientNetV2B3 algorithms. The results of the test comparisons are shown in Table 2. Compared with the general attention mechanisms proposed in recent years, such as SE, CBAM, and CoorDatt, the GAB attention mechanism is improved to some extent. The GABNet classification algorithm has demonstrated superior performance over Xception and EfficientNetV2B3, without the use of transfer learning, in terms of parameter efficiency and various classification metrics. This confirms the effectiveness of the GAB attention mechanism and the GABNet classification algorithm.

4.4.1. Validity of the GAB and GABNet

In this paper, Xception and EfficientNetV2B3 are regarded as the basic classification modules, in which the input height and width of the Xception classification framework are 299. To achieve the best classification effect for EfficientNetV2B3 (Tan and Le, 2021), the height and width of the input image provided to this classification model are scaled to 300. To verify the effectiveness of the GAB attention mechanism, we test whether the GAB has any influence on the Xception and EfficientNetV2B3 classification algorithms based on the same dataset and the same algorithm. Under the limited model, the effects of Xception and EfficientNetV2B3 integrated with GAB module is 1.4 and 0.5% higher than those obtained by models

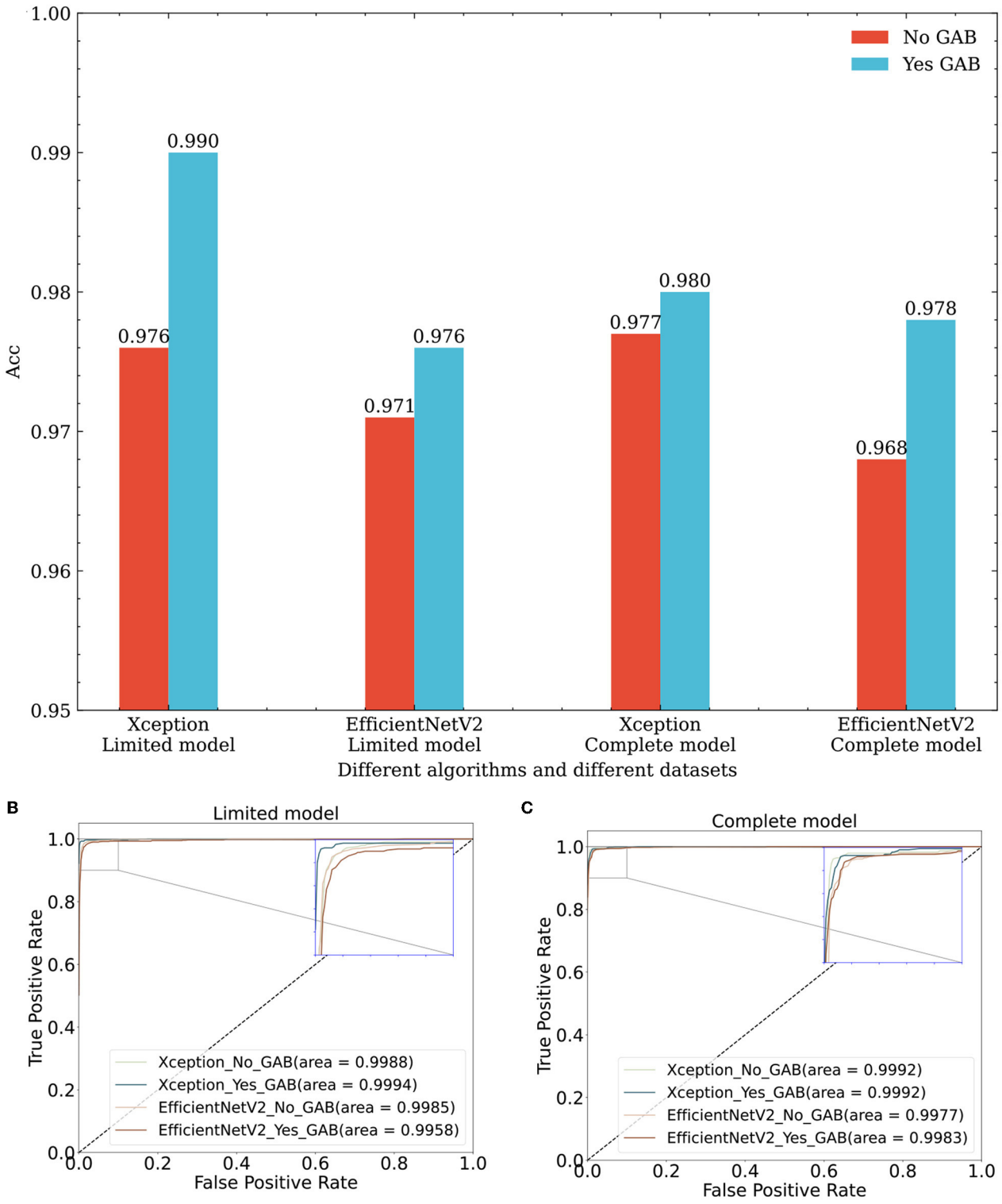


FIGURE 6
Performance comparison of classification algorithms with and without GAB on different datasets. **(A)** Effect of GAB on classification accuracy across diverse datasets and algorithms. **(B)** Effect of GAB on ROC of different algorithms in the limit model dataset. **(C)** Effect of GAB on ROC of different algorithms in the complete model dataset. Xception_No_GAB, Xception_Yes_GAB represent Xception without GAB and with GAB respectively; EfficientNetV2_No_GAB, EfficientNetV2_Yes_GAB represent EfficientNetV2 without GAB and with GAB respectively.

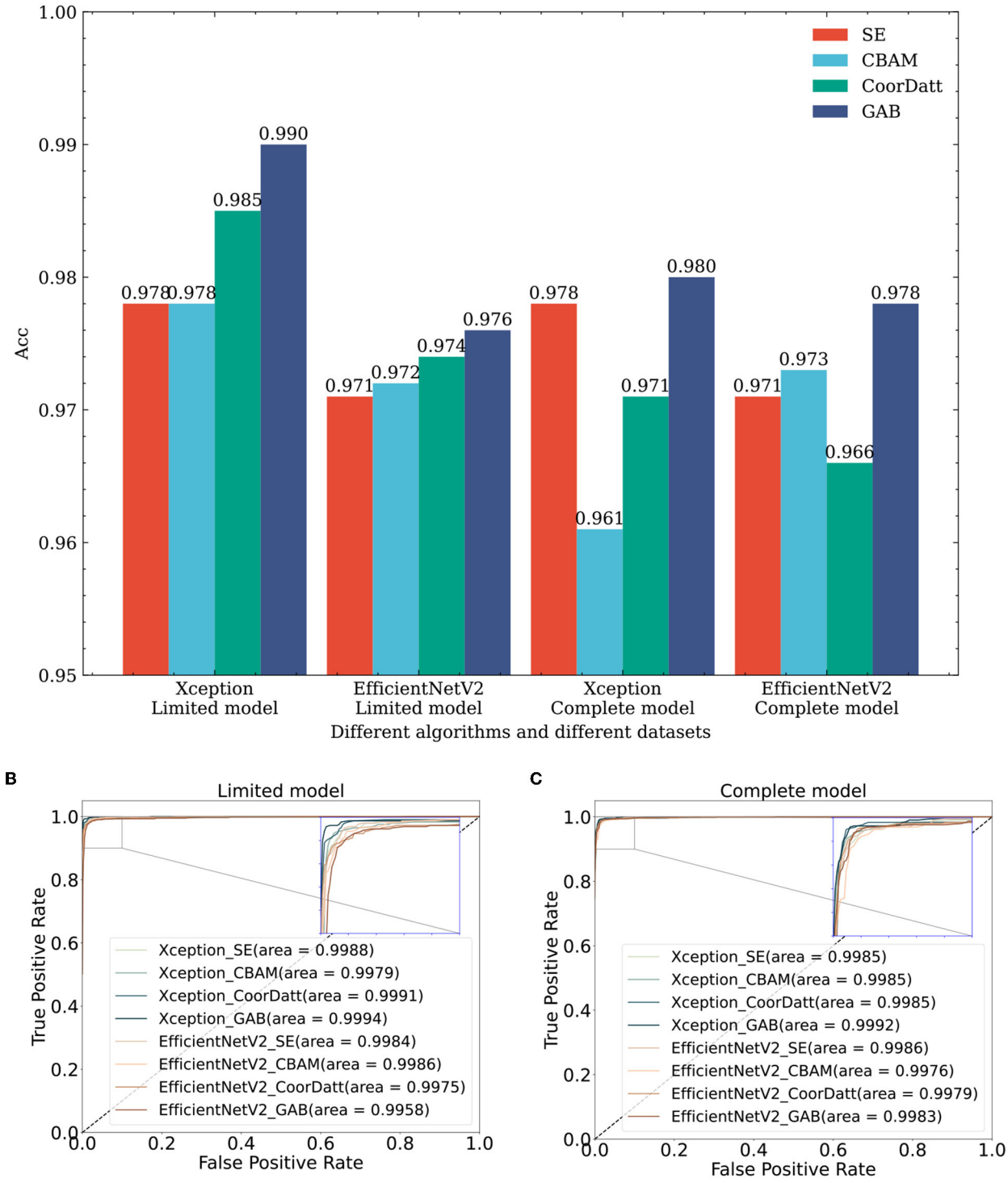


FIGURE 7
The impact of attention mechanisms on classification performance: a comparative study across multiple algorithms and datasets. **(A)** Effect of attention mechanisms on classification accuracy across diverse datasets and algorithms. **(B)** Effect of attention mechanisms on ROC of different algorithms in the limit model dataset. **(C)** Effect of attention mechanisms on ROC of different algorithms in the complete model dataset. Xception_SE, Xception_CBAM, Xception_CoorDatt and Xception_GAB represent the fusion of Xception with four different attention mechanisms, SE, CBAM, CoorDatt and GAB, respectively; EfficientNetV2_SE, EfficientNetV2_CBAM, EfficientNetV2_CoorDatt, and EfficientNetV2_GAB represent the fusion of EfficientNetV2 with four different attention mechanisms, SE, CBAM, CoorDatt and GAB, respectively.

TABLE 3 Evaluation of different algorithms on retinal OCT classification.

Algorithm	Acc	Recall	Specificity	F1	AUC	Parameters (M)	Test time
ResNet50[a] (He et al., 2016)	0.941	0.941	0.9803	0.9411	0.9923	23.595	4.81 s ± 315 ms
InceptionV3[a] (Szegedy et al., 2016)	0.96	0.96	0.9867	0.9599	0.9947	21.811	**4.78 s ± 277 ms**
Xception[a] (Chollet, 2017)	0.95	0.95	0.9833	0.9501	0.9961	20.87	4.81 s ± 285 ms
EfficientNetV2B3[a] (Tan and Le, 2021)	0.928	0.928	0.976	0.9273	0.9929	12.937	5.39 s ± 204 ms
Huang[a] (Huang et al., 2019)	0.884	0.846	N/A	N/A	N/A	N/A	N/A
GABNet[a]	**0.965**	**0.965**	**0.9883**	**0.965**	**0.9969**	**9.361**	7.26 s ± 353 ms
FN-F1-OCT[b] (Ai et al., 2022)	0.985	0.985	0.995	0.985	0.99	99.717	18.1 s ± 831 ms
FN-Weight-OCT[b] (Ai et al., 2022)	0.984	0.984	0.995	0.984	0.99	99.717	15.6 s ± 419 ms
FN-Auto-OCT[b] (Ai et al., 2022)	0.987	0.987	0.996	0.987	0.991	99.774	15.8 s ± 451 ms
Kermany[b] (Kermany et al., 2018)	0.961	0.961	0.987	0.961	0.99	N/A	N/A
Hwang[b] (Hwang et al., 2019)	0.9693	N/A	N/A	N/A	N/A	N/A	N/A
Sinha[b] (Sinha et al., 2023)	0.944	0.944	0.9815	0.9448	N/A	N/A	N/A
EfficientNetV2B3+GAB[b]	0.978	0.978	0.9927	0.9781	0.9983	**18.281**	5.53 s ± 94.5 ms
Xception+GAB[b]	**0.99**	**0.99**	**0.9967**	**0.99**	**0.9994**	30.354	**4.88 s ± 240 ms**

[a]Non-transfer learning methods.
[b]Transfer learning methods.
Value in bold means the best of the same class.
"N/A" Means that the metric was not displayed in the comparison article.

without the GAB module, respectively. Under the complete model, Xception and EfficientNetV2B3 with the GAB module achieve 0.3 and 1% improvements, respectively, over the models without the GAB module.

The GAB attention mechanism is composed of DSC operations, so it can be used as a general module and can be seamlessly connected to any network feature map. The evaluation results obtained by different algorithms on different datasets (lines 2, 6, 8, 12, 16, 20, 22, and 26 in Table 2 and Figure 6) are improved to a certain extent, thus proving the effectiveness of the GAB algorithm.

To verify the efficiency of the GAB attention mechanism, we conduct tests on the same dataset with the same algorithm, as shown in rows 3–6, 9–12, 17–20, and 23–26 in Table 2 and Figure 7, where the sample size of each dataset category is balanced 1,000 for the limited model dataset. In comparison of different attention modules, such as SE, CBAM, CoorDatt and GAB, the Acc, F1 value and AUC results show that they all provide certain improvements over models without any attention mechanisms. Notably, the GAB attention module proposed in this paper provides improvements of up to 1.4%. Under EfficientNetV2, in the same comparison, the GAB module is 0.5% better than the original algorithm. Compared to SE, the CBAM and CoorDatt, the GAB module can effectively learn the main differences between the categories when the sample size of the dataset is small, which can maximally improve the classification effect of the model. In parallel, for the complete model dataset, the dataset is the complete dataset provided by UCSD with a large sample size but significant imbalances between categories. The Xception algorithm in the complete model with four attention mechanisms is reduced compared to that in the limited model. The unbalanced training dataset of the model has an impact, resulting in the model having inconsistent recognition abilities for various categories and being biased toward categories with large sample sizes. The reason for that may be the Xception model has some limitations regarding its prediction ability for cases with small sample sizes. EfficientNetV2B3 produces little difference between the prediction results obtained on the two different datasets, indicating that it has little influence on the dataset differences and has some resistance to interference.

Similarly, to verify the effectiveness of the GABNet algorithm, comparison experiments are conducted on different classification algorithms with same datasets. For this paper, the comparison models Xception and EfficientNetV2B3 were chosen. These

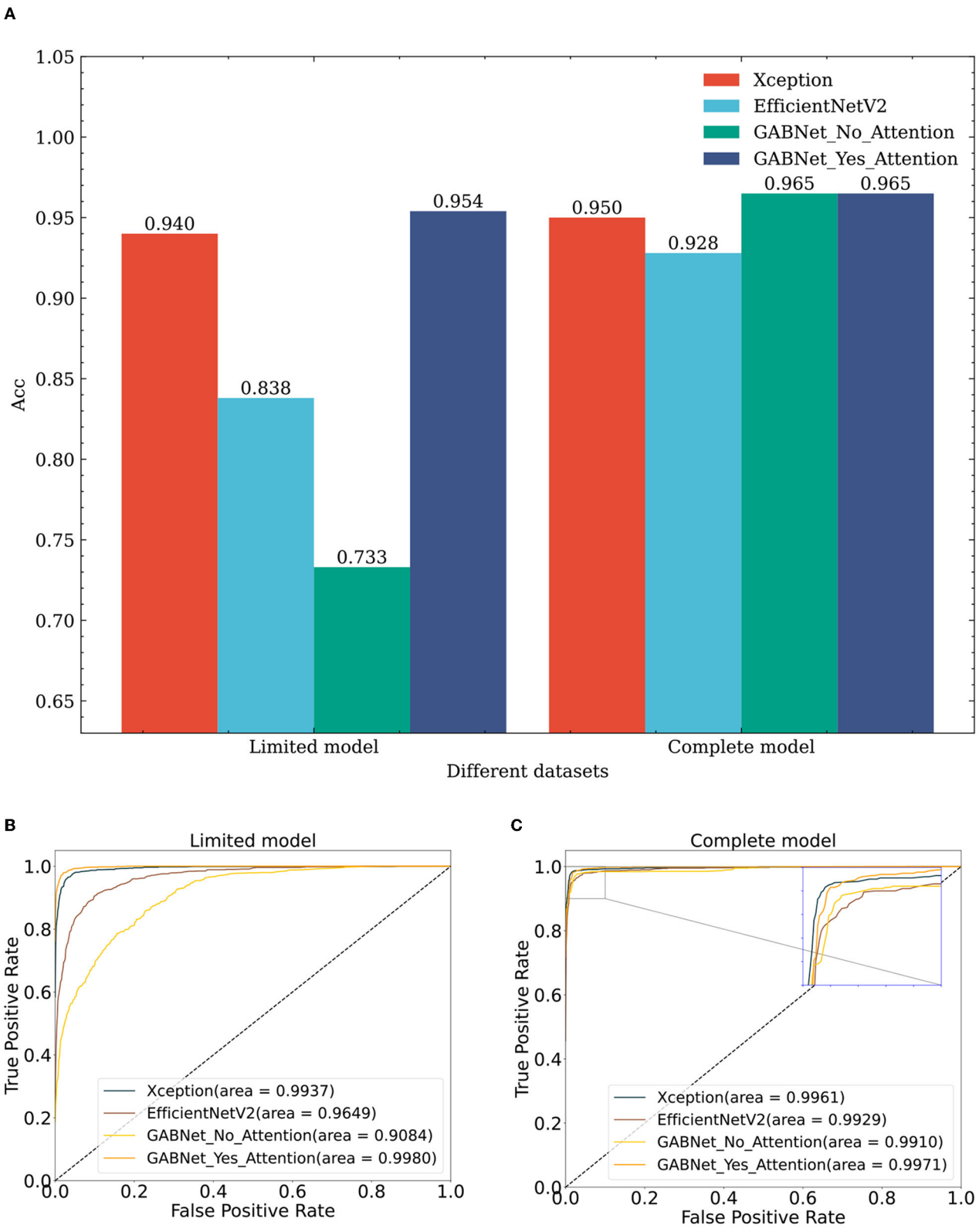


FIGURE 8
Comparing classification results of different algorithms on diverse datasets. **(A)** Effect of different algorithms on classification accuracy on different datasets. **(B)** Effect of different algorithms on ROC obtained on limit model dataset. **(C)** Effect of different algorithms on ROC obtained on complete model dataset. GABNet_No_GAB, GABNet_Yes_GAB represent GABNet without GAB and with GAB respectively.

TABLE 4 Evaluation of retinal OCT classification algorithms on external dataset.

Classification algorithm	Acc	Recall	Specificity	F1	AUC
GABNet	0.8047	0.8047	0.9332	0.7988	0.9654
EfficientNetV2+GAB	0.8594	0.8594	0.9531	0.8556	0.9727
Xception+GAB	**0.9141**	**0.9141**	**0.9723**	**0.915**	**0.9914**

Value in bold means the best of the same class.

TABLE 5 Evaluation of DR classification algorithms on fundus photograph dataset.

Classification algorithm	Transfer learning	Attention mechanism	Acc	Recall	Specificity	F1	AUC
VGG19 (Wu and Hu, 2019)	-	-	0.51	N/A	N/A	N/A	N/A
ResNet50 (Wu and Hu, 2019)	-	-	0.49	N/A	N/A	N/A	N/A
Inception V3 (Wu and Hu, 2019)	-	-	0.61	N/A	N/A	N/A	N/A
DR-IIXRN (Ai et al., 2021)	-	-	0.793	0.7933	0.8778	0.7602	0.7602
Xception	NO	NO	0.7479	0.7479	0.8307	0.6808	0.7307
	YES	NO	0.7901	0.7901	0.8767	0.7539	0.8093
	YES	YES	0.7939	0.7939	0.8711	0.7487	0.7942
EfficientNetV2B3	NO	NO	0.7358	0.7358	0.8271	0.6628	0.7127
	YES	NO	0.797	0.797	0.8776	0.7555	0.8025
	YES	YES	**0.804**	**0.804**	**0.8831**	**0.7653**	**0.8109**
GABNet	-	NO	0.6877	0.6876	0.807	0.6242	0.5987
	-	YES	0.7607	0.7607	0.8398	0.6954	0.743

Value in bold means the best of the same class.
"N/A" Means that the metric was not displayed in the comparison article.
"-" Means that the algorithm does not have this feature or property.

models used their respective network architectures without utilizing "ImageNet"-based initialization of the weight parameters. Instead, they were initialized with random values for their weight parameters. The number of parameters required by the proposed GABNet in this paper is significantly lower, as can be observed from the numbers of network parameters presented in Table 3. Specifically, the GABNet requires 0.44 and 0.72 times fewer parameters compared to Xception and EfficientNetV2B3, respectively, indicating a substantial reduction in the number of parameters. From a dataset perspective (Figure 8; lines 1, 7, 13–15, 21, and 27–28 in Table 2), if all three comparison algorithms use random initialization parameters, the results obtained with the complete model are much better than those of the limited model, where random network parameter initialization requires more image data to be fitted for optimization purposes. Under the limited model, GABNet improves by 1.4% over Xception and by 11.6% over EfficientNetV2B3. The small sample size in the training dataset has a large impact on the nonmigratory learning abilities of Xception and EfficientNetV2B3, and their training processes have difficulty in reaching saturation, so their prediction effects drop sharply. In contrast, GABNet still maintains a high prediction accuracy in this situation. With a larger sample size of the complete model dataset, all four algorithm models have been improved, especially GABNet without the attention module, which has increased its accuracy by 23.2%. The GABNet model with the attention module achieves an improved prediction effect with the completed model over that obtained with the limited model, but the improvement is not very significant, indicating that GABNet is able to learn the main differences between the categories even when the sample size is small.

To conduct an ablation experiment on the GABNet network, the enhancement effect of the attention module in GABNet is verified. For the test effect in Figure 8, GABNet improves by 22.1% on the basis of the limited model. GABNet can effectively maintain the stability of the classification index when the sample size of the given dataset is not large. In view of the difficulty in obtaining medical data at present, GABNet has application potential in certain scenarios.

The UCSD retinal OCT dataset is by far the largest open source dataset in terms of data volume for retinal classification, and thus a large number of researchers have conducted research tests on this dataset. To compare the effectiveness of the GAB and GABNet techniques proposed in this paper, we collect and download the network models that have been tested on the UCSD retinal OCT dataset over the past five years, as shown in Table 3. To ensure fairness in comparison, we conducted evaluation synthetically on Complete model and Limited model datasets using NVIDIA Geforce RTX 3060Ti with 8GB memory and a batch size of 10. We selected the model with higher accuracy for further analysis.

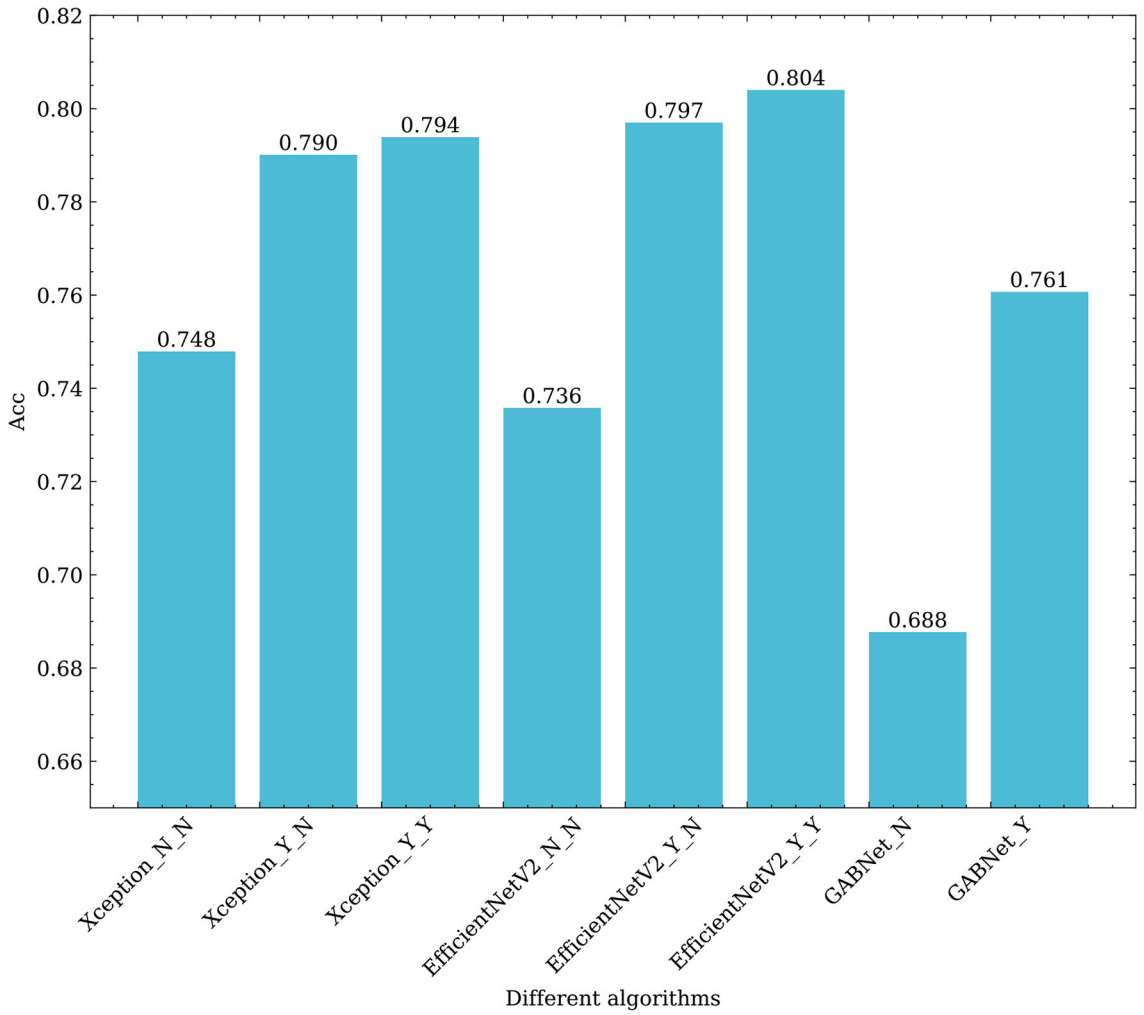


FIGURE 9
Effects of different algorithms on the classification results obtained for retinal fundus images. Xception_N_N, Xception_Y_N, Xception_Y_Y respectively represent Xception as the base classifier, no transfer learning and no GAB attention mechanism; Using transfer learning when not using GAB attention mechanism; Using transfer learning and use the GAB attention mechanism. EfficientNetV2_N_N, EfficientNetV2_Y_N, EfficientNetV2_Y_Y respectively represent EfficientNetV2 as the base classifier, no transfer learning and no GAB attention mechanism; Using transfer learning when not using GAB attention mechanism; Using transfer learning and use the GAB attention mechanism. GABNet_N,GABNet_Y represent GABNET without attention mechanism and with attention mechanism respectively.

As shown in Table 3, "Parameters (M)" represents the total number of parameters in the model, and "Test Time" represents the mean ± sd time obtained by running 5 tests per round for 10 rounds on the test dataset.

We compared non-transfer learning algorithms with transfer learning algorithms. In the comparison of non-transfer learning algorithms (Table 3, rows 1-6), the GABNet algorithm had a certain disadvantage in testing time due to its wider branch compared to other algorithms. However, with the improvement of computing performance, we believe that this gap can be further reduced. Specifically, in terms of classification performance, GABNet showed extremely strong feature extraction and classification capabilities in the comparison of non-transfer learning algorithms. In the comparison of transfer learning algorithms (Table 3, rows 7-14), the accuracy of Xception and EfficientNetV2B3 algorithms with GAB attention mechanism was greatly improved. The Ai et al. (2022) method used ensemble learning with strong feature extraction, but the Xception+GAB method proposed in this paper significantly reduced the number of required parameters while improving the evaluation metrics, indicating the improved effectiveness of the GAB attention mechanism compared to the Ai et al. (2022) method. Compared to the transfer learning algorithms proposed in the literature, the accuracy of GABNet was slightly lower, which is also the focus of future work in this paper. The algorithm will be trained on the ImageNet dataset to obtain a network with strong feature extraction capabilities, and then transfer learning will be performed on this dataset.

4.4.2. External validation dataset extension

This study retrospectively collected 256 retinal OCT images from Beijing Chaoyang Hospital, Capital Medical University.

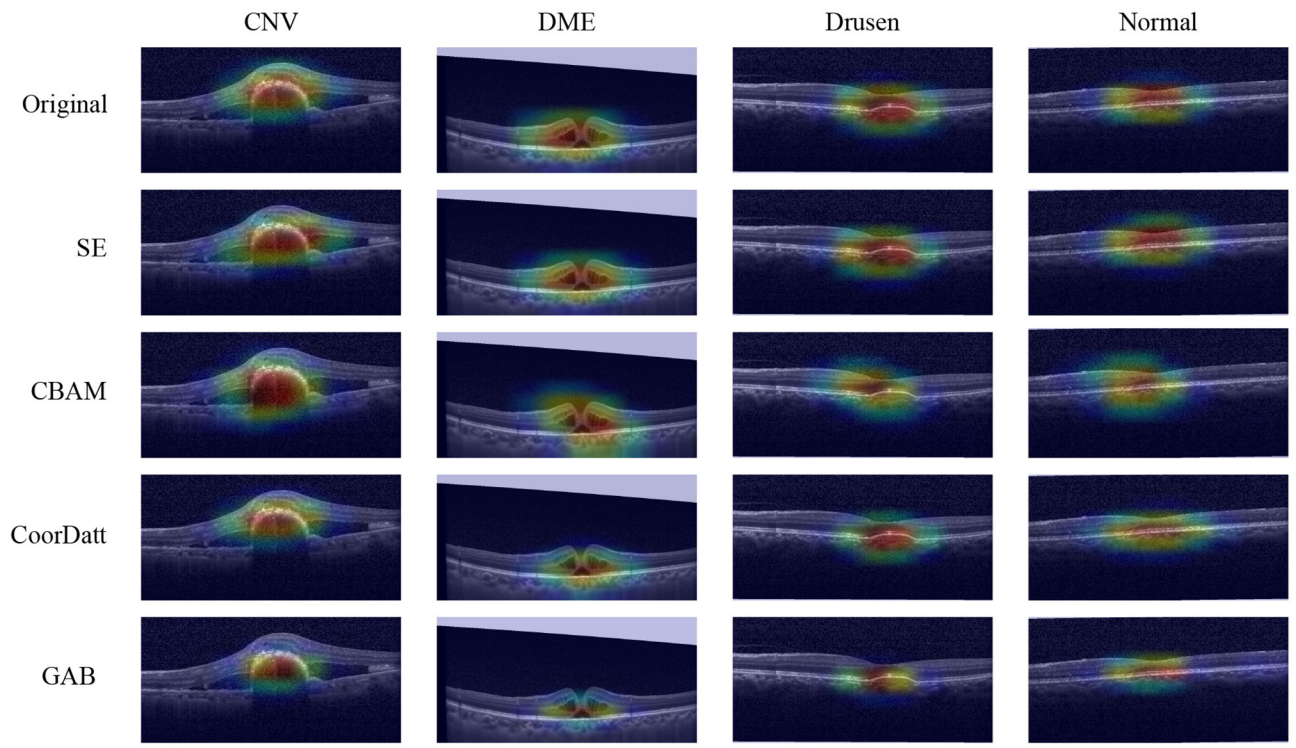


FIGURE 10
Model visualization maps of Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen and Normal by different attention mechanisms. The highlighted areas in red spectrum are considered as important in making its diagnosis.

The dataset comprises four categories: CNV, DME, Drusen, and Normal, with sample sizes of 59, 94, 20, and 83, respectively. The Medical Ethics Review Board of Beijing Chaoyang Hospital approved the retrospective study. We directly used the trained algorithm model to predict all image data. Similarly, we employed the preprocessing method used for the internal testing set to process the external validation dataset. The evaluation metrics for each category in the external validation dataset are shown in Table 4. As the OCT images used for training the network models were generated by a "Spectralis OCT" device, while the external dataset images were acquired from a "Cirrus HD" device, there were certain discrepancies observed in the OCT image characteristics produced by the two devices. Specifically, the Spectralis OCT equipment is known to produce images with higher clarity compared to the Cirrus HD-OCT, which might have resulted in a decline in the evaluation metrics during model testing. This is undoubtedly one of the future research directions that we plan to pursue. We aim to evaluate the generalization ability of our proposed algorithms by utilizing other locally available high-resolution image databases.

4.4.3. Extension to other types of datasets

To verify the performance of GAB and GABNet on other application scenario datasets, we selected a DR dataset for validation (Kaggle: https://www.kaggle.com/c/diabetic-retinopathy-detection/data). The dataset is taken from the Diabetic Retinopathy Detection Competition, a data modeling and data analysis competition platform. A total of 35,126 image samples are available and categorized into five categories: normal, mild, moderate, severe and proliferative DR. The sample sizes for each category are 25810, 2443, 5292, 873 and 708, respectively. Due to the severe imbalance in the dataset, this paper performs preprocessing using the data preprocessing scheme mentioned by (Ai et al., 2021). On this dataset, we use Xception, EfficientNetV2B3 and GABNet for testing, the results of which are shown in Table 5 and Figure 9.

First, it can be seen from Figure 9 that transfer learning can greatly improve the classification ability of the associated algorithm. The accuracy of Xception and EfficientNetV2B3 are improved by 4.2 and 6.1%, respectively. This demonstrates that the efficacy of feature extraction by each network is significantly enhanced through the "ImageNet" pre-training approach, which involves utilizing large datasets and ensuring a balance across various categories. Using the GAB attention mechanism, the accuracy of Xception and EfficientNetV2B3 are improved by 0.4 and 0.7%, respectively, which shows that the attention mechanism proposed in this paper is effective and efficient. On the other hand, GABNet can also improve the classification accuracy by 3% over that of Xception and EfficientNetV2B3 without transfer learning. This shows that GABNet has certain advantages not only in retinal OCT disease classification but also in DR grading detection by fundus photograph, indicating a good generalization ability.

As shown in Table 5, it is evident that the Xception+GAB, EfficientNetV2B3+GAB, and GABNet approaches presented in this study yield enhanced accuracy compared to the methods proposed by Wu and Hu (2019) and Ai et al. (2021). While the fusion

network utilized in Ai et al. (2021) demonstrated robust feature extraction, its parameter count was relatively high. In this paper, EfficientNetV2B3 and GABNet achieve improved classification correctness rates while reducing the number of required network parameters, proving that the GAB attention mechanism has a strong feature-weighted optimization function.

4.4.4. Model visualization and interpretation

To compare the differences of contributing regions between the proposed GAB attention mechanism and other attention mechanisms, a heatmap is created for each image using a visualization method, i.e., gradient-weighted class activation mapping (Grad-CAM). In the heatmaps (Figure 10), the most relevant category discriminating regions are highlighted in red. The fundamental purpose of heatmap generation is to construct an image that reveal the subregions of the original image to identify areas contributing to the algorithm's determination of the diagnosis. In this study, the Xception classifier is selected as the underlying model, and one image per category is arbitrarily chosen for feature visualization from the No Attention Mechanism, SE, CBAM, CoorDatt, and GAB methods.The contributing areas depicted in the heatmaps reveal that the GAB method highlights smaller regions compared to the Original, SE, CBAM, and CoorDatt methods, focusing predominantly on the lesion areas while disregarding other irrelevant regions. This substantiates the efficacy of the GAB attention mechanism. The locations highlighted by the GAB heatmap are partially consistent with human experts' experience which means good interpretation of this model.

5. Conclusion

This study presents a novel and effective Global Attention Block (GAB) for feedforward CNNs. The GAB is a versatile module that can be easily integrated into any CNN to improve its classification performance. Compared to commonly used attention mechanisms in current research, the GAB is shown to better focus on lesion locations in retinal OCT images, leading to improved classification results. Based on the GAB, a lightweight classification network model called GABNet is proposed, which demonstrates superior performance while also having a smaller number of parameters. Our future work includes testing the proposed algorithm on a larger set of locally sourced clinical image databases and optimizing it accordingly to improve the performance. Additionally, the use of a larger image classification database, such as ImageNet, is planned for training, and the transfer learning of the obtained GABNet classification model to more application scenarios will be expected to verify the algorithm's robustness.

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request. All deep learning methods are implemented by using TensorFlow (https://tensorflow.google.cn/). The custom script for this study will be available at https://github.com/YHHAZ/GABNet. Correspondence and requests for data materials should be addressed to Yaping Lu (luyaping@sinopharm.com).

Author contributions

Conceptualization and methodology: ZA and XH. Validation: ZA, XH, QW, BH, YL, and FZ. Funding acquisition: FZ, XH, HW, and YL. Writing—original draft: ZA and YL. Data collection: HW, CS, JF, and YT. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in parts by a grant from the National Natural Science Foundation of China to FZ (81902861), a grant from the National Natural Science Foundation of China (32000485) and the Beijing Chaoyang Hospital Science and Technology Innovation Fund (22kcjjyb-11), both awarded to XH, a grant from the National Natural Science Foundation of China to HW (62006161), and the Sinopharm Genomics Technology Co., Ltd. The funder Sinopharm Genomics Technology Co., Ltd. had the following involvement with the study: design, collection, analysis, interpretation of data, the writing of this article, and the decision to submit it for publication.

Acknowledgments

The numerical calculations in this paper were performed on the supercomputing system in the Supercomputing Center of Wuhan University.

Conflict of interest

ZA, QW, and YL are employees of Sinopharm Genomics Technology Co., Ltd. BH are employees of AI-Farm (Nanjing) Big Data Services Co., Ltd.

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

Publisher's note

References

Aditya, A., Zhou, L., Vachhani, H., Chandrasekaran, D., and Mago, V. (2021). "Collision Detection: An Improved Deep Learning Approach Using SENet and ResNext," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (Melbourne, Australia) 2075–2082. doi: 10.1109/SMC52423.2021.9659265

Ai, Z., Huang, X., Fan, Y., Feng, J., Zeng, F., Lu, Y., et al. (2021). Detection algorithm of diabetic retinopathy based on deep ensemble learning and attention mechanism. *Front. Neuroinform.* 15, 778552. doi: 10.3389/fninf.2021.778552

Ai, Z., Huang, X., Feng, J., Wang, H., Tao, Y., Zeng, F., et al. (2022). FN-OCT: disease detection algorithm for retinal optical coherence tomography based on a fusion network. *Front. Neuroinform.* 16, 876927. doi: 10.3389/fninf.2022.876927

Chen, J. L. (2007). Clinical applications and new developments of optical coherence tomography: an evidence-based review. *Clin. Exper. Optom.* 90, 317–335. doi: 10.1111/j.1444-0938.2007.00151.x

Chen, Y., Zhang, X., Chen, W., Li, Y., and Wang, J. (2020). Research on Recognition of Fly Species Based on Improved RetinaNet and CBAM. *IEEE Access.* 8, 102907–102919. doi: 10.1109/ACCESS.2020.2997466

Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions," in *Proceedings-30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Honolulu, Hawaii, USA) 1800–1807. doi: 10.1109/CVPR.2017.195

Chougrad, H., Zouaki, H., and Alheyane, O. (2018). Deep Convolutional Neural Networks for breast cancer screening. *Comput. Methods Programs Biomed.* 157, 19–30. doi: 10.1016/j.cmpb.2018.01.011

Dai, Q., Xie, Y., Xu, J., Xia, Y., Sheng, C., Tian, C., et al. (2022). "Tunnel crack identification based on improved YOLOv5," in *2022 7th International Conference on Automation, Control and Robotics Engineering (CACRE)* (Xi'an, China) 302–307. doi: 10.1109/CACRE54574.2022.9834211

Das, U. N. (2016). Diabetic macular edema, retinopathy and age-related macular degeneration as inflammatory conditions. *Arch. Med. Sci.* 12, 1142–1157. doi: 10.5114/aoms.2016.61918

Das, V., Dandapat, S., and Bora, P. K. (2021). Automated Classification of Retinal OCT Images Using a Deep Multi-Scale Fusion CNN. *IEEE Sens. J.* 21, 23256–23265. doi: 10.1109/JSEN.2021.3108642

Deng, J., Ma, Y., Li, D. A., Zhao, J., Liu, Y., and Zhang, H. (2020). Classification of breast density categories based on SE-Attention neural networks. *Comput. Methods Programs Biomed.* 193, 105489. doi: 10.1016/j.cmpb.2020.105489

Fang, L., Jin, Y., Huang, L., Guo, S., Zhao, G., Chen, X., et al. (2019). Iterative fusion convolutional neural networks for classification of optical coherence tomography images. *J. Vis. Commun. Image Represent.* 59, 327–333. doi: 10.1016/j.jvcir.2019.01.022

Farag, M. M., Fouad, M., and Abdel-Hamid, A. T. (2022). Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module. *IEEE Access.* 10, 38299–38308. doi: 10.1109/ACCESS.2022.3165193

He, K., Zhang, X., and Ren, S. J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Las Vegas, Nevada, USA) 770–778. doi: 10.1109/CVPR.2016.90

Hou, Q., and Zhou, D. J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 13708–13717. doi: 10.1109/CVPR46437.2021.01350

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L. C., Tan, M., et al. (2019). "Searching for mobileNetV3," in *Proceedings of the IEEE International Conference on Computer Vision* (Seoul, Korea) 1314–1324. doi: 10.1109/ICCV.2019.00140

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Huang, L., He, X., Fang, L., Rabbani, H., and Chen, X. (2019). Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. *IEEE Signal Process Lett.* 26, 1026–1030. doi: 10.1109/LSP.2019.2917779

Hwang, D.-., K., Hsu, C-, C., Chang, K.-., J., Chao, D., Sun, C-, H., and Jheng, Y.-., C., et al. (2019). Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics.* 9, 232. doi: 10.7150/thno.28447

Kassem, M. A., Hosny, K. M., and Fouad, M. M. (2020). Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning. *IEEE Access.* 8, 114822–114832. doi: 10.1109/ACCESS.2020.3003890

Kayadibi, I., and Güraksin, G. E. (2023). An early retinal disease diagnosis system using OCT images via CNN-based stacking ensemble learning. *Int. J. Multiscale Comput. Eng.* 21, 1–25. doi: 10.1615/IntJMultCompEng.2022043544

Kermany, D. (2018). Large dataset of labeled optical coherence tomography (OCT) and Chest X-Ray images. *Cell.* 172, 1122–1131. doi: 10.1016/j.cell.2018.02.010

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., et al. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell.* 172, 1122–1131.

Lee, J. A., Cheng, J., Xu, G., Ong, E. P., Lee, B. H., Wong, D. W. K., et al. (2015). "Registration of color and OCT fundus images using low-dimensional step pattern analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Munich, Germany) 214–221. doi: 10.1007/978-3-319-24571-3_26

Li, R., Wang, S., Wang, Z., and Zhang, L. (2021). "Breast cancer X-ray image staging: Based on efficient net with multi-scale fusion and cbam attention," in *Journal of Physics: Conference Series* (Zhuhai, China) 012006. doi: 10.1088/1742-6596/2082/1/012006

Li, X., Shen, X., Zhou, Y., Wang, X., and Classification, L., i. T. Q. (2020). of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLoS ONE.* 15, e0232127. doi: 10.1371/journal.pone.0232127

Liang, Y., Zhao, Z., Wang, H., Cao, Y., Huang, T., Medjadba, Y., et al. X. (2020). "Unbalanced Geologic Body Classification of Hyperspectral Data Based on Squeeze and Excitation Networks at Tianshan Area," in *International Geoscience and Remote Sensing Symposium (IGARSS)* (Waikoloa, HI, USA) 6981–6984. doi: 10.1109/IGARSS39084.2020.9323795

Liu, Y. Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J. S., Rehg, J. M., et al. (2011). Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Med. Image Anal.* 15, 748–759. doi: 10.1016/j.media.2011.06.005

Luo, Y., and Wang, Z. (2021). "An Improved ResNet Algorithm Based On CBAM," in *Proceedings - 2021 International Conference on Computer Network, Electronic and Automation, ICCNEA 2021* (Xi'an, China) 121–125. doi: 10.1109/ICCNEA53019.2021.00036

Narayan Das, N., Kumar, N., Kaur, M., Kumar, V., and Singh, D. (2022). Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays. *IRBM.* 43, 114–119. doi: 10.1016/j.irbm.2020.07.001

Shi, C., Lin, L., Sun, J., Su, W., Yang, H., Wang, Y. A., et al. (2022). "Lightweight YOLOv5 Transmission Line Defect Detection Method Based on Coordinate Attention," in *IEEE 6th Information Technology and Mechatronics Engineering Conference, ITOEC 2022* (Chongqing, China) 1779–1785. doi: 10.1109/ITOEC53115.2022.9734540

Sinha, B. B., Wary, A., and Dhanalakshmi, R. K. (2023). "5-transfer learning-based detection of retina damage from optical coherence tomography images," in *Computational Methods and Deep Learning for Ophthalmology*, eds. D. J. Hemanth (New York: Academic Press) 71–88. doi: 10.1016/B978-0-323-95415-0.00002-4

Sotoudeh-Paima, S., Jodeiri, A., Hajizadeh, F., and Soltanian-Zadeh, H. (2022). Multi-scale convolutional neural network for automated AMD classification using retinal OCT images. *Comput. Biol. Med.* 144, 105368. doi: 10.1016/j.compbiomed.2022.105368

Srinivasan, P. P., Kim, L. A., Mettu, P. S., Cousins, S. W., Comer, G. M., Izatt, J. A., et al. (2014). Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express.* 5, 3568–3577. doi: 10.1364/BOE.5.003568

Szegedy, C., Vanhoucke, V., Ioffe, S., and Shlens, J. Z. (2016). "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Las Vegas, Nevada, USA) 2818–2826. doi: 10.1109/CVPR.2016.308

Tan, M., and Le, Q. (2021). "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning* (PMLR) 10096–10106.

Wang, W., Tan, X., Zhang, P., and Wang, X. (2022). A CBAM based multiscale transformer fusion approach for remote sensing image change detection. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 15, 6817–6825. doi: 10.1109/JSTARS.2022.3198517

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "A CBAM Convolutional block attention module," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Munich, Germany) 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, Y., and Hu, Z. (2019). "Recognition of diabetic retinopathy basedon transfer learning," in *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (IEEE) 398–401. doi: 10.1109/ICCCBDA.2019.8725801

Xiang, X., Tian, D., Lv, N., and Yan, Q. (2022). FCDNet: a change detection network based on full-scale skip connections and coordinate attention. *IEEE Geosci. Remote Sens. Lett*. 19, 1–5. doi: 10.1109/LGRS.2022.3184179

XinSheng, Z., Yu, W. (2022). Industrial character recognition based on improved CRNN in complex environments. *Comput. Ind*. 142, 103732. doi: 10.1016/j.compind.2022.103732

Yan, W., and Hua, Y. (2020). "Deep Residual SENet for Foliage Recognition," in *Transactions on Edutainment XVI* (Berlin, Heidelberg) 92–104. doi: 10.1007/978-3-662-61510-2_9

Zha, M., Qian, W., Yi, W., and Hua, J. (2021). A lightweight yolov4-based forestry pest detection method using coordinate attention and feature fusion. *Entropy* 23, 1587. doi: 10.3390/e23121587

Zhang, L., Wang, J., Li, B., Liu, Y., Zhang, H., Duan, Q. A., et al. (2022). MobileNetV2-SENet-based method for identifying fish school feeding behavior. *Aquac. Eng*. 99, 102288. doi: 10.1016/j.aquaeng.2022.102288

Zrenner, E. (2002). Will retinal implants restore vision? *Science*. 295, 1022–1025. doi: 10.1126/science.1067996

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 08 June 2023
DOI 10.3389/fnins.2023.1195188

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Jiaqi Wang,
Xiang Yang No.1 People's Hospital, China
Yan Tong,
Renmin Hospital of Wuhan University, China
Chen-Xing Qi,
Renmin Hospital of Wuhan University, China

*CORRESPONDENCE
Bing-Lin Huang
✉ huangbinglin2009@163.com

RECEIVED 28 March 2023
ACCEPTED 04 May 2023
PUBLISHED 08 June 2023

CITATION
Li D-J, Huang B-L and Peng Y (2023) Comparisons of artificial intelligence algorithms in automatic segmentation for fungal keratitis diagnosis by anterior segment images. *Front. Neurosci.* 17:1195188. doi: 10.3389/fnins.2023.1195188

Comparisons of artificial intelligence algorithms in automatic segmentation for fungal keratitis diagnosis by anterior segment images

Dong-Jin Li[1], Bing-Lin Huang[2]* and Yuan Peng[3]

[1]Health Management Center, The First People's Hospital of Jiujiang City, Jiujiang, Jiangxi, China, [2]College of Clinical Medicine, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China, [3]Department of Ophthalmology, The Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China

Purpose: This study combines automatic segmentation and manual fine-tuning with an early fusion method to provide efficient clinical auxiliary diagnostic efficiency for fungal keratitis.

Methods: First, 423 high-quality anterior segment images of keratitis were collected in the Department of Ophthalmology of the Jiangxi Provincial People's Hospital (China). The images were divided into fungal keratitis and non-fungal keratitis by a senior ophthalmologist, and all images were divided randomly into training and testing sets at a ratio of 8:2. Then, two deep learning models were constructed for diagnosing fungal keratitis. Model 1 included a deep learning model composed of the DenseNet 121, mobienet_v2, and squeezentet1_0 models, the least absolute shrinkage and selection operator (LASSO) model, and the multi-layer perception (MLP) classifier. Model 2 included an automatic segmentation program and the deep learning model already described. Finally, the performance of Model 1 and Model 2 was compared.

Results: In the testing set, the accuracy, sensitivity, specificity, F1-score, and the area under the receiver operating characteristic (ROC) curve (AUC) of Model 1 reached 77.65, 86.05, 76.19, 81.42%, and 0.839, respectively. For Model 2, accuracy improved by 6.87%, sensitivity by 4.43%, specificity by 9.52%, F1-score by 7.38%, and AUC by 0.086, respectively.

Conclusion: The models in our study could provide efficient clinical auxiliary diagnostic efficiency for fungal keratitis.

KEYWORDS

anterior segment images, artificial intelligence, automatic segmentation, fungal keratitis, diagnosis

Introduction

Fungal keratitis, also known as keratomycosis, is a common blinding eye disease (Thomas et al., 2005). The main manifestations are corneal infiltration, rough corneal edge, and "satellite" lesions (Mahmoudi et al., 2017). Patients often suffer from eye injury, require eye surgery, must wear contact lenses, and suffer from other diseases caused by organic substances (especially plants; Ali Shah et al., 2017). According to statistics, every year ~1–14 million people are infected with fungal keratitis worldwide, of which 75% of patients might

be blind in one eye and 60% of patients might be blind even after treatment (Brown et al., 2021), which results in a huge burden to families and society. Therefore, early diagnosis and treatment of fungal keratitis is necessary. However, at present, fungal keratitis diagnosis depends mainly on traditional microbial culture (Sadik et al., 2022), which takes considerable time and cannot provide a basis for early treatment. At present, the diagnosis of fungal corneal ulcer is mainly based on confocal microscopy of corneal culture. Fungal corneal ulcer can cause corneal perforation and fungal endophthalmitis. Thus, accurate and rapid early diagnosis of fungal keratitis is important.

Recently, artificial intelligence (AI), especially machine learning (ML), has been applied in the field of ophthalmology (Lee et al., 2020) and has a significant role in corneal disease diagnosis (Siddiqui et al., 2020). A corneal ulcer can be diagnosed by anterior segment photography. At the same time, artificial intelligence technology has shown better diagnostic efficiency in medical image processing. Moreover, ML based on the deep neural network (DNN) is called deep learning and is considered the most advanced ML (LeCun et al., 2015; Litjens et al., 2020). Huang et al. used the deep learning model built by different convolutional neural networks (CNNs) to evaluate 580 patients to help distinguish bacterial keratitis (BK) and fungal keratitis quickly in clinical practice and found that DenseNet 161 in CNN has the best performance. This deep learning model can improve the recognition rate significantly between the two kinds of keratitis and provide better accuracy for clinical diagnosis (Hung et al., 2021). Additionally, Li et al. compared the classification ability of AlexNet, DenseNet 121, and InceptionV3 algorithms for 48,530 slit lamp images of different keratitis and found that DenseNet 121 had the best classification performance (Li et al., 2021). AI is used widely in the field of keratitis diagnosis, and algorithms, such as DenseNet 161 and DenseNet 121, have high performance in deep learning models. However, most of the existing AI-assisted diagnosis of fungal keratitis methods compare the performance of different single algorithm models. The application of the deep learning model built by integrating these different algorithms in the diagnosis of keratitis is relatively rare. In contrast to the abovementioned research, Ghosh et al. used three deep learning models constructed by VGG19, DenseNet 121, and RestNet50 to separate fungal keratitis and BK and then compared the results of each model and ensemble learning. Finally, ensemble learning had the largest area under the precision-recall curve (AUPRC) compared with any single architecture model, and they believed that ensemble learning can improve the performance of assisted diagnosis of diseases significantly (Ghosh et al., 2022). Therefore, the ensemble learning model composed of multiple algorithms is more accurate. Ensemble learning is a kind of fusion technology that is a fusion at the model level and belongs to late fusion. Early fusion is also named feature-level fusion, which emphasizes the data combination before the classification. The final feature vector consists of the features extracted from heterogeneous signals, and early fusion should put the final feature vector into the classifier alone (Zhang et al., 2017).

Currently, prior AI studies have mainly focused on the diagnosis of viral keratitis and bacterial keratitis. Most previous studies used traditional machine learning or deep learning based on original slit lamp images. No studies have investigated the early fusion method for fungal keratitis. Moreover, previous studies were based mostly on whole anterior segment images. However, the area outside the keratitis lesion might affect the performance of models. Therefore, it is necessary to segment the lesion area from the images. Manual segmentation is tedious, time-consuming, and user-dependent (Wang et al., 2016), and automatic segmentation can be faster but might not have the same accuracy as manual segmentation (Wang et al., 2016; Huang et al., 2019). Thus, we hypothesized that the early fusion method for fungal keratitis with automatic and manual segmentation may show better diagnostic and sorting efficiency.

Therefore, this study combines automatic segmentation and manual fine-tuning with an early fusion method to provide efficient clinical auxiliary diagnostic efficiency for fungal keratitis. In detail, we developed two AI platforms with a deep transfer-learning algorithm and multi-feature fusion for fungal keratitis and non-fungal keratitis; one is based on an automatic segmentation method, whereas the other is based on a manual segmentation method.

Materials and methods

Study design

To realize the automated diagnosis of fungal keratitis, two deep learning models were constructed. Model 1 only included a deep learning model which was composed of the DenseNet 121 (The idea of Dense Connection is used, that is, every layer is connected with all the previous layers, so that the model has better information transmission and reuse ability in feature extraction. DenseNet 121 refers to the fact that the model has 121 layers), mobienet_v2 (this is a lightweight Convolution neural network model, mainly is the depth of Separable Convolution (Depthwise Separable Convolution) and Linear Bottleneck (Linear Bottleneck) technology, such as small parameters, run fast) and squeezentet1_0 models (Squeezentet1_0 is another lightweight convolutional neural network model, which is composed of a Squeeze layer and an Expand layer. It also runs fast with fewer parameters) which are common convolutional neural network models are used for image classification and object detection. The least absolute shrinkage and selection operator (LASSO) model and multi-layer perception (MLP) classifier. Model 2 included the automatic segmentation program and the deep learning model as described above. The deep learning pipeline of our study is shown in Figure 1.

Establishment of a dataset and image preprocessing of anterior segment images

We collected 423 high-quality anterior segment images of keratitis in the Department of Ophthalmology of the Declaration of Helsinki and were approved by the Medical Ethics Committee of the affiliated Hospital of Jiangxi University of Traditional Chinese Medicine from February 2020 to September 2023. The inclusion criteria of fungal keratitis are as follows: the corneal scrape was examined with 10% potassium hydroxide wet tablet bacteria or

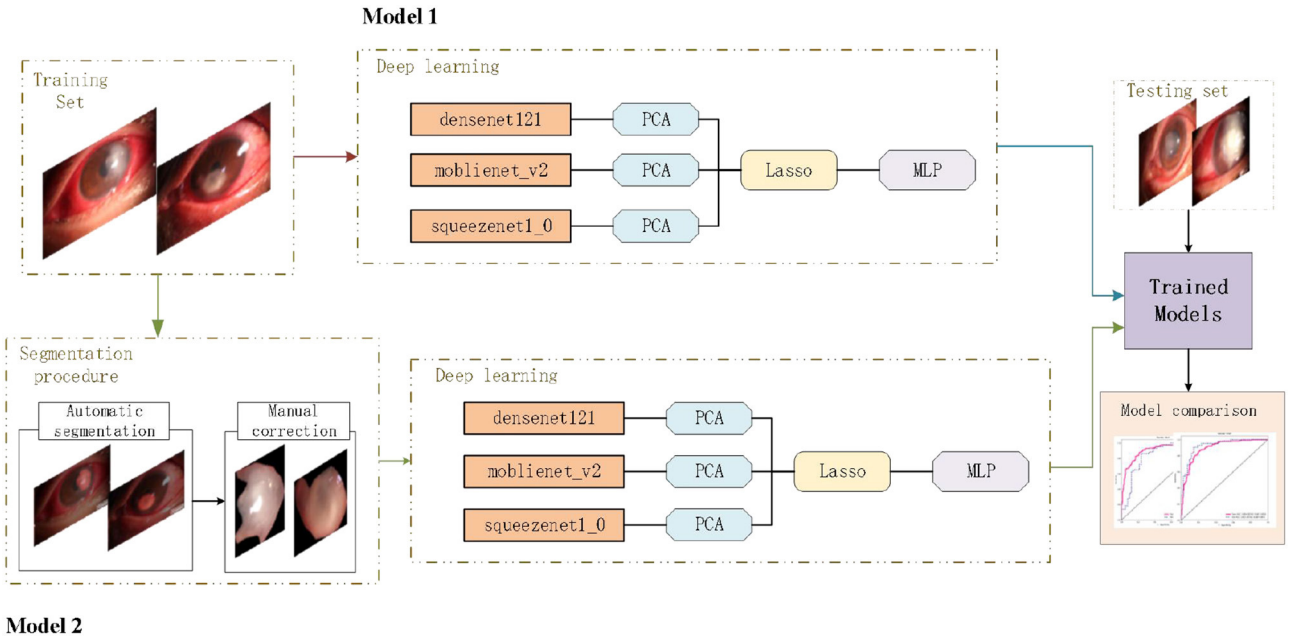


FIGURE 1
Deep learning pipeline. The two deep learning models were developed separately, and the performances of different models were compared based on a new testing set.

cornea, necrotic tissue, and pus in the potato culture medium to see bacteria falling growth and to make clinical manifestations such as ulcer surface with moss-like bad dead tissue, satellite foci, feathery edges, and furrow pits that can be seen around ulcer depression, and focal stromal infiltration dense, that may be accompanied by stromal abscess. The cornea is often pasted with a white mushy posterior corneal deposit (KP), anterior room pus color white matter thick, longer use of antibiotics, or cortical stimulation of patients with ineffective vegetarian treatment or ulcer aggravation.

To protect patient privacy, identifiable information was removed. Then, the images were divided into the fungal keratitis group and non-fungal keratitis group by a senior ophthalmologist, and all the images were randomly divided into training and testing sets at the ratio 8:2. The dataset contained a training set and a testing set, where the training set contained 168 fungal keratitis images and 170 non-fungal keratitis images, and the testing set contained 42 fungal keratitis images and 43 non-fungal keratitis images. This study was approved by the Ethics Committee of Jiangxi Province Peoples Hospital and adhered to the Declaration of Helsinki and the ARVO statement on human subjects.

Establishment of the automatic segmentation model

First, based on the anterior segment images, a senior ophthalmologist used the LabelMe software (https://github.com/wkentaro/labelme) to annotate the keratitis lesions area as the region of interest (ROI), respectively. The ROI of each image was annotated as "label 0" or "label 1." "Label 0" was defined as the fungal keratitis lesions area. "Label 1" was defined as the non-fungal keratitis lesions area. The FCNResnet50 which was a fully convolutional network based on ResNet50 was used to extract the ROI masks. First, the original images in the training set were used to train the FCNResnet50 model, and the obtained optimal parameters were then applied to the whole anterior segment images to get the automatic segmentation mask. Then, the manual segmentation mask annotated by the senior ophthalmologist was used as the gold standard. After the segmentation errors were adjusted, the final mask was obtained. Based on the final masks, the keratitis lesions area was segmented.

Establishment of the deep learning diagnostic model

The fungal keratitis detection was defined as a binary classification problem, with a label of 0 or 1 indicating that the image was fungal keratitis or non-fungal keratitis. This classification task was performed by the deep learning diagnostic models composed of DenseNet 121, mobienet_v2, squeezentet1_0, the least absolute shrinkage and selection operator (LASSO) model, and the multi-layer perception (MLP) classifier. In the training set, first, we used three models to extract features of the penultimate layers of the network and principal components analysis (PCA) in feature dimensionality reduction. Then, the features after dimensionality reduction were fusioned by channel concat which meant that the layer stacked features from each branch together. The LASSO logistic regression algorithm was used to select the optimal features. Finally, the optimal feature set was input into the MLP classifier to establish the final diagnostic model. In the

testing set, the 5-fold cross-validation was performed for parameter optimization. The selected features and the best parameters were applied for model evaluation.

Comparison and validation of diagnostic models

In Model 1, the original images in the training set were used to train the deep learning diagnostic model, and the original images in the testing set were used to validate this model. In Model 2, first, the original images in the training set were used to train the automatic segmentation model, then after manual fine-tuning, the keratitis lesions area segmented from the original images which were in the training set was used to train the deep learning diagnostic model. Finally, the original images in the testing set were used to validate Model 2.

To compare the performance of the two models, the receiver operating characteristic (ROC) curve was performed in this study to analyze the diagnostic ability of each model. The decision curve analysis (DCA) was used to evaluate the net benefit of the models for clinical decisions. The highest curve at any given threshold probability is the optimal decision-making strategy to maximize the net benefit (Gao et al., 2022). The gradient-weighted class activation mapping (Grad-CAM) was used for the visual verification of the diagnostic results of this method. The heatmap images created by the Grad-CAM indicated where the deep learning model was focused.

Statistical analysis

For the automatic segmentation model, we used the pixel-level classification accuracy, the average intersection-over-union (IOU), and dice coefficient to evaluate the performance. The pixel-level classification accuracy was the percentage of correctly classified pixels out of the total pixels in each image, and IoU evaluated precision by calculating the overlap between the prediction and target variables (Mahmoudi et al., 2017; Larsen et al., 2021). The dice coefficient is a set similarity measure function, the higher the dice coefficient, the better the segmentation effect (Li et al., 2020). For the deep learning diagnostic model, we measured the accuracy, sensitivity, specificity, and F1-score from the training set and testing set. We also plotted the DCA curves and the ROC curves from the two models. The area under the curve (AUC) with a 95% confidence interval (95% CI) of each model which was in the training set and testing set was calculated. All the methods were implemented in Python language using Python 3.9.7 version.

Results

Performance of automatic segmentation model

The pixel-level classification accuracy was 96.2%. The average IoU score was 81.3%. The mean dice score was 89%. The diagram of the segmentation image effect of the keratitis lesions area is shown in Figure 2.

Comparing the different deep learning diagnostic models in diagnosing fungal keratitis

In the testing set, the accuracy, sensitivity, specificity, and F1-score of Model 1 reached 77.65, 86.05, 76.19, and 81.42%, respectively. For Model 2, which is based on the segmentation images, the accuracy improved by 6.87%, sensitivity by 4.43%, specificity by 9.52%, and F1-score by 7.38%, respectively, as shown in Table 1.

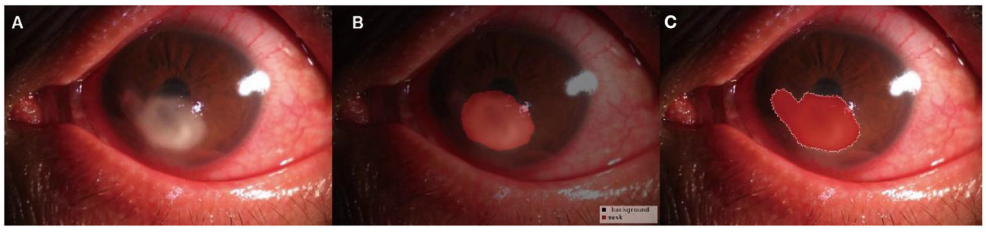


FIGURE 2
Diagram of the segmentation image effect of the keratitis lesions area: the original images **(A)**, the automatic segmentation results **(B)**, and the manual segmentation result **(C)**. The red area indicated the segmented lesion area of keratitis.

TABLE 1 Performance comparison of Model 1 and Model 2.

Model name	Train/test	Accuracy	AUC	95% CI	Sensitivity	Specificity	F1-score
Model 1	Train	82.84%	0.905	(0.874–0.937)	82.94%	83.33%	82.94%
	Test	77.65%	0.839	(0.751–0.927)	86.05%	76.19%	81.42%
Model 2	Train	81.71%	0.894	(0.861–0.928)	85.38%	79.76%	83.21%
	Test	84.52%	0.925	(0.869–0.981)	90.48%	85.71%	88.80%

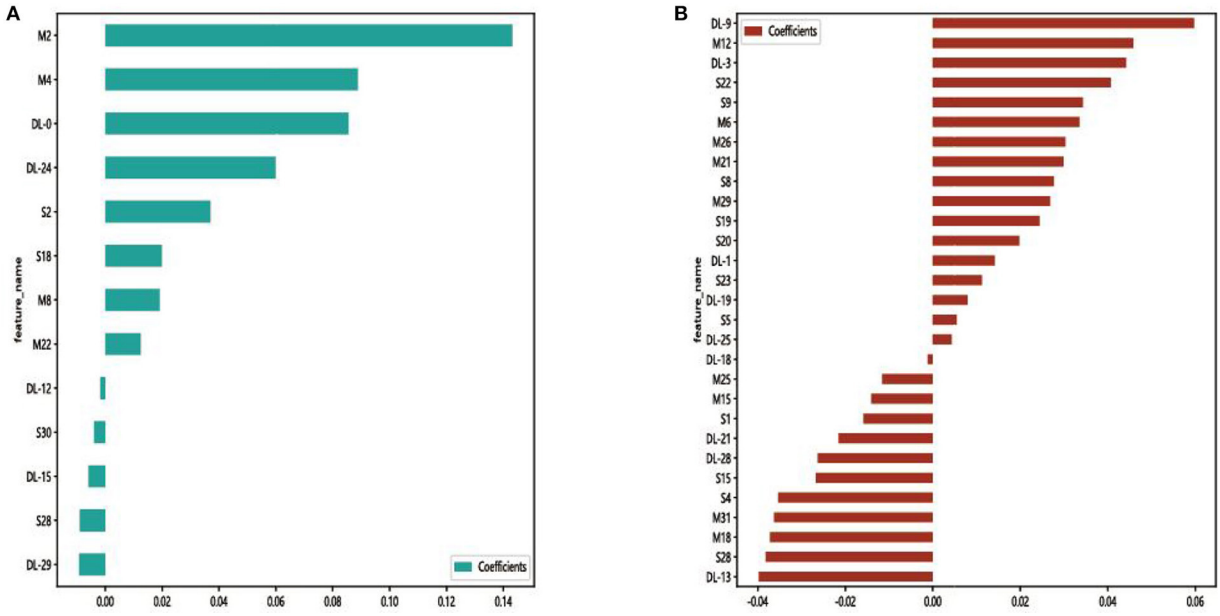


FIGURE 3
Selected features and their coefficient values in Model 1 **(A)** and Model 2 **(B)**. DL, M, and S indicate that the feature was from the DenseNet 121, mobienet_v2, and squeezentet1_0 models, respectively.

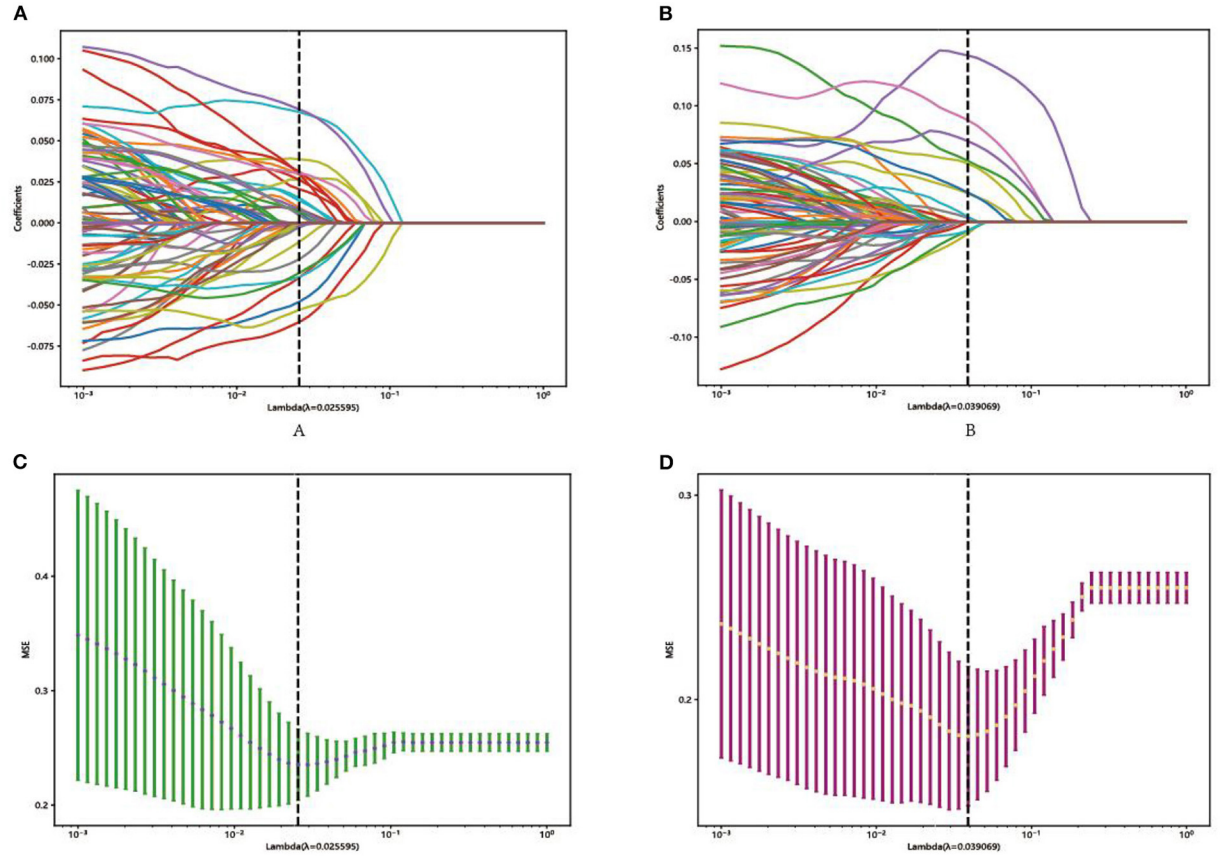


FIGURE 4
Representative LASSO coefficient distribution map, Model 1 **(A)** and Model 2 **(B)**. Selection of features based on the LASSO regression model, Model 1 **(C)**, and Model 2 **(D)**.

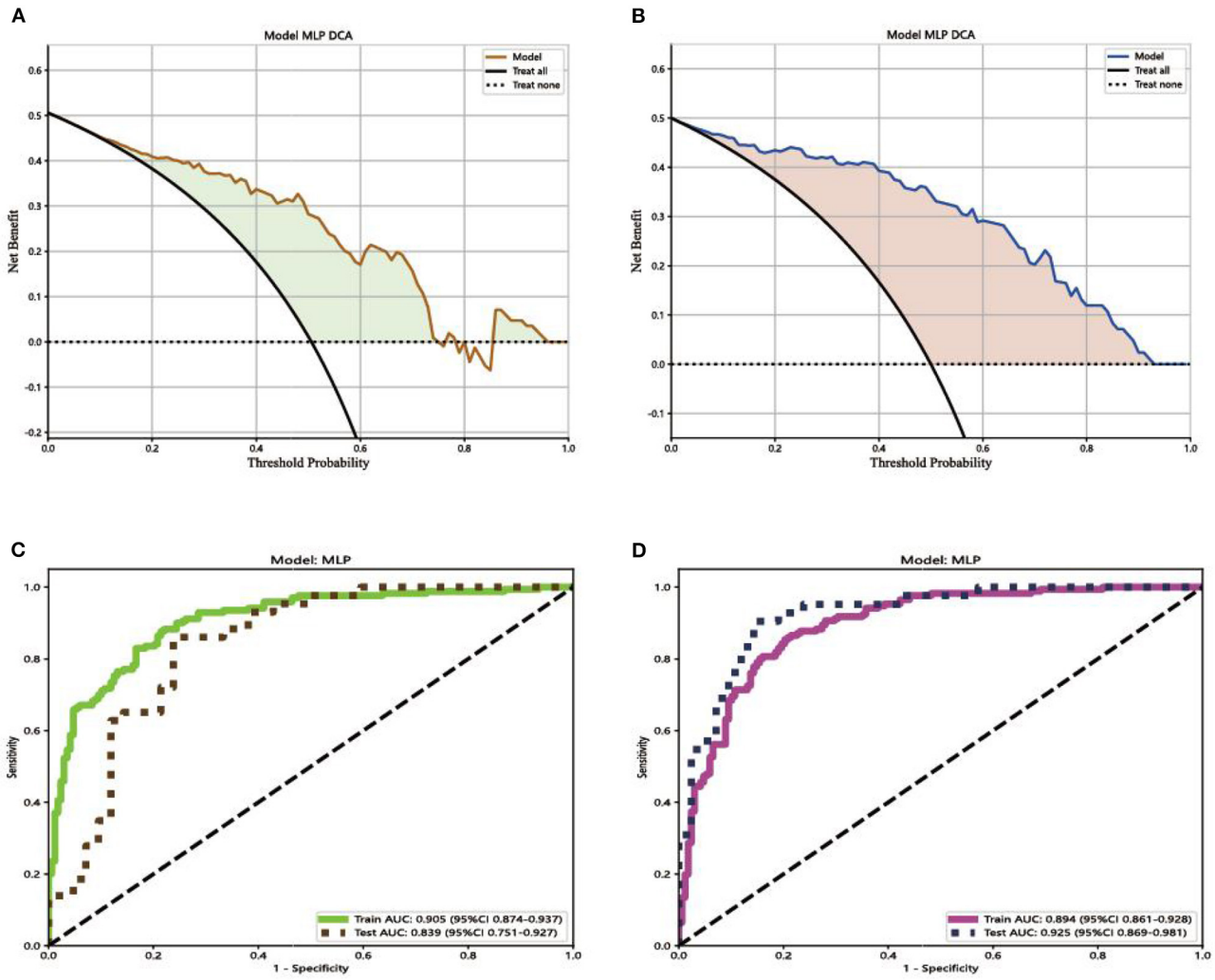


FIGURE 5
Decision curves and receiver operating characteristic (ROC) curves for the different models. **(A)** The net benefit of Model 1 in making a correct diagnosis of fungal keratitis; **(B)** the net benefit of Model 2 in making a correct diagnosis of fungal keratitis. The x-axis is the threshold probability, and the y-axis measures the net benefit. "Treat none" indicates that all samples were negative without intervention and the net benefit was 0. "Treat all" indicates that all samples were positive with intervention. **(C)** ROC curve for the different Model 1; **(D)** ROC curve for Model 2. AUC indicates the area under the curve of ROC.

In each model, a total of 50,176 features, 62,720 features, and 43,265 features were extracted from the DenseNet 121, mobienet_v2, and squeezentet1_0 models separately. After dimension reduction and channel concat, a total of 93 features were retained. After screening with the Lasso model, 13 features and 29 features were left for further classification in Model 1 and Model 2, respectively. The selected features and their coefficient values are shown in Figure 3. The LASSO screening process is shown in Figure 4.

Comparing the results of ROC curves, in the testing set, Model 1 achieved an AUC of 0.839 (95% CI 0.751–0.927). Model 2 achieved the highest AUC of 0.925 (95% CI 0.869–0.981). Compared with the result of the DCA curve analysis, Model 2 would substantially benefit in diagnosing fungal keratitis when the threshold probability was between 0 and 90% in the test set, which received a higher net benefit than Model 1, as shown in Figure 5.

Visualization of the deep learning process

We used Grad-CAM to locate the important region for the classification. The results of the heat map displayed the areas which Model 1 likely focused on and were located in the keratitis lesions area but covered the surrounding normal corneal tissues. The areas that Model 2 likely focused on were located in the center of the keratitis lesions area, as shown in Figure 6.

Discussion

Delayed diagnosis remains the main reason for the poor prognosis of deteriorating lesions (Wei et al., 2023). Therefore, this study developed an automatic diagnosis model to provide efficient clinical auxiliary diagnostic efficiency for fungal keratitis. As far

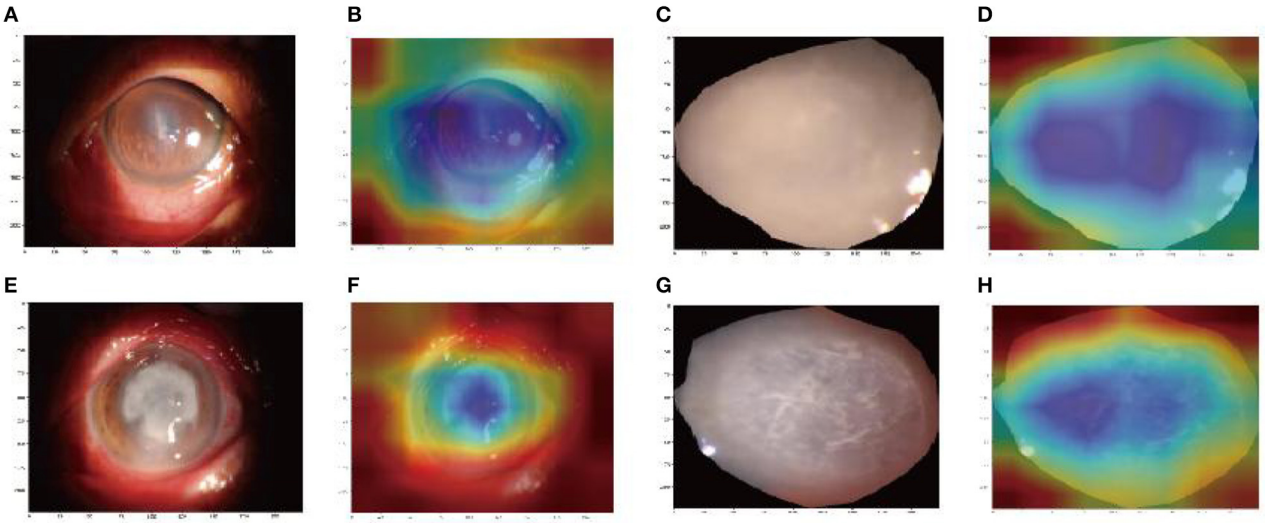


FIGURE 6
Image region heat maps based on Grad-CAM: the original images of Model 1 **(A, C)** and Model 2 **(E, G)**. The heat maps correspond to the original image of Model 1 **(B, D)** and Model 2 **(F, H)**. The blue areas were that the models likely focused on.

as we know, this is the first AI-assisted diagnostic model that combines automatic segmentation and manual fine-tuning with an early fusion method for fungal keratitis (FK) diagnosis.

Comparing the performance of Model 1 with Model 2 in our study, the accuracy, sensitivity, specificity, F1-score, and AUC of Model 2 were all significantly higher than that of Model 1 in the test set. In previous research, Hung et al. (2021) used U square Net (U^2 Net) to crop the image of the cornea and various CNN for identifying BK and FK. The DenseNet 161 model had an accuracy of 65.8%, which was the highest among all the models. The performance of their models is far below ours, indicating the limitation of single features in classification (Geng et al., 2017). Zhang et al. (2022) used a CNN to classify infectious keratitis. The highest accuracy and AUC of individual models was 77.11%. After the fusion of ResNext101_32x16d and DenseNet 169 models, although the accuracy was improved by 0.6%, this result is still lower than that of our Model 2. This is probably owing to the precise segmentation of the keratitis lesion area in our model, which ruled out the interference of the background. The performance of Model 1 is lower than Model 2 in our study, which also confirmed this notion. From this, the performance of models could be further improved by combining the segmentation model with the fusion method.

The decision curve analysis was implemented to evaluate the clinical usefulness of the model for diagnosing FK. The decision curve of a model is compared with extreme cases that include all patients or none. A model can be recommended for clinical use if its net benefit is greater than treating all and no patients (Du et al., 2021). Comparing the result of Model 1 with Model 2, the two models were both better than extreme cases (none and all) in the test set. Model 2 has greater potential for clinical application. Comparing the results of ROC curves, in the testing set, Model 1 achieved an AUC of 0.839 (95% CI 0.751–0.927). Model 2 achieved the highest AUC of 0.925 (95% CI 0.869–0.981). Compared with the result of the DCA curve analysis, Model 2 would substantially benefit in diagnosing fungal keratitis when the threshold probability was between 0 and 90% in the test set. Through the multi-feature transfer learning method combined with an automatic or manual segmentation algorithm, the resulting automatic segmentation platform can diagnose FK more quickly, whereas the resulting manual segmentation platform can diagnose FK more accurately.

Another strength of this study is the Grad-CAM introduction. Deep learning models are usually regarded as black boxes because the information regarding which features are important cannot be interpreted easily from the model (Wang et al., 2019). In our study, the heatmap images of Grad-CAM highlighted the important areas in corneal ulcer images used for AI diagnosis, which interprets the deep learning process effectively.

Our study also has certain limitations. First, the sample size included in this study was small. Second, the diagnosis of fungal keratitis is not entirely accurate, and some subjects lack laboratory tests. Third, the study only diagnosed FK and did not distinguish between different types of keratitis. Finally, accuracy needs to be improved. Therefore, in future studies, we will attempt to introduce transform learning to identify different keratitis types.

Conclusion

In this study, we combined automatic segmentation and manual fine-tuning with the early fusion method for FK diagnosis which provides efficient clinical auxiliary diagnostic efficiency for fungal keratitis. Through the multi-feature transfer learning method combined with an automatic or manual segmentation algorithm, the resulting automatic segmentation platform can diagnose FK more quickly, whereas the resulting manual segmentation platform can diagnose FK more accurately.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

D-JL, B-LH, and YP contributed to data collection, conducted statistical analyses, and wrote the manuscript. All authors read and approved the final manuscript, contributed to the article, and approved the submitted version.

Acknowledgments

We acknowledge the assistance provided by the National Natural Science Foundation of China Youth Science Fund Project (No. 82104935), Natural Science Foundation of Jiangxi Province (No. 20202BAB2060743), and Science and Technology Research Project of Jiangxi Provincial Department of Education (No. GJJ190647).

Conflict of interest

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

Publisher's note

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

References

Ali Shah, S. I. A., Shah, S. A., Rai, P., Katpar, N. A., Abbasi, S. A., and Soomro, A. A. (2017). Visual outcome in patients of keratomycosis, at a tertiary care centre in Larkana, Pakistan. *J. Pak. Med. Assoc.* 67, 1035–1038.

Brown, L., Leck, A. K., Gichangi, M., Burton, M. J., and Denning, D. W. (2021). The global incidence and diagnosis of fungal keratitis. *Lancet Infect. Dis.* 21, e49–e57. doi: 10.1016/S.1473-3099(20)30448-5

Du, L., Feng, L., Bi, S., Zhang, L., Tang, J., Zhong, L., et al. (2021). Probability of severe postpartum hemorrhage in repeat cesarean deliveries: A multicenter retrospective study in China. *Sci. Rep.* 11, 8434. doi: 10.1038/s41598-021-87830-7

Gao, W., Wang, W., Song, D., Yang, C., Zhu, K., Zeng, M., et al. (2022). A predictive model integrating deep and radiomics features based on gadobenate dimeglumine-enhanced MRI for postoperative early recurrence of hepatocellular carcinoma. *Radiol. Med.* 127, 259–271. doi: 10.1007/s11547-021-01445-6

Geng, Z., Tang, F., Ding, Y., Li, S., and Wang, X. (2017). Noninvasive continuous glucose monitoring using a multisensor-based glucometer and time series analysis. *Sci. Rep.* 7, 12650. doi: 10.1038/s41598-017-13018-7

Ghosh, A. K., Thammasudjarit, R., Jongkhajornpong, P., Attia, J., and Thakkinstian, A. (2022). Deep learning for discrimination between fungal keratitis and bacterial keratitis: DeepKeratitis. *Cornea* 41, 2830. doi: 10.1097/ICO.0000000000002830

Huang, C., Tian, J., Yuan, C., Zeng, P., He, X., Chen, H., et al. (2019). Fully automated segmentation of lower extremity deep vein thrombosis using convolutional neural network. *BioMed Res. Int.* 2019, 3401683. doi: 10.1155/2019/3401683

Hung, N., Shih, A. K. Y., Lin, C., Kuo, M. T., Hwang, Y. S., Wu, W. C., et al. (2021). Using slit-lamp images for deep learning-based identification of bacterial and fungal keratitis: Model development and validation with different convolutional neural networks. *Diagn. Basel Switz.* 11, 71246. doi: 10.3390/diagnostics11071246

Larsen, A., Hanigan, I., Reich, B. J., Qin, Y., Cope, M., Morgan, G., et al. (2021). A deep learning approach to identify smoke plumes in satellite imagery in near real-time for health risk communication. *J. Expo. Sci. Environ. Epidemiol.* 31, 170–176. doi: 10.1038/s41370-020-0246-y

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

Lee, J., Kim, J. S., Lee, H. J., Kim, S. J., Kim, Y. K., Park, K. H., et al. (2020). Discriminating glaucomatous and compressive optic neuropathy on spectral-domain optical coherence tomography with deep learning classifier. *Br. J. Ophthalmol.* 104, 1717–1723. doi: 10.1136/bjophthalmol-2019-314330

Li, Y., Wang, G., Li, M., Li, J., Shi, L., and Li, J. (2020). Application of CT images in the diagnosis of lung cancer based on finite mixed model. *Saudi J. Biol. Sci.* 27, 1073–1079. doi: 10.1016/j.sjbs.2020.02.022

Li, Z., Jiang, J., Chen, K., Zheng, Q., Liu, X., Weng, H., et al. (2021). Development of a deep learning-based image quality control system to detect and filter out ineligible slit-lamp images: A multicenter study. *Comput. Methods Programs Biomed.* 203, 106048. doi: 10.1016/j.cmpb.2021.106048

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2020). Deep learning in medical image analysis. *Adv. Exp. Med. Biol.* 1213, 3–21. doi: 10.1007/978-3-030-33128-3_1

Mahmoudi, S., Masoomi, A., Ahmadikia, K., Tabatabaei, S. A., Soleimani, M., Rezaie, S., et al. (2017). Fungal keratitis. *J. Fr. Ophtalmol.* 40, e307–e313. doi: 10.1016/j.jfo.2017.08.001

Sadik, N., Elzeiny, S. M., Ali, Y. E., and Sobeih, D. (2022). Fungal keratitis in the Egyptian delta: Epidemiology, risk factors and microbiological diagnosis. *Ophthal. Epidemiol.* 29, 1914667. doi: 10.1080/09286586.2021.1914667

Siddiqui, A. A., Ladas, J. G., and Lee, J. K. (2020). Artificial intelligence in cornea, refractive, and cataract surgery. *Curr. Opin. Ophthalmol.* 31, 253–260. doi: 10.1097/ICU.0000000000000673

Thomas, P. A., Leck, A. K., and Myatt, M. (2005). Characteristic clinical features as an aid to the diagnosis of suppurative keratitis caused by filamentous fungi. *Br. J. Ophthalmol.* 89, 1554–1558. doi: 10.1136/bjo.2005.076315

Wang, D., Cui, C., Ding, X., Xiong, Z., Zheng, M., Luo, X., et al. (2019). Improving the virtual screening ability of target-specific scoring functions using deep learning methods. *Front. Pharmacol.* 10, 924. doi: 10.3389/fphar.2019.00924

Wang, L., Gao, Y., Shi, F., Li, G., Chen, K. C., Tang, Z., et al. (2016). Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Med. Phys.* 43, 336. doi: 10.1118/1.4938267

Wei, Z., Wang, S., Wang, Z., Zhang, Y., Chen, K., Gong, L., et al. (2023). Development and multi-center validation of machine learning model for early detection of fungal keratitis. *EBioMedicine* 88, 104438. doi: 10.1016/j.ebiom.2023.104438

Zhang, P., Wang, X., Chen, J., and You, W. (2017). Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload. *Sensors* 17, 2315. doi: 10.3390/s17102315

Zhang, Z., Wang, H., Wang, S., Wei, Z., Zhang, Y., Wang, Z., et al. (2022). Deep learning-based classification of infectious keratitis on slit-lamp images. *Ther. Adv. Chronic Dis.* 13, 20406223221136070. doi: 10.1177/20406223221136071

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 June 2023
DOI 10.3389/fnins.2023.1160904

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Miriam Welgampola,
The University of Sydney, Australia
Kemar E. Green,
Johns Hopkins Medicine, United States

*CORRESPONDENCE
Haibo Li
✉ haibo021@sues.edu.cn

RECEIVED 07 February 2023
ACCEPTED 22 May 2023
PUBLISHED 09 June 2023

CITATION
Li H and Yang Z (2023) Torsional nystagmus recognition based on deep learning for vertigo diagnosis.
Front. Neurosci. 17:1160904.
doi: 10.3389/fnins.2023.1160904

Torsional nystagmus recognition based on deep learning for vertigo diagnosis

Haibo Li* and Zhifan Yang

College of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China

Introduction: Detection of torsional nystagmus can help identify the canal of origin in benign paroxysmal positional vertigo (BPPV). Most currently available pupil trackers do not detect torsional nystagmus. In view of this, a new deep learning network model was designed for the determination of torsional nystagmus.

Methods: The data set comes from the Eye, Ear, Nose and Throat (Eye&ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder. The dataset contains 24521 nystagmus videos. All torsion nystagmus videos were annotated by the ophthalmologist of the hospital. 80% of the data set was used to train the model, and 20% was used to test.

Results: Experiments indicate that the designed method can effectively identify torsional nystagmus. Compared with other methods, it has high recognition accuracy. It can realize the automatic recognition of torsional nystagmus and provides support for the posterior and anterior canal BPPV diagnosis.

Discussion: Our present work complements existing methods of 2D nystagmus analysis and could improve the diagnostic capabilities of VNG in multiple vestibular disorders. To automatically pick BPV requires detection of nystagmus in all 3 planes and identification of a paroxysm. This is the next research work to be carried out.

KEYWORDS

torsional nystagmus, deep learning, classification and identification, convolution network, benign paroxysmal positional vertigo

1. Introduction

The vestibular system informs us of three-dimensional (3D) head position in space. Vestibular asymmetry creates a hallucination of head movement and therefore generates a compensatory slow phase eye movement and a quick phase that returns the eye closer to its starting position. Nystagmus is an involuntary oscillating eye movement that accompanies vestibular disorders (Leigh and Zee, 2015). The involvement of a specific vestibular end organ can be identified by the nystagmus trajectory (Jiang et al., 2018). Nystagmus can be summarized into two types: pathological nystagmus and physiological nystagmus. A variety of diseases, such as BPPV, Meniere's disease and vestibular neuritis, are all associated with pathological nystagmus (Newman et al., 2019). Pathological nystagmus arises from asymmetries in the peripheral or central vestibular system. Physiological nystagmus can be generated by rotational or thermal stimulation of the vestibular system (Henriksson, 1956; Cohen et al., 1977). BPPV is usually

accompanied by nystagmus, which provoked by changes of the head position relative to gravity (Lim et al., 2019). To diagnose different types of BPPV, clinicians inspect the directional and velocity characteristics of positional nystagmus during provocative testing. Among them, for BPPV with the highest incidence rate in the posterior semicircular canal, the typical torsional nystagmus was regarded as an important diagnostic factor. So, nystagmus examination is very important for the diagnosis of BPPV (Wang et al., 2014).

Some nystagmus can be observed by doctors with naked eyes or Frenzel googles that allow for better visualization of nystagmus at the bedside. However, this diagnostic method was easily affected by the subjective experience of doctors (Slama et al., 2017). Not all pathological nystagmus was visible to the naked eye, since visual fixation suppresses peripheral spontaneous nystagmus. The other method is objective. This method usually uses electronystagmography (ENG) or video nystagmography (VNG) to record eye movements. The ENG method (Costa et al., 1995; Cesarelli et al., 1998) places sensors around the eyelid. Benign positional nystagmus arising from stimulation stimulation of one or more semicircular canals produces horizontal, vertical and torsional eye movements in the plane of that canal. The potential difference measured by the sensors is related to the horizontal and vertical movement of eyes. The velocity and frequency of eye movements can be obtained through potential difference analysis. BPPV also demonstrates a crescendo decrescendo velocity profile, the identification of which could assist with separation of BPV from its mimics. However, this method is vulnerable to electromagnetic interference, in which case the measured information is not accurate enough. The VNG methods generally uses infrared camera to obtain nystagmus video. The frequency and amplitude of nystagmus were obtained by analyzing the motion information of pupil in video (Eggers et al., 2019). Recognizing 3D eye movement trajectory assists in identifying the canal of origin in patients with BPPV.

Under normal test conditions, VNG system includes a series of visual and dynamic function tests (Halmágyi et al., 2001; Newman-Toker et al., 2008). At present, some researchers (Buizza et al., 1978; Van Beuzekom and Van Gisbergen, 2002; Akman et al., 2006) have done some related work on how to use technical means to detect nystagmus. Most of the proposed methods can not fully recognize nystagmus automatically. Some parts or stages of these methods need human intervention, for example, a recognition method of nystagmus proposed by Buizza et al. (1978). In this method, doctors need to calibrate the direction of phase change. Akman et al. proposed a method to detect the period of nystagmus (Akman et al., 2006). The confirmation of the end point of nystagmus still needs to be further improved with this method. Van et al. proposed a nystagmus recognition method with VNG technology (Van Beuzekom and Van Gisbergen, 2002). This method requires researchers to manually confirm the two endpoints of the phase and remove interference factors such as noise from videos.

The research work stated above can be summarized as invasive and non-invasive (Newman et al., 2019). Invasive methods, such as electromagnetic coil method, mainly embed hardware equipment into human eyes, which leads to direct contact between equipment and human eyes. This will cause direct or potential harm to human eye health. The non-invasive detection methods were mainly gaze description methods based on video image processing. These methods detect and locate the pupil based on the contours of the eyes, which were greatly improved in comfort and accuracy.

The Non-invasive inspection methods can be combined with artificial intelligence (AI) methods. At present, AI technology is developing rapidly. Deep Learning has promoted the development of Computer Vision (He et al., 2016; Mane and Mangale, 2018; Kim and Ro, 2019; Cong et al., 2020), Natural Language Processing (Karpathy and Fei-Fei, 2014; Sundermeyer et al., 2015; Young et al., 2018) and other technologies. The development of deep learning technology also provides the possibility for medical intelligent aided diagnosis. For example, CT images of thoracic nodules were analyzed to determine whether there was a tumor in the chest (Anthimopoulos et al., 2016; Setio et al., 2016). Other medical applications include automatic analysis of skin disease images (Rathod et al., 2018; Wu et al., 2019), automatic analysis of fundus disease images (Ting et al., 2017; Sertkaya et al., 2019) and automatic analysis of tumor pathological sections (Tra et al., 2016; Lavanyadevi et al., 2017), etc. A variety of algorithms based on deep learning were integrated into the innovative diagnosis and treatment system (Litjens et al., 2017). For example, Google used neural network to analyze diabetic retinopathy, and its analysis results were similar to those of human experts (Gulshan et al., 2016). In other application fields, deep learning has been applied to motion detection in videos and achieved good recognition results (Saha et al., 2016). Therefore, the recognition of nystagmus can be tried by using the method of deep learning. At present, many scholars have begun to use artificial intelligence methods to identify nystagmus (Zhang et al., 2021; Lu et al., 2022; Wagle et al., 2022). From the experimental results, the deep learning method can be used to detect nystagmus, and the recognition accuracy can be further improved.

This paper mainly focuses on a torsional nystagmus recognition method based on deep learning. With the development of deep learning technology, this paper proposed an automatic recognition method of torsional nystagmus based on deep learning technology to help doctors make rapid diagnosis.

2. Materials and methods

2.1. Detail of data sources

The data set of this paper comes from Eye, Ear, Nose and Throat (Eye & ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder with the model of VertiGoggles R ZT-VNG-II, which was provided by Shanghai Zhiting Medical Technology Co., Ltd. Eye movement recorder was used to record and save the patient's eye movements video. The video format is MP4. The size of video frame is 640×480 and the frame rate is 60fps. The data set include 26,931 nystagmus videos from 1,236 patients. After removing the abnormal and disturbed data, the remaining 24,521 videos were used as the data set. The length of each nystagmus video was not required to be exactly equal. The length of video in the data set was reduced to 6–10 s. The data were from patients with BPPV. The videos were monocular, including left and right eyes. All data were annotated by four ophthalmologists according to the motion characteristics of torsional nystagmus. 80% of the data were used for training and 20% for verification.

The doctors recruited eligible subjects in the otolaryngology clinic or vestibular function examination room. For patients who complained of positional vertigo, bilateral Dix-Hallpike test was performed first, and then bilateral Roll test was performed. Each body position change was rapid, but not exceeded the patient's tolerance. In case of atypical

symptoms such as hearing loss, severe headache, limb sensation or movement disturbance, consciousness disturbance, ataxia, etc., corresponding audiological or imaging examination was carried out first to eliminate other inner ear or central lesions. After judging that the conditions for enrollment were met, the subjects themselves signed the informed consent form and collected their basic information and contact information. The doctor collected nystagmus videos of patients in the whole process of Dix-Hallpike test suspension sitting position and Epley method reduction. The Epley reposition method maintained each position for at least 30 s until the nystagmus disappears. After the restoration, the subjects rested for 15 min, and then performed Dix-Hallpike test again. The negative person indicated that the restoration was successful. If the first reset failed, the doctor performed the reset again and collected the nystagmus video of the second reset.

2.2. Network model structure and classification process

Ethical statement. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by Ethics Committee of the Eye, Ear, Nose and Throat Hospital affiliated to Fudan University (approval number: 2020518). Written informed consent was obtained from all enrolled patients.

In order to recognize nystagmus automatically by deep learning, a recognition model as shown in Figure 1 was designed in this paper. Firstly, the nystagmus video was sent to the sequence layer in the model for processing. The output video frame sequence was transmitted to the input of the sequence folding layer. Secondly, the motion characteristics of each frame in the video was extracted independently by convolution operation. Thirdly, the extracted features were restored to the sequence structure after passing through the sequence unfolding layer and flattening layer. At the same time, the output was transformed into vector sequences. Finally, the obtained vector sequences were classified by using Bi-directional Long Short-Term Memory (BiLSTM) layer and output layer. The functions of each part of the network model are introduced as follows.

2.3. Converting video into video sequence and sequence folding

Firstly, a single video was processed to obtain the relevant parameters of the video, such as the height, width, number of channels and frames of the video. Then the video was cropped. This paper adopted the longest edge of the cropped video and adjusted its size to obtain a 224 × 224 fixed size. In order to enable the feature extraction network to obtain the features of single frame, a sequence folding layer was constructed to convert sequences into images. The sequence folding layer converts a batch of image sequences into a batch of images. The sequence unfolding layer restores the sequence structure of the input data after the sequence was folded.

2.4. Feature extraction

Feature extraction was mainly completed by five modules. The first module includes convolution layers and the maximum pooling layer. Convolution layer: the kernel size is 7 × 7; the step of sliding window is 2; the number of output channels is 64. Pooling layer: the window size is 3 × 3; the step of sliding window is 2; the output channel number is 64.

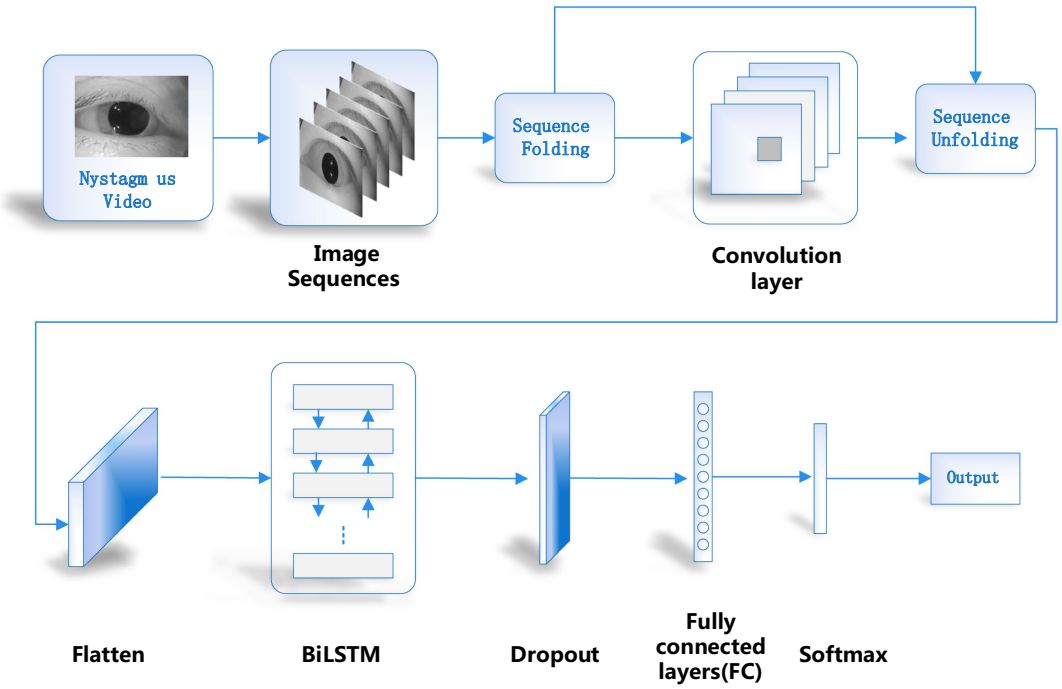


FIGURE 1
Network model structure of the proposed method.

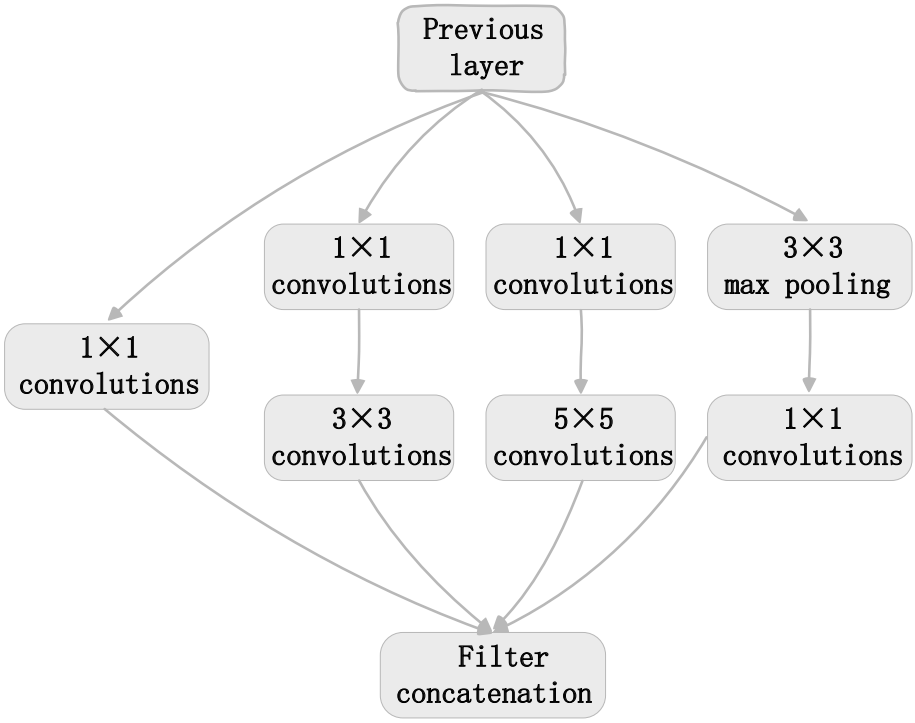


FIGURE 2
Inception structure.

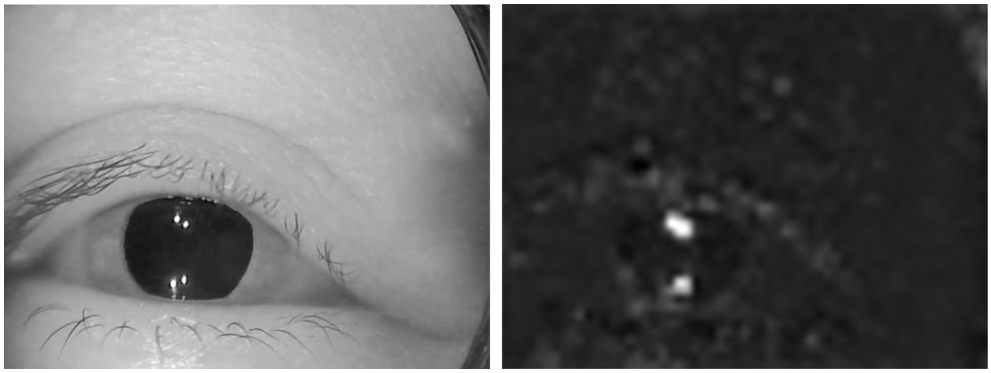


FIGURE 3
An example of extracted features. **(A)** Original image. **(B)** Extracted features.

The second module has two convolution layers and a maximum pool layer. Convolution layer: the kernel size is 3×3; the step of sliding window is 1; the output channel number is 192. Pooling layer: the window size is 3×3; the step of sliding window is 2; the output channel number is 192. The third module has two Inception modules in series, followed by a maximum pool layer. Figure 2 illustrates the structure of the Inception module. The Inception module adopts the idea of network in network (NIN). It extracts the local features of the image by using multiple convolution kernels with different scales. Each branch in the Inception module adopts 1×1 convolution kernel. It can effectively improve the receptive field of convolution kernel and reduce the dimension to accelerate the network calculation and strengthen the real-time performance. As can be seen from Figure 2, the Inception module has four main components: 1×1, 3×3, 5×5 convolution and 3×3 pooling. An example of extracted features in the four components of the inception module was shown in Figure 3. The main purpose of this structure is to extract the multi-scale information through a variety of convolution kernels of different sizes, and then fuse them, so as to have better image representation ability. In practice, using 3×3 and 5×5 convolution directly will lead to too much calculation. So, 1×1 convolution layer should be concatenated in front. The nonlinearity of the network can be increased at the same time.

The numbers of channels output by 4 lines of the first Inception are 64, 128, 32 and 32. The total number of output channels is the

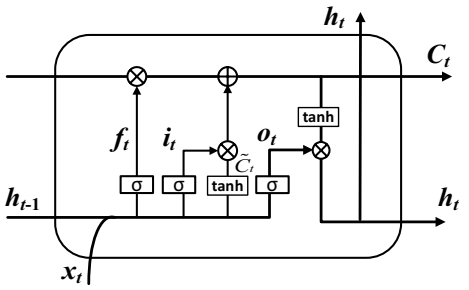


FIGURE 4
Structure of one-way branching model.

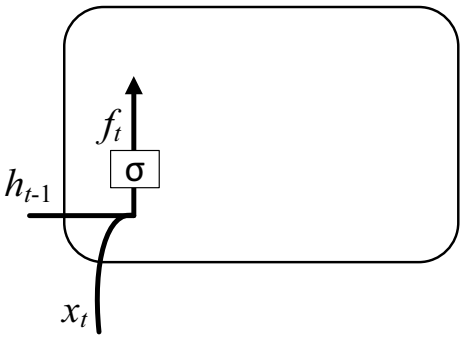


FIGURE 5
Computation of the forgetting gate.

accumulation of the four lines, which is 256. The numbers of channels output by 4 lines of the second Inception are 128, 192, 96 and 64 respectively, and the total number of output channels is 480. Pooling layer: the window size is 3×3; the step of sliding window is 2; the output channel number is 480.

The fourth module has five Inception blocks in series, followed by a maximum pool layer. The numbers of channels output by 4 lines of the first Inception are 192, 208, 48 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the second Inception are 160, 224, 64 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the third Inception are 128, 256, 64 and 64 respectively, and the total number of output channels is 512. The numbers of channels output by 4 lines of the fourth Inception are 112, 288, 64 and 64 respectively, and the total number of output channels is 528. The numbers of channels output by 4 lines of the fifth Inception are 256, 320, 128 and 128 respectively, and the total number of output channels is 832. Pooling layer: the window size is 3×3; the step of sliding window is 2; the output channel number is 832.

The fifth module has two Inception blocks in series, followed by a pooling layer. The output channel number of 4 lines are 256, 320, 128 and 128, respectively, in the first Inception, and the total number of output channels is 832. The numbers of channels output by 4 lines of the second Inception are 384, 384, 128 and 128 respectively, and the total number of output channels is 1,024. The pooling layer adopts global average pooling and the convolution layer with height and width of 1 is obtained. The number of output channels is 1,024.

2.5. Recovering sequence structure

The sequence structure was deleted by the sequence folding layer. So, the sequence structure should be restored after feature extraction. The recovery task of sequence structure was completed by sequence unfolding layer. The sequence unfolding layer takes the minibatchsize output information of the sequence folding layer as the minibatchsize input information of the sequence unfolding layer. The output of the sequence unfolding layer was reconstructed into vector sequences. The spatial dimension of the tensor was flatted to channel dimension. Flatten layer flattens input spatial dimension into a single channel. This layer retains the observation dimension (N) and sequence dimension (S) after flattening.

2.6. Sequence classification

Long Short-Term Memory (LSTM) model can record the relationship between elements in a spatial distance. This memory function can be realized by training LSTM model. But one disadvantage of LSTM model is that the order of memorizing information can only be from front to back. In order to better classify the types of nystagmus, this paper uses BiLSTM to solve this problem. BiLSTM is composed of two LSTMs with opposite directions. Figure 4 shows the structure of the one-way branching model in BiLSTM. In the figure, x_t, o_t, C_t, f_t, h_t, $\tilde{C}_t$ and i_t represent input vector, output gate, cell state, forgetting gate, hidden layer state, temporary cell state and memory gate, respectively.

The classification calculation process was completed by the following steps. Step 1: The discarded information was determined by calculating the forgetting gate. The input is the hidden layer state h_{t-1} at time t-1 and the input vector x_t at time t. The output is the value f_t of the forgetting gate at time t. As shown in Figure 5.

The input of h_{t-1} and x_t were calculated to obtain a forgetting gate output f_t through the sigmoid function, and its expression is shown in Equation (1).

$$f_t = \sigma\left(w_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

Where $f_t \in [0,1]$ (0 indicates to discard the information completely, and 1 indicates to retain the information completely); σ indicates the activation function; w_f represents a learnable connection vector; x_t is input; b_f represents the offset value.

Step 2: The retained information was determined by calculating the memory gate. The input is the hidden layer state h_{t-1} at time t-1 and the input vector x_t at time t. The output is the value i_t of the memory gate at time t and the value $\tilde{C}_t$ of the temporary cell state at time t. As shown in Figure 6. The value of the memory gate was obtained after that the hidden layer state value at time t-1 and the input vector at time t pass through the sigmoid activation function. The value of the temporary cell state was obtained after the hidden layer state value at time t-1 and the input vector at time t pass through the tanh activation function. The output values of two activation functions were multiplied to obtain the value of the input gate. The corresponding equation can be written as:

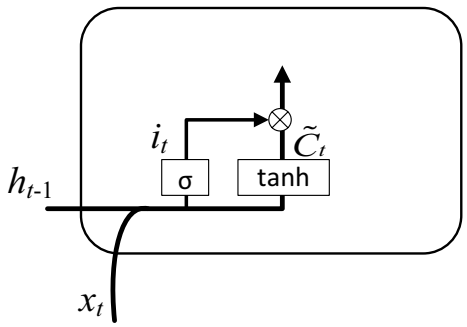


FIGURE 6
Calculation of the input gate.

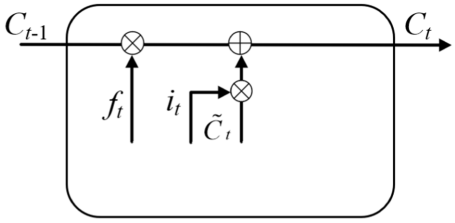


FIGURE 7
Calculation of the current cell state.

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$\tilde{C}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{3}$$

Where: *tanh* represents the activation function; W_i and W_c represent the learnable connection vectors; b_i and b_c represent the offset values.

Step 3: The cell state C_t was obtained through the joint action of forgetting gate and input gate. The input is the memory gate i_t at time t, the forgetting gate f_t at time t, the temporary cell state $\tilde{C}_t$ at time t and the cell state C_{t-1} at time t-1. The output is the cell state C_t at time t. As shown in Figure 7. The corresponding equation can be written as:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Step 4: The value of the output gate and the value of the hidden layer state were determined by calculation. The input is the hidden layer state h_{t-1} at time t-1, the input vector x_t at time t and the cell state C_t at time t. The output is the value o_t of the output gate at time t and the value h_t of the hidden layer at time t. As shown in Figure 8. The value of output gate was obtained after that the hidden layer state value at time t-1 and the input vector at time t pass through the sigmoid activation function. The value of hidden layer state was obtained after that the output gate value at time t and the cell state at time t pass through the tanh activation function. The corresponding expression can be written as:

$$o_t = \sigma\left(W_O \cdot \left[h_{t-1}, x_t\right] + b_O\right) \tag{5}$$

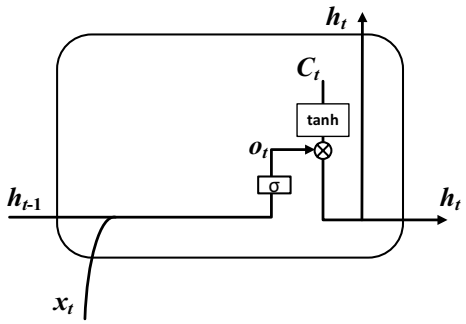


FIGURE 8
Calculation of hidden layer.

$$h_t = o_t * \tanh\left(C_t\right) \tag{6}$$

W_O and b_O represent learnable connection vectors and offset values, respectively.

Through the above steps, we can get the corresponding sequence, which is $\{h_0, h_1, \cdots, h_{n-1}\}$. BiLSTM consists of two branches in different directions mentioned above. The parameters on each branch are independent of the other branch. One branch can only fit time-related data from one direction. BiLSTM has two branches in opposite directions so that it can capture patterns that one branch may ignore. The structure can be seen in Figure 9.

If the hidden layer state sequence calculated by one branch of BiLSTM was represented by h^r, the hidden layer state sequence of the other branche in the opposite direction was represented by h^l. The final output result is as follow:

$$h_t = \alpha h^r + \beta h^l \tag{7}$$

$$y_t = \sigma\left(h_t\right) \tag{8}$$

Where α, β are constants and $\alpha + \beta = 1$. σ is the activation function.

After that the output results pass through the classification layer, the type results of nystagmus recognition can be obtained. The output layer includes dropout layer, full connection layer, softmax layer and classification layer.

3. Results

3.1. Model training and verification process

The data set of this paper comes from Eye, Ear, Nose and Throat (Eye & ENT) Hospital of Fudan University. In the process of data acquisition, the infrared videos were obtained from eye movement recorder with the model of VertiGoggles R ZT-VNG-II, which was provided by Shanghai Zhiting Medical Technology Co., Ltd. Eye movement recorder was used to record and save the patient's eye movements video. The video format is MP4. The size of video frame is 640×480 and the frame rate is 60fps. The

data set include 26,931 nystagmus videos from 1,236 patients. After removing the abnormal and disturbed data, the remaining 24,521 videos were used as the data set. All data were annotated by four ophthalmologists according to the motion characteristics of torsional nystagmus. 80% of the data were used for training and 20% for verification. The model training and verification process are shown in Figure 10.

Figure 10 shows that the classification accuracy of the training and verification process tends to be stable with the increase of iterations. The average accuracy after stabilization is shown in Table 1.

The loss during training and verification is shown in Figure 11.

As can be seen from Figure 11, with the increase of the number of iterations, the loss in the training and verification process has decreased to a stable state. In order to further evaluate the designed method. Figure 12 shows Area Under Curve (AUC).

3.2. Results of feature extraction by different methods

In addition, we also study the impact of different methods to extract video frame features on the classification effect. We use eight Fire modules to extract the video frame features, and the structure of other parts remains unchanged. This method is named method 2. The structure of Fire module (Kim and Kim, 2020) is shown in Figure 13.

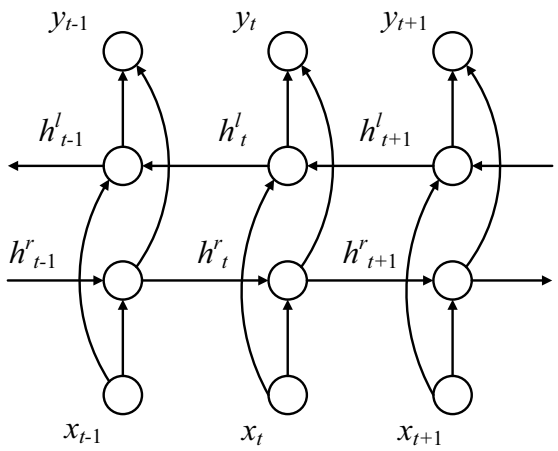


FIGURE 9
Structure of BiLSTM.

It can be seen from Figure 12 that the proposed method can identify torsional nystagmus more accurately. In addition, sensitivity⊕0.912 and specificity⊕0.946.

The same data set was used for training and verification. Figure 14 shows the classification accuracy of training and verification by method 2.

As can be seen from Figure 14, the classification accuracy tends to be stable with the increase of iterations, whether in the training process or verification process. Figure 15 shows the Loss during the training and verification process by method 2 using the same data set.

As can be seen from Figure 15, with the increase of iterations, the Loss of method 2 decreased to a stable state, whether in the training process or verification process. Method 2 was compared with the method proposed in this paper. The comparison results of classification accuracy in training set are shown in Figure 16.

As can be seen from Figure 16, the recognition accuracy of two methods tends to be stable with the increase of iterations. The average accuracy after stabilization is shown in Table 2.

It can be seen from Table 2 that the proposed method has high recognition accuracy in training process. The recognition accuracy of two methods in verification set is shown in Figure 17.

Figure 17 shows that the recognition accuracy of two methods in verification set tends to be stable with the increase of iterations. The average recognition accuracy of two methods after stabilization is shown in Table 3.

It can be seen from Table 3 that the proposed method has high recognition accuracy in verification set.

4. Discussion

The proposed method was compared with Zhang's method (Zhang et al., 2021) and Zhou's method (Zhou et al., 2022). The same data set was used for training and verification, respectively. The recognition accuracy of different methods in training set is shown in Figure 18.

As can be seen from Figure 18, the recognition accuracy of all methods in training set tends to be stable with the increase of

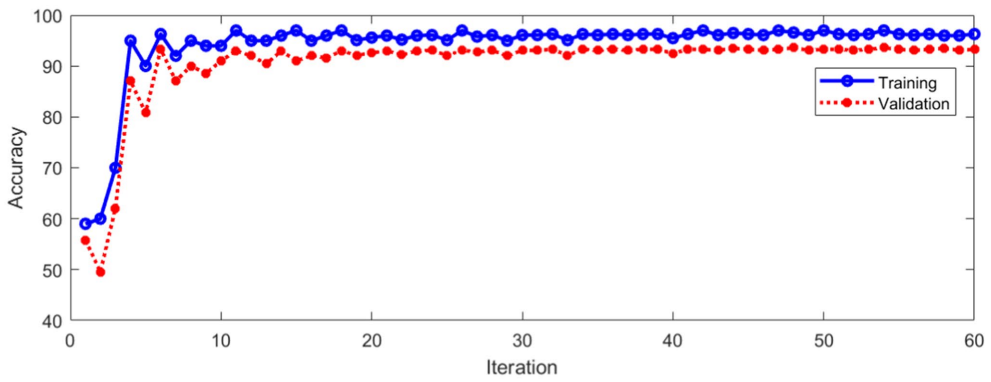


FIGURE 10
Classification accuracy of training and validation process.

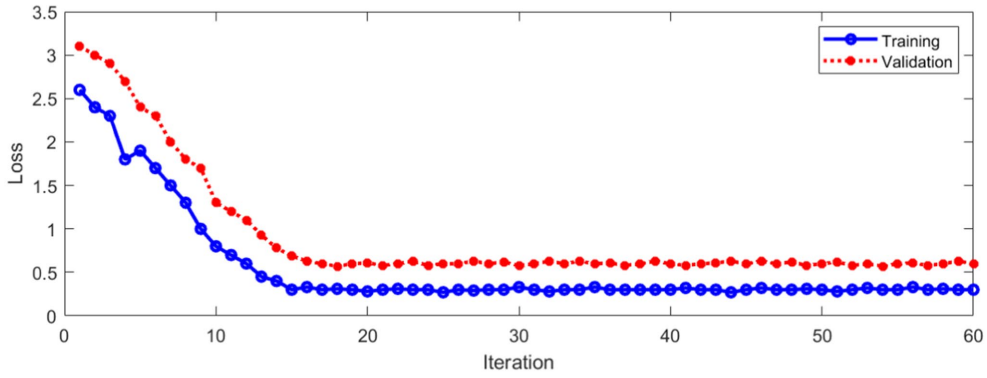


FIGURE 11
Loss during training and verification.

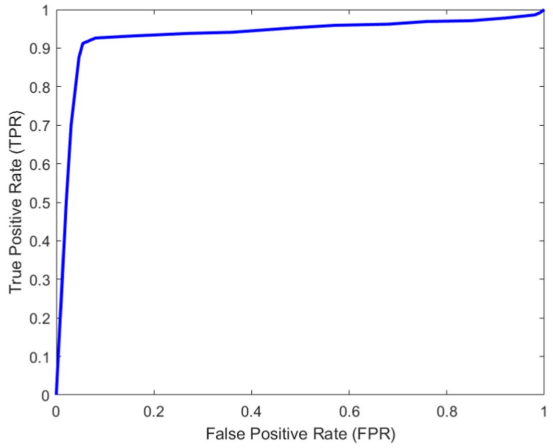


FIGURE 12
AUC curve.

TABLE 1 Recognition accuracy in training and verification stage.

Recognition accuracy	Stage
96.1%	Training
92.9%	Validation

TABLE 2 Recognition accuracy of two methods in training set.

Recognition accuracy	Method
96.1%	The proposed method
89.9%	Method 2

TABLE 3 Recognition accuracy of two methods in verification set.

Recognition accuracy	Method
92.9%	The proposed method
87.4%	Method 2

TABLE 4 Recognition accuracy of different methods in training set.

Recognition accuracy	Method
96.1%	The proposed method
91.2%	Zhang's method
93.9%	Zhou's method

TABLE 5 Recognition accuracy of different methods in verification set.

Recognition accuracy	Method
92.9%	The proposed method
89.4%	Zhang's method
90.7%	Zhou's method

TABLE 6 Statistical comparisons of the variable performance accuracy.

Method	Precision	Recall
The proposed method	94.3%	91.2%
Zhang's method	90.1%	87.6%
Zhou's method	91.9%	88.4%

iterations. The average recognition accuracy of different methods after stabilization is shown in Table 4.

It can be seen from Table 4 that the proposed method has high recognition accuracy in training set. The recognition accuracy of different methods in verification set is shown in Figure 19.

Figure 19 shows that the recognition accuracy of different methods in verification set tends to be stable with the increase of iterations. The average recognition accuracy of different methods after stabilization is shown in Table 5.

It can be seen from Table 5 that the proposed method in this paper has high recognition accuracy in verification set. This shows that the proposed method has a good effect in torsional nystagmus recognition. In addition, other statistical comparisons of the variable performance accuracy across models are shown in Table 6.

It can be seen from Table 6 that the proposed method has high precision and recall rate, which indicates that the recognition

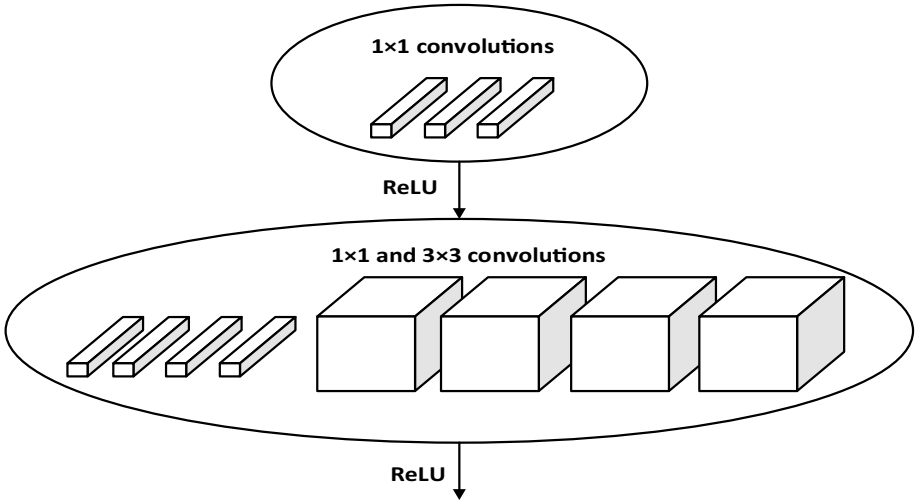


FIGURE 13
Fire module structure.

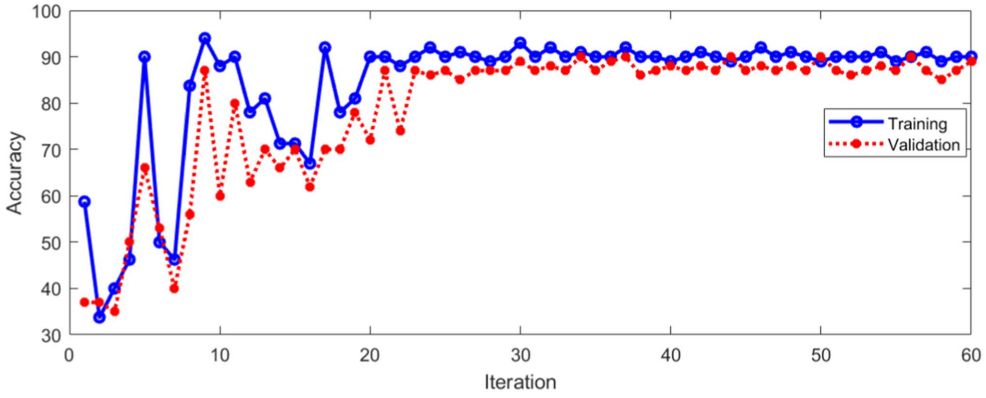


FIGURE 14
Classification accuracy of method 2.

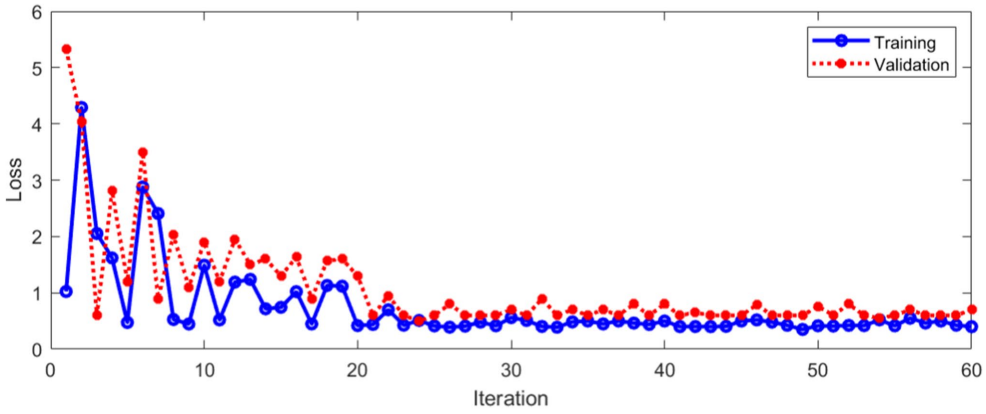


FIGURE 15
Loss during training and verification with method 2.

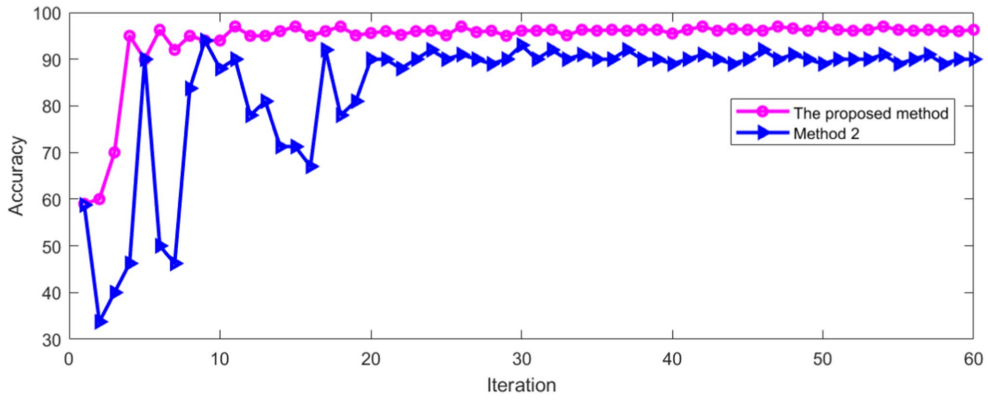


FIGURE 16
Comparison of accuracy in training set.

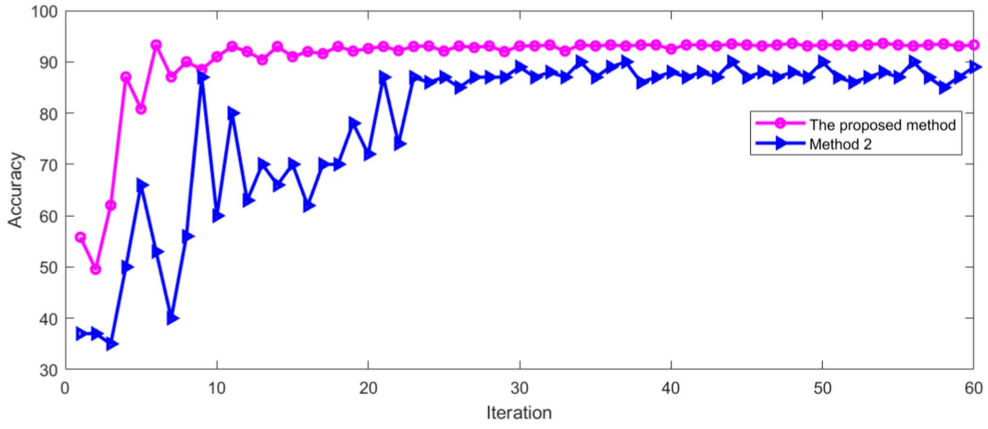


FIGURE 17
Comparison of accuracy in verification set.

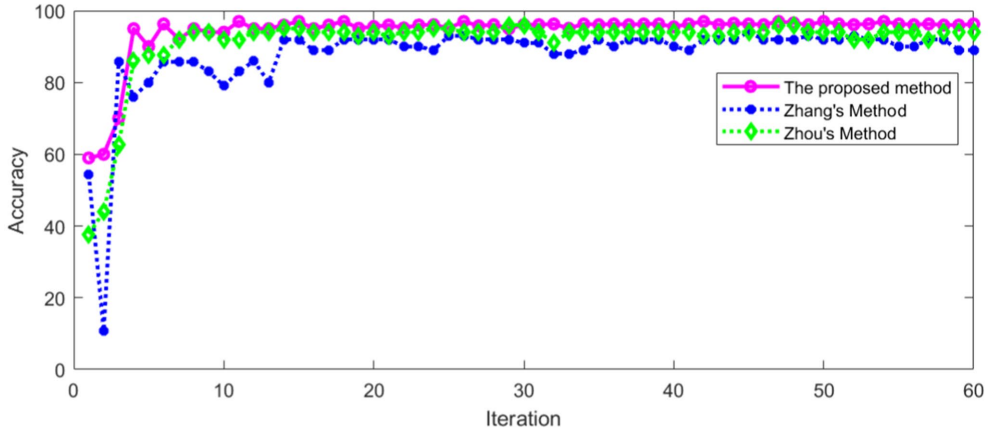


FIGURE 18
Comparison of accuracy in training set.

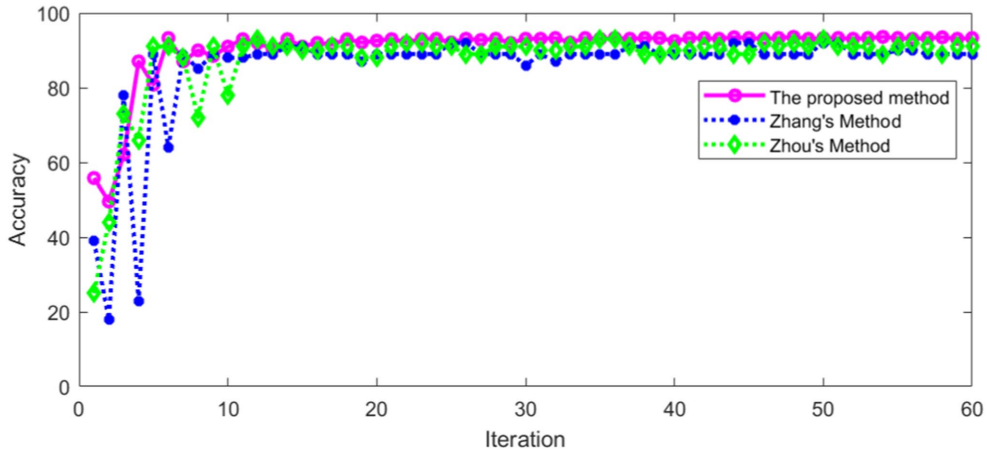


FIGURE 19
Comparison of accuracy in verification set.

performance of the algorithm is better than other methods. In addition, we also compared with the method proposed by Wagle et al. (2022), and the recognition accuracy of their method is 82.7%, which is lower than the proposed method.

Based on the real data of a large number of clinical patients, the characteristics and types of torsional nystagmus were intelligently recognized through the deep learning algorithm. The diagnosis of BPPV can be accurately predicted by combining the body position information, so as to realize the intelligent diagnosis and treatment of BPPV, improve the diagnosis efficiency and reduce the pain of patients. It is expected to comprehensively improve the diagnosis and treatment capacity of medical institutions at all levels for typical BPPV patients.

5. Conclusion

In this paper, a recognition model of torsional nystagmus was proposed based on deep learning network. From the experimental results, the nystagmus recognition model used convolution neural network to extract the frame features of the video sequence, and classified the obtained vector sequence, which can effectively identify torsional nystagmus. This shows that the recognition of torsional nystagmus can be accomplished by using deep learning network models with different structures. Although these changes in nystagmus are very complex for clinicians, they are indeed extractable features for deep learning. Once these specific nystagmus classification features are obtained, computer-aided clinical screening and classification of typical diseases can widely benefit patients with vertigo disease and help improve the diagnosis efficiency of vertigo disease. Compared with the existing methods, the proposed method further improved the recognition accuracy. In the future, we will label the slow phase velocity (SPV) of the nystagmus, so that we can analyze the performance of the model according to the SPV of the nystagmus. The development of an accurate torsion detection method has implications for correct interpretation of nystagmus overall. BPV is not the only disorder producing torsional nystagmus: stroke, vestibular migraine can present with torsional nystagmus; vestibular neuritis and Menieres disease can also generate horizontal torsional nystagmus. Our present work complements existing methods of 2D nystagmus analysis and could improve the diagnostic capabilities of VNG in multiple vestibular disorders. To automatically pick BPV requires detection of nystagmus in all 3 planes and identification of a paroxysm. This is the next research work to be carried out.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by Ethics Committee of the Eye, Ear, Nose and Throat Hospital affiliated to Fudan University (Approval number: 2020518). Written informed consent was obtained from all enrolled patients.

Author contributions

HL wrote the main manuscript. ZY prepared the Figures 16, 17. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Shanghai Hospital Development Center, grant number SHDC2020CR3050B.

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

Akman, O. E., Broomhead, D. S., Clement, R. A., and Abadi, R. V. (2006). Nonlinear time series analysis of jerk congenital nystagmus. *J. Comput. Neurosci.* 21, 153–170. doi: 10.1007/s10827-006-7816-4

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., and Mougiakakou, S. G. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Med. Imaging* 35, 1207–1216. doi: 10.1109/TMI.2016.2535865

Buizza, A., Schmid, R., Zanibelli, A., Mira, E., and Semplici, P. (1978). Quantification of vestibular nystagmus by an interactive computer program. *ORL J. Otorhinolaryngol. Relat. Spec.* 40, 147–159. doi: 10.1159/000275399

Cesarelli, M., Bifulco, P., and Loffredo, L. (1998). "EOG baseline oscillation in congenital nystagmus" in *VIII Mediterranean conference on medical biological engineering and computing-MEDICON* (Lemesos, Berlin: Springer), 14–17.

Cohen, B., Matsuo, V., and Raphan, T. (1977). Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus. *J. Physiol.* 270, 321–344. doi: 10.1113/jphysiol.1977.sp011955

Cong, W., Zhang, J., Niu, L., Liu, L., Ling, Z., Li, W., et al. (2020). "Dovenet: deep image harmonization via domain verification" in *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Seattle, WA. NJ: IEEE), 8394–8403.

Costa, M. H., Tavares, M. C., Richter, C. M., and Castagno, L. A. (1995). "Automatic analysis of electronystagmographic signals" in *38th Midwest symposium on circuits and systems. Proceedings* (Amsterdam, NJ: IEEE), 1349–1352.

Eggers, S. D., Bisdorff, A. R., Von, B. M., Zee, D. S., Kim, J., Pérez-Fernández, N., et al. (2019). Classification of vestibular signs and examination techniques: nystagmus and nystagmus-like movements. *J. Vestib. Res.* 29, 57–87. doi: 10.3233/VES-190658

Gulshan, V., Peng, L. H., Coram, M., Stumpe, M. C., Wu, D. J., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. doi: 10.1001/jama.2016.17216

Halmágyi, G. M., Aw, S. T., Cremer, P. D., Curthoys, I. S., and Todd, M. J. (2001). Impulsive testing of individual semicircular canal function. *Ann. N. Y. Acad. Sci.* 942, 192–200. doi: 10.1111/j.1749-6632.2001.tb03745.x

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition" in *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (Las Vegas, NV, NJ: IEEE), 770–778.

Henriksson, N. G. (1956). Speed of slow component and duration in caloric nystagmus. *Acta Otolaryngol.* 46, 3–29. doi: 10.3109/00016485609120817

Jiang, Y., He, J., Xiao, Y., Li, D., and Shen, Y. (2018). "Nystagmus signal feature extraction and tracking for diagnosis of the vestibular system" in *14th IEEE international conference on signal processing* (Beijing, NJ: IEEE), 308–311.

Karpathy, A., and Fei-Fei, L. (2014). Deep visual-semantic alignments for generating image descriptions. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 664–676. doi: 10.1109/TPAMI.2016.2598339

Kim, H., and Kim, G. (2020). Single image super-resolution using fire modules with asymmetric configuration. *IEEE Signal Process. Lett.* 27, 516–519. doi: 10.1109/LSP.2020.2980172

Kim, J. U., and Ro, Y. M. (2019). "Attentive layer separation for object classification and object localization in object detection" in *2019 IEEE international conference on image processing (ICIP)* (Taipei, NJ: IEEE), 3995–3999.

Lavanyadevi, R., Machakowsalya, M., Nivethitha, J., and Kumar, A. N. (2017). "Brain tumor classification and segmentation in MRI images using PNN" in *2017 IEEE international conference on electrical, instrumentation and communication engineering (ICEICE)* (Karur, NJ: IEEE), 1–6.

Leigh, RJ, and Zee, DS. *The neurology of eye movements. 5th Edn.* New York: Oxford University Press (2015).

Lim, E., Park, J. H., Jeon, H. J., Kim, H., Lee, H., Song, C., et al. (2019). Developing a diagnostic decision support system for benign paroxysmal positional vertigo using a deep-learning model. *J. Clin. Med.* 8:633. doi: 10.3390/jcm8050633

Litjens, G. J., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005

Lu, W., Li, Z., Li, Y., Li, J., Chen, Z., Feng, Y., et al. (2022). A deep learning model for three-dimensional nystagmus detection and its preliminary application. *Front. Neurosci.* 16:930028. doi: 10.3389/fnins

Mane, S., and Mangale, S. (2018). "Moving object detection and tracking using convolutional neural networks" in *2018 second international conference on intelligent computing and control systems (ICICCS)* (Madurai, India, NJ: IEEE), 1809–1813.

Newman, J. L., Phillips, J. S., Cox, S. J., Fitzgerald, J. E., and Bath, A. (2019). Automatic nystagmus detection and quantification in long-term continuous eye-movement data. *Comput. Biol. Med.* 114:103448. doi: 10.1016/j.compbiomed.2019.103448

Newman-Toker, D. E., Kattah, J. C., Alvernia, J. E., and Wang, D. Z. (2008). Normal head impulse test differentiates acute cerebellar strokes from vestibular neuritis. *Neurology* 70, 2378–2385. doi: 10.1212/01.wnl.0000314685.01433.0d

Rathod, J., Wazhmode, V., Sodha, A., and Bhavathankar, P. (2018). "Diagnosis of skin diseases using convolutional neural networks" in *2018 second international conference on electronics, communication and aerospace technology (ICECA)* (Coimbatore, NJ: IEEE), 1048–1051.

Saha, S., Singh, G., Sapienza, M., Torr, P. H., and Cuzzolin, F. (2016). Deep learning for detecting multiple space-time action tubes in videos. *arXiv preprint arXiv*:1608.01529. doi: 10.5244/C.30.58

Sertkaya, M. E., Ergen, B., and Togacar, M. (2019, 2019). "Diagnosis of eye retinal diseases based on convolutional neural networks using optical coherence images" in *23rd international conference electronics* (Palanga, NJ: IEEE), 1–5.

Setio, A. A., Ciompi, F., Litjens, G. J., Gerke, P. K., Jacobs, C., Riel, S. J., et al. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* 35, 1160–1169. doi: 10.1109/TMI.2016.2536809

Slama, A. B., Mouelhi, A., Sahli, H., Manoubi, S., Mbarek, C., Trabelsi, H., et al. (2017). A new preprocessing parameter estimation based on geodesic active contour model for automatic vestibular neuritis diagnosis. *Artif. Intell. Med.* 80, 48–62. doi: 10.1016/j.artmed.2017.07.005

Sundermeyer, M., Ney, H., and Schlüter, R. (2015). From feedforward to recurrent LSTM neural networks for language modeling. *IEEE/ACM Trans. Audio, Speech, Language Process.* 23, 517–529. doi: 10.1109/TASLP.2015.2400218

Ting, D. S., Cheung, C. Y., Lim, G., Tan, G. S., Quang, N. D., Gan, A. T., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318, 2211–2223. doi: 10.1001/jama.2017.18152

Tra, P. N. H., Hai, N. T., and Mai, T. T. (2016). "Image segmentation for detection of benign and malignant tumors" in *2016 international conference on biomedical engineering (BME-HUST)* (Hanoi, NJ: IEEE), 51–54.

Van Beuzekom, A. D., and Van Gisbergen, J. A. M. (2002). Interaction between visual and vestibular signals for the control of rapid eye movements. *J. Neurophysiol.* 88, 306–322. doi: 10.1152/jn.2002.88.1.306

Wagle, N., Morkos, J., Liu, J., Reith, H., Greenstein, J., Gong, K., et al. (2022). aEYE: a deep learning system for video nystagmus detection. *Front. Neurol.* 13:963968. doi: 10.3389/fneur.2022.963968

Wang, H., Yu, D., Song, N., Su, K., and Yin, S. (2014). Delayed diagnosis and treatment of benign paroxysmal positional vertigo associated with current practice. *Eur. Arch. Otorhinolaryngol.* 271, 261–264. doi: 10.1007/s00405-012-2333-8

Wu, Z., Zhao, S., Peng, Y., He, X., Zhao, X., Huang, K., et al. (2019). Studies on different CNN algorithms for face skin disease classification based on clinical images. *IEEE Access* 7, 66505–66511. doi: 10.1109/ACCESS.2019.2918221

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* 13, 55–75. doi: 10.1109/MCI.2018.2840738

Zhang, W., Wu, H., Liu, Y., Zheng, S., Liu, Z., Li, Y., et al. (2021). Deep learning based torsional nystagmus detection for dizziness and vertigo diagnosis. *Biomed. Signal Process. Control* 68:102616. doi: 10.1016/J.BSPC.2021.102616

Zhou, H., Zhao, X., Gao, Y., Li, H., and Cheng, R. (2022). Research on nystagmus video classification algorithm based on attention mechanism. *Laser Optoelectron. Prog.* 59:1617001. doi: 10.3788/lop202259.1617001

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/fnins.2023.1227081

OPEN ACCESS

EDITED BY
Xin Huang,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Jiaqi Wang,
Xiang Yang No. 1 People's Hospital, China
Chen-Xing Qi,
Renmin Hospital of Wuhan University, China
Yu Lin Zhong,
Jiangxi Provincial People's Hospital, China

*CORRESPONDENCE
Xiao-rong Wu
✉ wxr98021@126.com

†These authors have contributed equally to this work

RECEIVED 22 May 2023
ACCEPTED 04 July 2023
PUBLISHED 20 July 2023

CITATION
Ji Y, Wang Y-y, Cheng Q, Fu W-w, Huang S-q, Zhong P-p, Chen X-l, Shu B-l, Wei B, Huang Q-y and Wu X-r (2023) Machine learning analysis reveals aberrant dynamic changes in amplitude of low-frequency fluctuations among patients with retinal detachment. *Front. Neurosci.* 17:1227081. doi: 10.3389/fnins.2023.1227081

Machine learning analysis reveals aberrant dynamic changes in amplitude of low-frequency fluctuations among patients with retinal detachment

Yu Ji[1†], Yuan-yuan Wang[2†], Qi Cheng[1], Wen-wen Fu[1], Shui-qin Huang[1], Pei-pei Zhong[1], Xiao-lin Chen[1], Ben-liang Shu[1], Bin Wei[1], Qin-yi Huang[1] and Xiao-rong Wu[1*]

[1]Department of Ophthalmology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China, [2]Department of Radiology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

Background: There is increasing evidence that patients with retinal detachment (RD) have aberrant brain activity. However, neuroimaging investigations remain focused on static changes in brain activity among RD patients. There is limited knowledge regarding the characteristics of dynamic brain activity in RD patients.

Aim: This study evaluated changes in dynamic brain activity among RD patients, using a dynamic amplitude of low-frequency fluctuation (dALFF), k-means clustering method and support vector machine (SVM) classification approach.

Methods: We investigated inter-group disparities of dALFF indices under three different time window sizes using resting-state functional magnetic resonance imaging (rs-fMRI) data from 23 RD patients and 24 demographically matched healthy controls (HCs). The k-means clustering method was performed to analyze specific dALFF states and related temporal properties. Additionally, we selected altered dALFF values under three distinct conditions as classification features for distinguishing RD patients from HCs using an SVM classifier.

Results: RD patients exhibited dynamic changes in local intrinsic indicators of brain activity. Compared with HCs, RD patients displayed increased dALFF in the bilateral middle frontal gyrus, left putamen (Putamen_L), left superior occipital gyrus (Occipital_Sup_L), left middle occipital gyrus (Occipital_Mid_L), right calcarine (Calcarine_R), right middle temporal gyrus (Temporal_Mid_R), and right inferior frontal gyrus (Frontal_Inf_Tri_R). Additionally, RD patients showed significantly decreased dALFF values in the right superior parietal gyrus (Parietal_Sup_R) and right paracentral lobule (Paracentral_Lobule_R) [two-tailed, voxel-level $p < 0.05$, Gaussian random field (GRF) correction, cluster-level $p < 0.05$]. For dALFF, we derived 3 or 4 states of ALFF that occurred repeatedly. There were differences in state distribution and state properties between RD and HC groups. The number of transitions between the dALFF states was higher in the RD group than in the HC group. Based on dALFF values in various brain regions, the overall accuracies of SVM classification were 97.87, 100, and 93.62% under three different time windows; area under the curve values were 0.99, 1.00, and 0.95, respectively. No correlation was found between hamilton anxiety (HAMA) scores and regional dALFF.

Conclusion: Our findings offer important insights concerning the neuropathology that underlies RD and provide robust evidence that dALFF, a local indicator of brain activity, may be useful for clinical diagnosis.

KEYWORDS

retinal detachment, brain region, resting-state functional magnetic resonance imaging, dynamic amplitude of low-frequency fluctuation, sliding window, k-means clustering method, support vector machine

1. Introduction

Retinal detachment (RD) constitutes the separation of the neurosensory retina from the retinal pigment epithelium. There are multiple types of RD, among which rhegmatogenous is the most common (Steel, 2014). According to a recent report, there are 42 cases of RD per 100,000 people in Germany each year (Gerstenberger et al., 2021). In the early stage of RD, patients often experience acute-onset floaters, flashes of light, and visual field defects (Lumi et al., 2015). When the range of RD invades the macular area, vision decreases to light sensitivity or blindness. Many factors are associated with the occurrence of RD, including high myopia, eye trauma, cataract surgery, a history of retinal tears, and a family history of RD (Verhoekx et al., 2021). Because the retina and optic nerve are regarded as extensions of the central nervous system (CNS), they can be used as windows for assessment of CNS abnormalities (Vujosevic et al., 2023). Accordingly, a link may exist between RD and the CNS.

Currently, the diagnosis of RD mainly relies on optical coherence tomography and B-scan ultrasonography; the results of these examinations can help to identify the type of RD and extent of detachment (Ibrar et al., 2021). However, such examinations only explore ocular visual function in RD patients; it has been unclear whether CNS abnormalities exist in such patients. Recently, the increasing use of resting-state functional magnetic resonance imaging (rs-fMRI) to explore intrinsic brain activity has provided important information concerning the pathological mechanisms involved in RD (Huang et al., 2017; Kang et al., 2019). RD patients reportedly have aberrant functional connectivity (FC) density (Shao et al., 2021) and percent amplitude of fluctuation (Yang et al., 2021) values in various brain regions. Furthermore, Su et al. (2018) discovered that RD patients have altered FC in their default mode network. Thus far, research has mainly focused on changes in the static brain activity of RD patients; there has been a belief that functional interactions among brain regions remain unchanged in time during the whole MRI scan, which is obviously not objective. Chang and Glover (2010) found that when the relationships of the posterior cingulate gyrus were measured over time, FC differed throughout the brain, indicating that relationships among brain regions dynamically fluctuate over time. Since then, there has been increasing evidence that brain activity characteristics exhibit dynamic temporal variation (Sporns, 2011; Abrams et al., 2013). Accordingly, we presumed that analyses of dynamic brain activity would provide insights concerning altered neural mechanisms in RD patients.

The amplitude of low-frequency fluctuations (ALFF) method is useful for measurements of local brain activity. Previous studies have shown that low-frequency oscillations (<0.08 Hz) of blood oxygen level-related (BOLD) signals in the human brain are physiologically significant; such oscillations may represent spontaneous local neural activity (Biswal et al., 1995; Raichle, 2011; Gehrig et al., 2019). Zang et al. (2007) developed the ALFF index and used it to explore the regional intensity of spontaneous fluctuations in BOLD signals. Because the ALFF is calculated under the assumption that the data display temporal stability throughout the acquisition period, it excludes temporal variation in BOLD signals during fMRI scanning. Dynamic ALFF (dALFF) offers a new approach to dynamic brain activity analysis that involves a combination of ALFF and sliding-window methodologies. The dALFF analysis technique has been successfully used to evaluate dynamic changes in brain activity among patients with diabetic retinopathy (Huang et al., 2021), primary dysmenorrhea involving chronic menstrual pain (Gui et al., 2021), and transient ischemic attack (Ma et al., 2021). Additionally, k-means clustering method can cluster the dALFF values of all subjects under different sliding time Windows into several states, so as to better describe the working mode of human brain during the whole scanning time process. Finally, the support-vector machine (SVM) is a supervised machine learning technique that seeks to maximize the margin to sort input points into classes in a high-dimensional space (Pereira et al., 2009). The method of combining SVM and dALFF to analyze changes in brain activity in various diseases has been a research hotspot in recent years, such as comitant exotropia (Chen et al., 2022) and active thyroid-associated ophthalmopathy (Wen et al., 2023) and SVM has high accuracy in distinguishing patients from healthy populations. In the present study, we tested two hypotheses: (1) RD patients exhibit greater temporal variability compared with healthy controls (HCs); and (2) dALFF values are sensitive biomarkers that can distinguish RD patients from HCs.

2. Participants and methods

2.1. Participants

From January 2023 to April 2023, 23 RD patients and 24 HCs were enrolled in this study. All participants were matched for age and sex; they all were examined in the same clinic and provided written informed consent to participate in the study. All experimental procedures were conducted in accordance with the Declaration of Helsinki, and the study protocol was approved by the Medical Ethics Committee of the First Affiliated Hospital of Nanchang University (Jiangxi Province, China).

The inclusion criteria for RD patients were (1) idiopathic RD involving one or two retinal tears, (2) RD affecting one or two quadrants, and (3) absence of any ocular illness (e.g., cataracts, glaucoma, optic neuritis, or maculopathy) in both eyes. The exclusion criteria for RD patients were (1) recurrent RD or recurrence after RD repair surgery, (2)RD caused by high myopia, (3) ocular trauma-related RD, (4) serious complications associated

with RD (e.g., proliferative vitreoretinopathy, vitreous hemorrhage, or macular degeneration), (5) a history of laser treatment or surgery, (6) cardiovascular diseases (e.g., heart disease or hypertension), and (7) psychiatric disorders and cerebral infarction.

According to age, sex, and educational background, HCs were chosen at random from Nanchang City. The inclusion criteria for HCs were the absence of eye diseases and major illnesses (e.g., neurological illness or cerebral infarction); the presence of uncorrected vision or visual acuity better than 1.0; and the completion of MRI-related tests, optical coherence tomography, ultrasonography, and other ophthalmic examinations.

2.2. fMRI data acquisition

Rs-fMRI data were collected at the Department of Radiology in the First Affiliated Hospital, Nanchang University, China, using a 3 T MR scanner (Siemens, Erlangen, Germany) equipped with an 8-channel phased-array head coil. The following parameters were used to capture 240 resting-state volumes over 8-min: field of view, 240 mm × 240 mm; repetition time, 2,000 ms; echo time, 40 ms; flip angle, 90°; matrix, 64 × 64; slice thickness, 4 mm; and gap, 1 mm. Thirty axial slices were included in each brain volume. The following three-dimensional MRI parameters were used to acquire high-resolution T1-weighted images of each participant: repetition time, 1,900 ms; echo time, 2.26 ms; flip angle, 9°; field of view, 240 mm × 240 mm; matrix, 256 × 256; number of sagittal slices, 176; and slice thickness, 1 mm.

2.3. fMRI data preprocessing

All data preprocessing was conducted using SPM12 and RESTplus (Jia et al., 2019) version 1.25 running in matlab2017b. The data preprocessing steps were as follows: (1) Data collation and classification. (2) Conversion of file format from DICOM to NIFTI. (3) Removal of the first 10 time points. (4) Slice timing correction. (5) Head movement correction. (6) Normalization (standardization of individual space to Montreal Neurological Institute [MNI] standard space). (7) Spatial smoothing. (8) Detrending. (9) Regression of nuisance covariates.

2.4. dALFF variance computing

The Time Dynamic Analysis toolbox in RESTplus version 1.25 was used to calculate dynamic metrics. Appropriate window length is essential for dynamic analysis, and sliding windows have key roles in the assessment of dynamic spontaneous brain activity. Previous studies showed that an excessively short window length can increase dALFF signal instability, whereas an excessively long window length does not adequately reflect dynamic temporal changes in dALFF (Leonardi and Van De Ville, 2015; Li et al., 2019). To minimize subjective error caused by a single window length, we used a window length and step size of 1TR for 30TR (60 s), 50TR (100 s), and 80TR (160 s) to calculate the dALFF for each participant. For each participant's window-based ALFF map, we calculated the mean and standard deviation of each voxel, then determined the appropriate coefficient of variation (CV = standard deviation/mean). Further statistical analyses were conducted using the CV maps.

2.5. Statistical analysis

One-sample t-tests were used for statistical analyses of the CVALFF maps of RD patients and HCs. Two-sample *t*-tests were used to assess differences in CVALFF maps between groups. The Gaussian random field (GRF) method was used to correct for multiple comparisons and regressed covariates of age and sex (two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

2.6. Clustering analysis

To determine the dALFF occurrence state, a k-means method was used to the dALFF values for each participant. The k-means algorithm combines together data that is related in "k" ways, ensuring that the total of squares within clusters is as small as possible (Zhang et al., 2018). The Manhattan (L1) distance function method was performed to assess the reoccurrence over time in patterns of ALFF. All of the dALFF windows were clustered using the clustering centroids for the departure points.

2.7. Support vector machine analysis

To determine whether changes in dynamic metrics can be used as diagnostic indicators of RD, we investigated possible diagnostic indices using Gaussian radial basis function kernel SVMs and the LIBSVM software package (Pereira et al., 2009). The steps were as follows: (1) region of interest signal values were extracted from all differential brain regions; (2) a NII file was created to mask differential brain regions; and (3) the radial basis function of the grid search optimization algorithm was used to calculate parameters.

3. Results

3.1. Demographic characteristics

This study included 23 RD patients (12 men and 11 women; mean age, 51.70 ± 19.37 years) and 24 HCs (11 men and 13 women; mean age, 50.46 ± 14.55 years). Demographic characteristics are shown in Table 1.

3.2. Differences in dALFF values

Figure 1 shows the spatial distribution of dALFF values between RD patients and HCs at a sliding window size of 30TR. In the bilateral middle frontal gyrus, left putamen, left superior occipital gyrus, and right calcarine, dALFF values were significantly higher in RD patients than in HCs (Figure 1; Table 2; two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

Figure 2 shows the spatial distribution of dALFF values between RD patients and HCs at a sliding window size of 50TR. In the right middle temporal gyrus, left middle frontal gyrus, right calcarine, left putamen, left middle occipital gyrus, and right inferior frontal gyrus, dALFF values were significantly higher in RD patients than in HCs. Conversely, in the right superior parietal gyrus, dALFF values were significantly lower in RD patients than in HCs (Figure 2; Table 3; two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

TABLE 1 Demographic characteristics of RD patients and HCs.

Characteristic	RD patients	HCs	*p*-value
Men/women	12/11	11/13	0.664[x]
Age (years, mean ± SD)	51.70 ± 19.37	50.46 ± 14.55	0.225[t]
Duration of detachment (days)	15 (7, 90)[a]	N/A	N/A
IOP (mmHg, mean ± SD)	14.52 ± 4.62	N/A	N/A
Axial length of eye (mm, mean ± SD)	24.36 ± 1.88	N/A	N/A
Corneal endothelial cell count (mm^2, mean ± SD)	2,317 ± 512.24	N/A	N/A
HAMA score	3.89 (2, 5)[a]	N/A	N/A

[x]Data were obtained using Pearson's Chi-square tests; [t]Data were obtained using two-sample *t*-tests; [a]Median (interquartile range); SD, standard deviation; RD, retinal detachment; HCs, healthy controls; IOP, intraocular pressure; HAMA, Hamilton Anxiety Scale; N/A, not applicable.

Figure 3 shows the spatial distribution of dALFF values between RD patients and HCs at a sliding window size of 80TR. In the left middle frontal gyrus, dALFF values were significantly higher in RD patients than in HCs. Conversely, in the right paracentral lobule, dALFF values were significantly lower in RD patients than in HCs (Figure 3; Table 4; two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

3.3. Clustered dALFF states

Figure 4 depicts the overall state transition mode of all participants at sliding window sizes of 30TR, 50TR, and 80TR. The frequency and mean dwell time are shown for different states, along with the probability of transition between states.

3.4. Mean weighted dALFF values

The mean altered dALFF values between RD patients and HCs at sliding window sizes of 30TR, 50TR, and 80TR are shown in Figure 5.

3.5. SVM classification results

Figure 6A shows that the total accuracy and area under the curve score of SVM classification, based on dALFF with a sliding window size

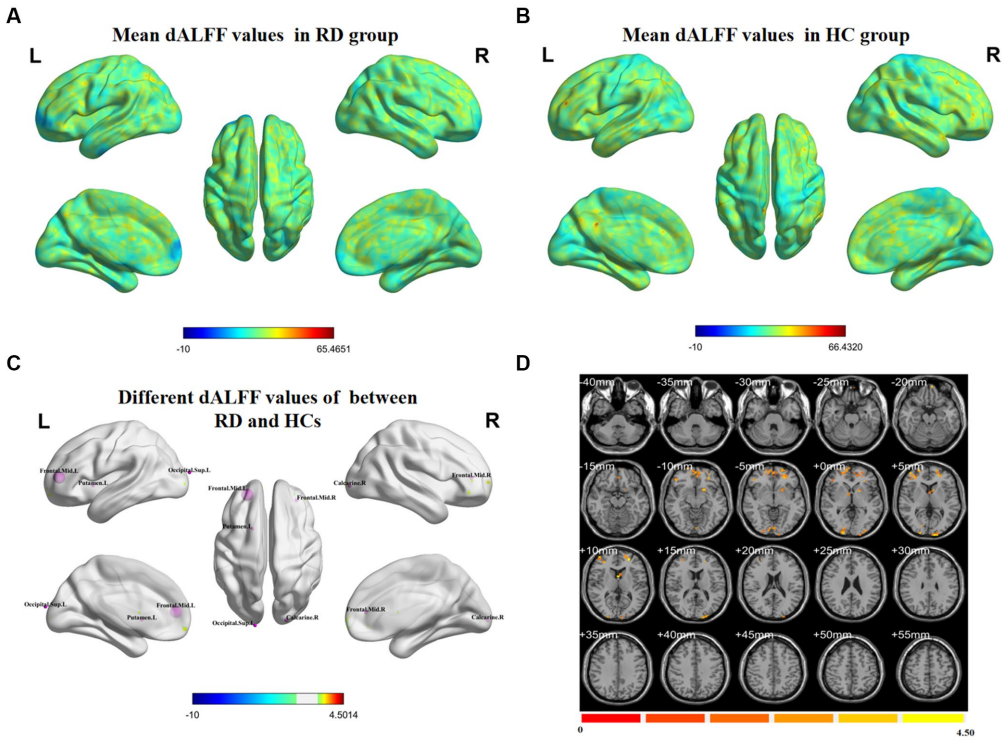


FIGURE 1
Spatial distributions of dALFF values in **(A)** RD patients (one-sample t-test) and **(B)** HCs (one-sample *t*-test). **(C,D)** Spatial distributions of dALFF values between RD patients and HCs (two-sample *t*-test). HCs, healthy controls; RD, retinal detachment; dALFF, dynamic amplitude of low-frequency fluctuation; L, left; R, right (two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

10.3389/fnins.2023.1227081

of 30TR, were 97.87% and 0.99, respectively. Figure 6B shows that the total accuracy and area under the curve score of SVM classification, based on dALFF with a sliding window size of 50TR, were 100% and 1.00, respectively. Figure 6C shows that the total accuracy and area under the curve score of SVM classification, based on dALFF with a sliding window size of 80TR, were 93.62% and 0.95, respectively. These results indicate that the dALFF may be useful in the clinical diagnosis of RD.

TABLE 2 Significant differences in dALFF values between RD patients and HCs at a sliding window size of 30TR.

Brain region	BA	Peak *t*-score	MNI coordinates (x, y, z)	Cluster size (voxels)
Frontal_Mid_L	–	4.5014	−24, 48, 9	232
Putamen_L	–	4.1322	−18, 9, 0	59
Frontal_Mid_R	–	3.5413	33, 42, 9	63
Occipital_Sup_L	–	4.0016	−15, −102, 15	56
Calcarine_R	–	3.595	21, −96, 0	67

dALFF, dynamic amplitude of low-frequency fluctuation; HCs, healthy controls; RD, retinal detachment; BA, Brodmann area; MNI, Montreal Neurological Institute; Frontal_Mid_L, left middle frontal gyrus; Putamen_L, left putamen; Frontal_Mid_R, right middle frontal gyrus; Occipital_Sup_L, left superior occipital gyrus; Calcarine_R, right calcarine.

3.6. Correlation analysis

Pearson and Spearman correlation analyses revealed that region of interest values in positive brain regions identified in RD patients at sliding window sizes of 30TR, 50TR, and 80TR were not correlated with HAMA scores (Table 5).

4. Discussion

This study showed that RD patients exhibited increased dALFF values in the bilateral middle frontal gyrus, left putamen, left superior occipital gyrus, left middle occipital gyrus, right calcarine, right middle temporal gyrus, and right inferior frontal gyrus. Additionally, RD patients showed significantly decreased dALFF values in the right superior parietal gyrus and right paracentral lobule (two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$). Using k-means clustering, three or four dALFF states were identified among all subjects. There were differences in state distribution and state properties between RD and HC groups. The number of transitions between the dALFF states was higher in the RD group than in the HC group. Based on dALFF values in various brain regions, the overall accuracies of SVM classification were 97.87, 100, and 93.62% under three different time windows; area under the curve values were 0.99, 1.00, and 0.95, respectively. No correlation was found between HAMA scores and regional dALFF.

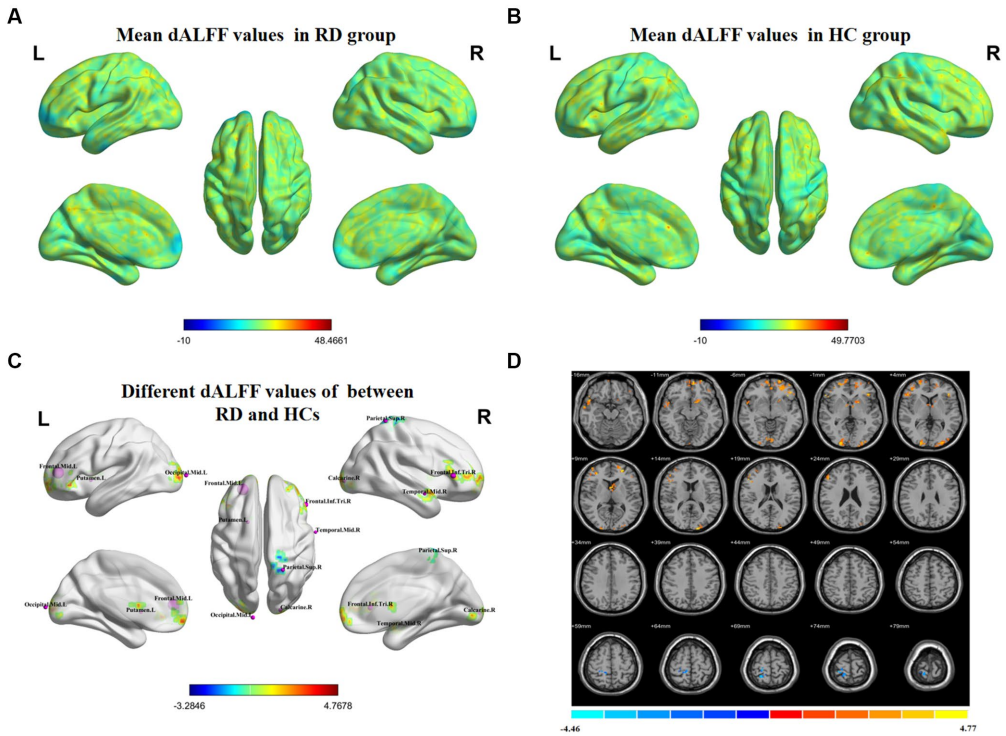


FIGURE 2
Spatial distributions of dALFF values in **(A)** RD patients (one-sample t-test) and **(B)** HCs (one-sample *t*-test). **(C,D)** Spatial distributions of dALFF values between RD patients and HCs (two-sample *t*-test). HCs, healthy controls; RD, retinal detachment; dALFF, dynamic amplitude of low-frequency fluctuation; L, left; R, right (two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

4.1. Differences in dALFF variability

The frontal lobe, which constitutes approximately one-third of the cerebral cortex in humans, is a region of the brain that develops more slowly than other regions. Its wide-ranging and complex functions encompass nearly all cognitive neuropsychological activities. In this study, we found that the RD group had significantly higher dALFF values in the bilateral middle frontal gyrus and right inferior frontal gyrus, which are important for high-level cognition (e.g., executive function and working memory; Zheng et al., 2015). The decreased central vision, visual distortion, and reduced field of vision in RD patients may cause some motor execution impairment and poor memory. Kawashima et al. (2021) found that functional activation of the bilateral dorsolateral prefrontal cortex and middle frontal gyrus was reduced in Parkinson's disease patients, presumably in relation to the pathophysiology of working memory disorder in such patients. Through voxel-based morphometry analysis, Zhao et al. (2021) demonstrated that gray matter volume was decreased in the bilateral middle frontal gyrus of patients with major depressive disorder, which may influence executive function in those patients. Chen et al. (2017) also revealed that FC between the right dorsolateral prefrontal cortex and right hippocampus was significantly reduced in long-term breast cancer survivors who had been treated with tamoxifen, implying significant defects in working memory and overall executive function. The above studies indicated that dALFF values in the bilateral middle frontal gyrus and right inferior frontal gyrus are significantly higher in RD patients than in HCs. Accordingly, we speculate that RD patients display motor execution impairment and poor working memory. In such patients, motor function and working memory are maintained before visual decline through compensatory mechanisms involving the bilateral middle frontal gyrus and right inferior frontal gyrus.

TABLE 3 Significant differences in dALFF values between RD patients and HCs at a sliding window size of 50TR.

Brain region	BA	Peak t-score	MNI coordinates (x, y, z)	Cluster size (voxels)
Temporal_Mid_R	21	3.8444	63, −3, −15	52
Frontal_Mid_L	–	4.7678	−24, 48, 9	249
Calcarine_R	–	4.1172	21, −96, 0	73
Putamen_L	–	3.7194	−18, 9, 0	58
Occipital_Mid_L	–	4.4051	−12, −105, 6	73
Frontal_Inf_Tri_R	–	3.5666	51, 30, 6	101
Parietal_Sup_R	5	−3.2846	24, −48, 69	69

dALFF, dynamic amplitude of low-frequency fluctuation; HCs, healthy controls; RD, retinal detachment; BA, Brodmann area; MNI, Montreal Neurological Institute; Temporal_Mid_R, right middle temporal gyrus; Frontal_Mid_L, left middle frontal gyrus; Calcarine_R, right calcarine; Putamen_L, left putamen; Occipital_Mid_L, left middle occipital gyrus; Frontal_Inf_Tri_R, right inferior frontal gyrus; Parietal_Sup_R, right superior parietal gyrus.

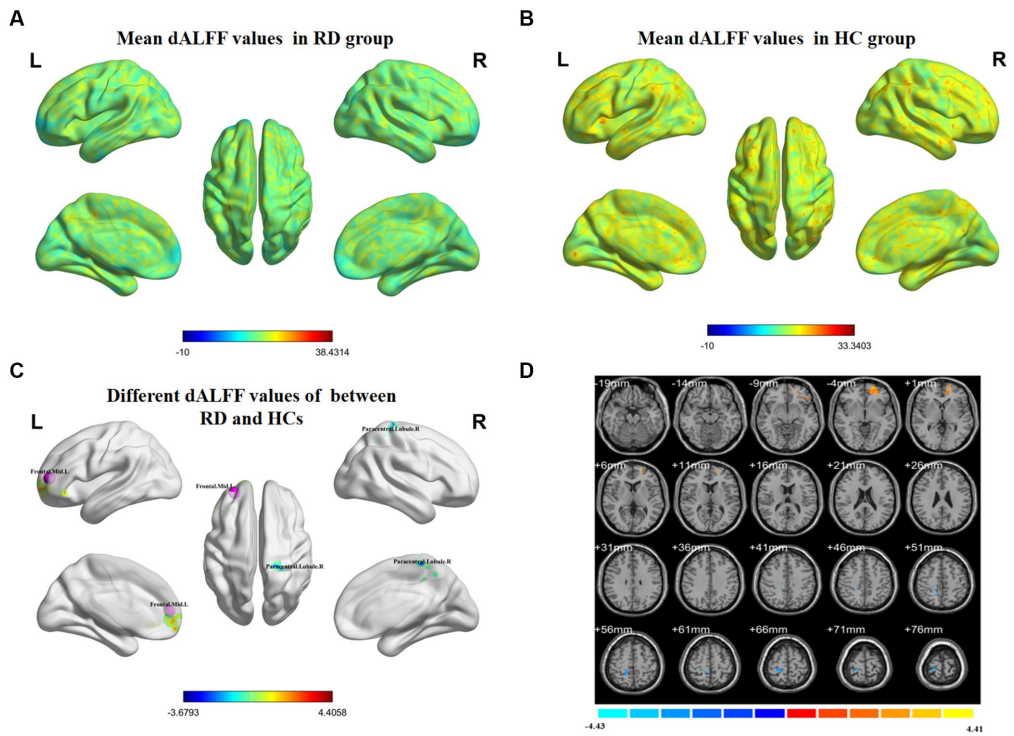


FIGURE 3
Spatial distributions of dALFF values in **(A)** RD patients (one-sample t-test) and **(B)** HCs (one-sample *t*-test). **(C,D)** Spatial distributions of dALFF values between RD patients and HCs (two-sample *t*-test). HCs, healthy controls; RD, retinal detachment; dALFF, dynamic amplitude of low-frequency fluctuation; L, left; R, right (two-tailed, voxel-level $p < 0.05$; GRF correction, cluster-level $p < 0.05$).

TABLE 4 Significant differences in dALFF values between RD patients and HCs at a sliding window size of 80TR.

Brain region	BA	Peak t-score	MNI coordinates (x, y, z)	Cluster size (voxels)
Frontal_ Mid_L	–	4.4058	−27, 51, 9	133
Paracentral_ Lobule_R	–	−3.6793	9, −36, 63	74

dALFF, dynamic amplitude of low-frequency fluctuation; HCs, healthy controls; RD, retinal detachment; BA, Brodmann area; MNI, Montreal Neurological Institute; Frontal_Mid_L, left middle frontal gyrus; Paracentral_Lobule_R, right paracentral lobule.

The putamen, a key component of the basal ganglia, is important for motor regulation (Romero et al., 2008; Vicente et al., 2012). We observed increased dALFF values in the left putamen of RD patients. Huang et al. (2018) found that the gray matter volume in the left putamen was increased among high myopia (HM) patients, suggesting that HM causes structural alterations in the bilateral putamen; this finding is consistent with the compensatory motor function observed in HM patients. Tong et al. (2021) revealed an increase in regional homogeneity in the left putamen among patients with iridocyclitis; they speculated that iridocyclitis causes functional changes in the putamen, which may lead to compensatory motor function. Hu et al. (2022) also demonstrated that voxel-based

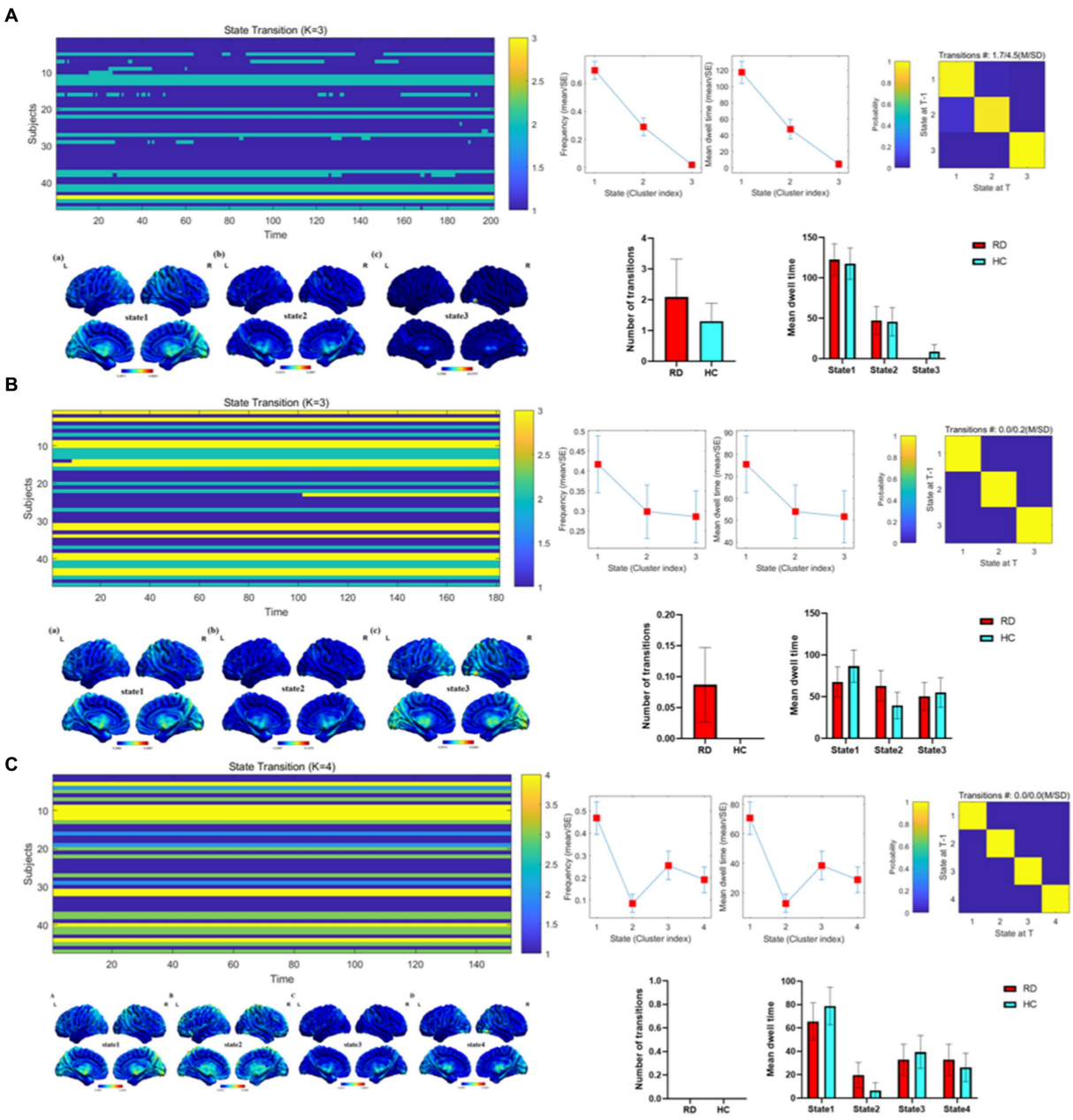


FIGURE 4
Temporal properties of dALFF patterns between RD and HC groups at sliding window sizes of 30TR **(A)**, 50TR **(B)**, and 80TR **(C)**.

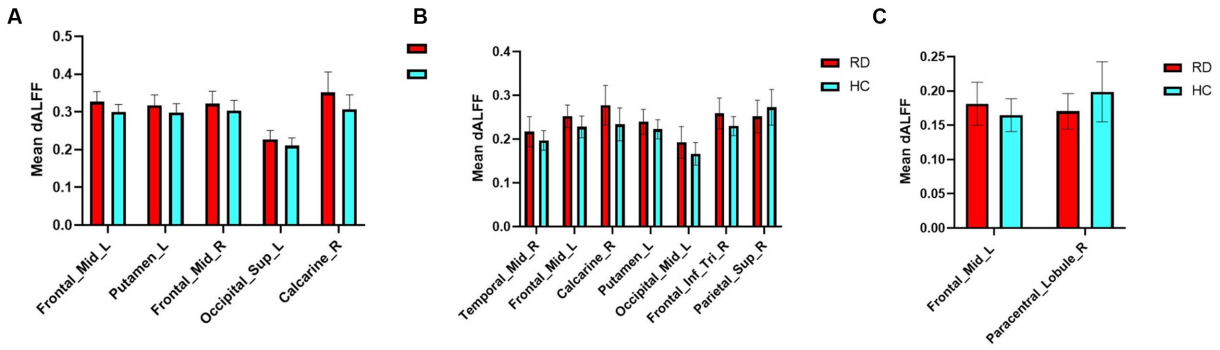


FIGURE 5
Mean weighted dALFF values of RD patients and HCs in altered brain regions at sliding window sizes of 30TR **(A)**, 50TR **(B)**, and 80TR **(C)**.

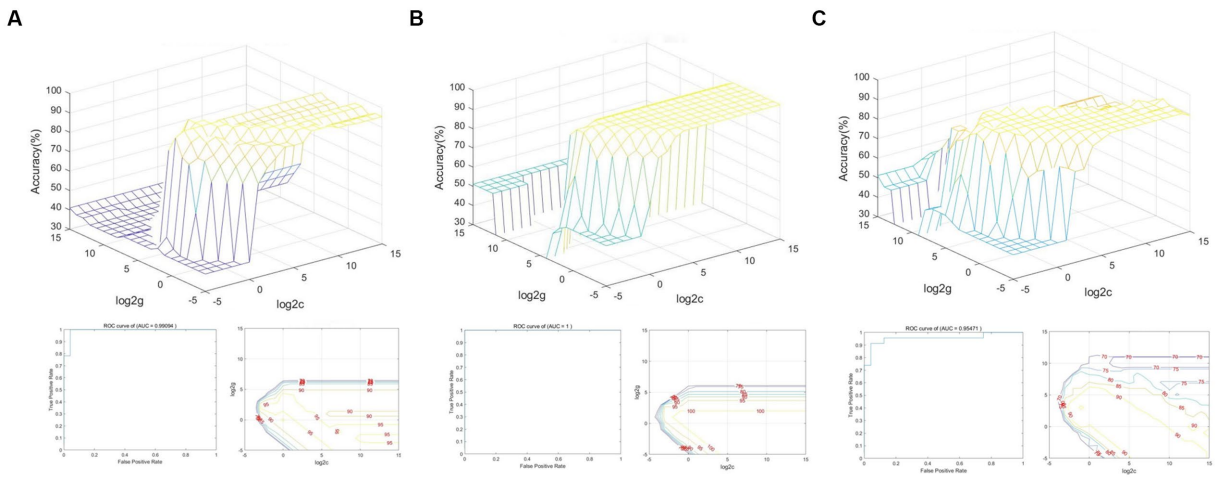


FIGURE 6
SVM classification of RD patients and HCs based on altered brain regions identified using sliding window sizes of 30TR **(A)**, 50TR **(B)**, and 80TR **(C)**.

morphometry values were decreased in the left putamen of female menopausal dry eye patients, indicating that these patients may exhibit cognitive or motor impairments. Because RD causes sudden vision loss, patients cannot rapidly adapt to monocular vision, and their motor regulation is partially decreased relative to the pre-injury state; this phenomenon may also explain a portion of the increase in dALFF in the left putamen of RD patients. Therefore, we suspect that the increased dALFF in the left putamen compensates for the decrease in motor regulation among RD patients.

The occipital lobe, located in the posterior cerebral cortex, is responsible for visual perception. In this study, we found that dALFF values were increased in the left superior occipital gyrus, left middle occipital gyrus, and right calcarine in RD patients. Shao et al. (2021) reported increased FC density values in the left inferior occipital gyrus of RD patients, which may influence the brain's effectiveness and accuracy in terms of processing visual digital information. Huang et al. (2017) also demonstrated that RD patients had decreased regional homogeneity in the right occipital lobe; they speculated that the decreased regional homogeneity reflected diminished synchrony among local brain regions, consistent with altered function in the primary visual cortex of RD patients. Wu et al. (2021) revealed that voxel-mirrored homotopic connectivity (VMHC) values in the bilateral calcarine were lower in bronchial asthma patients than in HCs. They suggested that the reduced VMHC values represent aberrant visual network function in asthma patients, leading to changes in visual function. The calcarine divides the occipital lobe into the cuneus above and lingual gyrus below; the primary visual cortex is located on both sides of the calcarine. Because the detached portion of the retina in RD patients cannot perceive light stimuli, the occipital lobe receives weaker visual signals, which may explain the increased dALFF values in the left superior occipital gyrus, left middle occipital gyrus, and right calcarine in RD patients. We suspect that these elevated dALFF values represent a compensatory mechanism by which the brain attempts to cope with vision loss in RD patients.

The temporal lobe, located below the lateral fissure, is divided into the superior temporal, middle temporal, and inferior temporal gyri; the inferior temporal gyrus is mainly involved in language comprehension (Dronkers et al., 2004). We found that dALFF values in the right middle temporal gyrus were increased in RD patients. Yu et al. (2008) reported that FC between the left superior temporal gyrus and middle temporal gyrus was decreased in patients with early blindness, whereas Huang et al. (2019) showed that degree centrality

TABLE 5 Correlations between HAMA scores and brain regions with significant differences in RD patients at sliding window sizes of 30TR, 50TR, and 80TR.

Sliding window size	Brain region	Normality test (p-value)	Pearson correlation analysis (p-value)	Spearman correlation analysis (p-value)
30TR	Frontal_Mid_L	0.016	–	0.572
	Putamen_L	0.795	0.696	–
	Frontal_Mid_R	0.006	–	0.360
	Occipital_Sup_L	0.893	0.735	–
	Calcarine_R	0.084	0.821	–
50TR	Temporal_Mid_R	0.392	0.887	–
	Frontal_Mid_L	0.334	0.870	–
	Calcarine_R	0.258	0.673	–
	Putamen_L	0.898	0.961	–
	Occipital_Mid_L	0.020	–	0.683
	Frontal_Inf_Tri_R	0.009	–	0.734
	Parietal_Sup_R	0.185	0.407	–
80TR	Frontal_Mid_L	0.564	0.354	–
	Paracentral_Lobule_R	0.071	0.671	–

(DC) in the left inferior temporal gyrus was increased in patients with advanced monocular blindness. They speculated that the increase in DC compensated for vision loss in patients with advanced monocular blindness. Qi et al. (2022) showed that VMHC values in the bilateral medial temporal gyrus were decreased in patients with thyroid-associated ophthalmopathy, which may reflect diminished visual processing and attention in such patients. To our knowledge, there have been no reports of language comprehension problems among RD patients. We suspect that RD patients undergo a long period of visual improvement from the initial detachment until postoperative recovery; they may experience some reductions in the ability to learn and perceive external things, which could affect language comprehension. Thus, the elevated dALFF value in the right temporal gyrus may represent a compensatory mechanism for the decrease in language comprehension.

Finally, our study showed that dALFF values were decreased in the right superior parietal gyrus and right paracentral lobule in RD patients. The parietal gyrus, located above the medial parietal sulci, is involved in the transmission of visual information and the integration of visual movement (Caminiti et al., 1996; Iacoboni and Zaidel, 2004). Tan et al. (2018) found that the DC of the right superior parietal gyrus was decreased in comitant exotropia strabismus patients; this decreased DC may reflect functional impairment of the right superior parietal gyrus, which would explain eye movement dysfunction in such patients. In a previous study, we found that FC between the left V1 and L-SPG was increased in HM patients; we speculated that this increase in FC was a compensatory response to prevent impaired top-down control of visual attention in HM patients (Ji et al., 2023). Considering the previous findings, the decreased dALFF values in the right superior parietal gyrus of RD patients may reflect functional impairment in this brain area, which would explain why RD patients experience visual impairment. Moreover, the paracentral lobule extends from the lateral surface of the anterior and posterior central dorsal gyrus to the medial surface, which is closely associated with cognitive impairment (Mascalchi et al., 2014). Kim et al. (2019) found that changes in subnetworks, such as the paracentral lobule, were associated with cognitive scores in patients with subjective cognitive decline. However, Liang et al. (2020) showed that gray matter volume in the paracentral lobule was increased in patients with subjective cognitive decline; they suggested that the increased gray matter volume in the paracentral lobule represents a compensatory mechanism, although it is unclear whether the mechanism is associated with cognitive function. This notion is consistent with our findings that the dALFF value of the right paracentral lobule was decreased in RD patients, although we found no correlation between this brain region and HAMA scores. Therefore, we speculate that the decreased dALFF value of the right paracentral lobule in RD patients reflects inhibition of this brain region. However, there is no clear evidence of diminished cognitive function in RD patients.

In our study, the overall accuracies of SVM classification were 97.87, 100, and 93.62% under three different time windows; area under the curve values were 0.99, 1.00, and 0.95, respectively. Thus, dALFF may offer sensitive biomarkers for distinguishing patients with RD from HCs.

4.2. Differences in metrics of the dALFF states

Importantly, there were significant differences in the temporal characteristics of dALFF states between the two groups. Our results showed that the number of transitions in RD patients was higher than that in HC group at both 30TR and 50TR, while the number of transitions in RD patients and HC group was zero at 80TR. At the same time, compared with the HC group, RD patients at 30TR and 50TR showed 3 different time states, and RD patients at 80TR showed 4 different time states. In other words, during the entire resting state MRI scanning period, the brain's working mode can be divided into three states at 30TR and 50TR, and the brain's working mode can be divided into four states at 80TR. For patients

with RD, state 1 accounted for a larger proportion of these states, suggesting that state 1 may represent a pattern of major brain activity in individuals with RD. In general, the mean dwell time and number of transitions are used as parameters in dynamic pattern analysis to describe state properties that represent brain functional activity and can be reconfigured during illness (Xu et al., 2023). Previous studies have found that patients with diabetic retinopathy exhibit three different temporal states, with state 1 occupying a larger proportion, while patients with diabetic retinopathy have a lower number of transitions than those in the HC group (Huang et al., 2021). Zhao et al. (2021) proposed that an increase in the number of transitions is associated with a decrease in the efficiency of information flow in brain networks. Therefore, we speculate that visual dysfunction leads to a decrease in the efficiency of information flow, which increases the number of conversions in patients with RD. This result also suggests that the whole brain integration of visual functions is abnormal.

5. Limitations

This study had some important limitations. First, it included a small number of RD patients. The lack of correlation between region of interest values in positive brain regions and HAMA scores may have been related to the small sample size. Second, the data were frequently affected by unavoidable factors in the fMRI environment (e.g., heartbeat, muscle beat, and respiratory motion).

6. Conclusion

In this study, we used the dALFF method, k-means clustering method and an SVM classification approach to explore dynamic changes in spontaneous brain activity among RD patients. Our findings offer important insights regarding the neuropathology that underlies RD and provide robust evidence that dALFF, a local indicator of brain activity, may be useful for clinical diagnosis.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of the First Affiliated Hospital of Nanchang University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YJ is responsible for writing manuscript. Y-yW is in charge of proofreading and refining the manuscript's wording. QC, W-wF, S-qH, P-pZ, X-lC, B-lS, BW, and Q-yH contributed to data collection and statistical analyses. YJ and Y-yW designed the protocol and contributed to the MRI analysis. YJ, Y-yW, and X-rW designed the study, oversaw all clinical aspects of study conduct, and manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

We acknowledge the assistance provided by the National Nature Science Foundation of China (grant no. 82160207), Key projects of Jiangxi Youth Science Fund (no. 20202ACBL216008), Science and Technology Plan of Jiangxi Provincial Health and Health Commission (202130156), and Postgraduate Innovation Special Fund Project in Jiangxi Province (YC2022—s198).

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

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., et al. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12060–12065. doi: 10.1073/pnas.1302982110

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Caminiti, R., Ferraina, S., and Johnson, P. B. (1996). The sources of visual information to the primate frontal lobe: a novel role for the superior parietal lobule. *Cereb. Cortex* 6, 319–328. doi: 10.1093/cercor/6.3.319

Chang, C., and Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage* 50, 81–98. doi: 10.1016/j.neuroimage.2009.12.011

Chen, X., He, X., Tao, L., Li, J., Wu, J., Zhu, C., et al. (2017). The working memory and dorsolateral prefrontal-hippocampal functional connectivity changes in long-term survival breast Cancer patients treated with tamoxifen. *Int. J. Neuropsychopharmacol.* 20, 374–382. doi: 10.1093/ijnp/pyx008

Chen, R. B., Ye, S. Y., Pei, C. G., and Zhong, Y. L. (2022). Altered temporal dynamics of the amplitude of low-frequency fluctuations in Comitant exotropia patients. *Front. Hum. Neurosci.* 16:944100. doi: 10.3389/fnhum.2022.944100

Dronkers, N. F., Wilkins, D. P., van Valin, R. D., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92, 145–177. doi: 10.1016/j.cognition.2003.11.002

Gehrig, J., Michalareas, G., Forster, M. T., Lei, J., Hok, P., Laufs, H., et al. (2019). Low-frequency oscillations code speech during verbal working memory. *J. Neurosci.* 39, 6498–6512. doi: 10.1523/JNEUROSCI.0018-19.2019

Gerstenberger, E., Stoffelns, B., Nickels, S., Münzel, T., Wild, P. S., Beutel, M. E., et al. (2021). Incidence of retinal detachment in Germany: results from the Gutenberg health study. *Ophthalmologica* 244, 133–140. doi: 10.1159/000513080

Gui, S. G., Chen, R. B., Zhong, Y. L., and Huang, X. (2021). Machine learning analysis reveals abnormal static and dynamic low-frequency oscillations indicative of long-term menstrual pain in primary dysmenorrhea patients. *J. Pain Res.* 14, 3377–3386. doi: 10.2147/JPR.S332224

Hu, M. Y., Zhang, L. J., Kang, M., Pan, Y. C., Ge, Q. M., Li, Q. Y., et al. (2022). Brain activity in different brain areas of patients with dry eye during the female climacteric period according to voxel-based morphometry. *Front. Neurol.* 13:879444. doi: 10.3389/fneur.2022.879444

Huang, X., Hu, Y., Zhou, F., Xu, X., Wu, Y., Jay, R., et al. (2018). Altered whole-brain gray matter volume in high myopia patients: a voxel-based morphometry study. *Neuroreport* 29, 760–767. doi: 10.1097/WNR.0000000000001028

Huang, X., Li, D., Li, H. J., Zhong, Y. L., Freeberg, S., Bao, J., et al. (2017). Abnormal regional spontaneous neural activity in visual pathway in retinal detachment patients: a resting-state functional MRI study. *Neuropsychiatr. Dis. Treat.* 13, 2849–2854. doi: 10.2147/NDT.S147645

Huang, X., Li, H. J., Peng, D. C., Ye, L., Yang, Q. C., Zhong, Y. L., et al. (2019). Altered brain network centrality in patients with late monocular blindness: a resting-state fMRI study. *Arch. Med. Sci.* 15, 1301–1307. doi: 10.5114/aoms.2019.87133

Huang, X., Wen, Z., Qi, C. X., Tong, Y., and Shen, Y. (2021). Dynamic changes of amplitude of low-frequency fluctuations in patients with diabetic retinopathy. *Front. Neurol.* 12:611702. doi: 10.3389/fneur.2021.611702

Iacoboni, M., and Zaidel, E. (2004). Interhemispheric visuo-motor integration in humans: the role of the superior parietal cortex. *Neuropsychologia* 42, 419–425. doi: 10.1016/j.neuropsychologia.2003.10.007

Ibrar, A., Panayiotis, M., and Mohamed, E. A. (2021). Recognising and managing retinal detachments. *Br. J. Hosp. Med. (Lond.)* 82, 1–11. doi: 10.12968/hmed.2021.0145

Ji, Y., Huang, S. Q., Cheng, Q., Fu, W. W., Zhong, P. P., Chen, X. L., et al. (2023). Exploration of static functional connectivity and dynamic functional connectivity alterations in the primary visual cortex among patients with high myopia via seed-based functional connectivity analysis. *Front. Neurosci.* 17:1126262. doi: 10.3389/fnins.2023.1126262

Jia, X. Z., Wang, J., Sun, H. Y., Zhang, H., Liao, W., Wang, Z., et al. (2019). RESTplus: an improved toolkit for resting-state functional magnetic resonance imaging data processing. *Sci Bull (Beijing).* 64, 953–954. doi: 10.1016/j.scib.2019.05.008

Kang, H. H., Shu, Y. Q., Yang, L., Zhu, P. W., Li, D., Li, Q. H., et al. (2019). Measuring abnormal intrinsic brain activities in patients with retinal detachment using amplitude of low-frequency fluctuation: a resting-state fMRI study. *Int. J. Neurosci.* 129, 681–686. doi: 10.1080/00207454.2018.1554657

Kawashima, S., Shimizu, Y., Ueki, Y., and Matsukawa, N. (2021). Impairment of the visuospatial working memory in the patients with Parkinson's disease: an fMRI study. *BMC Neurol.* 21:335. doi: 10.1186/s12883-021-02366-7

Kim, D., Lee, S., Choi, M., Youn, H. C., Suh, S., Jeong, H. G., et al. (2019). Diffusion tensor imaging reveals abnormal brain networks in elderly subjects with subjective cognitive deficits. *Neurol. Sci.* 40, 2333–2342. doi: 10.1007/s10072-019-03981-6

Leonardi, N., and Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage* 104, 430–436. doi: 10.1016/j.neuroimage.2014.09.007

Li, J., Duan, X., Cui, Q., Chen, H., and Liao, W. (2019). More than just statics: temporal dynamics of intrinsic brain activity predicts the suicidal ideation in depressed patients. *Psychol. Med.* 49, 852–860. doi: 10.1017/S0033291718001502

Liang, L., Zhao, L., Wei, Y., Mai, W., Duan, G., Su, J., et al. (2020). Structural and functional hippocampal changes in subjective cognitive decline from the community. *Front. Aging Neurosci.* 12:64. doi: 10.3389/fnagi.2020.00064

Lumi, X., Hawlina, M., Glavač, D., Facskó, A., Moe, M. C., Kaarniranta, K., et al. (2015). Ageing of the vitreous: from acute onset floaters and flashes to retinal detachment. *Ageing Res. Rev.* 21, 71–77. doi: 10.1016/j.arr.2015.03.006

Ma, H., Huang, G., Li, M., Han, Y., Sun, J., Zhan, L., et al. (2021). The predictive value of dynamic intrinsic local metrics in transient ischemic attack. *Front. Aging Neurosci.* 13:808094. doi: 10.3389/fnagi.2021.808094

Mascalchi, M., Ginestroni, A., Toschi, N., Poggesi, A., Cecchi, P., Salvadori, E., et al. (2014). The burden of microstructural damage modulates cortical activation in elderly subjects with MCI and leuko-araiosis. A DTI and fMRI study. *Hum. Brain Mapp.* 35, 819–830. doi: 10.1002/hbm.22216

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007

Qi, C. X., Wen, Z., and Huang, X. (2022). Reduction of interhemispheric homotopic connectivity in cognitive and visual information processing pathways in patients with thyroid-associated Ophthalmopathy. *Front. Hum. Neurosci.* 16:882114. doi: 10.3389/fnhum.2022.882114

Raichle, M. E. (2011). The restless brain. *Brain Connect.* 1, 3–12. doi: 10.1089/brain.2011.0019

Romero, M. C., Bermudez, M. A., Vicente, A. F., Perez, R., and Gonzalez, F. (2008). Activity of neurons in the caudate and putamen during a visuomotor task. *Neuroreport* 19, 1141–1145. doi: 10.1097/WNR.0b013e328307c3fc

Shao, Y., Yang, L., Zhu, P. W., Su, T., Zhou, X. Z., Li, B., et al. (2021). Functional connectivity density alterations in middle-age retinal detachment patients. *Brain Behav.* 11:e01783. doi: 10.1002/brb3.1783

Sporns, O. (2011). The non-random brain: efficiency, economy, and complex dynamics. *Front. Comput. Neurosci.* 5:5. doi: 10.3389/fncom.2011.00005

Steel, D. (2014). Retinal detachment. *BMJ Clin Evid.* 2014:0710.

Su, T., Shu, Y. Q., Liu, K. C., Ye, L., Chen, L. L., Shi, W. Q., et al. (2018). Functional connectivity of paired default mode network subregions in retinal detachment. *Transl. Vis. Sci. Technol.* 7:15. doi: 10.1167/tvst.7.6.15

Tan, G., Dan, Z. R., Zhang, Y., Huang, X., Zhong, Y. L., Ye, L. H., et al. (2018). Altered brain network centrality in patients with adult comitant exotropia strabismus: a resting-state fMRI study. *J. Int. Med. Res.* 46, 392–402. doi: 10.1177/0300060517715340

Tong, Y., Huang, X., Qi, C. X., and Shen, Y. (2021). Disrupted neural activity in individuals with Iridocyclitis using regional homogeneity: a resting-state functional magnetic resonance imaging study. *Front. Neurol.* 12:609929. doi: 10.3389/fneur.2021.609929

Verhoekx, J., van Overdam, K. A., Gishti, O., van Leeuwen, R., and Crama, N. (2021). Acute onset of floaters, even without flashes, is an urgent ophthalmic warning sign. *Ned. Tijdschr. Geneeskd.* 165:D5850.

Vicente, A. F., Bermudez, M. A., Romero Mdel, C., Perez, R., and Gonzalez, F. (2012). Putamen neurons process both sensory and motor information during a complex task. *Brain Res.* 1466, 70–81. doi: 10.1016/j.brainres.2012.05.037

Vujosevic, S., Parra, M. M., Hartnett, M. E., O'Toole, L., Nuzzi, A., Limoli, C., et al. (2023). Optical coherence tomography as retinal imaging biomarker of neuroinflammation/neurodegeneration in systemic disorders in adults and children. *Eye (Lond.)* 37, 203–219. doi: 10.1038/s41433-022-02056-9

Wen, Z., Kang, Y., Zhang, Y., Yang, H., Zhao, Y., Huang, X., et al. (2023). Disrupted dynamic amplitude of low-frequency fluctuations in patients with active thyroid-associated ophthalmopathy. *Front. Cell Dev. Biol.* 11:1174688. doi: 10.3389/fcell.2023.1174688

Wu, Y. J., Rao, J., Huang, X., Wu, N., Shi, L., Huang, H., et al. (2021). Impaired interhemispheric synchrony in bronchial asthma. *Int J Gen Med.* 14, 10315–10325. doi: 10.2147/IJGM.S343269

Xu, L., Wei, H., Sun, Z., Chu, T., Li, M., Liu, R., et al. (2023). Dynamic alterations of spontaneous neural activity in post-stroke aphasia: a resting-state functional magnetic resonance imaging study. *Front. Neurosci.* 17:1177930. doi: 10.3389/fnins.2023.1177930

Yang, Y. C., Li, Q. Y., Chen, M. J., Zhang, L. J., Zhang, M. Y., Pan, Y. C., et al. (2021). Investigation of changes in retinal detachment-related brain region activities and functions using the percent amplitude of fluctuation method: a resting-state functional magnetic resonance imaging study. *Neuropsychiatr. Dis. Treat.* 17, 251–260. doi: 10.2147/NDT.S292132

Yu, C., Liu, Y., Li, J., Zhou, Y., Wang, K., Tian, L., et al. (2008). Altered functional connectivity of primary visual cortex in early blindness. *Hum. Brain Mapp.* 29, 533–543. doi: 10.1002/hbm.20420

Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002

Zhang, W., Li, S., Wang, X., Gong, Y., Yao, L., Xiao, Y., et al. (2018). Abnormal dynamic functional connectivity between speech and auditory areas in schizophrenia patients with auditory hallucinations. *Neuroimage Clin.* 19, 918–924. doi: 10.1016/j.nicl.2018.06.018

Zhao, L., Wang, D., Xue, S. W., Tan, Z., Wang, Y., and Lian, Z. (2021). Aberrant state-related dynamic amplitude of low-frequency fluctuations of the emotion network in major depressive disorder. *J. Psychiatr. Res.* 133, 23–31. doi: 10.1016/j.jpsychires.2020.12.003

Zhao, S., Xia, Y., Huang, Y., Zou, H., Wang, X., Chen, Z., et al. (2021). The correlation between thyroid function, frontal gray matter, and executive function in patients with major depressive disorder. *Front Endocrinol (Lausanne).* 12:779693. doi: 10.3389/fendo.2021.779693

Zheng, J. O., Yu, L., Huang, D. H., Cao, X. L., Chen, Z. R., and Ye, W. (2015). Reduced self-regulation of cerebrum contributes to executive impairment in patients with temporal lobe epilepsy. *Int. J. Clin. Exp. Med.* 8, 9610–9619.

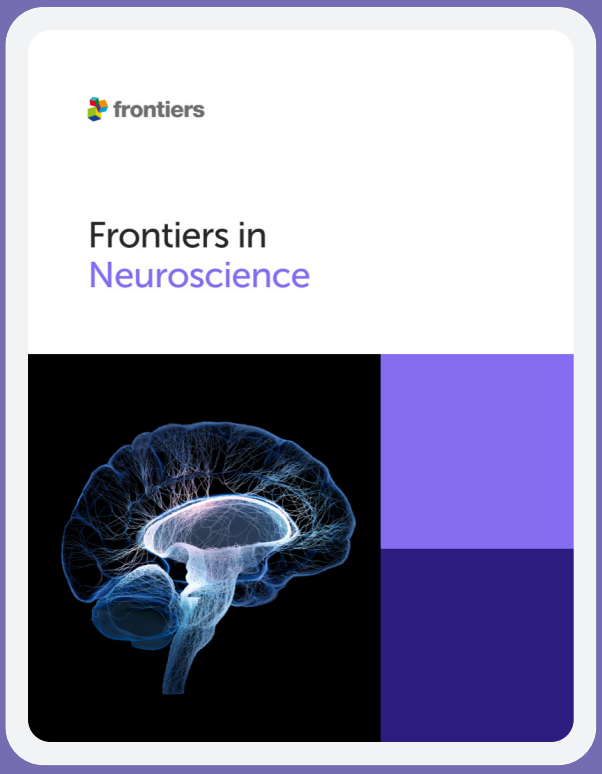
Frontiers in Neuroscience

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