frontiers | Research Topics

Efficient artificial intelligence (AI) in ophthalmic imaging

Edited by
Yanda Meng, Yalin Zheng, Haoyu Chen and Meng Wang

Published in
Frontiers in Medicine

FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-5882-9
DOI 10.3389/978-2-8325-5882-9

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

Efficient artificial intelligence (AI) in ophthalmic imaging

Topic editors

Yanda Meng — University of Exeter, United Kingdom
Yalin Zheng — University of Liverpool, United Kingdom
Haoyu Chen — The Chinese University of Hong Kong, China
Meng Wang — Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore

Citation

Meng, Y., Zheng, Y., Chen, H., Wang, M., eds. (2025). *Efficient artificial intelligence (AI) in ophthalmic imaging*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5882-9

Table of contents

04 Editorial: Efficient artificial intelligence (AI) in ophthalmic imaging
Yanda Meng, Meng Wang, Haoyu Chen and Yalin Zheng

06 Automated evaluation of retinal hyperreflective foci changes in diabetic macular edema patients before and after intravitreal injection
Xingguo Wang, Yanyan Zhang, Yuhui Ma, William Robert Kwapong, Jianing Ying, Jiayi Lu, Shaodong Ma, Qifeng Yan, Quanyong Yi and Yitian Zhao

19 Deep learning-based estimation of axial length using macular optical coherence tomography images
Jing Liu, Hui Li, You Zhou, Yue Zhang, Shuang Song, Xiaoya Gu, Jingjing Xu and Xiaobing Yu

26 Exploring large language model for next generation of artificial intelligence in ophthalmology
Kai Jin, Lu Yuan, Hongkang Wu, Andrzej Grzybowski and Juan Ye

35 Neighbored-attention U-net (NAU-net) for diabetic retinopathy image segmentation
Tingting Zhao, Yawen Guan, Dan Tu, Lixia Yuan and Guangtao Lu

47 DBPF-net: dual-branch structural feature extraction reinforcement network for ocular surface disease image classification
Cheng Wan, Yulong Mao, Wenqun Xi, Zhe Zhang, Jiantao Wang and Weihua Yang

57 Artificial intelligence-assisted management of retinal detachment from ultra-widefield fundus images based on weakly-supervised approach
Huimin Li, Jing Cao, Kun You, Yuehua Zhang and Juan Ye

67 Deep learning based retinal vessel segmentation and hypertensive retinopathy quantification using heterogeneous features cross-attention neural network
Xinghui Liu, Hongwen Tan, Wu Wang and Zhangrong Chen

77 Segmentation of retinal microaneurysms in fluorescein fundus angiography images by a novel three-step model
Jing Li, Qian Ma, Mudi Yao, Qin Jiang, Zhenhua Wang and Biao Yan

87 AI-driven generalized polynomial transformation models for unsupervised fundus image registration
Xu Chen, Xiaochen Fan, Yanda Meng and Yalin Zheng

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 17 December 2024
DOI 10.3389/fmed.2024.1523647

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Jodhbir Mehta,
Singapore National Eye Center, Singapore

*CORRESPONDENCE
Yanda Meng
✉ y.m.meng@exeter.ac.uk

RECEIVED 06 November 2024
ACCEPTED 02 December 2024
PUBLISHED 17 December 2024

CITATION
Meng Y, Wang M, Chen H and Zheng Y (2024) Editorial: Efficient artificial intelligence (AI) in ophthalmic imaging. *Front. Med.* 11:1523647. doi: 10.3389/fmed.2024.1523647

COPYRIGHT
© 2024 Meng, Wang, Chen and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Efficient artificial intelligence (AI) in ophthalmic imaging

Yanda Meng[1,2]*, Meng Wang[3], Haoyu Chen[4] and Yalin Zheng[5]

[1]Department of Computer Science, University of Exeter, Exeter, United Kingdom, [2]Department of Cardiovascular and Metabolic Medicine, University of Liverpool, Liverpool, United Kingdom, [3]Centre for Innovation and Precision Eye Health, National University of Singapore, Singapore, Singapore, [4]Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, China, [5]Department of Eye and Vision Science, University of Liverpool, Liverpool, United Kingdom

KEYWORDS

efficient AI, ophthalmic imaging, retinal, ocular, OCT, color fundus

Editorial on the Research Topic
Efficient artificial intelligence (AI) in ophthalmic imaging

Techniques like optical coherence tomography (OCT), fundus photography, and fluorescein angiography produce a wealth of visual data, offering a detailed view of the eye's structure and function. These imaging tools are essential for identifying and tracking the progression of conditions such as age-related macular degeneration, diabetic retinopathy, and glaucoma. While ophthalmologists are well-trained in reading these images, manually analyzing large datasets can be slow, prone to mistakes, and can vary between observers. Recently, AI tools have shown great promise in supporting ophthalmologists, helping to speed up and improve the accuracy of their diagnoses. This Research Topic comprises nine original research articles covering several different topics using efficient AI, including diabetic macular edema, large language models, macular axial length measurement, ocular surface disease diagnosis, diabetic retinopathy, retinal detachment management, retinal vessel, and microaneurysms segmentation, fundus image registration. A summary of these articles is presented as follows.

Wang et al. introduced an automated framework leveraging deep learning advancements to extract twelve 3D parameters from segmented hyperreflective foci in optical coherence tomography (OCT) images. This development is crucial for understanding various ocular diseases.

Retinal vessels are vital biomarkers for detecting conditions like hypertensive retinopathy. Manual identification is labor-intensive and time-consuming. Liu X. et al. addressed this by proposing a heterogeneous feature cross-attention neural network for retinal vessel segmentation in color fundus images.

Image registration aligns multiple images from different viewpoints or spaces, which is essential in vision applications. Chen et al. introduced an AI-driven approach to unsupervised fundus image registration using a Generalized Polynomial Transformation (GPT) model. Trained on a large synthetic dataset, GPT simulates diverse polynomial transformations.

Microaneurysms, early indicators of diabetic retinopathy, are challenging to detect due to low contrast and similarity to retinal vessels in fluorescein fundus angiography (FFA)

images. Li J. et al. presented a model for automatic microaneurysm detection to address these challenges.

Retinal detachment (RD) is a common sight-threatening condition in emergency departments. Early postural intervention based on detachment regions can improve visual prognosis. Li H. et al. developed a weakly supervised model using 24,208 ultra-widefield fundus images to localize and outline anatomical RD regions.

The increasing prevalence of diabetic retinopathy-related (DR-related) diseases among younger individuals poses a significant threat to eye health. Zhao et al. proposed the Neighbored Attention U-Net (NAU-Net) to balance identification performance and computational cost for DR fundus image segmentation.

Pterygium, an ocular surface disease characterized by fibrovascular overgrowth invading the cornea, requires accurate diagnosis. Wan et al. proposed a dual-branch network reinforced by a PFM block (DBPF-Net) for the four-way classification of ocular surface diseases, utilizing a conformer model backbone.

Axial length (AL) is significant for defining the eye's refractive status and is associated with retinal and macular complications. Excessive AL elongation, often over 26.0 mm, increases the risk of posterior segment complications. Liu J. et al. developed deep learning models using macular OCT images to estimate ALs in eyes without maculopathy.

Jin et al. discussed the promising role of large language models (LLMs) in shaping AI's future in ophthalmology. By leveraging AI, ophthalmologists can access information, enhance diagnostic accuracy, and provide better patient care. Despite challenges, ongoing AI advancements and research pave the way for next-generation AI-assisted ophthalmic practices.

Author contributions

YM: Writing – original draft, Writing – review & editing. MW: Writing – review & editing. HC: Writing – review & editing. YZ: Writing – review & editing.

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

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.

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 October 2023
DOI 10.3389/fmed.2023.1280714

Check for updates

OPEN ACCESS

EDITED BY
Meng Wang,
Agency for Science,
Technology and Research (A*STAR), Singapore

REVIEWED BY
Yi Zhou,
Soochow University, China
Ao Cheng,
Soochow University, China

*CORRESPONDENCE
Yitian Zhao
✉ yitian.zhao@nimte.ac.cn
Quanyong Yi
✉ quanyong__yi@163.com

†These authors have contributed equally to this work

RECEIVED 21 August 2023
ACCEPTED 21 September 2023
PUBLISHED 06 October 2023

CITATION
Wang X, Zhang Y, Ma Y, Kwapong WR, Ying J, Lu J, Ma S, Yan Q, Yi Q and Zhao Y (2023) Automated evaluation of retinal hyperreflective foci changes in diabetic macular edema patients before and after intravitreal injection. *Front. Med.* 10:1280714. doi: 10.3389/fmed.2023.1280714

COPYRIGHT
© 2023 Wang, Zhang, Ma, Kwapong, Ying, Lu, Ma, Yan, Yi and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automated evaluation of retinal hyperreflective foci changes in diabetic macular edema patients before and after intravitreal injection

Xingguo Wang[1,2†], Yanyan Zhang[3†], Yuhui Ma[2], William Robert Kwapong[4], Jianing Ying[5], Jiayi Lu[1,2], Shaodong Ma[2], Qifeng Yan[2], Quanyong Yi[3]* and Yitian Zhao[1,2]*

[1]Cixi Biomedical Research Institute, Wenzhou Medical University, Ningbo, China, [2]Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China, [3]The Affiliated Ningbo Eye Hospital of Wenzhou Medical University, Ningbo, China, [4]Department of Neurology, West China Hospital, Sichuan University, Chengdu, China, [5]Health Science Center, Ningbo University, Ningbo, China

Purpose: Fast and automated reconstruction of retinal hyperreflective foci (HRF) is of great importance for many eye-related disease understanding. In this paper, we introduced a new automated framework, driven by recent advances in deep learning to automatically extract 12 three-dimensional parameters from the segmented hyperreflective foci in optical coherence tomography (OCT).

Methods: Unlike traditional convolutional neural networks, which struggle with long-range feature correlations, we introduce a spatial and channel attention module within the bottleneck layer, integrated into the nnU-Net architecture. Spatial Attention Block aggregates features across spatial locations to capture related features, while Channel Attention Block heightens channel feature contrasts. The proposed model was trained and tested on 162 retinal OCT volumes of patients with diabetic macular edema (DME), yielding robust segmentation outcomes. We further investigate HRF's potential as a biomarker of DME.

Results: Results unveil notable discrepancies in the amount and volume of HRF subtypes. In the whole retinal layer (WR), the mean distance from HRF to the retinal pigmented epithelium was significantly reduced after treatment. In WR, the improvement in central macular thickness resulting from intravitreal injection treatment was positively correlated with the mean distance from HRF subtypes to the fovea.

Conclusion: Our study demonstrates the applicability of OCT for automated quantification of retinal HRF in DME patients, offering an objective, quantitative approach for clinical and research applications.

KEYWORDS

diabetic macular edema, hyperreflective foci, optical coherence tomography, artificial intelligence, deep learning

1. Introduction

Diabetic retinopathy (DR) is one of the most common complications of diabetes (1). With about 1 in every 10 diabetic patients developing visual impairment due to DR (2). One of the leading causes of visual impairment in DR patients is diabetic macular edema (DME) (3). It is suggested that in DR patients, disruption of the blood-retina barrier leads to increased fluid leakage within the retina, resulting in the development of DME (4) ultimately resulting in visual loss.

In recent decades, advances in high-resolution fundus imaging techniques have led to the discovery of specific imaging features of retinal diseases, which may serve as diagnostic, predictive, and prognostic biomarkers for this disease (5). Optical coherence tomography (OCT) is an imaging tool that can help in the visualization of the intra-retinal layers. Due to its non-invasiveness, affordability and high resolution, this imaging tool is suggested as the gold standard for the diagnosis and monitoring of DME (6). 'Hyperreflective foci' (HRF) is a term denoting any hyperreflective lesion, focal or dotted appearance, seen at any retinal layer on OCT images (7). Reports suggest that HRF is associated with lipid extravasation (7), microglia cells (8), migrating retinal pigment epithelium (RPE) cells (9), degenerated photoreceptor cells, and visual prognosis (10), increasing its clinical significance. In the last decade, it was shown that the presence of HRF was associated with DME, and several more recent studies have indicated that HRF could serve as a promising biomarker for investigating DME, due to its association with the soluble cluster of differentiation 14 (CD14) pro-inflammatory cytokine expressed by glial cells, monocytes, and macrophages (8, 11).

However, manual annotation of HRF in OCT is time-consuming, and sometimes excessively subjective. With the rapid development of computer science, there is great potential for automatic segmentation and quantification of HRF in OCT images, with benefits for clinical practice. The segmentation algorithms for HRF can be categorized into two primary groups: traditional segmentation algorithms and deep learning-based segmentation methods. Traditional HRF segmentation approaches usually require manual parameter tuning and extensive prior knowledge. Okuwobi et al. (12) employed an automated grow-cut algorithm for HRF segmentation. It is difficult for traditional automated methods to perform accurate HRF segmentation due to boundary blurring and speckle noise within HRF images. Okuwobi et al. (13) introduced another component tree-based method to segment HRF by extracting the extreme regions from the connected areas. Still, the method is complicated and relies on handcrafted features. Deep learning techniques have achieved significant success in medical image segmentation. Yu et al. (14) modified GoogLeNet for HRF segmentation in DR using pixel-level predictions of small image patches. However, this method partially addresses the class imbalance issue, leading to the mis-segmentation of large blood vessels or low-contrast backgrounds as HRF. Xie et al. (15) modified 3D-UNet for HRF segmentation, introducing denoised and enhanced OCT images as a dual-channel input and dilation convolution in the final layer of the encoder to expand the receptive field. Nevertheless, this approach overlooks false positive outcomes caused by high-frequency noise in the NFL/GCL and IS/OS layers. Yao et al. (16) modified U-Net for HRF segmentation, enhancing gradient propagation by replacing ordinary convolution blocks with dual residual modules and integrating adaptive modules within the bottleneck layer to fuse local features and global dependencies. However, this network ignores the inappropriateness of employing deformable convolutions for the segmentation of HRF due to its small size and lack of shape information. Wei et al. (17) preprocessed images using Non-local means (NLM) filters and adopted a patch-based segmentation approach, employing a lightweight network for automated HRF segmentation. This network relies on the patch-based method, which further diminishes the limited semantic information inherent in HRF.

In this study, we presented a deep learning-based framework for the quantitative analysis of HRF in OCT images. Specifically, the main contributions of our article can be summarized as follows:

- We achieve excellent HRF segmentation performance by combining nnU-Net (18) adaptability with the advanced long-range feature-capturing abilities of channel and spatial attention modules.
- Using the proposed method, we extracted 12 parameters to characterize HRF morphology and distribution, showing significant differences in volume and amount among the three HRF sub-types in retinal OCT images.
- Using the extracted 12 HRF parameters, we evaluated changes in HRF before and after treatment and their correlation with central macular thickness (CMT) improvement.

2. Materials and methods

This is a retrospective, longitudinal study conducted at the Affiliated Ningbo Eye Hospital of Wenzhou Medical University (Ningbo, China) from November 2020 to July 2022. This study was approved by the ethics committee of the Affiliated Ningbo Eye Hospital of Wenzhou Medical University (ID: 20210327A), and informed written consent was obtained from each participant involved in our study according to the Declaration of Helsinki.

2.1. DME participants

Type 2 diabetes mellitus (DM) patients were recruited and diagnosed by an endocrine specialist. Demographic and clinical information from all patients such as age, gender, duration of DM, and systolic/diastolic blood pressure were recorded. All patients had an extensive ophthalmic examination, involving slit-lamp biomicroscopy, and assessment of intraocular pressure, axial length, and visual acuity. The inclusion criteria of our patients are as follows: 1. Diagnosed with type 2 DM; 2. Age > 18 years; 3. Macular edema, defined clinically and by a retinal thickness of >250 μm in the central subfield (19); 4. Could cooperate with OCT imaging. Exclusion criteria were as follows: 1. Myopia; 2. Presence of media opacities; 3. Inability to cooperate with OCT imaging.

2.2. OCT image acquisition

3D retinal imaging was performed using the OCT tool (Spectralis HRA + OCT; Heidelberg Engineering, Heidelberg, Germany, software version V6.16.2). This imaging equipment has a scanning protocol of 40,000 A-scans/s (20), with an axial resolution of 3.9 μm and a lateral

resolution of 11.4 μm in high-speed mode. We acquired OCT images covering fovea-centered regions of $4.5 \times 4.5\,mm^2$, with 384 B-scans, and $6 \times 6\,mm^2$, with 512 B-scans. OCT images showing retinal abnormalities such as age macular degeneration (AMD), severe cataract, and glaucoma; images with signal quality less than 7; or with OCT artifacts present, were excluded. OCT data displayed in our study followed the OSCAR-IB quality criteria (21) and APOSTEL recommendation (22). Patients were excluded if their CMT did not increase after treatment with anti-vascular endothelial growth factor (anti-VEGF).

2.3. HRF and retinal layers segmentation

We introduced an automatic tool for HRF analysis in OCT images. A deep learning-based approach was employed for precise segmentation of HRF, boundaries of inner retina (IR) and outer retina (OR) in OCT images. The resulting segmentations are then used to calculate HRF parameters.

2.3.1. Hyperreflective foci segmentation

HRF was defined as discrete and well-defined lesions distributed between the in-ternal limiting membrane (ILM) and retinal pigmented epithelium (RPE), with similar reflectivity to the RPE layer (8). Considering that the most HRFs cross 2–4 B-scans (15), we randomly selected 8 consecutive B-scans from each OCT volume for manual annotations of HRF. Two senior ophthalmologists made manual annotations of HRF on 140 OCT volumes, and their consensus was defined as the ground-truth. 112 OCT volumes were randomly selected for training: the rest were used for validation. The best-performing model during training was then used for the evaluation of HRF segmentation in intact OCT data from all participants, across 22 OCT volumes from 11 eyes and a total of 9,216 B-scans. Figure 1A shows the automated segmentation results indicating HRF. Section 2.4 gives a detailed description of the proposed approach.

2.3.2. Inner and outer retinal layers segmentation

The distribution of HRF in the IR and OR, and their downward shift, have been previously studied (23). The IR region is defined as the region between the upper boundary of the ILM and the upper boundary of the outer plexiform layer (OPL), while the OR region is defined as the region between the upper boundary of the OPL and the lower boundary of the RPE (24). The whole retinal layer (WR) region is then defined as including both IR and OR. When HRF cross the upper boundary of OPL, they are considered located in the OR region. The IR and OR boundaries of 1,120 OCT images randomly selected from the training and validation dataset in section 2.3.1 were manually annotated by a senior ophthalmologist (Y.Y.Z). We used 896 images for training and the rest for validation. The evaluation dataset is also the same as in section 2.3.1. Figure 1B illustrates an example of IR and OR segmentation in OCT images.

2.4. Methods

2.4.1. Network architecture

In this research, we modified the nnU-Net, to, respectively, perform two segmentation tasks: hyperreflective foci segmentation and retinal layer segmentation. The framework comprises a basic U-Net architecture library that includes 2D and 3D version.

For the retinal layer segmentation task, we modified the 2D version of nnU-Net as the underlying network topology. The network architecture is shown in Figure 2. The network consists of six symmetric encoder-decoder layers with skip connections, which provides detailed features from the encoder to the decoder. A 384×384 patch with 3 channels is first input to one 3×3 convolution with stride 1 to obtain the low-level feature map with 32 channels. In the encoder, each layer contains two 3×3 convolutions with stride 1 followed by one 3×3 down-convolution with stride 2. In the decoder, each layer contains a 2×2 up-convolution with stride 2 followed by two 3×3 convolutions with stride 1. Finally, the feature map of the last decoder layer is fed into one 1×1 convolution with stride 1 to output the segmentation map.

Convolutional neural networks with U-Net structure have higher inductive bias, but lack the ability to capture long-distance dependent features. Inspired by CS2-Net (25), we embed a spatial and channel attention (SCA) module integrating channel attention and spatial attention mechanisms at the bottleneck layer. Specifically, the features output by the encoder are fed into two sub-modules of SCA in parallel. Spatial Attention Block (SAB) aggregates features at each spatial location to correlate similar features, while Channel Attention Block (CAB) enhances the contrast of each channel feature. The spatial attention matrix models the spatial relationship between pixel features. The acquisition of intra-class spatial association can be expressed as follows:

$$S_{(x,y)} = \exp\left(Q_y^T \cdot K_x\right) / \sum_{x'=1}^{N} \exp\left(Q_y^T \cdot K_{x'}\right) \quad (1)$$

where $S_{(x,y)}$ represents the influence of the y position on the x position. N represents the number of features. T denotes matrix transposition. Q_y and K_x represent two new feature maps generated from input features, representing the vertical and horizontal directions of structural features. The channel attention matrix enhances similar channel features and reduces different channel features, which can be expressed as follows:

$$C_{(x,y)} = \exp\left(F_x \cdot F_y^T\right) / \sum_{x'=1}^{C} \exp\left(F_{x'} \cdot F_y^T\right) \quad (2)$$

where $C_{(x,y)}$ represents the association between the features of the x-channel and y-channel. C denotes the number of channels. T represents matrix transpose. F_x and F_y represent the original input features.

For the HRF segmentation task, we have extended the modified nnU-Net from 2D to 3D. To achieve this, we have replaced all the 2D operations in both the encoder and decoder modules with 3D ones. Additionally, we have incorporated the 3D version of the SCA module into the bottleneck layer of the network. The detailed network architecture is illustrated in Figure 3.

All convolutions in the encoder and decoder adopt the form of Convolution-InstanceNorm-LeakyReLU, which are different from that in the vanilla architecture. Specifically, LeakyReLU (negative slope = 0.01) is used instead of ReLU, and instance normalization (26) is used instead of batch normalization (27). To train the network, the framework adopts a combination of dice coefficient loss and cross-entropy loss:

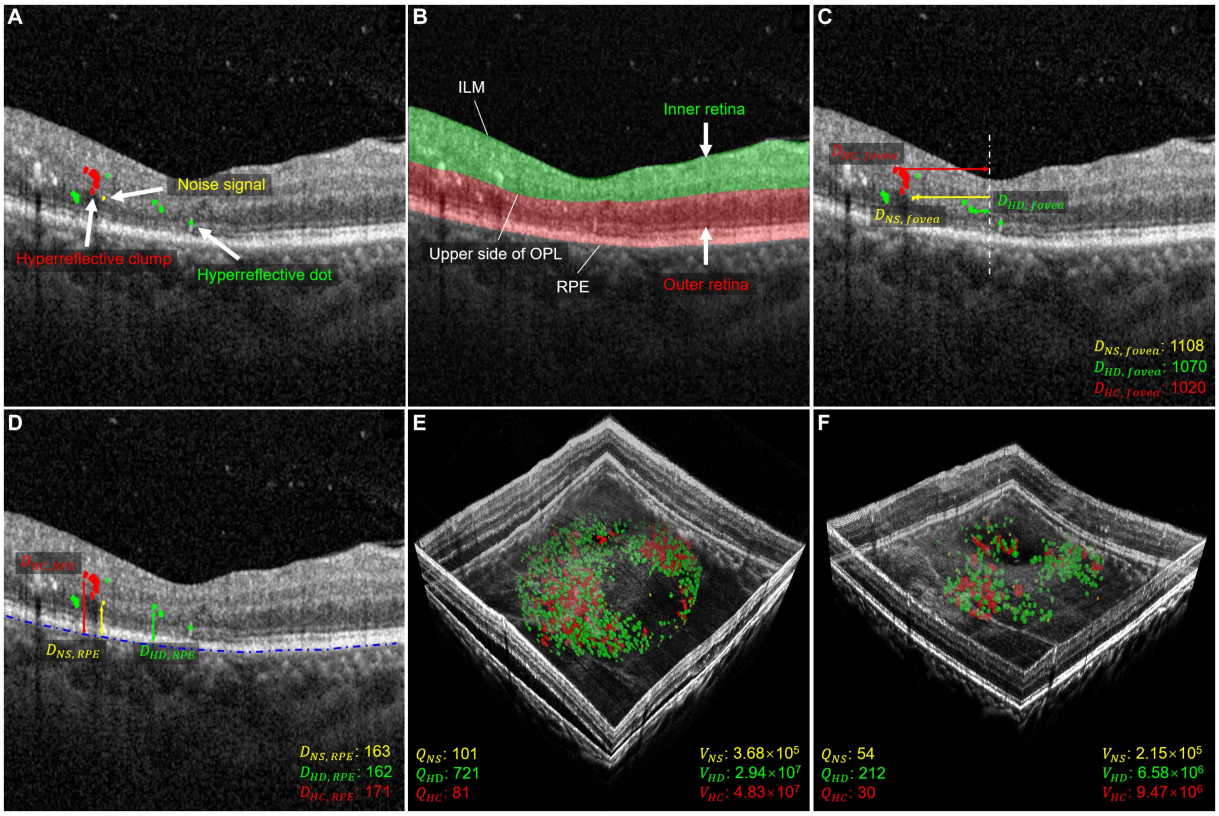

FIGURE 1
Morphology and distribution-related parameters used in quantitative measurements. **(A)** Shows the segmentation of HRF with NS in yellow, HD in green, and HC in red. **(B)** Shows the segmentation of the retina, with the inner layer in green and the outer layer in red. **(C)** Shows the distance parameter for the foveal direction of HRF. The distance between NS and fovea is shown in yellow, the distance between HD and fovea is shown in green, and the distance between HC and fovea is shown in red. The distance is measured in μm. **(D)** Shows the distance parameters between HRF and RPE. The distance between NS and RPE is shown in yellow, the distance between HD and RPE is shown in green, and the distance between HC and RPE is shown in red. The distance is measured in μm. **(E,F)** illustrate three-dimensional volume-rendered optical coherence tomography at the initial visit and six months after the initial visit, with NS in yellow, HD in green, and HC in red. For this case, the number parameters (Q_{NS}, Q_{HD}, Q_{HC}), and volume parameters (V_{NS}, V_{HD}, and V_{HC}) decreased.

$$\mathcal{L}_{total} = \mathcal{L}_{dice} + \mathcal{L}_{CE} \tag{3}$$

The dice loss formula used here is a variant of that used in Drozdzal et al. (28), and it is implemented as follows:

$$\mathcal{L}_{dc} = \left(-2/|K|\right) \cdot \sum_{k\in K}\left[\sum_{i\in I} u_i^k v_i^k \Big/ \left(\sum_{i\in I} u_i^k + \sum_{i\in I} v_i^k\right)\right] \tag{4}$$

where $u \in R^{I\times K}$ denotes the softmax output of the network, $v \in R^{I\times K}$ denotes the one-hot encoding of the ground truth, I represent the number of pixels in a training batch and K represents the number of categories.

2.5. Definitions of quantitative parameters

In this study, we analyzed changes in HRF's morphology and distribution in OCT images before and after IVI treatment. A previous study limited the maximum diameter range of HRF to 20–50 μm, which excludes the other two signals (8), refers to HRF <20 μm and > 50 μm, respectively. These signals were considered small noise signals (NS) in OCT images, and as hyperreflective clumps (HC) that appear as hard exudates in fundus images, respectively. By contrast, our study included all three types of these HRFs, allowing us to comprehensively investigate their differences in terms of number, volume, and spatial distribution. To this end, we first divided HRF into three types: NS, hyperreflective dots (HD), and HC, which are, respectively, defined as simply connected regions with a diameter range 0–20 μm, 20–50 μm, and greater than 50 μm. We then focused on 12 parameters that describe the distribution and morphological characteristics of these HRF in the retinal regions to be analyzed, as depicted in Figures 1C–F. Figures 1E,F demonstrate a 3D volume reconstruction case before and after IVI treatment. Following previous studies (8, 29), we selected a circular range of 3 mm in diameter, centered on the central macular region, for assessment of horizontal B-scans across the macular region. This region was used for analysis to ensure consistency in the region of interest across all participants.

2.5.1. Morphology-related parameters

- *Noise Signal Quantity* (Q_{NS}): Number of NS within the analyzed region.

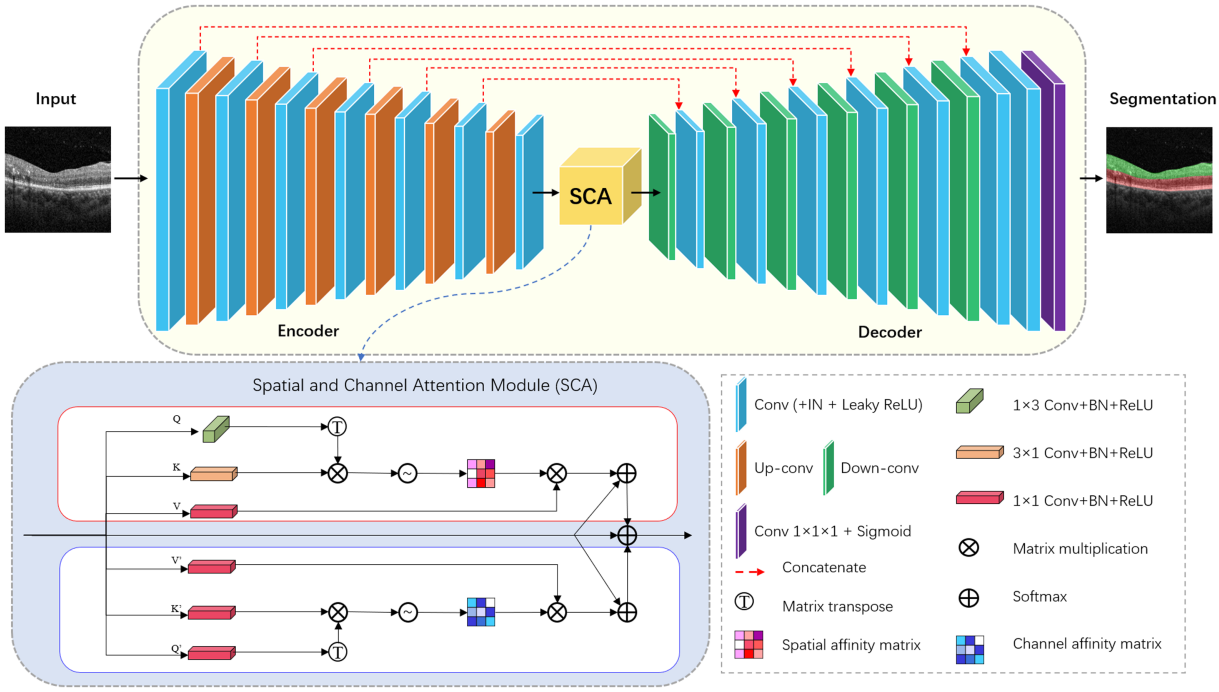

FIGURE 2
Architecture of modified nnU-Net (2D version).

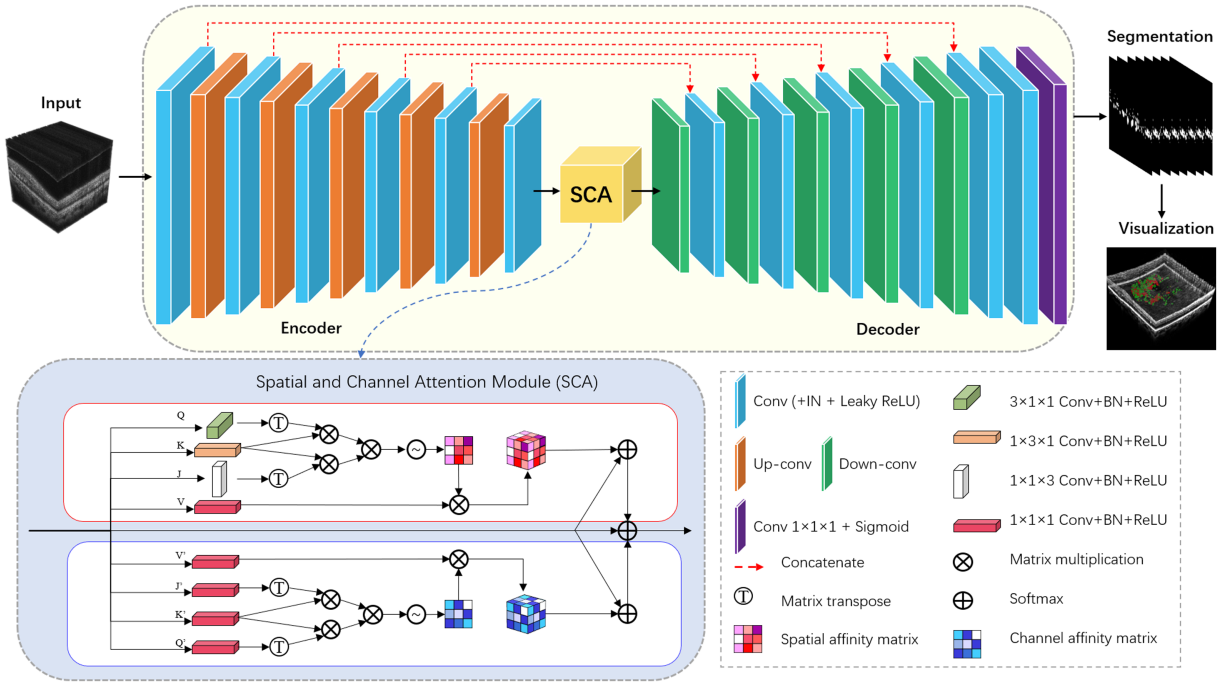

FIGURE 3
Architecture of modified nnU-Net (3D version).

- *Hyperreflective Dots Quantity* (Q_{HD}): Number of HD within the analyzed region.
- *Hyperreflective Clumps Quantity* (Q_{HC}): Number of HC within the analyzed region.
- *Noise Signal Volume* (V_{NS}): Volume of NS within the analyzed region in μm^3.
- *Hyperreflective Dots Volume* (V_{HD}): Volume of HD within the analyzed region in μm^3.

- *Hyperreflective Clumps Volume (V_{HC})*: Volume of HC within the analyzed region in μm^3.

2.5.2. Distribution-related parameters

- *Distance between noise signal and fovea ($D_{NS,fovea}$)*: Distance between NS and fovea, indicating average distance of NS pixels from the foveal center in μm.
- *Distance between hyperreflective dots and fovea ($D_{HD,fovea}$)*: Distance between HD and fovea, indicating average distance of HD pixels from the foveal center in μm.
- *Distance between hyperreflective clumps and fovea ($D_{HC,fovea}$)*: Distance between HC and fovea, indicating average distance of HC pixels from the foveal center in μm.
- *Distance between noise signal and RPE ($D_{NS,RPE}$)*: Distance between NS and RPE, indicating average distance of NS pixels from RPE in μm.
- *Distance between hyperreflective dots and RPE ($D_{HD,RPE}$)*: Distance between HD and RPE, indicating average distance of HD pixels from RPE in μm.
- *Distance between hyperreflective clumps and RPE ($D_{HC,RPE}$)*: Distance between HC and RPE, indicating average distance of HC pixels from RPE in μm.

2.6. Statistical analysis

All statistical analysis was performed using version 18.0 of SPSS software (SPSS, Inc., Chicago, IL, USA). Continuous variables were expressed as mean ± standard deviation (SD) for normal data; and median and interquartile ranges (IQR) for skewed data. Categorical variables were presented as frequencies. To compare the differences among different subtypes of HRF and the differences in HRF parameters before and after treatment, the Wilcoxon signed-rank test was used, and the results were expressed as the median (IQR). To investigate the correlation between the improvement in CMT and given parameters of HRF, Spearman's rank correlation coefficients were calculated using a non-parametric test for linear correlation. A significance level of $p < 0.05$ (two-sided test) was adopted to express statistical significance.

3. Results

3.1. Experimental results

3.1.1. Implementation details

The proposed model was implemented in PyTorch using an NVIDIA GeForce 3,090 GPU with 24GB memory. The training process involved 500 epochs, and employed the following settings: Adam optimization, with an initial learning rate of 0.01; a batch size of 2 for HRF segmentation; and a batch size of 1 for retinal layer segmentation. To enhance training stability, we adopted a poly learning rate policy, with a momentum of 0.9.

3.1.2. Evaluation metrics

To quantitatively assess the proposed network's segmentation performance, we employ the following metrics. The Dice Similarity Coefficient (DSC) quantifies the agreement between HRF manually annotated by expert ophthalmologists and those automatically segmented by the proposed network, which can be defined as:

$$DSC = \frac{2TP}{FP + FN + 2TP} \tag{5}$$

We also assess our method using Intersection over Union (IOU), precision, recall, and F1-Score, defined as:

$$IOU = \frac{TP}{FP + FN + TP} \tag{6}$$

$$Precision = \frac{TP}{FP + TP} \tag{7}$$

$$Recall = \frac{TP}{FN + TP} \tag{8}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

where TP indicates true positives, FP indicates false positives, TN indicates true negatives, and FN indicates false negatives.

3.1.3. Comparison of different segmentation methods

In order to evaluate the effectiveness of the proposed network, we selected several state-of-the-art neural networks for comparison, including FCN (30), U-Net (31), U-Net++ (32), Res U-Net (33), 3D U-Net (34), SW-3DUNet (15), SANet (16), DBR-Net (17). The evaluation metrics utilized include the DSC, IOU, precision, recall, and F1 Score, as detailed in Table 1. We show that the proposed network outperforms other methods regarding DSC, IOU, and precision. Although the proposed method has a slightly lower recall rate than U-Net, when we consider both precision and recall comprehensively, the proposed method outperforms in terms of the F1 Score.

As seen in Figure 4, our proposed network outperforms at identifying complete HRF regions and avoiding errors in segmentation when compared to other methods in the task of HRF segmentation for DME diseases. This indicates that the proposed network can effectively extract detailed HRF features and analyze them, by combining robust pre-processing capabilities from the baseline network and embedding spatial and channel attention modules. As a result, there is a notable improvement in segmentation effectiveness.

3.1.4. Ablation experiment

To demonstrate the effectiveness of the channel attention and spatial attention modules, we compared our proposed method with the baseline method and two variants.

TABLE 1 Comparison of different segmentation methods.

Method	DSC (%)	IOU (%)	Recall (%)	Precision (%)	F1 Score (%)
FCN	59.31 ± 9.30	44.60 ± 9.32	66.07 ± 7.93	57.69 ± 9.69	59.31 ± 9.30
U-Net	62.12 ± 8.91	47.84 ± 9.42	**68.89 ± 8.44**	60.44 ± 8.64	62.12 ± 8.91
U-Net++	61.60 ± 9.26	47.48 ± 9.73	67.43 ± 6.60	60.37 ± 11.26	61.60 ± 9.26
Res U-Net	63.93 ± 8.82	50.32 ± 9.31	62.71 ± 9.20	71.37 ± 7.66	63.93 ± 8.82
3D U-Net	61.45 ± 8.64	49.41 ± 10.18	58.64 ± 12.91	65.58 ± 8.73	61.45 ± 8.64
SW-3DUNet	51.18 ± 9.87	36.26 ± 9.08	60.68 ± 12.10	47.62 ± 10.17	51.18 ± 9.87
SANet	64.33 ± 8.68	50.59 ± 9.31	63.09 ± 8.23	70.90 ± 8.40	64.33 ± 8.68
DBR-Net	51.14 ± 7.57	36.71 ± 6.77	48.88 ± 7.69	59.99 ± 6.03	51.14 ± 7.57
Proposed method	**66.83 ± 9.06**	**56.33 ± 7.08**	61.12 ± 11.9	**82.31 ± 6.39**	**66.83 ± 9.06**

[1]The variable was expressed as the mean ± standard deviation (SD). The bold value represents the optimal result for the column of indicators.

- Baseline + SAB: We removed the CAB from this variant to assess its contribution.
- Baseline + CAB: We removed the SAB from this variant to assess its contribution.
- Baseline: We removed both SAB and CAB to evaluate their combined contribution.

Table 2 presents the experimental results for our proposed method, the baseline, and its two variants. Compared to the results of our proposed method, the variant without SAB exhibited reductions of 0.31% in DSC, 4.29% in IOU, 0.7% in recall, 0.4% in precision, and 0.31% in F1 Score. The variant without CAB showed reductions of 1.15% in DSC, 5% in IOU, 0.55% in recall, 0.75% in precision, and 1.55% in F1 Score. Removing both SAB and CAB resulted in reductions of 1.7% in DSC, 5.65% in IOU, and 2.91% in recall. Although there was a slight increase of 0.41% in precision, there was a decrease of 1.7% in F1 Score. The experimental results above demonstrate the rationality and effectiveness of embedding spatial and channel attention modules in the bottleneck layer of the baseline model.

3.2. Quantitative parameter evaluation

We enrolled 47 eyes from 26 patients with DME, acquired with OCT (Spectralis HRA + OCT), and a total of 11 eyes from 8 patients were included in this study. We excluded 36 eyes from 18 patients from the analysis. One eye of one patient was excluded due to poor OCT image quality (motion artifacts on OCT images): 22 eyes of 11 patients were excluded due to lack of follow-up records; and 13 eyes from 7 patients were excluded due to no improvement in CMT after anti-VEGF or dexamethasone IVI treatment. The characteristics and clinical information of our study participants are displayed in Table 3. Two sets of OCT data were included for each eye, one at baseline, and one at follow-up, for a total of 9,472 OCT B-scans included in the study.

3.2.1. Parametric comparison of baseline hyperreflective foci

Table 4 compares 12 quantitative parameters of HRF, classified by different diameter sizes in WR at baseline. Among the morphology-related parameters, significant differences were observed between Q_{NS} and Q_{HD}, Q_{HD} and Q_{HC}, Q_{NS} and Q_{HC}, V_{NS} and V_{HD}, and V_{NS} and V_{HC} (all $p = 0.003$). No significant differences were found between V_{HD} and V_{HC} ($p = 0.131$). Among the distance-related parameters, the results showed no significant differences between the HRF classified according to their diameter size.

3.2.2. Parametric comparison of follow-up hyperreflective foci

Due to the retrospective design of the study, OCT examinations were not performed at regular intervals. To avoid bias related to the duration of follow-up, only two consecutive follow-up visits with improvement in CMT were selected for each eye. The longitudinal study included 11 eyes from 8 patients, with a follow-up of 1.9 ± 1.6 months (range 1 to 6, median 1). During study period, all eyes were treated with intravitreal injections: 91% (10/11) of eyes received anti-VEGF injections and 9% (1/11) of eyes received dexamethasone injections. The number of intravitreal injections was 1.4 ± 0.7 (range 1 to 3, median 1).

We assessed whether changes in HRF were significant at two consecutive follow-up visits in the presence of improved CMT. Table 5 showed the comparison of the 12 quantitative parameters of HRF in WR, IR, and OR between the pre-IVI and post-IVI stages. In WR, Q_{HD}, V_{NS}, V_{HD}, $D_{NS,RPE}$, $D_{HD,RPE}$, $D_{HC,RPE}$ and CMT were significantly reduced in post-IVI compared with pre-IVI ($p = 0.003$, $p = 0.033$, $p = 0.003$, $p = 0.026$, $p = 0.008$, $p = 0.004$, $p = 0.003$, respectively). There were no significant changes in the other six quantitative parameters between the two phases. In IR, $D_{NS,RPE}$ and $D_{HD,RPE}$ were significantly reduced in post-IVI compared with pre-IVI ($p = 0.016$, $p = 0.016$, respectively). There were no significant changes in the other 10 quantitative parameters between the two phases. In OR, Q_{HD}, V_{NS}, V_{HD}, $D_{HD,RPE}$, $D_{HC,RPE}$ were significantly reduced in post-IVI compared to pre-IVI ($p = 0.006$, $p = 0.047$, $p = 0.006$, $p = 0.004$, $p = 0.004$, respectively). There were no significant changes in the other seven quantitative parameters between the two phases.

3.2.3. Correlation between follow-up CMT changes and baseline hyperreflective foci

We assessed whether there was a significant correlation between improvement in CMT at two consecutive follow-up visits and baseline HRF. Table 6 shows the correlation between the 12 quantitative parameters of baseline HRF in WR, IR, OR, and the percentage of

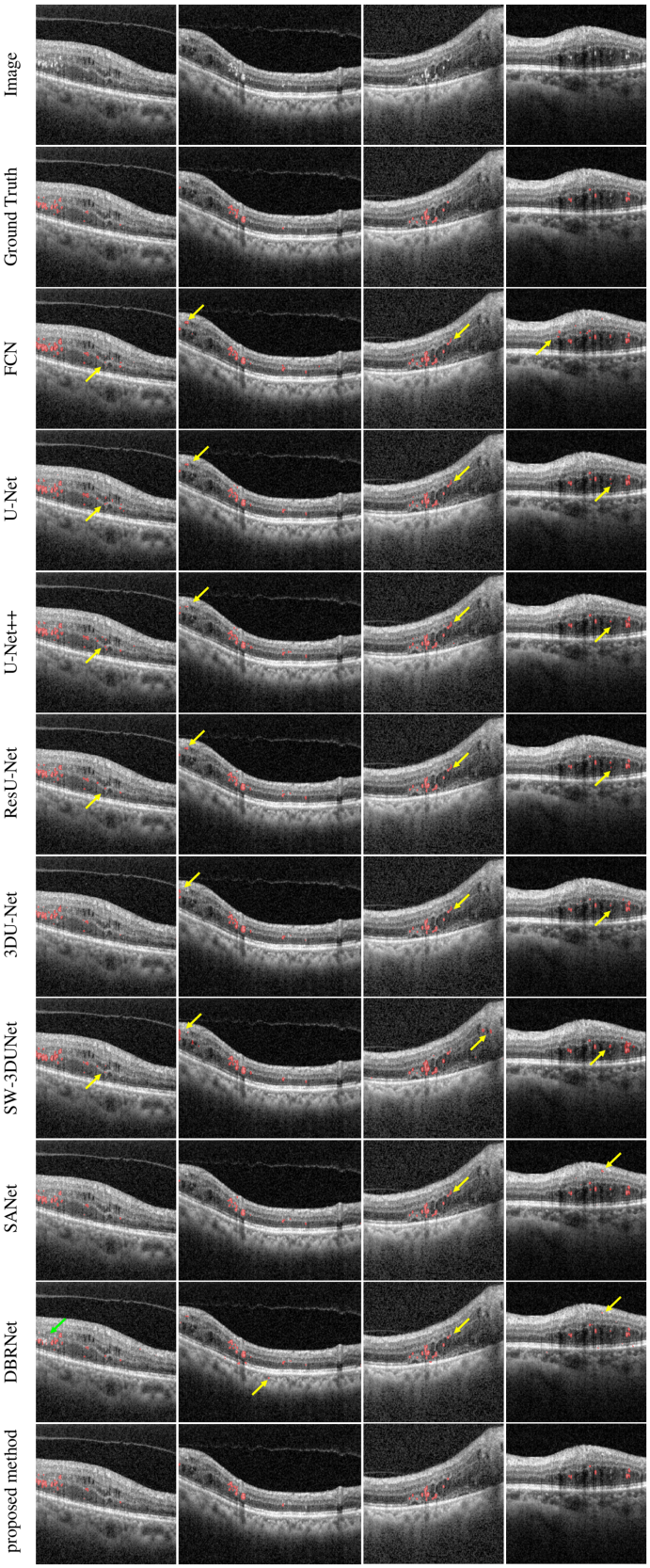

FIGURE 4
Comparison between the proposed SW-3DUNet and other methods. Yellow and green arrows represent the regions of over-segmentation and under-segmentation. B-scans in the second column are taken from fovea-centered regions of $6 \times 6\,mm^2$, while B-scans in the other columns are taken from fovea-centered regions of $4.5 \times 4.5\,mm^2$.

TABLE 2 Ablation experiment.

Method	DSC (%)	IOU (%)	Recall (%)	Precision (%)	F1 score (%)
Baseline	65.13 ± 9.75	50.68 ± 9.89	58.31 ± 12.93	**82.72 ± 5.68**	65.13 ± 9.75
Baseline + SAB	65.68 ± 9.62	51.33 ± 9.65	59.57 ± 12.41	81.56 ± 6.19	65.68 ± 9.62
Baseline + CAB	66.52 ± 9.55	52.04 ± 10.10	60.42 ± 12.37	81.91 ± 6.26	66.52 ± 9.55
Proposed method	**66.83 ± 9.06**	**56.33 ± 7.08**	**61.12 ± 11.9**	82.31 ± 6.39	**66.83 ± 9.06**

[1]The variable was expressed as the mean ± standard deviation (SD). The bold value represents the optimal result for the column of indicators.

TABLE 3 Demographics.

Characteristics	
Number of patients	8
Number of eyes	11
Age, mean ± SD (range), years	48.4 ± 11.0 (33 to 61)
Male gender, *n* (%)	4 (50)
Type 2 diabetes, *n* (%)	8 (100)
Diabetes duration, mean (IQR), years	7.9 (5 to 10)
DR severity, *n* (%)	
Severe non-proliferative DR	10 (91)
Proliferative DR	1 (9)
Baseline BCVA, mean ± SD (range), LogMar	0.59 ± 0.17 (0.4 to 0.8)
Final BCVA, mean ± SD (range), LogMar	0.57 ± 0.16 (0.3 to 0.8)

[1]The retrospective nature of this study resulted in missing data, including the duration of diabetes for one patient, baseline BCVA for one eye, and follow-up BCVA for another eye.
[2]DR, diabetic retinopathy; IQR, interquartile range; BCVA, best-corrected visual acuity; LogMar, logarithm of the minimal angle of resolution.

CMT improvement ($\Delta CMT(\%)$). In WR, significant positive correlations were shown between baseline $D_{NS,fovea}$, $D_{HD,fovea}$, $D_{HC,fovea}$, and $\Delta CMT(\%)$ ($p=0.015$, $p=0.016$, $p<0.001$, respectively). There was no significant correlation between the other nine baseline quantitative parameters and $\Delta CMT(\%)$. In OR, significant positive correlations were shown between baseline $D_{NS,fovea}$, $D_{HD,fovea}$, $D_{HC,fovea}$, and $\Delta CMT(\%)$ ($p=0.006$, $p=0.019$, $p<0.001$, respectively). There was no significant correlation between the other nine baseline quantitative parameters and $\Delta CMT(\%)$. In IR, there was no significant correlation between all 12 baseline quantitative parameters and $\Delta CMT(\%)$.

4. Discussion

The given study aims to quantify HRF in OCT images as part of a retrospective study on patients with DME at baseline and follow-up. Previous studies relied on manual counting methods to quantify HRF, which is time-consuming and less reliable. In examining HRF as a potential biomarker, the existing body of literature has been inconsistent (23, 35–42), which may be due to variations in the OCT tool used, image quality, and manual segmentation of HRF. To address these challenges, our study employed artificial intelligence techniques for quantifying HRF, thereby overcoming some limitations of previous studies. With artificial intelligence, retinal images can be analyzed in a completely new way. We showed that the three subtypes of HRF were significantly different in volume and number on retinal OCT images, with HC pre-dominating in volume and HD in number. We also showed that the mean distance from HRF to RPE was reduced after IVI treatment compared to before IVI treatment. In addition, we showed that eyes with less HRF in the center of the macula showed greater reduction in macular edema after IVI treatment. These findings validate previous findings and suggest new insights, emphasizing the potential of deep learning as a powerful tool for analyzing baseline and follow-up HRF in DME patients.

4.1. Differences between baseline HRF parameters

Statistical analysis indicated significant disparities in both the number and volume parameters of baseline HRF subtypes. Our study validated HRF discrimination based on diameter range by analyzing baseline HRF morphological parameters. A previous study used 20 μm and 50 μm diameters to differentiate HRF subtypes (8), found a positive correlation between the number of HRF subtypes in the 20–50 μm range and the levels of CD14, without discussing the other two subtypes. Our study revealed significant differences among the three subtypes. Our findings corroborated previous studies showing that smaller HRFs merge into larger HRF (7), and show differential treatment responses (23). Furthermore, our study observed different responses to IVI treatment in the number and volume of the smallest HRF subtype, which may include microglia cells, whose activation decreased with treatment (43).

4.2. Follow-up findings

We showed the mean distance from HRF to RPE was significantly reduced after IVI treatment. By studying the distribution parameters of HRF in a longitudinal analysis of two consecutive follow-ups, our study indicated the tendency of HRF to migrate from the inner retina to the outer retina after IVI treatment; similar to our findings, Pemp et al. showed that DME uptake triggered the downward migration of HRF (44) into the outer retina. Notably, despite the lack of response to IVI treatment, the largest diameter HRF subtype exhibited a significant reduction in mean distance to the RPE in both OR and WR. This finding was consistent with Marmor's mechanistic model of retinal fluid movement (45), which postulated fluid flowed across the retina due to intraocular pressure, choroidal osmolarity, and active fluid uptake by the RPE. The migration of partial HRF was impeded by narrow channels on the ELM, composed of zonular adhesions between Müller cells and photoreceptor inner segments. Consequently,

TABLE 4 Parametric comparisons of hyperreflective foci of different diameters.

Variable, in WR	Morphology-related parameters		Variable, in WR	Distribution-related parameters	
	Pre-IVI, $n = 11$	P		Pre-IVI, $n = 11$	P
Q_{NS} Q_{HD}	60 (26–101) 239 (76–411)	0.003	$D_{NS,fovea}$ $D_{HD,fovea}$	991.35 (959.52–1072.49) 982.67 (921.75–1028.42)	0.062
Q_{HD} Q_{HC}	239 (76–411) 20 (8–37)	0.003	$D_{HD,fovea}$ $D_{HC,fovea}$	982.67 (921.75–1028.42) 985.87 (909.17–1093.25)	1.0
Q_{NS} Q_{HC}	60 (26–101) 20 (8–37)	0.003	$D_{NS,fovea}$ $D_{HC,fovea}$	991.35 (959.52–1072.49) 985.87 (909.17–1093.25)	0.286
V_{NS} V_{HD}	2.88×10^5 (1.14×10^5–3.68×10^5) 9.22×10^6 (2.56×10^6–1.30×10^7)	0.003	$D_{NS,RPE}$ $D_{HD,RPE}$	216.45 (204.38–239.87) 213.52 (188.28–230.87)	0.424
V_{HD} V_{HC}	9.22×10^6 (2.56×10^6–1.30×10^7) 1.18×10^7 (1.53×10^6–2.86×10^7)	0.131	$D_{HD,RPE}$ $D_{HC,RPE}$	213.52 (188.28–230.87) 213.86 (202.93–273.82)	0.110
V_{NS} V_{HC}	2.88×10^5 (1.14×10^5–3.68×10^5) 1.1810^7 (1.53×10^6–2.86×10^7)	0.003	$D_{NS,RPE}$ $D_{HC,RPE}$	216.45 (204.38–239.87) 213.86 (202.93–273.82)	0.062

[1]The variable was expressed as the median (IQR); the p value was obtained by Wilcoxon Signed Rank Test. [2] WR, the whole retinal layer; IVI, intravitreal injection; Q_{NS}, noise signal quantity; Q_{HD}, hyperreflective dots quantity; Q_{HC}, hyperreflective clump quantity; V_{NS} noise signal volume; e; V_{HD}, hyperreflective dots volume; V_{HC}, hyperreflective clump volume; $D_{NS,fovea}$, distance between noise signal and fovea; $D_{HD,fovea}$, distance between hyperreflective dots and fovea; $D_{HC,fovea}$, distance between hyperreflective clumps and fovea; $D_{NS,RPE}$, distance between noise signal and RPE; $D_{HD,RPE}$, distance between hyperreflective dots and RPE; $D_{HC,RPE}$, distance between hyperreflective clumps and RPE.

this fraction of HRF aggregated in front of the ELM, forming the HRF isoform with the largest diameter, supporting the study conducted by Bolz et al. (7). However, no evidence was found to indicate migration of HRF toward or away from the fovea after IVI treatment, suggesting that IVI treatment did not significantly impact the distribution of HRF in the direction of the fovea. Further studies are required to confirm this conjecture.

4.3. Correlations between HRF parameters and CMT improvement

We also showed that improvement in CMT resulting from IVI treatment was positively correlated with the mean distance from HRF to the fovea in OR and WR, and not significantly correlated with other parameters. We explored the correlation between the therapeutic effect of IVI on foveal edema and the quantitative parameters of HRF at baseline by examining the quantitative parameters of HRF at baseline and the percentage improvement in CMT at two consecutive follow-up visits. A previous study (46) performed two-dimensional quantification of hard exudates in OCT enface images, and found that the area of hard exudates in the fovea at baseline was inversely correlated with BCVA at the 12th month. Similar to the aforementioned report, we showed that in both OR and WR, the percentage of IVI treatment-induced improvement in CMT was inversely correlated with the concentration of HRF in the fovea at baseline, but independent of other quantified parameters. Notably, the concentration of the largest-diameter HRF subtype in the fovea was inversely correlated with the reduction in CMT (r_s =0.882, $p < 0.001$), which could explain why there was no correlation between the concentration of baseline HRF in the fovea in IR and the reduction in CMT, since the convergence of smaller HRF subtypes to larger HRF subtypes mainly occurs in OR according to the discussion above. We speculate that future studies on the differential distribution of HRF aggregated in the fovea may be able to verify whether it produces some physiological changes that affect the outcome of IVI treatment.

4.4. Limitations

Our study has certain limitations. Firstly, all participants were Chinese, and enrolled from a single medical center: larger and more various samples would be an advantage. More multi-center studies should therefore be conducted on larger cohorts to confirm the reproducibility of analysis on these parameters of HRF. Secondly, the enrollment criteria for this study only included cases with a positive response to injection therapy, rather than including refractory cases. It would be more convincing to recruit subjects with definitive treatment and make a long-term follow-up comparison. Thirdly, the study design lacked untreated blank controls to derive reasons for changes in parameters before and after treatment, while the small sample size and high homogeneity made the study findings more indicative of a pilot.

5. Conclusion

We introduced a deep learning-based approach to quantify hyperreflective foci in OCT images of DME patients. Our retrospective analysis of 11 eyes using this method showed that it effectively quantified baseline and follow-up changes in hyperreflective foci by extracting relevant geometric parameters. In this study, we were able to validate certain findings reported in prior research and uncover novel insights: for instance, our investigation revealed that the concentration of HRF in the fovea region may influence the efficacy of IVI treatment. We believe that accurate quantification and follow-up of HRF in OCT images at baseline and during treatment may enable clinicians to monitor DME disease progression, assess treatment response and identify patients who may benefit from a personalized approach to treatment.

TABLE 5 Comparison of parameters between pre-IVI and post-IVI.

	WR			IR			OR		
	Pre-IVI, $n = 11$	Post-IVI, $n = 11$	P	Pre-IVI, $n = 11$	Post-IVI, $n = 11$	P	Pre-IVI, $n = 11$	Post-IVI, $n = 11$	P
Q_{NS}	60 (26–101)	64 (24–84)	0.247	14 (9–36)	15 (9–20)	0.241	43 (21–94)	48 (15–63)	0.533
Q_{HD}	239 (76–411)	210 (71–232)	0.003	33 (20–75)	23 (12–49)	0.173	216 (59–318)	168 (55–209)	0.006
Q_{HC}	20 (8–37)	27 (13–40)	0.213	0 (0–1)	0 (0–1)	0.317	19 (8–37)	27 (12–38)	0.213
V_{NS}	2.88×10^5 (1.14×10^5–3.68×10^5)	2.47×10^5 (1.05×10^5–3.54×10^5)	0.033	5.32×10^4 (3.40×10^4–1.35×10^5)	5.27×10^4 (3.85×10^4–9.58×10^4)	0.374	2.16×10^5 (9.22×10^4–3.45×10^5)	1.94×10^5 (6.43×10^4–2.57×10^5)	0.047
V_{HD}	9.22×10^6 (2.56×10^6–1.30×10^7)	5.66×10^6 (2.55×10^6–8.45×10^6)	0.003	9.71×10^5 (5.74×10^5–2.04×10^6)	7.61×10^5 (2.86×10^5–1.32×10^6)	0.155	7.67×10^6 (1.84×10^6–1.05×10^7)	4.52×10^6 (2.10×10^6–7.69×10^6)	0.006
V_{HC}	1.18×10^7 (1.53×10^6–2.85×10^7)	9.97×10^6 (4.65×10^6–2.66×10^7)	0.929	0 (0–1.53×10^5)	0 (0–2.11×10^5)	0.345	1.18×10^7 (1.53×10^6–2.84×10^7)	9.97×10^6 (4.65×10^6–2.66×10^7)	0.929
$D_{NS,fovea}$	991.35 (959.52–1072.49)	1042.95 (1014.85–1064.91)	0.534	1103.94 (1057.83–1212.58)	1086.73 (1030.12–1237.50)	0.534	954.27 (906.40–1023.99)	1011.02 (960.74–1060.23)	0.594
$D_{HD,fovea}$	980.67 (921.75–1028.42)	1005.93 (973.22–1066.87)	0.131	1035.34 (919.43–1184.14)	1014.14 (956.53–1106.73)	0.79	1000.76 (888.70–1024.25)	1011.21 (948.34–1083.34)	0.131
$D_{HC,fovea}$	985.87 (909.17–1093.25)	970.50 (833.29–1020.72)	0.79	0 (0–909.00)	0 (0–664.33)	0.893	9983.01 (909.17–1093.25)	970.50 (833.29–1022.10)	0.79
$D_{NS,RPE}$	216.45 (204.38–239.87)	190.71 (160.61–202.84)	0.026	293.66 (260.66–322.31)	234.73 (197.79–246.13)	0.013	191.16 (179.96–214.40)	174.93 (147.46–185.65)	0.075
$D_{HD,RPE}$	213.52 (188.28–230.87)	177.60 (162.27–200.99)	0.008	278.50 (270.14–336.15)	242.80 (210.56–249.98)	0.016	196.62 (181.72–211.72)	172.52 (149.30–192.06)	0.004
$D_{HC,RPE}$	213.86 (202.93–273.82)	177.44 (167.83–204.70)	0.004	0 (0–274.30)	0 (0–242.46)	0.686	213.28 (202.93–273.82)	177.44 (167.83–204.70)	0.004
CMT	401 (383–490)	315 (253–367)	0.003						

[1]the variable was expressed as the median (IQR); the P value was obtained by Wilcoxon Signed Rank Test. [2]WR, the whole retinal layer; IR, the inner retinal layer; OR, the outer retinal layer; IVI, intravitreal injection; Q_{NS}, noise signal quantity; Q_{HD}, hyperreflective dots quantity; Q_{HC}, hyperreflective clump quantity; V_{NS}, noise signal volume; V_{HD}, hyperreflective dots volume; V_{HC}, hyperreflective clump volume; $D_{NS,fovea}$, distance between noise signal and fovea; $D_{HD,fovea}$, distance between hyperreflective dots and fovea; $D_{HC,fovea}$, distance between hyperreflective clumps and fovea; $D_{NS,RPE}$, distance between noise signal and RPE; $D_{HD,RPE}$, distance between hyperreflective dots and RPE; $D_{HC,RPE}$, distance between hyperreflective clumps and RPE; CMT, central macular thickness.

10.3389/fmed.2023.1280714

TABLE 6 Results of spearman correlation analysis.

$\Delta CMT(\%)$	Pre-IVI, $n = 11$					
	WR		IR		OR	
	r_S	P	r_S	P	r_S	P
Q_{NS}	0.355	0.285	0.118	0.729	0.273	0.17
Q_{HD}	0.509	0.110	0.150	0.629	0.464	0.151
Q_{HC}	0.600	0.051	−0.064	0.852	0.600	0.051
V_{NS}	0.355	0.285	0.091	0.790	0.300	0.370
V_{HD}	0.573	0.066	0.227	0.502	0.500	0.117
V_{HC}	0.527	0.090	−0.092	0.787	0.527	0.096
$D_{NS,fovea}$	0.709	0.015	0.182	0.593	0.764	0.006
$D_{HD,fovea}$	0.700	0.016	−0.236	0.484	0.691	0.019
$D_{HC,fovea}$	0.882	0.000	−0.035	0.919	0.882	0.000
$D_{NS,RPE}$	0.373	0.259	0.255	0.450	0.255	0.450
$D_{DF,RPE}$	0.200	0.555	0.336	0.312	0.309	0.355
$D_{HC,RPE}$	0.355	0.285	−0.185	0.586	0.355	0.285

[1]The p value was obtained by Spearman Rank Correlation analysis. r_S is the rank correlation coefficient to indicate the closeness and direction of correlation between the two variables seen in a linear correlation. [2]WR, the whole retinal layer; IR, the inner retinal layer; OR, the outer retinal layer; IVI, intravitreal injection; Q_{NS}, noise signal quantity; Q_{HD}, hyperreflective dots quantity; Q_{HC}, hyperreflective clump quantity; V_{NS}, noise signal volume; V_{HD}, hyperreflective dots volume; V_{HC}, hyperreflective clump volume; $D_{NS,fovea}$, distance between noise signal and fovea; $D_{HD,fovea}$, distance between hyperreflective dots and fovea; $D_{HC,fovea}$, distance between hyperreflective clumps and fovea; $D_{NS,RPE}$, distance between noise signal and RPE; $D_{HD,RPE}$, distance between hyperreflective dots and RPE; $D_{HC,RPE}$, distance between hyperreflective clumps and RPE; $\Delta CMT(\%)$ ratio of central macular thickness reduction to baseline central macular thickness.

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 humans were approved by the ethics committee of the Affiliated Ningbo Eye Hospital of Wenzhou Medical University (ID: 20210327A). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XW: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Formal analysis. YaZ: Conceptualization, Formal analysis, Resources, Writing – review & editing. YM: Writing – review & editing. WK: Conceptualization, Formal analysis, Writing – review & editing. JY: Investigation, Writing – review & editing. JL: Software, Writing – review & editing. SM: Conceptualization, Writing – review & editing. QiY: Conceptualization, Validation, Writing – review & editing. QuY: Conceptualization, Resources, Writing – review & editing. YiZ: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China (62272444); Zhejiang Provincial Natural Science Foundation (LR22F020008); Youth Innovation Promotion Association CAS (2021298). National Science Foundation Program of China (62302488); Zhejiang Provincial Natural Science Foundation of China (LQ23F010007).

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

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

References

1. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet*. (2010) 376:124–36. doi: 10.1016/S0140-6736(09)62124-3

2. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. (2012) 35:556–64. doi: 10.2337/dc11-1909

3. Jampol LM, Glassman AR, Sun JK. Evaluation and Care of Patients with diabetic retinopathy. *N Engl J Med*. (2020) 382:1629–37. doi: 10.1056/NEJMra1909637

4. Klein R, Lee KE, Danforth L, Tsai MY, Gangnon RE, Meuer SE, et al. The relationship of retinal vessel geometric characteristics to the incidence and progression of diabetic retinopathy. *Ophthalmology*. (2018) 125:1784–92. doi: 10.1016/j.ophtha.2018.04.023

5. Suciu CI, Suciu VI, Nicoara SD. Optical coherence tomography (angiography) biomarkers in the assessment and monitoring of diabetic macular edema. *J Diabetes Res*. (2020) 2020:6655021–10. doi: 10.1155/2020/6655021

6. Olson J, Sharp P, Goatman K, Prescott G, Scotland G, Fleming A, et al. Improving the economic value of photographic screening for optical coherence tomography-detectable macular oedema: a prospective, multicentre, UK study. *Health Technol Assess*. (2013) 17:1–142. doi: 10.3310/hta17510

7. Bolz M, Schmidt-Erfurth U, Deak G, Mylonas G, Kriechbaum K, Scholda C. Optical coherence tomographic hyperreflective foci: a morphologic sign of lipid extravasation in diabetic macular edema. *Ophthalmology*. (2009) 116:914–20. doi: 10.1016/j.ophtha.2008.12.039

8. Lee H, Jang H, Choi YA, Kim HC, Chung H. Association between soluble CD14 in the aqueous humor and Hyperreflective foci on optical coherence tomography in patients with diabetic macular edema. *Invest Ophthalmol Vis Sci*. (2018) 59:715–21. doi: 10.1167/iovs.17-23042

9. Curcio CA, Zanzottera EC, Ach T, Balaratnasingam C, Freund KB. Activated retinal pigment epithelium, an optical coherence tomography biomarker for progression in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. (2017) 58:BIO211–26. doi: 10.1167/iovs.17-21872

10. Uji A, Murakami T, Nishijima K, Akagi T, Horii T, Arakawa N, et al. Association between hyperreflective foci in the outer retina, status of photoreceptor layer, and visual acuity in diabetic macular edema. *Am J Ophthalmol*. (2012) 153:710–717.e1. doi: 10.1016/j.ajo.2011.08.041

11. Fragiotta S, Abdolrahimzadeh S, Dolz-Marco R, Sakurada Y, Gal-Or O, Scuderi G. Significance of Hyperreflective foci as an optical coherence tomography biomarker in retinal diseases: characterization and clinical implications. *J Ophthalmol*. (2021) 2021:1–10. doi: 10.1155/2021/6096017

12. Okuwobi IP, Fan W, Yu C, Yuan S, Liu Q, Zhang Y, et al. Automated segmentation of hyperreflective foci in spectral domain optical coherence tomography with diabetic retinopathy. *J Med Imag*. (2018) 5:014002:1. doi: 10.1117/1.JMI.5.1.014002

13. Okuwobi IP, Ji Z, Fan W, Yuan S, Bekalo L, Chen Q. Automated quantification of hyperreflective foci in SD-OCT with diabetic retinopathy. *IEEE J Biomed Health Inform*. (2019) 24:1125–36. doi: 10.1109/JBHI.2019.2929842

14. Yu C, Xie S, Niu S, Ji Z, Fan W, Yuan S, et al. Hyper-reflective foci segmentation in SD-OCT retinal images with diabetic retinopathy using deep convolutional neural networks. *Med Phys*. (2019) 46:4502–19. doi: 10.1002/mp.13728

15. Xie S, Okuwobi IP, Li M, Zhang Y, Yuan S, Chen Q. Fast and automated hyperreflective foci segmentation based on image enhancement and improved 3D U-net in SD-OCT volumes with diabetic retinopathy. *Transl Vis Sci Technol*. (2020) 9:21. doi: 10.1167/tvst.9.2.21

16. Yao C, Zhu W, Wang M, Zhu L, Huang H, Chen H, et al. SANet: a self-adaptive network for hyperreflective foci segmentation in retinal OCT images. *Med Imag*. (2021) 11596:809–15. doi: 10.1117/12.2580699

17. Wei J, Yu S, Du Y, Liu K, Xu Y, Xu X. Automatic segmentation of Hyperreflective foci in OCT images based on lightweight DBR network. *J Digit Imaging*. (2023) 36:1148–57. doi: 10.1007/s10278-023-00786-0

18. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

19. Zur D, Iglicki M, Busch C, Invernizzi A, Mariussi M, Loewenstein A, et al. OCT biomarkers as functional outcome predictors in diabetic macular edema treated with dexamethasone implant. *Ophthalmology*. (2018) 125:267–75. doi: 10.1016/j.ophtha.2017.08.031

20. Bosche F, Andresen J, Li D, Holz F, Brinkmann C. Spectralis OCT1 versus OCT2: time efficiency and image quality of retinal nerve Fiber layer thickness and Bruch's membrane opening analysis for Glaucoma patients. *J Curr Glaucoma Pract*. (2019) 13:16–20. doi: 10.5005/jp-journals-10078-1244

21. Tewarie P, Balk L, Costello F, Green A, Martin R, Schippling S, et al. The OSCAR-IB consensus criteria for retinal OCT quality assessment. *PLoS One*. (2012) 7:e34823. doi: 10.1371/journal.pone.0034823

22. Aytulun A, Cruz-Herranz A, Aktas O, Balcer LJ, Balk L, Barboni P, et al. APOSTEL 2.0 recommendations for reporting quantitative optical coherence tomography studies. *Neurology*. (2021) 97:68–79. doi: 10.1212/WNL.0000000000012125

23. Rübsam A, Wernecke L, Rau S, Pohlmann D, Müller B, Zeitz O, et al. Behavior of SD-OCT detectable hyperreflective foci in diabetic macular edema patients after therapy with anti-VEGF agents and dexamethasone implants. *J Diabetes Res*. (2021) 2021:1–13. doi: 10.1155/2021/8820216

24. de Moura J, Samagaio G, Novo J, Almuina P, Fernández MI, Ortega M. Joint diabetic macular edema segmentation and characterization in OCT images. *J Digit Imaging*. (2020) 33:1335–51. doi: 10.1007/s10278-020-00360-y

25. Mou L, Zhao Y, Fu H, Liu Y, Cheng J, Zheng Y, et al. CS2-net: deep learning segmentation of curvilinear structures in medical imaging. *Med Image Anal*. (2021) 67:101874. doi: 10.1016/j.media.2020.101874

26. Ulyanov D, Vedaldi A, Lempitsky V. (2016). Instance normalization: The missing ingredient for fast stylization. Available at: https://arxiv.org/abs/1607.08022

27. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *ICML*. (2015) 37:448–56.

28. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation[C]//international workshop on deep learning in medical image analysis In: *International workshop on large-scale annotation of biomedical data and expert label synthesis*. Eds. Carneiro G, Mateus D, Peter L, Bradley A, Manuel J, Tavares R. et al Cham: Springer (2016). 179–87.

29. Szeto SK, Hui VW, Tang FY, Yang D, Han Sun Z, Mohamed S, et al. OCT-based biomarkers for predicting treatment response in eyes with Centre-involved diabetic macular oedema treated with anti-VEGF injections: a real-life retina clinic-based study. *Br J Ophthalmol*. (2023) 107:525–33. doi: 10.1136/bjophthalmol-2021-319587

30. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:640–51. doi: 10.1109/TPAMI.2016.2572683

31. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *MICCAI*. (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

32. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation. Deep learn med image anal multimodal learn Clin Decis support. *PRO*. (2018) 4:3–11. doi: 10.1007/978-3-030-00889-5_1

33. Schlegl T, Bogunovic H, Klimscha S, Seeboeck P, Sadeghipour A, Gerendas BS, et al. (2018). Fully automated segmentation of hyperreflective foci in optical coherence tomography images. Available at: https://arxiv.org/abs/1805.03278

34. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *MICCAI*. (2016) 9901:424–32. doi: 10.1007/978-3-319-46723-8_49

35. Framme C, Schweizer P, Imesch M, Wolf S, Wolf-Schnurrbusch U. Behavior of SD-OCT-detected hyperreflective foci in the retina of anti-VEGF-treated patients with diabetic macular edema. *Invest Ophthalmol Vis Sci*. (2012) 53:5814–8. doi: 10.1167/iovs.12-9950

36. Vujosevic S, Torresin T, Bini S, Convento E, Pilotto E, Parrozzani R, et al. Imaging retinal inflammatory biomarkers after intravitreal steroid and anti-VEGF treatment in diabetic macular oedema. *Acta Ophthalmol*. (2017) 95:464–71. doi: 10.1111/aos.13294

37. Liu S, Wang D, Chen F, Zhang X. Hyperreflective foci in OCT image as a biomarker of poor prognosis in diabetic macular edema patients treating with Conbercept in China. *BMC Ophthalmol*. (2019) 19:157. doi: 10.1186/s12886-019-1168-0

38. Kang JW, Chung H, Chan KH. Correlation of optical coherence tomographic Hyperreflective foci with visual outcomes in different patterns of diabetic macular edema. *Retina*. (2016) 36:1630–9. doi: 10.1097/IAE.0000000000000995

39. Ceravolo I, Oliverio GW, Alibrandi A, Bhatti A, Trombetta L, Rejdak R, et al. The application of structural retinal biomarkers to evaluate the effect of intravitreal Ranibizumab and dexamethasone intravitreal implant on treatment of diabetic macular edema. *Diagnostics (Basel)*. (2020) 10:413. doi: 10.3390/diagnostics10060413

40. Schreur V, Altay L, van Asten F, Groenewoud JM, Fauser S, Klevering BJ, et al. Hyperreflective foci on optical coherence tomography associate with treatment outcome for anti-VEGF in patients with diabetic macular edema. *PLoS One*. (2018) 13:e0206482. doi: 10.1371/journal.pone.0206482

41. Narnaware SH, Bawankule PK, Raje D. Short-term outcomes of intravitreal dexamethasone in relation to biomarkers in diabetic macular edema. *Eur J Ophthalmol*. (2021) 31:1185–91. doi: 10.1177/1120672120925788

42. Vujosevic S, Berton M, Bini S, Casciano M, Cavarzeran F, Midena E. Hyperreflective retinal spots and visual function after anti-vascular endothelial growth factor treatment in center-involving diabetic macular edema. *Retina*. (2016) 36:1298–308. doi: 10.1097/IAE.0000000000000912

43. Grigsby JG, Cardona SM, Pouw CE, Muniz A, Mendiola AS, Tsin AT, et al. The role of microglia in diabetic retinopathy. *J Ophthalmol*. (2014) 2014:1–15. doi: 10.1155/2014/705783

44. Pemp B, Deák G, Prager S, Mitsch C, Lammer J, Schmidinger G, et al. Distribution of intraretinal exudates in diabetic macular edema during anti-vascular endothelial growth factor therapy observed by spectral domain optical coherence tomography and fundus photography. *Retina*. (2014) 34:2407–15. doi: 10.1097/IAE.0000000000000250

45. Marmor MF. Mechanisms of fluid accumulation in retinal edema. *Doc Ophthalmol*. (1999) 97:239–49. doi: 10.1023/a:1002192829817

46. Srinivas S, Verma A, Nittala MG, Alagorie AR, Nassisi M, Gasperini J, et al. Effect of intravitreal ranibizumab on intraretinal hard exudates in eyes with diabetic macular edema. *Am J Ophthalmol*. (2020) 211:183–90. doi: 10.1016/j.ajo.2019.11.014

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 17 November 2023
DOI 10.3389/fmed.2023.1308923

Check for updates

OPEN ACCESS

EDITED BY
Haoyu Chen,
The Chinese University of Hong Kong, China

REVIEWED BY
Yi-Ting Hsieh,
National Taiwan University Hospital, Taiwan
Mengxi Shen,
University of Miami Health System,
United States

*CORRESPONDENCE
Xiaobing Yu
✉ yuxiaobing1214@163.com

RECEIVED 07 October 2023
ACCEPTED 06 November 2023
PUBLISHED 17 November 2023

CITATION
Liu J, Li H, Zhou Y, Zhang Y, Song S, Gu X, Xu J and Yu X (2023) Deep learning-based estimation of axial length using macular optical coherence tomography images.
Front. Med. 10:1308923.
doi: 10.3389/fmed.2023.1308923

COPYRIGHT
© 2023 Liu, Li, Zhou, Zhang, Song, Gu, Xu and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning-based estimation of axial length using macular optical coherence tomography images

Jing Liu[1,2], Hui Li[1], You Zhou[3], Yue Zhang[1,2], Shuang Song[1], Xiaoya Gu[1], Jingjing Xu[3] and Xiaobing Yu[1,2]*

[1]Department of Ophthalmology, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China, [2]Graduate School of Peking Union Medical College, Beijing, China, [3]Visionary Intelligence Ltd., Beijing, China

Background: This study aimed to develop deep learning models using macular optical coherence tomography (OCT) images to estimate axial lengths (ALs) in eyes without maculopathy.

Methods: A total of 2,664 macular OCT images from 444 patients' eyes without maculopathy, who visited Beijing Hospital between March 2019 and October 2021, were included. The dataset was divided into training, validation, and testing sets with a ratio of 6:2:2. Three pre-trained models (ResNet 18, ResNet 50, and ViT) were developed for binary classification (AL ≥ 26 mm) and regression task. Ten-fold cross-validation was performed, and Grad-CAM analysis was employed to visualize AL-related macular features. Additionally, retinal thickness measurements were used to predict AL by linear and logistic regression models.

Results: ResNet 50 achieved an accuracy of 0.872 (95% Confidence Interval [CI], 0.840–0.899), with high sensitivity of 0.804 (95% CI, 0.728–0.867) and specificity of 0.895 (95% CI, 0.861–0.923). The mean absolute error for AL prediction was 0.83 mm (95% CI, 0.72–0.95 mm). The best AUC, and accuracy of AL estimation using macular OCT images (0.929, 87.2%) was superior to using retinal thickness measurements alone (0.747, 77.8%). AL-related macular features were on the fovea and adjacent regions.

Conclusion: OCT images can be effectively utilized for estimating AL with good performance via deep learning. The AL-related macular features exhibit a localized pattern in the macula, rather than continuous alterations throughout the entire region. These findings can lay the foundation for future research in the pathogenesis of AL-related maculopathy.

KEYWORDS

optical coherence tomography, axial length, artificial intelligence, deep learning, Grad-CAM

1 Introduction

Axial length (AL) is a widely discussed parameter, significant not only for defining the eye's refractive status but also due to its strong association with retinal and macular complications (1, 2). The excessive elongation of AL, often exceeding 26.0 mm, is the dominant cause of an increased risk of posterior segment complications, including vitreous liquefaction, choroidal atrophy, retinoschisis, macular hole, and macular choroidal neovascularization (3). These

complications are vision-threatening and often result in irreversible and permanent vision damage if left untreated (4). In the past, it has not been clear whether there are pre-existing differences in macular structure among eyes with prolonged AL prior to the development of maculopathies, except a few studies have reported that AL was positively associated with central retinal thickness, but negatively associated with peripheral retinal thickness farther from the macula (5–8).

Artificial intelligence, specifically deep learning, has exhibited significant potential in medical imaging diagnosis and interpretation (9, 10). Deep learning allows systems to acquire predictive characteristics directly from an extensive collection of labeled images, eliminating the necessity for explicit rules or manually designed features (11). In recent research, deep learning models have been developed that demonstrate precise estimation of AL or refractive error using color fundus photographs (12–14). Additionally, Yoo et al. (15) have introduced a deep learning model that predicts uncorrected refractive error by utilizing posterior segment optical coherence tomography images, suggesting a potential association between AL and the sectional structure of the retina. Considering that a long AL is a significant risk factor for complications that can potentially impair vision, investigating the alterations in macular structure resulting from prolonged AL prior to the onset of maculopathies holds immense significance in guiding the clinical management and prognosis of patients with long AL eyes (4). However, the application of deep learning to estimate AL based on macular OCT images remains unexplored.

Gradient-weighted class activation mapping (Grad-CAM), a commonly employed approach for visualizing models, utilizes the gradient details that flows into the final convolutional layer of a convolutional neural network (CNN) to construct a heat map that unveils the pivotal regions that are most relevant for the decision-making process (16). This study aimed to assess the capability of macular OCT images to estimate ALs of eyes without maculopathy using deep learning algorithms and visualize the cross-sectional alterations in macular structure resulting from the prolonged AL using Grad-CAM.

2 Materials and methods

2.1 Study design and overview

The data of this study were retrospectively collected from patients who visited the Department of Ophthalmology at Beijing Hospital between January 2019 and October 2021 and were scheduled for cataract surgery. Patients included in the study were required to be aged 18 years or older and have undergone macular OCT examination and AL measurement. Eyes with evident macular abnormalities, such as macular edema, epiretinal membrane, macular hole, macular retinoschisis, and macular neovascularization, were excluded. Furthermore, images of poor quality were also excluded. The study followed the principles of the Declaration of Helsinki and received approval from the institutional review board at Beijing Hospital. Given the retrospective nature of the study, the requirement for written informed consent was waived.

In this study, OCT scans were acquired using the Spectralis OCT device (Heidelberg Engineering, Germany). Images scanned with a stellate scan model centered on the fovea were selected for model development. This scanning model comprises six scans that traverse the fovea, each spanning a length of 6 mm. Moreover, retinal thickness in various subfields was recorded using OCT. The macular region was divided into 9 subfields by employing three concentric circles centered on the fovea, with diameters of 1 mm, 3 mm, and 6 mm. The average thickness of the innermost ring defined the central retinal thickness (CRT). Furthermore, the inner (1–3 mm) and outer (3–6 mm) rings were subdivided into superior, nasal, inferior, and temporal subfields, designated as the parafovea and perifovea, respectively. AL measurements were obtained from the IOL Master 700 (Carl Zeiss, Germany).

2.2 Deep learning model and its training

Figure 1 presents the data management and the flowchart for deep learning models in this study. Two classic CNN models, ResNet18 and ResNet50, along with a Transformer-based model called Vision Transformer (ViT), were introduced to establish the relationship. The detailed description of the models used in this study was presented in Supplementary material 1. In the ViT architecture, the number of encoder blocks was reduced to 6 to prevent overfitting. The input size for the vision transformer is fixed at 224 * 224 to ensure a fair comparison across all models. The SGD (Stochastic Gradient Descent) method serves as the optimizer for all three models. AL measurements obtained by IOL Master 700 (Carl Zeiss, Germany) served as the ground truth for AL prediction. The prediction task is divided into a regression task and a binary classification task by adjusting the dimension of the output result for comprehensive evaluation. To improve accuracy and efficiency, we implement a transfer learning strategy using models pretrained on ImageNet. The salient areas of the feature maps in the latter layers of these models are visualized using the Grad-CAM interpretability method, which illustrates the contribution of each pixel to the final decision.

During training, we employ multiple data augmentation methods to enhance the model's generalization ability. The random resize crop strategy is used to capture different parts of the image with varying scales. Furthermore, horizontal flipping, color jittering, gamma transformation, and random Gaussian noise are applied to augment the training samples for OCT data. Eventually, we implement the normalization to scale the training and testing input from 0 to 1. To expand the dataset, each case is considered independent and equipped with 6 OCT B-scans. This allows us to formulate a dataset with 2,664 images. The data split ratio for training, validation, and testing was 6:2:2, and the split was randomized based on the AL. The training set and validation set were combined, and a 10-fold cross-validation was conducted to demonstrate the reliability of the methods. In the 10-fold cross-validation, the training instances are divided into 10 equally-sized partitions with similar class distributions. Subsequently, each partition is sequentially employed as the test dataset for the classifier generated using the remaining nine partitions.

2.3 Statistical analysis

The classification task utilized the cross-entropy loss function, and various metrics such as sensitivity, specificity, area under the receiver operating characteristic curve (AUC), and accuracy were calculated

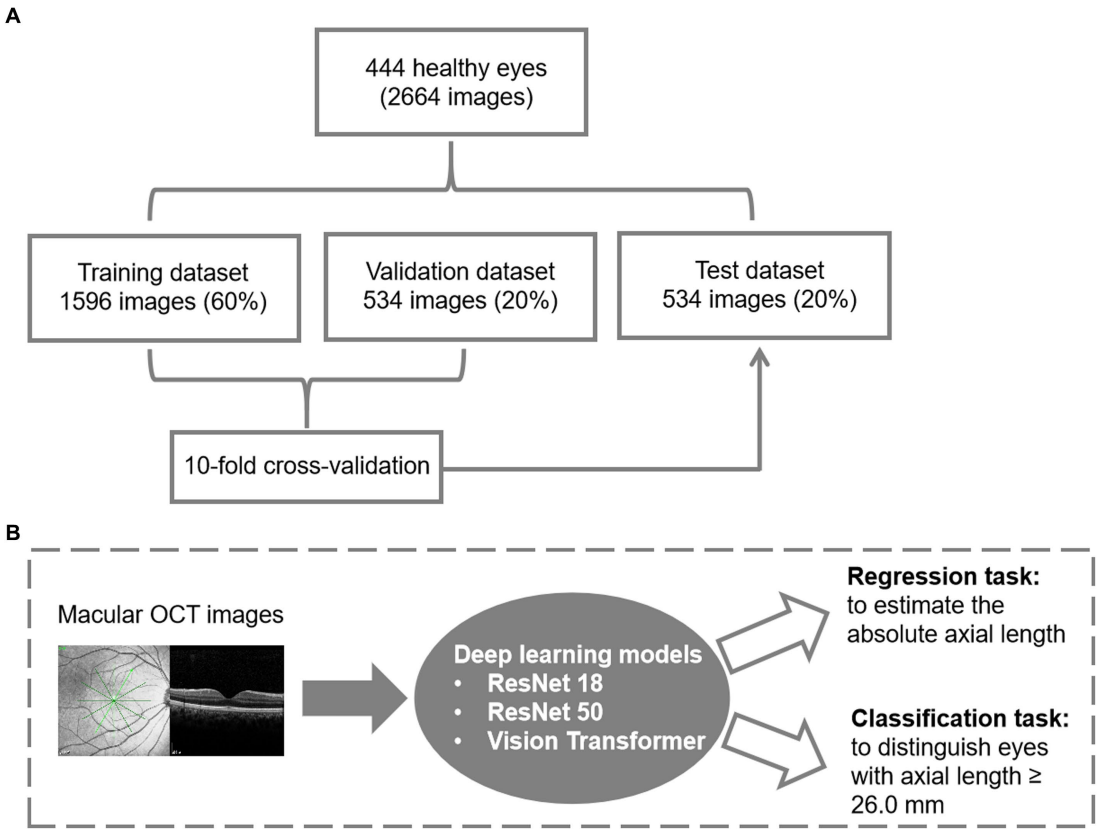

FIGURE 1
Datasets and the architecture of the deep learning model. **(A)** Data management for model development. **(B)** The flowchart for deep learning.

to evaluate performance. In the regression task, the MAELoss function was used as the loss function, and the mean absolute error (MAE) was used as the evaluation metric. The agreement between the actual and predicted AL was assessed using the Bland–Altman plot. The Y-axis represents the difference between the actual and predicted ALs, and the X-axis represents the average of the actual and predicted ALs. The mean difference (MD) and 95% limits of agreement (MD ± 1.96 standard deviations) were calculated to assess the agreement.

TABLE 1 Summary of the demographical characteristics of training, validation, and test data sets.

	Training set	Validation set	Test set
No. of eyes	266	89	89
No. of images	1,596	534	534
Age, year	69.36 ± 10.52	67.89 ± 10.65	69.21 ± 9.63
Sex, male, *n* (%)	121 (45.5%)	43 (48.3%)	41 (46.1%)
AL, mm	24.67 ± 2.19	24.73 ± 2.19	24.73 ± 2.33
AL < 26 mm	199	66	66
AL ≥ 26 mm	67	23	23

AL, axial length.

3 Results

3.1 AI models performance

Finally, a total of 2,664 images from 444 eyes (306 patients) were included in the model development. The mean age was 69.02 ± 10.37 years. Among 444 eyes, 113 eyes (25.5%) were high myopic without maculopathy (AL ≥ 26.0 mm). Finally, 266 eyes (1,596 images) were used for training (60%), 89 eyes (534 images) for validation (20%), and 89 eyes (534 images) for testing (20%). The mean age for the training, validation and testing set were 69.36 ± 10.52, 67.89 ± 10.65, and 69.21 ± 9.63 years, respectively. Demographic characteristics of each dataset are summarized in Table 1. Three models (ResNet 50, ResNet 18, and ViT) were developed for the binary classification task of distinguishing AL ≥ 26.0 mm from others. The 10-fold cross-validation results showed the robust performance and high discriminative power of all three models, as illustrated in Table 2. On the test dataset, ResNet 18, ResNet 50, and ViT achieved AUC (95% Confidence Interval [CI]) values of 0.918 (0.886–0.951), 0.929 (0.899–0.960), and 0.924 (0.892–0.955), respectively (as shown in Figure 2A). ResNet 50 and ResNet 18 had the same accuracy of 0.872 (95%CI, 0.840–0.899), which was the highest among the models. ResNet 50 also exhibited the highest performance, with a sensitivity of 0.804 (95%CI, 0.728–0.867) and specificity of 0.895 (95%CI, 0.861–0.923). Therefore, based on the classification results, particularly the AUC and accuracy, ResNet 50 was selected for further analyses.

The ResNet 50 model was employed for the regression task. The MAE for predicting AL on the test dataset was 0.83 mm (95%CI,

TABLE 2 Performance of deep learning models for binary task (axial length ≥ 26.0 mm).

	Mean results of 10-fold cross validation				Test set (95% CI)			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
ResNet 18	0.908 ± 0.048	0.898 ± 0.042	0.807 ± 0.107	0.997 ± 0.008	0.918 (0.886, 0.951)	0.872 (0.840, 0.899)	0.783 (0.704, 0.848)	0.902 (0.869, 0.929)
ResNet 50	0.932 ± 0.048	0.906 ± 0.033	0.920 ± 0.082	1.000	0.929 (0.899, 0.960)	0.872 (0.840, 0.899)	0.804 (0.728, 0.867)	0.895 (0.861, 0.923)
ViT	0.885 ± 0.075	0.884 ± 0.051	0.766 ± 0.151	1.000	0.924 (0.892, 0.955)	0.867 (0.836, 0.895)	0.693 (0.609, 0.769)	0.927 (0.897, 0.951)

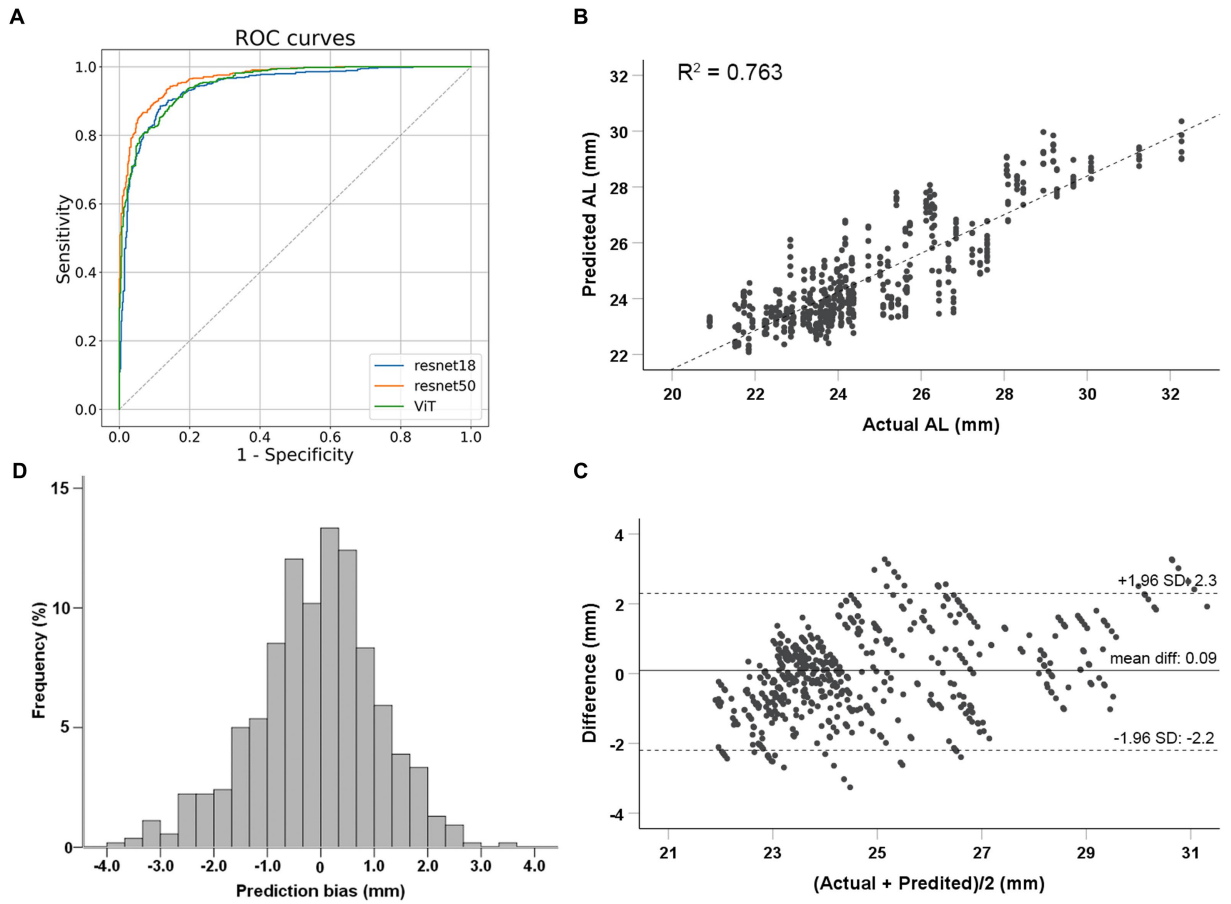

FIGURE 2
Performance evaluation of deep learning models. **(A)** Classification performance of deep learning models to identify eyes with axial lengths ≥26.0 mm in the test dataset. **(B)** Correlations between actual and predicted axial length using the ResNet 50 model. **(C)** Bland–Altman plots for the real and predicted axial length using ResNet 50 in test dataset. **(D)** Prediction bias frequency distribution.

0.72–0.95 mm). The predicted AL and actual AL had a linear relationship with an R^2 of 0.763 in the ResNet 50 model (Figure 2B). Bland–Altman plots revealed a bias of 0.09 mm, with 95% limits of agreement ranging from −2.2 to 2.3 mm (Figure 2C). Prediction bias of 64.8% of the test dataset was less than 1 mm error (Figure 2D); while a calculation of relative bias revealed that 73.1% of the testing difference was within the range of 5% error and 96.5% within 10% error.

3.2 Grad-CAM and model visualization

Grad-CAM was used to identify the regions within the original OCT images that the models relied on for their predictions. Figure 3 shows representative OCT images with their corresponding Grad-CAM from the test set, which were correctly predicted. The heat maps revealed that AL-related macular features exhibit a localized pattern in the macula, rather than continuous alterations throughout the entire region. Both the region of retina and choroid were highlighted in the heat maps. For eyes with ALs < 26.0 mm, the CNN models predominantly relied on the curvature and shape of the fovea, whereas for eyes with ALs ≥ 26.0 mm, the models relied on the regions flanking the fovea, where the most obvious retinal curvature changes.

3.3 Predicting AL based on retinal thicknesses

The macular thickness of the eyes from the test set was recorded. ROC analyses and linear regression analyses were performed to

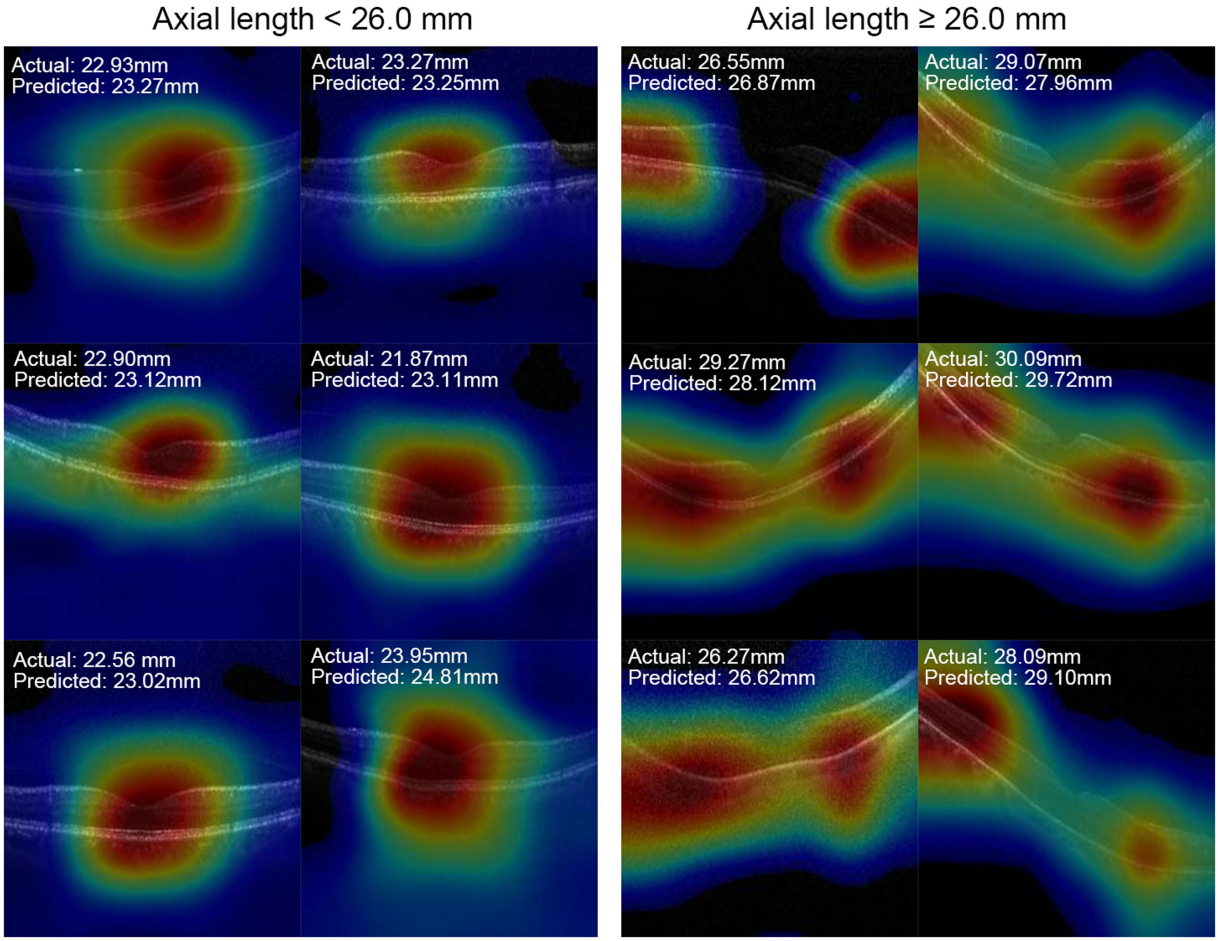

FIGURE 3
Representative OCT images and their heat map images.

predict AL based on retinal thickness in different macular regions. The largest AUC value, 0.747, was obtained for CRT. The highest accuracy in distinguishing long AL eyes was 77.8%, achieved by using retinal thickness measurements in the perifoveal (3–6 mm) nasal quadrant (Supplementary material 2). However, both the AUC and accuracy were lower compared to deep learning models that utilized OCT images ($p < 0.001$). Linear regression analyses showed that the MAE values were 1.78 ± 1.25 mm and 1.57 ± 1.29 mm when using CRT and retinal thickness measurements from all nine regions to predict AL, respectively. These biases were also higher than those observed in deep learning models ($p < 0.001$).

4 Discussion

The present study demonstrated that deep learning models using macular OCT images can accurately estimate AL and differentiate eyes with long AL. The Grad-CAM analysis revealed that the deep learning models primarily relied on the foveal and adjacent regions, as well as the subfoveal choroid for AL estimation. This deep learning model was designed to estimate the AL based on macular OCT images. This study established a significant association between AL and macular structure, demonstrating the AL-related changes in the macular structure as imaged by OCT. These findings provide a solid foundation for research on the pathogenesis of AL-related structural maculopathy.

Previous studies have used fundus photos to estimate AL via developing deep learning models. Dong et al. (12) and Jeong et al. (17) reported the use of CNN models to estimate AL based on 45 degrees fundus photographs, achieving MAE values of 0.56 mm (95% CI, 0.53–0.61 mm) and 0.90 mm (95% CI, 0.85–0.91 mm), and R^2 values of 0.59 (95% CI, 0.50–0.65) and 0.67 (95% CI, 0.58–0.87), respectively. Oh et al. (14) developed an AL estimation model using ultra-widefield funds photos with an MAE of 0.74 mm (95% CI, 0.71–0.78 mm) and an R^2 value of 0.82 (95% CI, 0.79–0.84). However, this study represents the first attempt to estimate AL using macular B-scan OCT images via deep learning. B-scan images provide cross-sectional views of the retina, offering improved visualization of retinal layers and their integrity (18). The theoretical foundation of this study lies in utilizing the potential alterations in macular structure associated with AL elongation to predict AL. Additionally, we also excluded the eyes with any maculopathy to investigate the changes in macular structure before the development of myopic maculopathy in eyes with long AL. In the current study, the MAE was found to be 0.83 mm (95% CI, 0.72–0.95 mm) and the R2 was 0.763 in the regression task, while the classification model achieved an accuracy of 0.872 (95% CI, 0.840–0.899) in identifying eyes with AL ≥ 26.0 mm. These findings suggest

that macular structure changes in eyes with long AL occur independently of OCT-detectable myopic maculopathy, which aligns with clinical observations of a higher risk of the prevalence and progression of myopic maculopathy in eyes with longer AL (19, 20).

The results showed that the accuracy of AL estimation using macular OCT images (87.2%) was superior to using retinal thickness measurements alone (77.8%) in the same study sample. This can be attributed to the detailed structural information available in B-scan images (18). The Grad-CAM analysis revealed that for eyes with ALs shorter than 26.0 mm, the deep learning models primarily relied on the fovea, while for eyes with AL greater than or equal to 26.0 mm, the models showed a preference for regions mainly on either side of the fovea. These findings are consistent with a deep learning model for AL estimation using color fundus photos reported by Dong et al. (12). In their study, the heat map analysis demonstrated that eyes with ALs shorter than 26.0 mm predominantly utilized signals from the foveal region in the fundus photos, while those with AL greater than 26 mm primarily relied on signals from the extrafoveal region (12).

Clinical studies have demonstrated that eyes with high myopia, characterized by an AL exceeding 26.0 or 26.5 mm, were more likely to develop traction maculopathy, such as macular hole and maculoschisis (4, 21, 22). Furthermore, Park et al. (23) found that the development of myopic traction maculopathy was associated with the foveal curvature, which were calculated based on the retinal pigment epithelium hyper-reflective line in OCT images including the fovea. Based on the visualization results obtained from our OCT-based AL estimation model, we speculate that the highlighted regions in the heat maps indicate areas where the changes in retinal curvature are most pronounced (24). In addition, our results also suggested that structural changes in the macula caused by axial elongation exhibit a localized pattern, primarily concentrated at the fovea and the areas where the retinal curvature changes the most significantly, rather than displaying continuous alterations throughout the entire region. Besides retina, the choroid from the corresponding regions were also highlighted in the heat maps. Previous studies have reported that AL was negatively associated with choroidal thickness in both young and elderly people (25, 26), indicating the choroidal atrophy with the elongation of AL. These findings can explain the involvement of choroid in the heat maps when predicting AL in this study. These findings will be helpful for further research on the pathogenesis and prevention of AL-related structural maculopathy.

Several limitations should be noted in this study. First, the sample size is relative small. To minimize the impact of potential sources of bias, we specifically enrolled subjects from a solitary ophthalmological clinic and utilized images acquired using the identical imaging machine. Consequently, the recruitment of additional samples was constrained. Advancements in model predictive performance can be expected when more samples are gathered and analyzed. Second, due to the limited number of eyes with short AL in this study, only two groups (whether AL longer than 26.0 mm) were defined in the classification model development. Nevertheless, this limitation is unlikely to undermine the overall findings, as the focus of this study was on the deep learning model's performance in distinguishing eyes with elongated AL. Third, it is important to note that we excluded eyes with OCT-detectable maculopathy as our aim was to identify AL-specific macular characteristics prior to the onset of myopic maculopathy. Therefore, caution should be exercised when generalizing these findings to eyes with existing maculopathy. Lastly, the current model was developed based on the macular B-scans centered on the fovea by the stellate 6-scan pattern, which scans from 6 different directions. Since OCT B-scans centered on the fovea exhibit the similar imaging pattern, it is very likely that the deep learning model developed in this study would be applicable to macular OCT B-scans scanned by other pattern centered on the fovea or OCT scans from different manufacturers. However, further research and verification are needed to validate the generalization of the model. Additionally, it's worth noting that this model was developed using adult eyes with a mean age of 69 years. Considering that the macula develops and axial length increases in children and teenagers, additional studies are required to develop models based on younger age groups.

5 Conclusion

This study developed a deep learning model using macular OCT images to estimate AL and identify eyes with long AL, achieving good performance. The AL-related macular features exhibit a localized pattern, primarily concentrated in the central fovea and adjacent regions, suggesting that these specific areas may serve as the initial sites for macular alterations caused by AL elongation. These findings have significant implications for further research on the pathogenesis of AL-related structural maculopathy.

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 humans were approved by institutional review board at Beijing Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the retrospective nature of the study.

Author contributions

JL: Conceptualization, Data curation, Methodology, Writing – review & editing, Investigation, Validation, Writing – original draft. HL: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. YoZ: Conceptualization, Investigation, Methodology, Writing – original draft, Software, Validation. YuZ: Conceptualization, Investigation, Methodology, Writing – original draft, Data curation. SS: Conceptualization, Data curation, Investigation, Methodology, Funding acquisition, Supervision, Writing – review & editing. XG: Conceptualization, Data curation, Methodology, Supervision, Writing – review & editing. JX: Conceptualization, Methodology, Supervision, Writing – review &

editing, Investigation, Software. XY: Conceptualization, Methodology, Supervision, Writing – review & editing, Data curation, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by National High Level Hospital Clinical Research Funding (BJ-2022-104 and BJ-2020-167).

Conflict of interest

YoZ and JX are employees of Visionary Intelligence Ltd., which is a Chinese AI startup dedicated to exploring the application of AI technology in the field of healthcare, particularly focusing on ophthalmology.

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1308923/full#supplementary-material

References

1. Foster PJ, Broadway DC, Hayat S, Luben R, Dalzell N, Bingham S, et al. Refractive error, axial length and anterior chamber depth of the eye in British adults: the EPIC-Norfolk eye study. *Br J Ophthalmol*. (2010) 94:827–30. doi: 10.1136/bjo.2009.163899

2. Xiao O, Guo X, Wang D, Jong M, Lee PY, Chen L, et al. Distribution and severity of myopic maculopathy among highly myopic eyes. *Invest Ophthalmol Vis Sci*. (2018) 59:4880–5. doi: 10.1167/iovs.18-24471

3. Jonas JB, Jonas RA, Bikbov MM, Wang YX, Panda-Jonas S. Myopia: histology, clinical features, and potential implications for the etiology of axial elongation. *Prog Retin Eye Res*. (2022) 96:101156. doi: 10.1016/j.preteyeres.2022.101156

4. Ruiz-Medrano J, Montero JA, Flores-Moreno I, Arias L, García-Layana A, Ruiz-Moreno JM. Myopic maculopathy: current status and proposal for a new classification and grading system (ATN). *Prog Retin Eye Res*. (2019) 69:80–115. doi: 10.1016/j.preteyeres.2018.10.005

5. Wong AC, Chan CW, Hui SP. Relationship of gender, body mass index, and axial length with central retinal thickness using optical coherence tomography. *Eye*. (2005) 19:292–7. doi: 10.1038/sj.eye.6701466

6. Wu PC, Chen YJ, Chen CH, Chen YH, Shin SJ, Yang HJ, et al. Assessment of macular retinal thickness and volume in normal eyes and highly myopic eyes with third-generation optical coherence tomography. *Eye*. (2008) 22:551–5. doi: 10.1038/sj.eye.6702789

7. Jiang Z, Shen M, Xie R, Qu J, Xue A, Lu F. Interocular evaluation of axial length and retinal thickness in people with myopic anisometropia. *Eye Contact Lens*. (2013) 39:277–82. doi: 10.1097/ICL.0b013e318296790b

8. Jonas JB, Xu L, Wei WB, Pan Z, Yang H, Holbach L, et al. Retinal thickness and axial length. *Invest Ophthalmol Vis Sci*. (2016) 57:1791–7. doi: 10.1167/iovs.15-18529

9. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

10. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cells*. (2018) 172:1122–31.e9. doi: 10.1016/j.cell.2018.02.010

11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539

12. Dong L, Hu XY, Yan YN, Zhang Q, Zhou N, Shao L, et al. Deep learning-based estimation of axial length and subfoveal choroidal thickness from color fundus photographs. *Front Cell Dev Biol*. (2021) 9:653692. doi: 10.3389/fcell.2021.653692

13. Zou H, Shi S, Yang X, Ma J, Fan Q, Chen X, et al. Identification of ocular refraction based on deep learning algorithm as a novel retinoscopy method. *Biomed Eng Online*. (2022) 21:87. doi: 10.1186/s12938-022-01057-9

14. Oh R, Lee EK, Bae K, Park UC, Yu HG, Yoon CK. Deep learning-based prediction of axial length using ultra-widefield fundus photography. *Korean J Ophthalmol*. (2023) 37:95–104. doi: 10.3341/kjo.2022.0059

15. Yoo TK, Ryu IH, Kim JK, Lee IS. Deep learning for predicting uncorrected refractive error using posterior segment optical coherence tomography images. *Eye*. (2022) 36:1959–65. doi: 10.1038/s41433-021-01795-5

16. Zhang Y, Hong D, McClement D, Oladosu O, Pridham G, Slaney G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J Neurosci Methods*. (2021) 353:109098. doi: 10.1016/j.jneumeth.2021.109098

17. Jeong Y, Lee B, Han J-H, Oh J. Ocular axial length prediction based on visual interpretation of retinal fundus images via deep neural network. *IEEE J Selected Topics Quant Elect*. (2020) 27:1–7. doi: 10.1109/JSTQE.2020.3038845

18. Sakamoto A, Hangai M, Yoshimura N. Spectral-domain optical coherence tomography with multiple B-scan averaging for enhanced imaging of retinal diseases. *Ophthalmology*. (2008) 115:1071–8.e7. doi: 10.1016/j.ophtha.2007.09.001

19. Fang Y, Yokoi T, Nagaoka N, Shinohara K, Onishi Y, Ishida T, et al. Progression of myopic maculopathy during 18-year follow-up. *Ophthalmology*. (2018) 125:863–77. doi: 10.1016/j.ophtha.2017.12.005

20. Hashimoto S, Yasuda M, Fujiwara K, Ueda E, Hata J, Hirakawa Y, et al. Association between axial length and myopic maculopathy: the Hisayama study. *Ophthalmol Retina*. (2019) 3:867–73. doi: 10.1016/j.oret.2019.04.023

21. Frisina R, Gius I, Palmieri M, Finzi A, Tozzi L, Parolini B. Myopic traction maculopathy: diagnostic and management strategies. *Clin Ophthalmol*. (2020) 14:3699–708. doi: 10.2147/OPTH.S237483

22. Cheong KX, Xu L, Ohno-Matsui K, Sabanayagam C, Saw SM, Hoang QV. An evidence-based review of the epidemiology of myopic traction maculopathy. *Surv Ophthalmol*. (2022) 67:1603–30. doi: 10.1016/j.survophthal.2022.03.007

23. Park UC, Ma DJ, Ghim WH, Yu HG. Influence of the foveal curvature on myopic macular complications. *Sci Rep*. (2019) 9:16936. doi: 10.1038/s41598-019-53443-4

24. Park SJ, Ko T, Park CK, Kim YC, Choi IY. Deep learning model based on 3D optical coherence tomography images for the automated detection of pathologic myopia. *Diagnostics*. (2022) 12:742. doi: 10.3390/diagnostics12030742

25. Ikuno Y, Kawaguchi K, Nouchi T, Yasuno Y. Choroidal thickness in healthy Japanese subjects. *Invest Ophthalmol Vis Sci*. (2010) 51:2173–6. doi: 10.1167/iovs.09-4383

26. Sato M, Minami S, Nagai N, Suzuki M, Kurihara T, Shinojima A, et al. Association between axial length and choroidal thickness in early age-related macular degeneration. *PLoS One*. (2020) 15:e0240357. doi: 10.1371/journal.pone.0240357

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 23 November 2023
DOI 10.3389/fmed.2023.1291404

Check for updates

OPEN ACCESS

EDITED BY
Haoyu Chen,
The Chinese University of Hong Kong, China

REVIEWED BY
Jo-Hsuan "Sandy" Wu,
University of California, San Diego,
United States
Jana Lipkova,
University of California, Irvine, United States

*CORRESPONDENCE
Juan Ye
✉ yejuan@zju.edu.cn

†These authors have contributed equally to this work

RECEIVED 09 September 2023
ACCEPTED 20 October 2023
PUBLISHED 23 November 2023

CITATION
Jin K, Yuan L, Wu H, Grzybowski A and Ye J (2023) Exploring large language model for next generation of artificial intelligence in ophthalmology.
Front. Med. 10:1291404.
doi: 10.3389/fmed.2023.1291404

COPYRIGHT
© 2023 Jin, Yuan, Wu, Grzybowski and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploring large language model for next generation of artificial intelligence in ophthalmology

Kai Jin[1†], Lu Yuan[2†], Hongkang Wu[1], Andrzej Grzybowski[3] and Juan Ye[1*]

[1]Eye Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2]Department of Ophthalmology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [3]Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland

In recent years, ophthalmology has advanced significantly, thanks to rapid progress in artificial intelligence (AI) technologies. Large language models (LLMs) like ChatGPT have emerged as powerful tools for natural language processing. This paper finally includes 108 studies, and explores LLMs' potential in the next generation of AI in ophthalmology. The results encompass a diverse range of studies in the field of ophthalmology, highlighting the versatile applications of LLMs. Subfields encompass general ophthalmology, retinal diseases, anterior segment diseases, glaucoma, and ophthalmic plastics. Results show LLMs' competence in generating informative and contextually relevant responses, potentially reducing diagnostic errors and improving patient outcomes. Overall, this study highlights LLMs' promising role in shaping AI's future in ophthalmology. By leveraging AI, ophthalmologists can access a wealth of information, enhance diagnostic accuracy, and provide better patient care. Despite challenges, continued AI advancements and ongoing research will pave the way for the next generation of AI-assisted ophthalmic practices.

KEYWORDS

artificial intelligence, large language model, ChatGPT, ophthalmology, diagnostic accuracy and efficacy

Introduction

The history of artificial intelligence (AI) in medicine dates back to the 1950s when researchers began to explore the use of computers to analyze medical data and make diagnostic decisions. However, past methods had limitations in accuracy and speed and still could not analyze unstructured medical data (1). Natural Language Processing (NLP) is a subfield of AI that focuses on enabling computers to understand, interpret, and generate human language. It involves the development of algorithms and models that can process and analyze unstructured text data. Large Language Models (LLM) refer to advanced artificial intelligence models, such as GPT-3 (Generative Pre-trained Transformer 3), that are built on transformer architecture. The transformer architecture is a deep learning model that efficiently captures context and dependencies in sequential data, making it a fundamental choice for natural language processing tasks and beyond. These models are trained on massive amounts of text data from the internet, enabling them to generate human-like text and perform a wide range of NLP tasks with remarkable accuracy and versatility. ChatGPT builds on the capabilities of large language models to

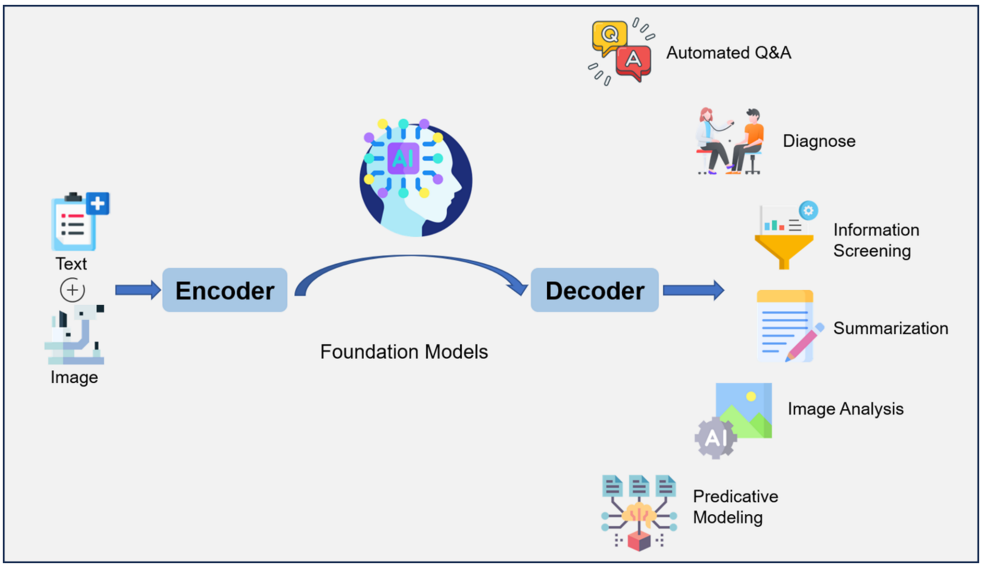

FIGURE 1
Workflow of large language model (LLM) for artificial intelligence (AI) in ophthalmology. Text (symptoms, medical history, etc.) and images (Optical coherence tomography, Fundus fluorescein angiography, etc.) are encoded and fed into a model that has been trained on a large amount of data, which can decode the relevant information required. LLM applications include automated question-answering, diagnose, information screening, summarization, image analysis, predictive modeling.

generate coherent and contextually relevant responses, making it well-suited for chatbot applications. It is designed to generate human-like responses to a wide range of prompts and questions and may enhance healthcare delivery and patients' quality of life (Figure 1) (2). The use of LLMs in healthcare offers several potential benefits.

ChatGPT and LLMs can be applied in various ways. They can serve as clinical documentation aids, helping with administrative tasks such as clinic scheduling, medical coding for billing, and generating preauthorization letters (3). LLMs can also be used as summarization tools, improving communication with patients and assisting in clinical trials. They can make processes such as curriculum design, testing of knowledge base, and continuing medical education more dynamic (4). LLMs can reduce the burden of administrative tasks for healthcare professionals, save time, and improve efficiency. They also have the potential to provide valuable clinical insights and support decision-making (5). This capability may help ophthalmologists enabling evidence-based decision-making and revolutionizing various aspects of eye care and research.

Method of literature search

For this review, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) guidelines.

Study selection and search strategy

We conducted a comprehensive literature search following the PRISMA guidelines. Searches were performed on PubMed and Google Scholar databases, spanning from January 2016 to June 2023. Keywords were selected from two distinct categories: ophthalmology-related terms (ophthalmology, eye diseases, eye disorders) and large language model-related terms (large language models, ChatGPT, natural language processing, chatbots). The search strategy involved the use of the following keywords: ("Ophthalmology" OR "Eye Diseases" OR "Eye Disorders") AND ("Large Language Models" OR "ChatGPT" OR "Natural Language Processing" OR "Chatbots"). The terms from each category were cross-referenced independently with terms from the other category.

Inclusion and exclusion criteria

We established specific inclusion criteria for article selection. The publication period considered research from January 2016 to June 2023 to ensure the inclusion of up-to-date findings. Initially, 6,130 articles were identified through titles and abstracts. We prioritized research quality and the application of Large Language Models (LLMs) in our selection process. Additionally, articles published prior to 2016 were included for historical context and those pertinent to closely related topics.

In the meantime, studies meeting the following criteria will be excluded: (1) duplicate literature previously included in the review, (2) irrelevant topics, where the article is unrelated to ophthalmology or the application of the large language model, (3) conference abstracts, and (4) non-original research, such as editorials, case reports or commentaries.

Abbreviations: AI, Artificial Intelligence; NLP, Natural Language Processing; LLM, Large Language Model; GPT, Generative Pre-trained Transformer; HER, Electronic Health Records; eAMD, exudative Age-related Macular Degeneration; DR, Diabetic Retinopathy; OCT, Optical Coherence Tomography; BERT, Bidirectional Encoder Representations from Transformers.

Language considerations

A comprehensive review was conducted primarily on English-language articles, totaling 6,130 papers. Furthermore, we evaluated 14 papers predominantly published in Chinese. For articles in languages such as French, Spanish, and German, we assessed their abstracts. This multilingual approach allowed us to comprehensively evaluate the literature. The primary inclusion criterion required research to specifically address the application of AI in ophthalmology and demonstrate a certain level of perceived quality.

Data extraction and analysis

Following a rigorous selection process, relevant data were extracted and analyzed from the selected articles. Key themes, trends, advancements, and challenges related to the utilization of LLMs in ophthalmology were systematically synthesized.

In accordance with the PRISMA guidelines, this review adhered to a structured and rigorous approach, encompassing a comprehensive literature search, meticulous inclusion criteria, language considerations, and thorough data extraction (Figure 2). A total of 108 articles were independently screened for eligibility by two reviewers (Kai Jin and Lu Yuan), including assessments of titles and abstracts, followed by full-text review. Any disagreements were resolved through discussion with a third author (Juan Ye). Ultimately, 108 studies were included in the review.

Results

We finally included 108 studies. The results (Table 1) encompass a diverse range of studies in the field of ophthalmology, highlighting the versatile applications of LLMs. The results reflect a wide spectrum of LLM applications, and subfields of interest in ophthalmology. They showcase the versatility of LLMs in addressing various aspects of automated question-answering (55 studies), diagnose (5 studies), information screening (27 studies), summarization (5 studies), image analysis (5 studies), predictive modeling (11 studies). Subfields encompass general ophthalmology (38 studies), retinal diseases (32 studies), anterior segment diseases (27 studies), glaucoma (6 studies), and ophthalmic plastics (5 studies) (Figure 3).

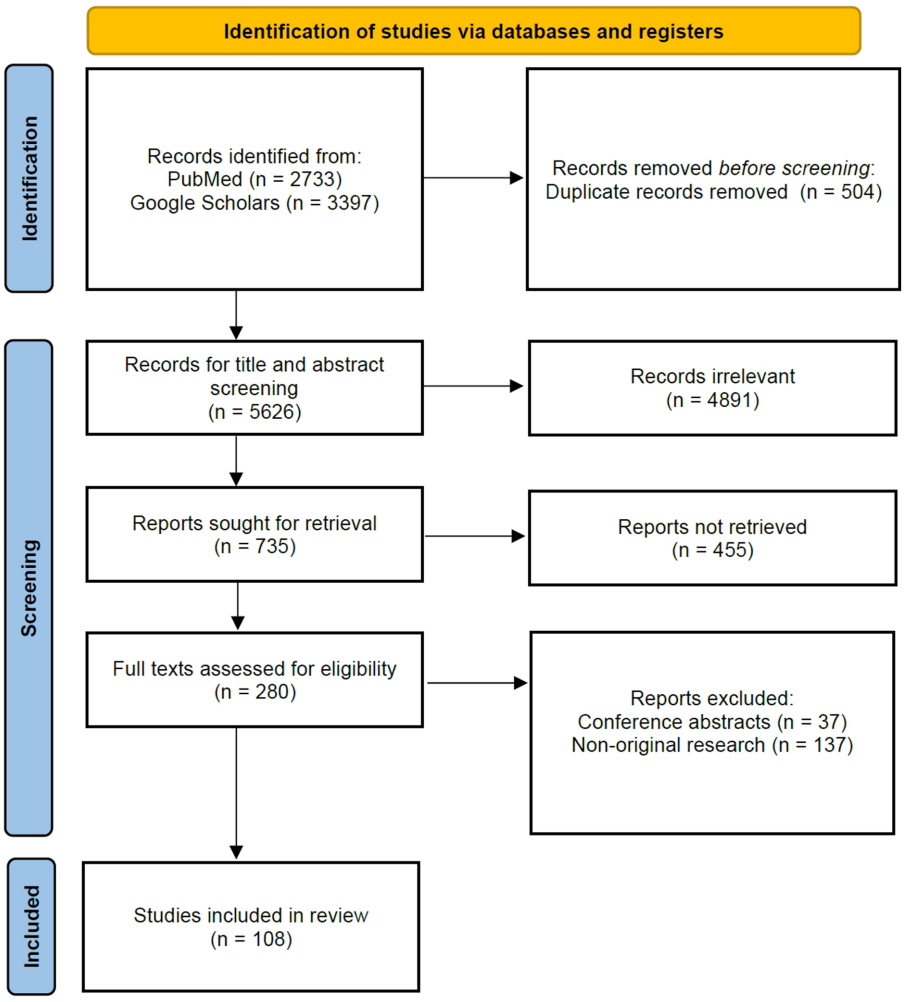

FIGURE 2
PRISMA 2020 flow diagram for this systematic review.

TABLE 1 Summary of representative current studies using LLM in ophthalmology.

Reference	Year	Publication	Subspeciality	Aim	Application	Approaches
Lin et al. (6)	2023	Eye	General ophthalmology	To compare the performance on a practice ophthalmology written examination	Automated question-answering	GPT-3.5, GPT-4
Antaki et al. (7)	2023	Ophthalmology Science	General ophthalmology	To evaluate the performance on ophthalmology questions	Automated question-answering	ChatGPT
Cai et al. (8)	2023	American Journal of Ophthalmology	General ophthalmology	To compare the performance on ophthalmology board-style questions.	Automated question-answering	Bing Chat, ChatGPT 3.5, and ChatGPT 4.0,
Mihalache et al. (9)	2023	JAMA Ophthalmology	General ophthalmology	To assess the performance on board certification exam in ophthalmology	Automated question-answering	ChatGPT
Bernstein et al. (10)	2023	JAMA Network Open	General ophthalmology	To generate ophthalmology advice	Automated question-answering	ChatGPT version 3.5
Ali et al. (11)	2023	Ophthalmic Plast Reconstr Surg	Lacrimal drainage disorders	To response to lacrimal drainage disorders	Automated question-answering	ChatGPT
Tsui et al. (12)	2023	Eye	Posterior vitreous detachment, retinal tear and detachment, ocular surface disease, exudative age-related macular degeneration (eAMD), and post-intravitreal injection pain and redness	To response to common ocular symptoms	Automated question-answering	ChatGPT
Potapenko et al. (13)	2023	Acta Ophthalmologica	Retinal diseases	To evaluate accuracy on patient information	Automated question-answering	ChatGPT
Momenaei et al. (14)	2023	Ophthalmology Retina	Retinal diseases	To evaluate the appropriateness and readability of the medical knowledge	Automated question-answering	ChatGPT-4
Waisberg et al. (15)	2023	Irish Journal of Medical Science	Anterior ischemic optic neuropathy	Fundus image analysis	Image analysis	GPT-4
Hu et al. (16)	2022	Transl Vis Sci Technol.	Glaucoma	To Predict Glaucoma Progression Requiring Surgery	Predictive Modeling	Pre-trained Transformers
Lee et al. (17)	2023	Ophthalmic Res	General ophthalmology	To assign procedural codes based on the surgical report	Predictive Modeling	Bidirectional Encoder Representations from Transformers (BERT)
Liu et al. (18)	2023	AMIA	Retinal vascular disease	To provide a diagnosis based on FFA reports	Summarization	GPT3.5-Turbo
Yu et al. (19)	2022	BMC Medical Informatics and Decision Making	Diabetic retinopathy	To Identify diabetic retinopathy-related clinical concepts and their attributes	Information screening	NLP(Extraction, Named entity recognition), DL, Pre-trained Transfomers
Valentín-Bravo et al. (20)	2023	Arch Soc Esp Oftalmol.	Vitreoretinal disease	To write a scientific article	Information screening	ChatGPT, DALL-E 2
Singh et al. (4)	2023	Clin Exp Ophthamol.	Dry eye disease	To conduct a literature review	Information screening	ChatGPT
Singh et al. (21)	2023	Seminars in Ophthalmology	Cornea, retina, glaucoma, pediatric ophthalmology, neuroophthalmology, and ophthalmic plastics surgery	To construct ophthalmic discharge summaries and operative notes	Information screening	ChatGPT
Rasmussen et al. (22)	2023	Graefe's archive for clinical and experimental ophthalmology	Vernal keratoconjunctivitis	To provided responses to patient and parent questions	Automated question-answering	ChatGPT
lim et al. (23)	2023	Ebiomedicine	Myopia	To deliver accurate responses to common myopia-related query	Automated question-answering	ChatGPT-3.5, ChatGPT-4.0, and Google Bard
Waisberg et al. (24)	2023	Annals of Biomedical Engineering	General ophthalmology	To write ophthalmic operative notes	Information screening	GPT-4

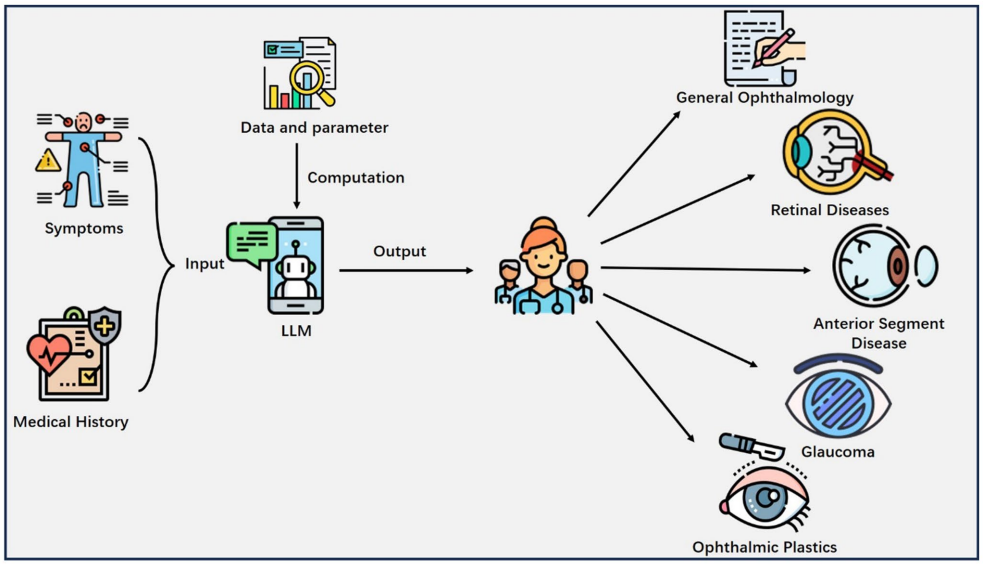

FIGURE 3
Major applications of LLM in Ophthalmology. The patient's information like symptoms, medical history and other health-related details are inputted into the LLM, which outputs valuable clinical insights to the physician and helps him or her make decisions.

General ophthalmology

The application of LLMs in ophthalmology is a rapidly growing field with promising potential, encompassing various aspects of patient care and clinical workflows. LLMs can analyze general ophthalmology patient data and medical records to recommend personalized diagnosis and treatment plans for individuals with specific eye conditions. Chatbots integrated with electronic health record (EHR) systems can access patient information to provide context-aware responses and support clinical decision-making.

The majority of current NLP applications in ophthalmology focus on extracting specific text, such as visual acuity, from free-text notes for the purposes of quantitative analysis (25). NLP also offers opportunities to develop search engines for data within free-text notes, clean notes, automated question-answering, and translating ophthalmology notes for other specialties or for patients. Low vision rehabilitation improves quality of life for visually impaired patients, free-text progress notes within the EHR using NLP provide valuable information relevant to predicting patients' visual prognosis (26). NLP with unstructured clinician notes supports low vision and blind rehabilitation for war veterans with traumatic brain injury based on veterans' needs rather than system-level factors (27, 28). This suggests that AI with NLP may be particularly important for the performance of predictive models in ophthalmology. Given the potential of LLMs in healthcare and the increasing reliance of patients on online information, it is important to evaluate the quality of chatbot-generated advice and compare it with human-written advice from ophthalmologists. The panel of ophthalmologists had a 61.3% accuracy in distinguishing between chatbot and human responses (10).

As chatbot technology is continually evolving, there are additional applications in general ophthalmology. The researchers evaluated the ability of the ChatGPT to respond to ocular symptoms by scripting 10 prompts reflective of common patient messages relating to various ocular conditions (12). These conditions included posterior vitreous detachment, retinal tear and detachment, ocular surface disease, exudative age-related macular degeneration (eAMD), and post-intravitreal injection pain and redness. The abilities of ChatGPT in constructing discharge summaries and operative notes were evaluated through a study conducted by Swati et al. (21). The study found that ChatGPT was able to construct ophthalmic discharge summaries and operative notes in a matter of seconds, with tailored responses based on the quality of inputs given. However, there were some limitations such as the presence of generic text and factual inaccuracies in some responses. The authors suggest that ChatGPT can be utilized to minimize the time spent on discharge summaries and improve patient care, but it should be used with caution and human verification. Another study aimed to assess the performance of an AI chatbot, ChatGPT, in answering practice questions for ophthalmology board certification examinations (9). ChatGPT correctly answered 46.4% of the questions, with the best performance in the category of general medicine (79%) and the poorest in retina and vitreous (0%). ChatGPT provided explanations and additional insight for 63% of questions but selected the same multiple-choice response as the most common answer provided by ophthalmology trainees only 44% of the time. The researchers compared the performance of several generative AI models on the ophthalmology board-style questions (6–8), including Bing Chat (Microsoft), ChatGPT 3.5 and 4.0 (OpenAI). Performance was compared with that of human respondents. Results showed that ChatGPT-4.0 and Bing Chat performed comparably to human respondents.

Existing electronic differential diagnosis support tools, like the Isabel Pro Differential Diagnosis Generator, have limitations in terms of structured input and context-specific language processing. In one study, ChatGPT identified the correct diagnosis in 9 out of 10 cases and had the correct diagnosis listed in all 10 of its lists of differentials (29). Isabel, on the other hand, identified only 1 out of 10 provisional diagnoses correctly, but included the correct diagnosis in 7 out of 10 of its differential diagnosis lists. The median position of the correct diagnosis in the ranked differential lists was 1.0 for ChatGPT versus 5.5 for Isabel.

Retinal diseases

Some studies evaluate the accuracy of an AI-based chatbot in providing patient information on common retinal diseases, including AMD, diabetic retinopathy (DR), retinal vein occlusion, retinal artery occlusion, and central serous chorioretinopathy.

In healthcare settings, when patients provide information about their medical history, symptoms, or other health-related details, there is the potential for miscommunication or misalignment between the patient's perspective and the physician's understanding of the situation. Traditional methods of obtaining patient information may lead to dissatisfaction if the information obtained misaligns with the physician's information (30). ChatGPT can improve patient satisfaction in terms of information provision by providing accurate and well-formulated responses to various topics, including common retinal diseases (13). This accessibility can be particularly beneficial when ophthalmologists are not readily available. Among retinal diseases, DR is a leading cause of blindness in adults, and there is increasing interest in developing AI technologies to detect DR using EHRs. Most AI-based DR diagnoses are focused on medical images, but there is limited research exploring the lesion-related information captured in the free text image reports. In Yu et al. (19) study, two state-of-the-art transformer-based NLP models, including BERT and RoBERTa, were examined and compared with a recurrent neural network implemented using Long short-term memory (LSTM) to extract DR-related concepts from clinical narratives. The results show that for concept extraction, the BERT model pretrained with the MIMIC III dataset outperformed other models, achieving the highest performance with F1-scores of 0.9503 and 0.9645 for strict and lenient evaluations, respectively. The findings of this study could have a significant impact on the development of clinical decision support systems for DR diagnoses.

Anterior segment disease

Anterior segment vision-threatening disease included the diagnosis of corneal ulcer, iridocyclitis, hyphema, anterior scleritis, or scleritis with corneal involvement. Patients with anterior segment diseases present a diagnostic challenge for many primary care physicians. The researchers developed a decision support tool to predict vision-threatening anterior segment disease using primary clinical notes based on NLP (31). The ultimate prediction model exhibited an area under the curve (AUC) of 0.72, with a 95% confidence interval ranging from 0.67 to 0.77. Using a threshold that achieved a sensitivity of 90%, the model demonstrated a specificity of 30%, a positive predictive value of 5.8%, and a high negative predictive value of 99%. One study evaluates the accuracy of responses provided by the ChatGPT to patient and parent questions on vernal keratoconjunctivitis (VKC), a complex and recurring disease primarily affecting children (22). The researchers formulated questions in four categories and assessed the chatbot's responses for information accuracy. The chatbot was found to provide both relevant and inaccurate statements. Inaccurate statements were particularly observed regarding treatment and potential side effects of medications. A comparative analysis of the performance of three LLMs, namely ChatGPT-3.5, ChatGPT-4.0, and Google Bard, was conducted in delivering accurate and comprehensive responses to common myopia-related queries. ChatGPT-4.0 demonstrated the highest accuracy, with 80.6% of responses rated as 'good', compared to 61.3% in ChatGPT-3.5 and 54.8% in Google Bard (23).

Glaucoma

Previous studies have developed predictive models for glaucoma progression, but uncertainty remains on how to integrate the information in free-text clinical notes, which contain valuable clinical information (32). Some studies aim to predict glaucoma progression requiring surgery using deep learning approaches on EHRs and natural language processing of clinical free-text notes. Sunil et al. presents an artificial intelligence approach to predict near-term glaucoma progression using clinical free-text notes and data from electronic health records (33). The authors developed models that combined structured data and text inputs to predict whether a glaucoma patient would require surgery within the following year. The model incorporating both structured clinical features and free-text features achieved the highest performance with an AUC of 0.899 and an F1 score of 0.745. Another study aims to fill the gap by developing a deep learning predictive model for glaucoma progression using both structured clinical data and natural language processing of clinical free-text notes from EHRs. The combination model showed the best AUC (0.731), followed by the text model (0.697) and the structured model (0.658) (34). Hu et al. (16) explored the use of transformer-based language models, specifically Bidirectional Encoder Representations from Transformers (BERT), to predict glaucoma progression requiring surgery using clinical free-text notes from EHRs. The results showed that the BERT models outperformed an ophthalmologist's review of clinical notes in predicting glaucoma progression. Michelle et al. (35) utilized an automated pipeline for data extraction from EHRs to evaluate the real-world outcomes of glaucoma surgeries, tube shunt surgery had a higher risk of failure (Baerveldt: Hazard Ratio (HR) 1.44, 95% CI 1.02 to 2.02; Ahmed: HR 2.01, 95% CI 1.28 to 3.17).

Ophthalmic plastics

In the study conducted by Mohammad et al. (11), ChatGPT's performance in providing information about primary acquired nasolacrimal duct obstruction and congenital nasolacrimal duct obstruction was evaluated. Regarding insights into the history and effectiveness of dacryocystorhinostomy surgery, ChatGPT was tested on this specific topic. Agreement among the three observers was high (95%) in grading the responses. The responses of ChatGPT were graded as correct for only 40% of the prompts, partially correct in 35%, and outright factually incorrect in 25%. Hence, some degree of factual inaccuracy was present in 60% of the responses, if we consider the partially correct responses.

Discussion

The newer generation of GPT models, exemplified by GPT-3 and beyond, differs from their predecessors through significantly larger model sizes, improved performance on various language tasks, enhanced few-shot learning abilities, and increased versatility, while also necessitating more substantial computational resources and raising ethical considerations.

Strengths

AI technology, such as online chat-based AI language models, has the potential to assist clinical workflows and augment patient education and communication about common ophthalmology diseases prevention queries (Table 1). GPT's medical subspecialty capabilities have improved significantly from GPT-3 to GPT-4. Both LLMs struggled with image-based and higher-order ophthalmology questions, perhaps reflecting the importance of visual analysis in ophthalmology. Given the ongoing advances in computer vision, it may be possible to address this limitation in future LLMs. There is room for improvement in medical conversational agents, as all models exhibited instances of hallucination, incorrect justification, or non-logical reasoning (36). Although ChatGPT 4.0 has demonstrated remarkable capabilities in a variety of domains, the presence of these errors raises concerns about the reliability of the system, especially in critical clinical decision making.

Ophthalmologists are starting to use ChatGPT to help with paperwork such as scientific articles, discharge summaries and operative notes (15, 24, 37). The scientific accuracy and reliability on certain topics were not sufficient to automatically generate scientifically rigorous articles. This was also objected to by some ophthalmologists (38). Firstly, operative notes are not general descriptions of surgical procedures and a specific patient has its own unique characteristics. Secondly, operative notes are legal documents and the surgeon is responsible for the accuracy and completeness of the notes. Thirdly, there is no evidence that GPT-4 can accurately capture the unique aspects of individual cases in the real world, such as intraoperative complications. Finally, the writing of operative notes requires a degree of clinical decision-making and clinical judgment that cannot be automated.

In a recent development, ChatGPT has emerged as an author or co-author of scientific papers in the field of ophthalmology (39, 40). This innovative inclusion has sparked discussions and garnered attention from the scientific community. The presence of ChatGPT as an author in scientific research reflects the evolving landscape of artificial intelligence's involvement in various domains, including ophthalmology, opening avenues for new perspectives and collaborative contributions.

Challenges

Despite the promising future, integrating LLMs into ophthalmology also poses several challenges that need to be addressed. Firstly, ensuring patient data privacy and maintaining the security of sensitive medical information will be critical (41). These models require vast amounts of data to achieve their potential, but data-sharing must be conducted responsibly and in compliance with strict ethical and legal guidelines (42, 43).

Another significant challenge is the potential for bias in the data used to train these language models (44). If the data used for training is not diverse enough, the models may exhibit biases that can lead to inaccurate or unfair recommendations, particularly when dealing with underrepresented populations. Efforts must be made to identify and mitigate these biases to ensure equitable and reliable outcomes for all patients.

Furthermore, there may be resistance or skepticism among some healthcare professionals towards adopting AI-driven technologies like LLMs. It will be crucial to address these concerns, provide proper training, and foster a collaborative environment where human experts and AI work together synergistically (45).

The interpretability and explainability of the decisions made by these models are another challenge. As they are often considered "black boxes," "understanding the reasoning behind their recommendations can be difficult," leading to potential mistrust from clinicians and patients (46). Developing methods to make the models more transparent and explainable will be essential for their widespread acceptance and adoption (47).

Lastly, the rapidly evolving nature of AI and language model technologies demands continuous updates and improvements. Staying up-to-date with the latest advancements and incorporating new knowledge into the models is essential to maintain their accuracy and relevance in the ever-changing field of ophthalmology.

While LLMs like ChatGPT offer tremendous potential in ophthalmology, addressing the challenges of AI hallucination and misinformation is paramount. It is essential to consider the broader societal implications, including patient trust, medical liability, ethical concerns, scientific integrity, health disparities, and regulatory oversight when integrating AI into ophthalmic practices. Responsible AI implementation and continuous monitoring are essential to harness the benefits of AI while minimizing potential risks. One concern in the use of LLMs for medical applications is the lack of reproducibility, as these generative models may not consistently provide the same answers, potentially impacting the reliability of their outputs in clinical settings. Addressing these challenges will be essential to fully realize the potential benefits of large language models in ophthalmology and to ensure their responsible and ethical implementation in patient care (48).

Future perspectives

The future perspectives of LLMs in ophthalmology hold tremendous promise for transforming the landscape of eye care and research (49). These advanced language models, powered by AI and NLP, are poised to revolutionize how ophthalmologists diagnose, treat, and manage various eye conditions. LLMs can be integrated with image analysis techniques to create multimodal AI systems. These systems can process both textual and visual information, enhancing their capabilities in ophthalmology. For instance, LLMs can analyze textual patient records and medical literature, while image analysis algorithms can interpret medical images such as fundus photographs. Through their ability to analyze vast amounts of medical literature, patient data, and diagnostic images, these models can provide more accurate and timely diagnoses, personalized treatment plans, and even predict disease progression. The combination of LLMs and image analysis can lead to more efficient and accurate decision-making in ophthalmic practice. Additionally, LLMs can be used as tools to support communication and knowledge exchange in the following ways. While LLMs themselves do not directly facilitate communication like human interaction, their capabilities can enhance and streamline information exchange and knowledge sharing among eye care professionals worldwide. As research and development in this field continue to progress, we can expect these language models to become

indispensable tools that enhance efficiency, accessibility, and ultimately improve patient outcomes in ophthalmology.

Limitations

This review acknowledges several potential limitations that may have affected the comprehensiveness and potential bias of the literature search and selection process. These limitations include publication bias, language bias due to the focus on English-language studies, potential database selection bias, the possibility of excluding relevant studies due to search term restrictions, the limited date range, and the predefined exclusion criteria that may have omitted relevant research. The review also recognizes the potential for missed references and acknowledges the subjectivity in reviewer bias, which could impact study inclusion. Moreover, the review underscores the importance of addressing these limitations to ensure a more comprehensive and balanced assessment of the field of AI in ophthalmology. Despite these potential constraints, the review provides valuable insights into the applications and challenges of AI in ophthalmology, but readers should consider these limitations when interpreting the findings and drawing conclusions from the review.

Author contributions

KJ: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. LY: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. HW: Formal analysis, Software, Validation, Writing – review & editing. AG: Supervision, Validation, Visualization, Writing – review & editing. JY: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work has been financially supported by Natural Science Foundation of China (grant number 82201195), and Clinical Medical Research Center for Eye Diseases of Zhejiang Province (grant number 2021E50007).

Acknowledgments

Thanks to all the peer reviewers and editors for their opinions and suggestions.

Conflict of interest

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

Publisher's note

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

References

1. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: current status and future perspectives. *Adv Ophthalmol Pract Res*. (2022) 2:100078. doi: 10.1016/j.aopr.2022.100078

2. Will ChatGPT transform healthcare? *Nat Med*. (2023) 29:505–6. doi: 10.1038/s41591-023-02289-5

3. Sharma P, Parasa S. ChatGPT and large language models in gastroenterology. *Nat Rev Gastroenterol Hepatol*. (2023) 20:481–2. doi: 10.1038/s41575-023-00799-8

4. Singhal K, Azizi S. Large language models encode clinical knowledge. *Nature*. (2023) 620:172–80. doi: 10.1038/s41586-023-06291-2

5. Arora A, Arora A. The promise of large language models in health care. *Lancet (London, England)*. (2023) 401:641. doi: 10.1016/S0140-6736(23)00216-7

6. Lin JC, Younessi DN, Kurapati SS, Tang OY, Scott IU. Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination. *Eye (London, England)*. (2023). doi: 10.1038/s41433-023-02564-2

7. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci*. (2023) 3:100324. doi: 10.1016/j.xops.2023.100324

8. Cai LZ, Shaheen A, Jin A, Fukui R, Yi JS, Yannuzzi N, et al. Performance of generative large language models on ophthalmology board style questions. *Am J Ophthalmol*. (2023) 254:141–9. doi: 10.1016/j.ajo.2023.05.024

9. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence Chatbot in ophthalmic knowledge assessment. *JAMA Ophthal*. (2023) 141:589–97. doi: 10.1001/jamaophthalmol.2023.1144

10. Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, et al. Comparison of ophthalmologist and large language model Chatbot responses to online patient eye care questions. *JAMA Netw Open*. (2023) 6:e2330320. doi: 10.1001/jamanetworkopen.2023.30320

11. Ali MJ. ChatGPT and lacrimal drainage disorders: performance and scope of improvement. *Ophthal Plast Reconstr Surg*. (2023) 39:221–5. doi: 10.1097/IOP.0000000000002418

12. Tsui JC, Wong MB, Kim BJ, Maguire AM, Scoles D. Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. *Eye (Lond)*. (2023). doi: 10.1038/s41433-023-02556-2

13. Potapenko I, Boberg-Ans LC, Stormly HM. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. *Acta Ophthalmol*. (2023). doi: 10.1111/aos.15661

14. Momenaei B, Wakabayashi T, Shahlaee A, Durrani AF, Pandit SA, Wang K, et al. Appropriateness and readability of chatgpt-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. (2023). 7:862–8.

15. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci*. (2023). doi: 10.1007/s11845-023-03377-8

16. Hu W, Wang SY. Predicting Glaucoma progression requiring surgery using clinical free-text notes and transfer learning with transformers. *Transl Vis Sci Technol*. (2022) 11:37. doi: 10.1167/tvst.11.3.37

17. Lee YM, Bacchi S, Macri C, Tan Y, Casson R, Chan WO. Ophthalmology operation note encoding with open-source machine learning and natural language processing. *Ophthalmic Res*. (2023) 66:928–39.

18. Liu X, Wu J, Shao A, Shen W, Ye P, Wang Y, et al. Transforming retinal vascular disease classification: a comprehensive analysis of chatgpt's performance and inference abilities on non-english clinical environment. *medRxiv* (2023). doi: 10.1101/2023.06.28.23291931

19. Yu Z, Yang X, Sweeting GL, Ma Y, Stolte SE, Fang R, et al. Identify diabetic retinopathy-related clinical concepts and their attributes using transformer-based

natural language processing methods. *BMC Med Inform Decis Mak*. (2022) 22:255. doi: 10.1186/s12911-022-01996-2

20. Valentín-Bravo FJ, Mateos-Álvarez E, Usategui-Martín R, Andrés-Iglesias C, Pastor-Jimeno JC, Pastor-Idoate S. Artificial intelligence and new language models in ophthalmology: complications of the use of silicone oil in vitreoretinal surgery. *Arch Soc Esp Oftalmol (Engl Ed)*. (2023) 98:298:303.

21. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol*. (2023) 38:503–7. doi: 10.1080/08820538.2023.2209166

22. Rasmussen MLR, Larsen AC, Subhi Y, Potapenko I. Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis. *Graefes Arch Clin Exp Ophthalmol*. (2023) 261:3041–3. doi: 10.1007/s00417-023-06078-1

23. Lim ZW, Pushpanathan K, Yew SME, Lai Y, Sun CH, Lam JSH, et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google bard. *EBioMedicine*. (2023) 95:104770. doi: 10.1016/j.ebiom.2023.104770

24. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4 and ophthalmology operative notes. *Ann Biomed Eng*. (2023). doi: 10.1007/s10439-023-03263-5

25. Chen JS, Baxter SL. Applications of natural language processing in ophthalmology: present and future. *Front Med*. (2022) 9:906554. doi: 10.3389/fmed.2022.1078403

26. Gui H, Tseng B, Hu W, Wang SY. Looking for low vision: predicting visual prognosis by fusing structured and free-text data from electronic health records. *Int J Med Inform*. (2022) 159:104678. doi: 10.1016/j.ijmedinf.2021.104678

27. Winkler SL, Finch D, Llanos I, Delikat J, Marszalek J, Rice C, et al. Retrospective analysis of vision rehabilitation for veterans with traumatic brain injury-related vision dysfunction. *Mil Med*. (2023) 188:e2982–6. doi: 10.1093/milmed/usad120

28. Winkler SL, Finch D, Wang X, Toyinbo P, Marszalek J, Rakoczy CM, et al. Veterans with traumatic brain injury-related ocular injury and vision dysfunction: recommendations for rehabilitation. *Optom Vis Sci*. (2022) 99:9–17. doi: 10.1097/OPX.0000000000001828

29. Balas M, Ing EB. Conversational ai models for ophthalmic diagnosis: comparison of chatgpt and the isabel pro differential diagnosis generator. *JFO Open Ophthalmol*. (2023) 1:100005. doi: 10.1016/j.jfop.2023.100005

30. Visser M, Deliens L, Houttekier D. Physician-related barriers to communication and patient- and family-centred decision-making towards the end of life in intensive care: a systematic review. *Crit Care*. (2014) 18:604. doi: 10.1186/s13054-014-0604-z

31. Singh K, Thibodeau A, Niziol LM, Nakai TK, Bixler JE, Khan M, et al. Development and validation of a model to predict anterior segment vision-threatening eye disease using primary care clinical notes. *Cornea*. (2022) 41:974–80. doi: 10.1097/ICO.0000000000002877

32. Salazar H, Misra V, Swaminathan SS. Artificial intelligence and complex statistical modeling in glaucoma diagnosis and management. *Curr Opin Ophthalmol*. (2021) 32:105–17. doi: 10.1097/ICU.0000000000000741

33. Jalamangala Shivananjaiah SK, Kumari S, Majid I, Wang SY. Predicting near-term glaucoma progression: an artificial intelligence approach using clinical free-text notes and data from electronic health records. *Front Med*. (2023) 10:1157016. doi: 10.3389/fmed.2023.1157016

34. Wang SY, Tseng B, Hernandez-Boussard T. Deep learning approaches for predicting Glaucoma progression using electronic health records and natural language processing. *Ophthalmol Sci*. (2022) 2:100127. doi: 10.1016/j.xops.2022.100127

35. Sun MT, Singh K, Wang SY. Real-world outcomes of Glaucoma filtration surgery using electronic health records: an informatics study. *J Glaucoma*. (2022) 31:847–53. doi: 10.1097/IJG.0000000000002122

36. Azamfirei R, Kudchadkar SR, Fackler J. Large language models and the perils of their hallucinations. *Crit Care*. (2023) 27:120. doi: 10.1186/s13054-023-04393-x

37. Asensio-Sánchez VM. Artificial intelligence and new language models in ophthalmology: complications of the use of silicone oil in vitreoretinal surgery. *Arch Soc Esp Oftalmol (Engl Ed)*. (2023) 98:298–303.

38. Lawson MLA. Artificial intelligence in surgical documentation: a critical review of the role of large language models. *Ann Biomed Eng*. (2023). doi: 10.1007/s10439-023-03282-2

39. Salimi A, Saheb H. Large language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol*. (2023) 254:177–81. doi: 10.1016/j.ajo.2023.06.004

40. Ali MJ, Singh S. ChatGPT and scientific abstract writing: pitfalls and caution. *Graefes Arch Clin Exp Ophthalmol*. (2023) 261:3205–6. doi: 10.1007/s00417-023-06123-z

41. Abdullah YI, Schuman JS, Shabsigh R, Caplan A, Al-Aswad LA. Ethics of artificial intelligence in medicine and ophthalmology. *Asia-Pacific J. Ophthalmol. (Phila Pa)*. (2021) 10:289–98. doi: 10.1097/APO.0000000000000397

42. Shen Y, Heacock L, Elias J. ChatGPT and other large language models are double-edged swords. *Radiology*. (2023) 307:e230163. doi: 10.1148/radiol.230163

43. Tom E, Keane PA, Blazes M, Pasquale LR, Chiang MF, Lee AY, et al. Protecting data privacy in the age of AI-enabled ophthalmology. *Transl Vis Sci Technol*. (2020) 9:36. doi: 10.1167/tvst.9.2.36

44. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. (2018) 178:1544–7. doi: 10.1001/jamainternmed.2018.3763

45. Dow ER, Keenan TDL, Lad EM, Lee AY, Lee CS, Loewenstein A, et al. From data to deployment: the collaborative community on ophthalmic imaging roadmap for artificial intelligence in age-related macular degeneration. *Ophthalmology*. (2022) 129:e43–59. doi: 10.1016/j.ophtha.2022.01.002

46. González-Gonzalo C, Thee EF, Klaver CCW, Lee AY, Schlingemann RO, Tufail A, et al. Trustworthy AI: closing the gap between development and integration of AI systems in ophthalmic practice. *Prog Retin Eye Res*. (2022) 90:101034. doi: 10.1016/j.preteyeres.2021.101034

47. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*. (2023) 613:612. doi: 10.1038/d41586-023-00191-1

48. Chou YB, Kale AU, Lanzetta P, Aslam T, Barratt J, Danese C, et al. Current status and practical considerations of artificial intelligence use in screening and diagnosing retinal diseases: vision academy retinal expert consensus. *Curr Opin Ophthalmol*. (2023) 34:403–13. doi: 10.1097/ICU.0000000000000979

49. Li JO, Liu H, Ting DSJ, Jeon S, Chan RVP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res*. (2021) 82:100900. doi: 10.1016/j.preteyeres.2020.100900

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 07 December 2023
DOI 10.3389/fmed.2023.1309795

Check for updates

OPEN ACCESS

EDITED BY
Yanda Meng,
University of Liverpool, United Kingdom

REVIEWED BY
Jianyang Xie,
University of Liverpool, United Kingdom
Xu Chen,
University of Cambridge, United Kingdom

*CORRESPONDENCE
Guangtao Lu
✉ luguangtao@wust.edu.cn

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

RECEIVED 08 October 2023
ACCEPTED 22 November 2023
PUBLISHED 07 December 2023

CITATION
Zhao T, Guan Y, Tu D, Yuan L and Lu G (2023) Neighbored-attention U-net (NAU-net) for diabetic retinopathy image segmentation. *Front. Med.* 10:1309795. doi: 10.3389/fmed.2023.1309795

COPYRIGHT
© 2023 Zhao, Guan, Tu, Yuan and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Neighbored-attention U-net (NAU-net) for diabetic retinopathy image segmentation

Tingting Zhao[1†], Yawen Guan[1†], Dan Tu[1], Lixia Yuan[2] and Guangtao Lu[3*]

[1]The Second Department of Internal Medicine, Donghu Hospital of Wuhan, Wuhan, China, [2]The Department of Ophthalmology, Donghu Hospital of Wuhan, Wuhan, China, [3]Precision Manufacturing Institute, Wuhan University of Science and Technology, Wuhan, China

Background: Diabetic retinopathy-related (DR-related) diseases are posing an increasing threat to eye health as the number of patients with diabetes mellitus that are young increases significantly. The automatic diagnosis of DR-related diseases has benefited from the rapid development of image semantic segmentation and other deep learning technology.

Methods: Inspired by the architecture of U-Net family, a neighbored attention U-Net (NAU-Net) is designed to balance the identification performance and computational cost for DR fundus image segmentation. In the new network, only the neighboring high- and low-dimensional feature maps of the encoder and decoder are fused by using four attention gates. With the help of this improvement, the common target features in the high-dimensional feature maps of encoder are enhanced, and they are also fused with the low-dimensional feature map of decoder. Moreover, this network fuses only neighboring layers and does not include the inner layers commonly used in U-Net++. Consequently, the proposed network incurs a better identification performance with a lower computational cost.

Results: The experimental results of three open datasets of DR fundus images, including DRIVE, HRF, and CHASEDB, indicate that the NAU-Net outperforms FCN, SegNet, attention U-Net, and U-Net++ in terms of Dice score, IoU, accuracy, and precision, while its computation cost is between attention U-Net and U-Net++.

Conclusion: The proposed NAU-Net exhibits better performance at a relatively low computational cost and provides an efficient novel approach for DR fundus image segmentation and a new automatic tool for DR-related eye disease diagnosis.

KEYWORDS

image semantic segmentation, deep learning, diabetic retinopathy, neighbored-attention U-net, fundus image

1 Introduction

Recently, as the number of patients with diabetes mellitus (DM) has increased greatly and they tend to be younger, an increasing number of people suffer from diabetic retinopathy (DR) (1, 2). As an eye disease, DR may cause visual impairment or even blindness if not diagnosed and treated in a timely manner (3). DR typically results in optic disc (OD) lesions. These lesions

involve abnormal changes in retinal blood flow, and the abnormalities primarily include microaneurysms (MA), hard exudates, soft exudates, hemorrhages (HA), neovascularization (NV), and macular edema (ME) (4). These changes in the OD can be captured and recorded in images using a DR screening device, and OD abnormalities can be easily distinguished by experienced doctors by analyzing the fundus images. However, the manual diagnosis of DR requires doctors to check numerous images, which is time-consuming, resource-intensive, and expensive.

Owing to the increasing development of computer vision technologies, deep learning methods, especially image identification technology including image classification and image semantic segmentation method, have been introduced for automatic diagnosis of DR. As a method of image identification technologies based on computer vision, image classification algorithms typically preprocess the images first using image processing technologies and then enhance or extract some features from the preprocessed images, including histograms of oriented gradients (HOG), higher-order spectra (HOS), and speeded-up robust features (SURF). The extracted features are input into an intelligent classifier model with a category or label. After the classifier is trained, it is used to predict a new DR fundus or other medical images. It outputs a category or label that represents the type of disease (5, 6). The commonly used classifiers include support vector machine (SVM), genetic algorithm (GA), and convolutional neural networks (CNN) (4, 7). Orfao and Haar (8) compared the performance of different classifiers, and their experimental results indicated that the radial basis function SVM (RBF-SVM) model obtained a higher accuracy and F1-score using the HOG feature of the green channel. Ghoushchi et al. (9) combined fuzzy C-mean (FCM) and GA algorithms to identify diabetic and nondiabetic eye images with a relatively high recognition rate. Li et al. (10) obtained the features of the DR1 and Messidor datasets using a fine-tuning CNN and used an SVM model to classify the images. Le et al. (11) first selected the feature using an adaptive particle-grey wolf optimization method and classified the image using a multilayer perceptron (MLP). Their comparative results showed that the new algorithm predicted the images with a higher accuracy.

Moreover, because the CNN model shows a powerful ability for image enhancement, various CNN models have been introduced into image feature selection. CNN models are typically connected by a series of convolutional, activation, pooling, dropping, and fully connected layers, and based on the architecture of the backbones of the CNN, various CNN models, including AlexNet, VGG, DenseNet, ResNet, MobileNet, are used for DR and other medical image segmentation. Shanthi and Sabeenian (12) used an AlexNet with four convolution layers and three pooling layers to augment the fundus images of the Messidor dataset and classified the severity using filtered data. Khan et al. (13) modified the architecture of VGG16 to improve the performance of DR image diagnosis and tested the identification performance using the Kaggle dataset. Kobat et al. (14) first separated the DR image into parts by resizing and dividing the original image and then trained DenseNet201 and SVM classifiers to augment and estimate the DR images, respectively. Al-Moosawi and Khudeyer (15) diagnosed four different categories of DR using a trained ResNet34 and compared the performances of different DL architectures. The identification results of the fundus images from APTOS 2019 and IDRiD showed that ResNet34 performed better in image feature enhancement. Moreover, considering its powerful target detection ability, the popular Yolo V3 model was introduced for automatic DR fundus image identification by Pal et al. (16). Similar studies have been conducted by Wang et al. (17), Das et al. (18), Mohamed et al. (19), and Santos et al. (20).

In contrast to the aforementioned image classification methods, image semantic segmentation methods detect and classify images at each pixel (21, 22). Therefore, after semantic image segmentation, the retinal blood vessels or other important structures of the DR or other medical images are augmented, and the lesion area is directly detected and located. Image semantic segmentation algorithms are derived from or based on CNN, and typical image semantic segmentation architectures are fully convolutional networks (FCN), SegNet, pyramid scene parsing networks (PSPNet), DeepLab, Unet, etc. (23, 24). To achieve a tradeoff between semantic and location information, Wang et al. (25) improved the original R-FCN by adding an upsampling unit in the common ResNet101 and used a feature pyramid network to generate a feature map with different feature map levels. Using the modified R-FCN, higher sensitivity and specificity for DR image segmentation were obtained. To increase the feature map resolution, the original SegNet used an encoder to obtain the feature maps and employed a decoder to up-sample the feature maps (26). SegNet was first proposed by Saha et al. (27) for road and indoor scene segmentation, and Ananda et al. (28) introduced SegNet for DR image segmentation. To make optimal use of the global feature in image segmentation tasks, a global pyramid pooling layer and certain new strategies were proposed in PSPNet and compared with FCN (29). Fang et al. (30) combined a phase-up-sampling module and PSPNet for fundus image segmentation. This improved model obtained higher intersection over union (IoU) and pixel accuracy than the native PSPNet. Chen et al. (31) introduced arouse convolution and a conditional random field (32) to strengthen the boundary details and finally obtained a better image segmentation effect. This architecture is known as DeepLab v1. To further improve the identification accuracy of the boundary, DeepLab v2 (33), DeepLab v3 (34), and DeepLab v3+ (35) were developed by modifying certain modules of the DeepLab v1 network. Some researchers have reviewed and compared the performances of other networks (36).

However, the performance of these image segmentation algorithms is affected by the number of training samples. In addition, datasets of medical images, particularly images of rare cases, are typically insufficient. Therefore, the U-Net was first reported by Ronneberger et al. (37) to improve the performance of small-sample image segmentation. U-Net uses a symmetric architecture to suppress the key image features by down-sampling and to extract low-level features by skip connection and up-sampling. It finally exhibits excellent performance by fusing all the features. Moreover, various variants of U-Net have been developed by modifying or adding modules to improve their accuracy. However, these variants typically achieve excellent performance by fusing multi-scale feature maps with dense links between the encoder and decoder, and as a result, they usually need the expense of computational and time costs. Therefore, to balance the identification performance and computational of the algorithm, a novel U-Net named neighboring attention U-Net is designed for DR fundus image semantic segmentation.

The paper is structured as follows: Section 2 summarizes and discusses the studies on U-Net and its variants. Section 3 introduces

the architecture and workflow of the proposed network. Section 4 provides the details of the datasets and compares the testing performances of the different networks. Section 5 summarizes the whole study.

2 Related previous works of U-Net family

Since the U-Net was first reported by Ronneberger et al. (37) in 2015, various variants of the U-Net have been developed and have displayed a wide and strong applicability for DR fundus, cell, lung, skin cancer, colorectal adenocarcinoma gland, and coronary artery image segmentation in the field of medicine. Figure 1 shows the structure of U-Net and its variants. Apart from the original U-Net, the U-Net family primarily includes attention U-Net, residual U-Net, residual-attention U-Net, recurrent residual convolutional neural network (RRCN) based on U-Net (R2U-Net), U-Net++, Nested U-Net, etc. As shown in Figure 1, in these variants, some modules are modified or added to further focus on their ability for image feature extraction and fusion at different levels.

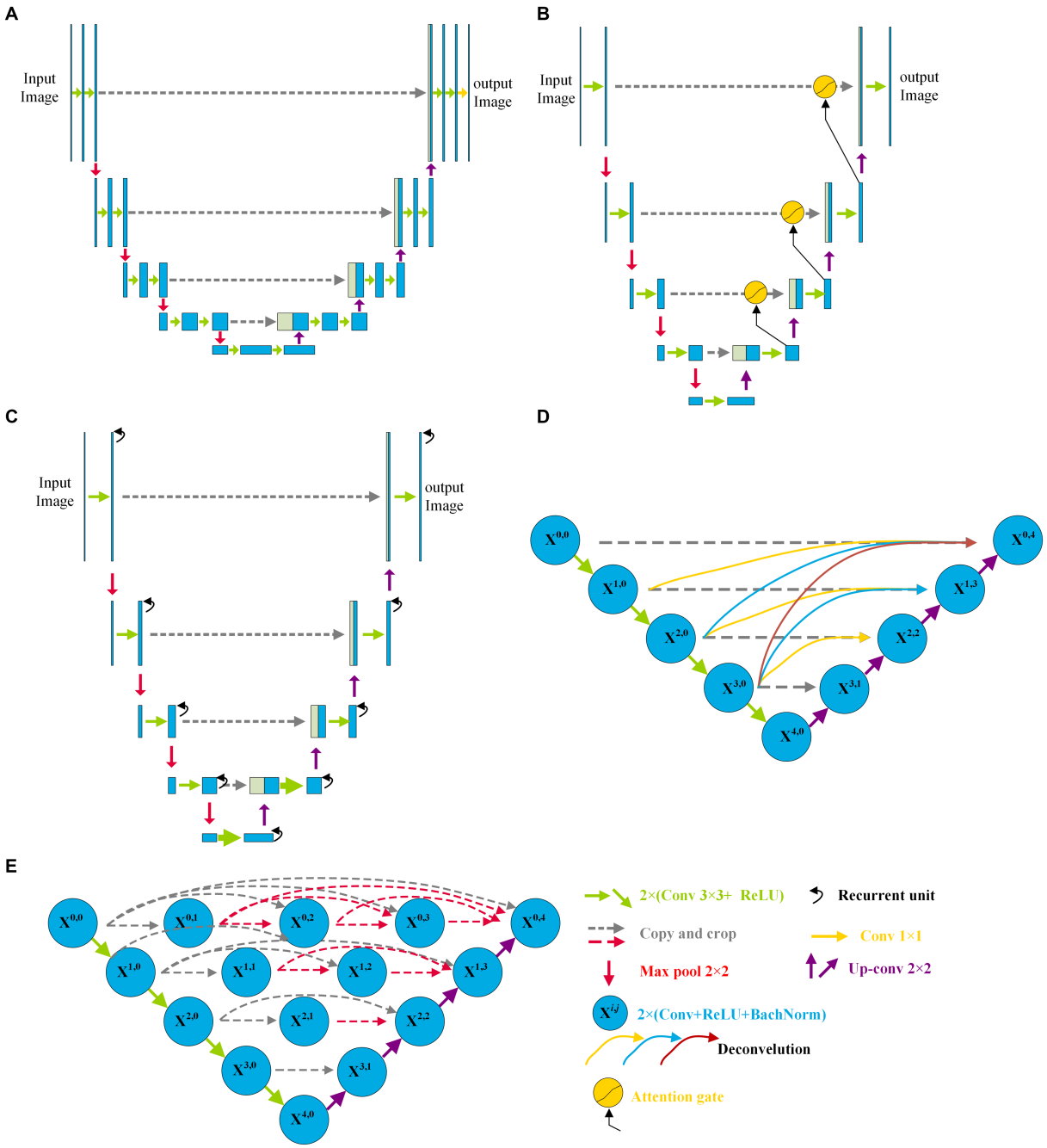

FIGURE 1
Architectures of some variants of U-Net: **(A)** U-Net, **(B)** Attention U-Net, **(C)** R2U-Net, and **(D)** CE-Net; **(E)** U-Net++.

Inspired by the concept of FCN, a new network with an encoder and decoder was designed in 2015, and it was called "U-Net" because of its symmetric architecture. As shown in Figure 1A, the U-Net encoder primarily consists of convolution, ReLU activation, and max pooling modules, whereas the decoder primarily consists of up-convolution, convolution, ReLU activation, and max pooling modules. Moreover, to join the features of the encoder, a cropping operation is performed at the corresponding levels in the decoding process. Owing to its innovation in image feature extraction and fusion at different levels, U-Net has displayed outstanding performance in medical image segmentation with a small sample size. Çiçek et al. (38) transformed a 2D U-Net model into a 3D U-Net for volumetric segmentation of biomedical images using 3D modules. To be more sensitive to the local region, Oktay et al. (39) added three attention gates before the copy and cropping operations in attention U-Net as shown in Figure 1B. The attention U-Net had a higher Dice score and a lower surface distance in the CT abdominal image segmentation. To simplify training and decrease the degradation of the U-Net model, Zhang et al. (40) introduced a residual mechanism into the architecture and designed a deep residual U-Net model for road image segmentation. The residual U-Net inherits the depth of the residual network and feature fusion ability at different levels. Combining the advantages of the residual, recurrent, and U-Net modules, Alom et al. (41) designed R2U-Net in 2019. In R2U-Net of Figure 1C, the introduction of RRCN modules further enhances the feature extraction ability at each pixel and increases its depth. Owing to the powerful abilities of the modules, R2U-Net displayed a better response than U-Net in various medical image segmentations. Considering that the pooling and convolution operations of U-Net typically result in a loss of feature resolution and spatial information, Gu et al. (42) designed a network called CE-Net shown in Figure 1D, based on U-Net. In addition to the encoder and decoder, CE-Net has a context extractor for dense atrous convolution and residual modules. The advantages of the proposed context extractor in CE-Net are compared and proven by segmenting different types of images. Moreover, to achieve high accuracy in medical image segmentation, Zhou et al. (43) nested different layers of U-Net by adding new skip pathways; therefore, this network is called U-Net++ or Nested U-Net. As shown in Figure 1E, in U-Net ++, the redesigned pathways mapped the feature maps of the encoder to the decoder; consequently, the feature maps of the two networks were fused. As the number of pathways increased significantly, the parameters of the model expanded, and the computational cost increased. The experimental test of CT image segmentation showed that it achieved an average IoU improvement of approximately 3%, and its total parameters increased by approximately 16.5% compared with U-Net. To balance the computational cost and segmentation performance, the AdaBoosted supervision mechanism was added to U-Net, and this architecture was called ADS_U-Net (44). In this model, deep supervision and performance-weighted combination were conducted to reduce the correlations between different feature maps and obtain excellent comprehensive performance in image segmentation and computation costs. Inspired by U-Net++, Li et al. (45) proposed a residual-attention U-Net++ in which the residual and attention modules were embedded into U-Net++. With the assistance of these two modules, the degradation was weakened and irrelevant features were filtered; therefore, the target feature was enhanced. As a result, the modified U-Net++ obtained higher IoU and Dice scores.

As shown in Figure 1, compared with the original architecture of U-Net, attention U-Net, U-Net++, and residual-attention U-Net++ have more links between the low- and high-dimensional feature maps, and these features of different levels are well combined, which filters the low-relevance features and boosts the target features. More complicated nested layers assist in improving the performance; however, they introduce a larger number of parameters and increase the computational cost. Therefore, to balance computational performance and cost, neighbored attention U-Net (NAU-Net) is proposed for DR and other medical image segmentation. In this new network, neighboring high- and low-dimensional feature maps are fused by an attention gate to filter the target features at a relatively low cost.

3 Methodology

3.1 Whole architecture of NAU-Net

Figure 2 shows the NAU-Net's structure. As shown in Figure 2, the NAU-Net adds four attention gates to map the feature maps of the encoder to the decoder at different levels. The inputs to the attention gate are the two neighboring feature maps of the encoder and decoder at the same level. Using these attention gates, similar feature maps are fused, and the target features are enhanced. Moreover, this network only uses neighboring layers and does not include the inner layers commonly used in U-Net++ and residual attention U-Net++. Consequently, the proposed network incurs a lower computational cost.

To fuse the feature maps conveniently and make the output size similar to the input image, the conventional kernel size is 3×3, and its stride and padding are one. After the convolution operation, the ReLU, batch normalization, and max pooling operations are performed. The maximum pooling is 2×2, and the stride is two. The up-convolution operation included up-sampling, 2×2 convolution with a stride and padding of one, batch normalization, and ReLU operations. Finally, a 3×3 convolution operation transfers the filtered image to one channel.

3.2 Neighbored feature maps fusion

As the convolutional layers of encoder increase, more and more detailed features of the target get loss. However, there is some similarity between the two-neighboring high-dimensional feature maps in the encoder, and this connection between the maps faraway gets weaker. Therefore, to enhance the common features in the maps with a relative low computation cost, only the two neighboring feature maps of the encoder are fused by an attention in NAU-Net.

Before the feature maps of the encoder and decoder are combined, the two neighboring feature maps of the encoder are fused. Because the dimensions of the two neighbored feature maps of the encoder are different, the lower-dimensional feature map is first filtered by an up-convolution operation and then fused by a concatenation operation. The entire fusion operation is shown in

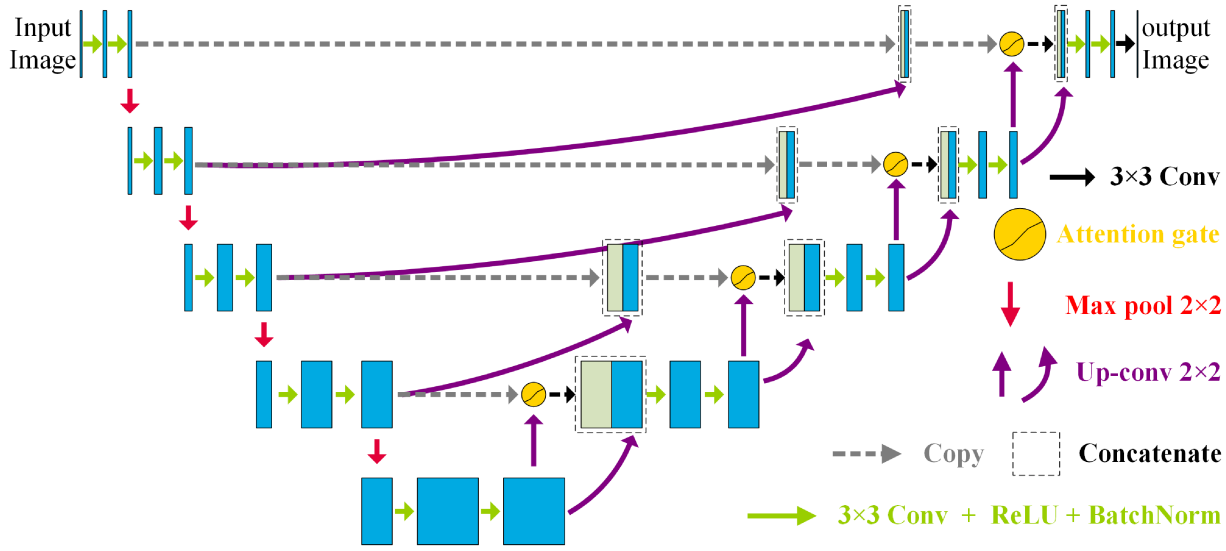

FIGURE 2
Structure of the NAU-Net.

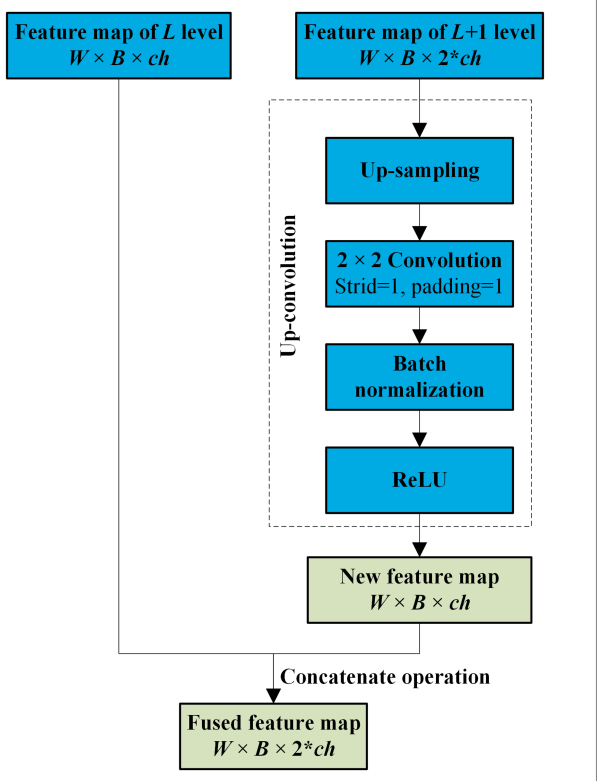

FIGURE 3
Neighbored feature maps fusion.

Figure 3. As shown in Figure 3, after the feature map of the L + 1 level with dimensions $W \times B \times 2 \times ch$ is processed by up-sampling, convolution, batch normalization, and ReLU sequentially, a new feature map with the same dimensions as the Lth feature map is obtained. Subsequently, the new feature map is concatenated with the Lth feature map.

3.3 Attention mechanism in NAU-Net

The high-dimensional feature maps of the encoder usually contain fine-grained features of the target, while the low-dimensional ones of the decoder contain coarse texture of the target. Therefore, to increase the identification accuracy, the multi-scale features of the target in low- and high-dimensional feature maps are extracted and fused by the attention mechanism in NAU-Net. Figure 4 shows the entire procedure for the attention mechanism in NAU-Net. In Figure 4, the low- and high-dimensional feature maps are inputted to a common attention gate, and the output $d_L^{'}$ of the attention gate is expressed as follows:

$$q_L = \sigma_1\left(w_g\left(e_L^{'}\right) + w_x\left(d_L\right)\right) \quad (1)$$

$$\alpha_L = w_\alpha\left(q_L\right) \quad (2)$$

$$d_L^{'} = \alpha_L d_L \quad (3)$$

where σ_1 represents the ReLU operation, d_L represents the feature map of decoder at the level L, w_g and w_x represent the plain convolution and batch normalization operations of the feature maps $e_L^{'}$ and d_L, respectively, α_L represents the attention coefficient, w_α represents the combining operation convolution, batch normal, and sigmoid activation. It is noteworthy that the kernel size of the attention gate convolution is 3×3 with a stride of 1.

3.4 Loss function

In this study, binary cross-entropy and Dice loss (BCE-Dice loss) are selected as loss functions to evaluate segmentation performance

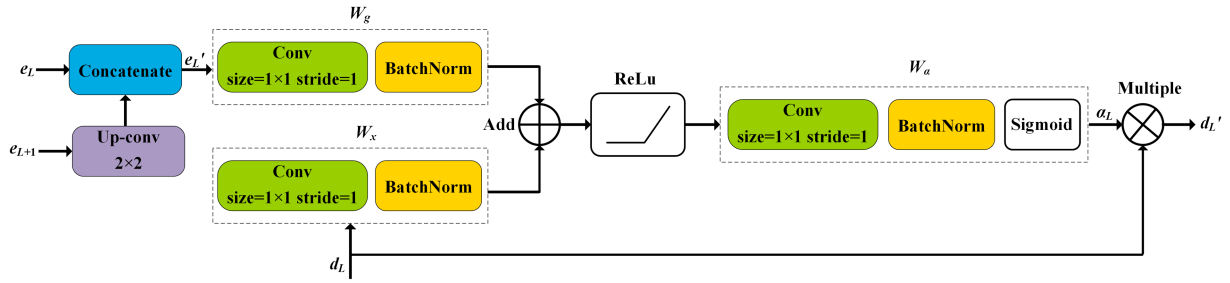

FIGURE 4
Attention gate in NAU-Net.

(46). The *i*th predicted image and its corresponding ground-truth image are p_i and g_i. The BCE-Dice loss is expressed as follows:

$$L_{BCE-Dice} = -\frac{1}{N}\sum_{i=1}^{N}\left(p_i \log\left(g_i\right) + \left(1-p_i\right)\log\left(1-g_i\right)\right) + \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{2TP_i}{2TP_i + FP_i + FN_i}\right) \quad (4)$$

where N is the total number of images and TP_i, FP_i, and FN_i are the true positives, false positives, and false negatives of the *i*th predicted image, respectively.

4 Experiments and results

To test and compare the performance of NAU-Net, datasets of DR fundus images, including digital retinal images for vessel extraction (DRIVE), high-resolution fundus (HRF), and CHASEDB, were tested. Figure 5 shows the fundus images from the three datasets. Moreover, the segmentation performance of NAU-Net was compared with FCN, SegNet and two variants of U-Net, namely attention U-Net and U-Net++, whose networks are similar to the proposed model. The proposed model and a few existing networks were established by using the PyTorch framework (version 1.10.0), and all experimental tests were conducted at the High-Performance Computing Center at Wuhan University of Science and Technology. All the tests were conducted on a computer with four NVIDIA Tesla V100S GPUs, and the memory capacity of each GPU board was 32 GB.

4.1 Datasets

4.1.1 DRIVE dataset

The total of 40 color DR fundus images from DRIVE were used in this study (47). The resolution of the images was 584 × 565 pixels per channel, and each image had three channels. The ratio of the training and testing split was 20:20. The ground truth of each image was manually segmented and marked by one or two different ophthalmological experts.

4.1.2 HRF dataset

The HRF (48) included 45 original DR fundus images, including 15 healthy, 15 DR, and 15 glaucomatous fundus images. All images were manually marked by experts. The image resolution was 3,504 × 2,336 pixels. Moreover, in this study, healthy and DR fundus images were imported; among them, 26 images were selected as the training set, and the remaining four were selected as the testing set.

4.1.3 CHASEDB dataset

The 28 color fundus images from CHASEDB (49) were also used to display the performance of NAU-Net. Each image contained 999 × 960 pixels and was marked by two independent experts. The training and testing sets contained 21 and seven images, respectively.

4.1.4 Evaluation metrics

To display and compare segmentation performance, some commonly used evaluation metrics, including the Dice score, IoU, accuracy (AC), and precision (PC), were introduced in this study. These four metrics are obtained as follows:

$$DC = \frac{2TP}{FP + FN + 2TP} \quad (5)$$

$$IoU = \frac{TP}{FP + FN + TP} \quad (6)$$

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$PC = \frac{TP}{TP + FP} \quad (8)$$

where TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively.

Moreover, the computational cost was evaluated by comparing the total number of parameters and GPU memory demands of the models.

4.2 Results

During the inference process, the Adam optimizer was selected, and its learning rate was adjusted using the CosineAnnealinLR scheduler. The maximum number of iterations was 10. The minimum learning rate of the scheduler was 0.0001. The total number of epochs was 140, and the batch size was selected as four. All images were

FIGURE 5
Fundus images of open datasets: **(A)** DRIVE, **(B)** Ground true image of **(A,C)** HRF; **(D)** Ground truth image of **(C)**, **(E)** CHASEDB, and **(F)** Ground truth image of **(E)**.

resized to 576 × 576 pixels before inference. Before the images were put into the model, they were preprocessed by normalization with parameters mean = [0.485, 0.456, 0.406] and std. = [0.229, 0.224, 0.225]. Moreover, the information of libraries used in this study is available at website https://github.com/Aynor007/MyNAUnet.

4.2.1 Computation cost comparison

To evaluate the computational cost of NAU-Net, the number of parameters, total memory demand, and complexity of different models, including FCN, SegNet, attention U-Net, U-Net++, and NAU-Net, were evaluated and compared. The model complexity was evaluated by the number of floating points (FLOPs) and multiple adds (MAdds), and it was calculated with the help of Torchstat 0.0.7. Table 1 lists the total number of parameters, total memory demand, number of FLOPs, and number of MAdds. Table 1 shows that the computational cos of the U-Net family is higher than other models including FCN and SegNet. It should be also noted that FCN and SegNet usually need a relative larger number of training samples to obtain a satisfactory identification accuracy, which finally results in a significant increase of the training cost.

Moreover, Table 1 also demonstrates that the number of parameters in NAU-Net is slightly higher than those of attention U-Net and U-Net++. The total memory of NAU-Net is 20.63% higher than that of attention U-Net and 9.53% lower than that of U-Net++. Moreover, the number of FLOPs in NAU-Net is 33.29% higher than that of attention U-Net, it is 35.68% lower than U-Net++. The number

of MAdds in NAU-Net is 33.31% higher than that of attention U-Net, which is 35.84% lower than that of U-Net++. To reduce the semantic gap between the low- and high-dimensional feature maps, a series of nested pathways are designed in the U-Net++, and as a result the computational cost accordingly increases. However, in the NAU-Net, only the neighboring high- and low-dimensional feature maps are linked. Moreover, since only the two feature maps with the same dimension are connected by an attention gate in the attention U-Net, the attention U-Net has less parameters than NAU-Net. Therefore, the computational cost of NAU-Net is between the cost of attention U-Net and U-Net++.

4.2.2 DRIVE image segmentation

Figure 6 shows three DR fundus images of the DRIVE dataset, their ground-truth images of retinal blood vessels, and the identification results of attention U-Net, U-Net++, and NAU-Net. Figure 6 clearly demonstrates that the proposed NAU-Net can identify some tiny and small retinal blood vessels of the DR fundus while the other two models detect less, which indicates that the fusion operation of the neighboring feature maps successfully extracts detailed features from the encoder, and therefore the proposed NAU-Net displays a better performance of tiny and small retinal blood vessel segmentation than attention U-Net and U-Net++.

Table 2 compares the segmentation performance of the DRIVE DR images obtained using the proposed NAU-Net and other 5 existing models including FCN, SegNet, attention U-Net and U-Net++. Table 2 clearly shows that the proposed NAU-Net obtained the maximum values of the Dice score, IoU, and accuracy for DR image segmentation of the DRIVE dataset among the five models. Since FCN and SegNet usually needs a relative larger number of training samples to obtain a satisfactory performance, their evaluation metrics are much lower than the models of U-Net family. Moreover, compared to attention U-Net and U-Net++, NAU-Net achieves a performance improvement from 0.18 to 7.10%, which indicates that the proposed

TABLE 1 Parameters, memory, FLOPs, and Madd of different models.

Models	Number of parameters (MB)	Total memory (GB)	FLOPs (G)	MAdds (G)
FCN	15.11	1.07	102.13	203.94
SegNet	29.44	1.14	203.02	405.73
Attention U-Net	34.88	6.06	337.26	673.77
U-Net++	36.63	8.08	698.94	1,400
NAU-Net	37.25	7.31	449.53	898.19

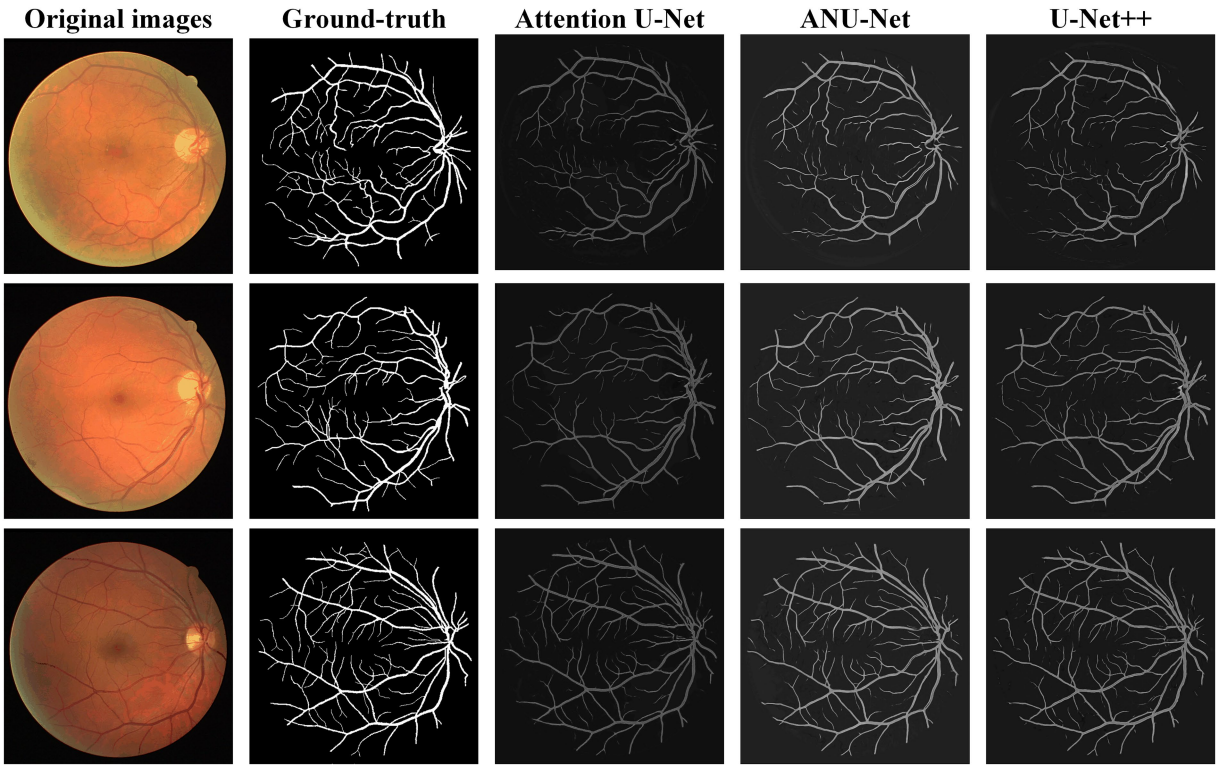

FIGURE 6
Image segmentation results of different models for DRIVE dataset.

NAU-Net has a stronger DR fundus image segmentation ability for the DRIVE dataset than the other two U-Net variants.

4.2.3 HRF image segmentation

Figure 7 shows three DR fundus images from the HRF dataset, their ground-truth images of retinal blood vessels, and the identification results of attention U-Net, U-Net++, and NAU-Net. Figure 7 shows that after the training, attention U-Net successfully detects most of the large vessels, while U-Net++ identifies some tiny retinal blood vessels that are not identified in the original image or ground truth. By contrast, the proposed NAU-Net correctly detects most of the vessels with the help of the fusion operation of the neighboring feature maps, including some tiny ones, which demonstrates that the proposed NAU-Net displays a better performance in retinal blood vessel segmentation than attention U-Net and U-Net++.

Table 3 compares the segmentation performances of the HRF DR images obtained using the proposed NAU-Net and other 5 existing models including FCN, SegNet, attention U-Net, and U-Net++. Similarly, Table 3 demonstrates that FCN and SegNet display a relative worse performance than the U-Net family. Table 3 also clearly shows that the proposed NAU-Net obtained the maximum values of the accuracy and precision for DR image segmentation of the HRF dataset among the attention U-Net, U-Net++, and NAU-Net, and its Dice and IoU are very close to the ones of U-Net++. Moreover, compared to attention U-Net and U-Net++, NAU-Net achieves a performance improvement from 0.28 to 9.19%, which indicates that NAU-Net has a stronger ability to DR fundus image segmentation for the HRF dataset than the other two U-Net variants, and the improvement of the proposed model is benefit to feature extraction.

4.2.4 CHASEDB image segmentation

Figure 8 shows three DR fundus images from CHASEDB, their ground-truth images of retinal blood vessels, and the identification results of attention U-Net, U-Net++, and NAU-Net. Table 4 lists the segmentation performance for the CHASEDB DR images obtained

TABLE 2 DRIVE DR image segmentation performance of NAU-Net and other models.

Models		Metrics (Mean ± Standard deviation)			
		Dice	IoU	Accuracy	Precision
FCN		0.614 ± 0.109	0.45 ± 0.095	0.940 ± 0.008	0.704 ± 0.082
SegNet		0.663 ± 0.111	0.505 ± 0.109	0.942 ± 0.028	0.743 ± 0.148
Attention U-Net		0.730 ± 0.145	0.592 ± 0.149	0.950 ± 0.039	0.799 ± 0.16
U-Net++		0.745 ± 0.147	0.609 ± 0.134	0.960 ± 0.013	0.820 ± 0.068
NAU-Net		0.750 ± 0.133	0.613 ± 0.126	0.962 ± 0.013	0.855 ± 0.055
Improvement (%)	Over Attention U-Net	2.64	3.50	1.21	7.10
	Over U-Net++	0.69	0.71	0.18	4.29

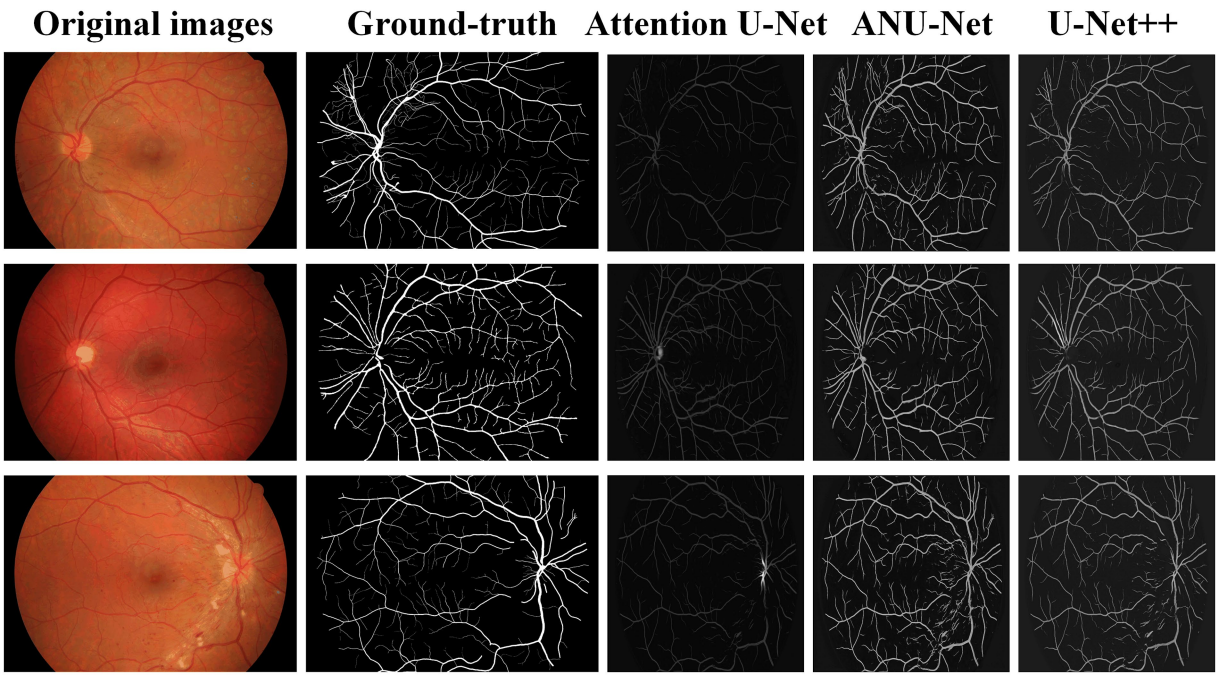

FIGURE 7
Image segmentation results of different models for HRF dataset.

TABLE 3 HRF DR image segmentation performance of NAU-Net and other models.

Models		Metrics (Mean ± Standard deviation)			
		Dice	IoU	Accuracy	Precision
FCN		0.547 ± 0.013	0.377 ± 0.012	0.924 ± 0.008	0.497 ± 0.047
SegNet		0.592 ± 0.096	0.427 ± 0.1	0.952 ± 0.003	0.78 ± 0.086
Attention U-Net		0.765 ± 0.035	0.621 ± 0.046	0.965 ± 0.003	0.76 ± 0.079
U-Net++		0.787 ± 0.044	0.651 ± 0.06	0.966 ± 0.006	0.733 ± 0.087
NAU-Net		0.786 ± 0.032	0.649 ± 0.044	0.969 ± 0.004	0.801 ± 0.108
Improvement (%)	Over Attention U-Net	2.74	4.49	0.37	5.39
	Over U-Net++	−0.09	−0.29	0.28	9.19

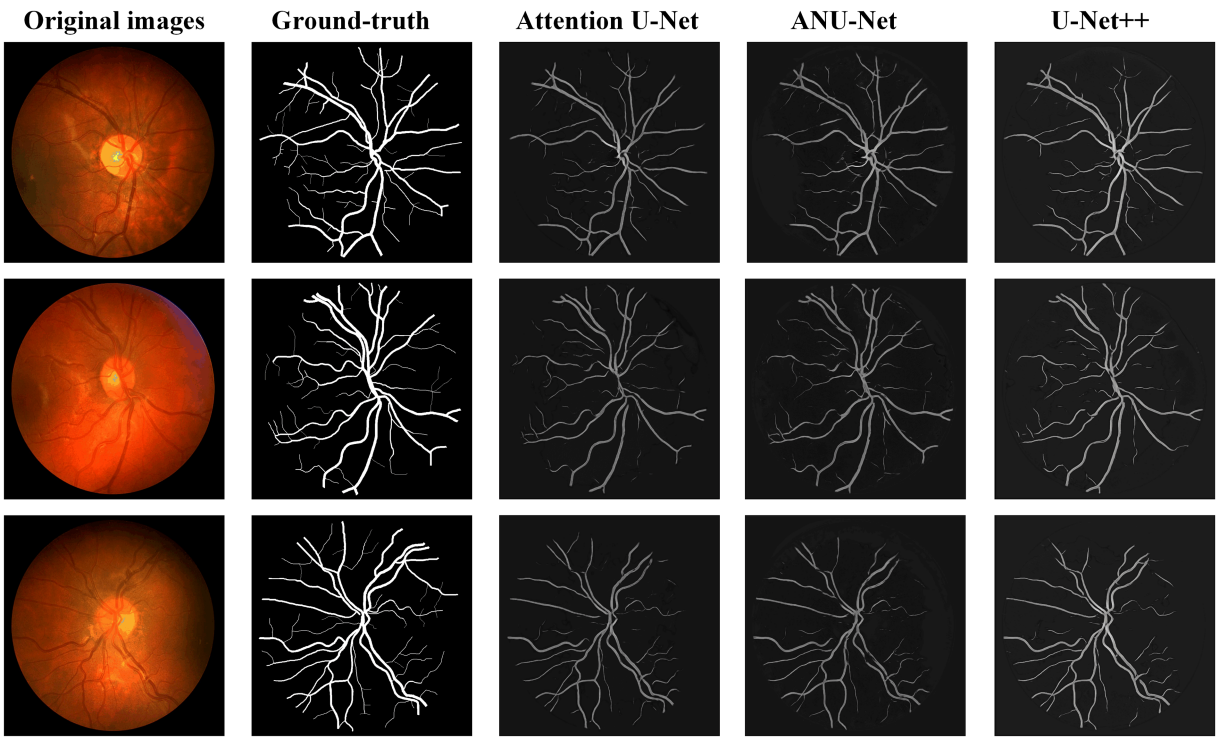

FIGURE 8
Image segmentation results of different models for CHASEDB dataset.

TABLE 4 CHASEDB DR image segmentation performance of NAU-Net and other models.

Models		Metrics (Mean ± Standard deviation)			
		Dice	IoU	Accuracy	Precision
FCN		0.625 ± 0.081	0.459 ± 0.082	0.944 ± 0.006	0.584 ± 0.051
SegNet		0.475 ± 0.157	0.325 ± 0.13	0.948 ± 0.004	0.765 ± 0.067
Attention U-Net		0.736 ± 0.104	0.592 ± 0.116	0.967 ± 0.006	0.800 ± 0.059
U-Net++		0.738 ± 0.085	0.591 ± 0.095	0.968 ± 0.005	0.839 ± 0.051
NAU-Net		0.755 ± 0.052	0.609 ± 0.063	0.969 ± 0.003	0.829 ± 0.058
Improvement (%)	Over Attention U-Net	2.54	2.86	0.19	3.60
	Over U-Net++	2.30	3.02	0.08	−1.18

using the proposed NAU-Net and other 5 existing models including FCN, SegNet, attention U-Net, and U-Net++. Table 4 clearly shows that the proposed NAU-Net obtains the maximum value of the Dice score, IoU, and accuracy for DR image segmentation of the CHASEDB

dataset among the five models, and U-Net++ achieves the highest precision. Moreover, compared to attention U-Net and U-Net++, NAU-Net improves the segmentation performance with an average increase of 0.08 to 3.60%, which demonstrates that NAU-Net has a stronger ability for image segmentation for the CHASEDB dataset than the other two U-Net variants.

5 Conclusion

In this study, to achieve a balance between identification performance and computational cost, a modified U-Net called NAU-Net is proposed for image segmentation of the DR fundus. In our new network, only the neighboring high- and low-dimensional feature maps of both the encoder and decoder are fused using four attention gates. With the help of this improvement, the common target features in the high-dimensional feature maps of encoder are enhanced, and they are also fused with the low-dimensional feature map of decoder by using these attention gates. Moreover, this network uses only neighboring layers and does not include inner layers commonly used in U-Net++. Consequently, the proposed network incurs a better identification performance with a lower computational cost. The experimental results of three open datasets of DR fundus images, including DRIVE, HRF, and CHASEDB, show that the proposed NAU-Net obtains higher scores for the Dice score, IoU, accuracy, and precision than FCN, SegNet, attention U-Net and U-Net++, while its computation cost is between the costs of the two models of attention U-Net and U-Net++. Therefore, the proposed NAU-Net exhibits better performance with a relatively low computational cost and provides an efficient novel method for DR fundus image segmentation and a new automatic tool for DR-related eye disease diagnosis. In future work, we will develop an end-to-end automatic diagnosis model that combines the proposed architecture with other classification models. Moreover, the architecture will be further improved for multitask image segmentation of DR fundus images with multiple types of lesions.

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

TZ: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. YG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. DT: Software, Validation, Writing – original draft. LY: Formal analysis, Writing – original draft. GL: Conceptualization, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Program of the Wuhan Municipal Health Commission (grant number: WX18D54) and the Program of the Health and Family Planning Commission of Hubei Province (grant number: WJ2019F006).

Acknowledgments

We would like to thank Editage (www.editage.com) for English language editing. We would also like to thank the High-Performance Computing Center at Wuhan University of Science and Technology for the support of the numerical calculation.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Publisher's note

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

References

1. Lin A, Xia H, Zhang A, Liu X, Chen H. Vitreomacular interface disorders in proliferative diabetic retinopathy: an optical coherence tomography study. *J Clin Med*. (2022) 11:11. doi: 10.3390/jcm11123266

2. Saleh E, Błaszczyński J, Moreno A, Valls A, Romero-Aroca P, De La Riva-Fernández S, et al. Learning ensemble classifiers for diabetic retinopathy assessment. *Artif Intell Med*. (2018) 85:50–63. doi: 10.1016/j.artmed.2017.09.006

3. Wang M, Lin T, Peng Y, Zhu W, Zhou Y, Shi F, et al. Self-guided optimization semi-supervised method for joint segmentation of macular hole and cystoid macular edema in retinal OCT images. *IEEE Trans Biomed Eng*. (2023) 70:2013–24. doi: 10.1109/TBME.2023.3234031

4. Shaukat N, Amin J, Sharif MI, Sharif MI, Kadry S, Sevcik L. Classification and segmentation of diabetic retinopathy: a systemic review. *Appl Sci*. (2023) 13:3108. doi: 10.3390/app13053108

5. Cen L-P, Ji J, Lin J-W, Ju S-T, Lin H-J, Li T-P, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun*. (2021) 12:4828. doi: 10.1038/s41467-021-25138-w

6. Meng Y, Preston FG, Ferdousi M, Azmi S, Petropoulos IN, Kaye S, et al. Artificial intelligence based analysis of corneal confocal microscopy images for diagnosing peripheral neuropathy: a binary classification model. *J Clin Med*. (2023) 12:12. doi: 10.3390/jcm12041284

7. Winder RJ, Morrow PJ, McRitchie IN, Bailie JR, Hart PM. Algorithms for digital image processing in diabetic retinopathy. *Comput Med Imaging Graph*. (2009) 33:608–22. doi: 10.1016/j.compmedimag.2009.06.003

8. Orfao J, Van Der Haar D. "A comparison of computer vision methods for the combined detection of glaucoma, diabetic retinopathy and cataracts." (2021). In: Medical image understanding and analysis: 25th annual conference (MIUA 2021). 30–42.

9. Ghoushchi SJ, Ranjbarzadeh R, Dadkhah AH, Pourasad Y, Bendechache M. An extended approach to predict retinopathy in diabetic patients using the genetic algorithm and fuzzy C-means. *Biomed Res Int*. (2021) 2021:5597222. doi: 10.1155/2021/5597222

10. Li X, Pang T, Xiong B, Liu W, Liang P, Wang T. "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification." (2017). 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 1–11.

11. Le TM, Vo TM, Pham TN, Dao SVT. A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. *IEEE Access*. (2021) 9:7869–84. doi: 10.1109/ACCESS.2020.3047942

12. Shanthi T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. *Comput Electr Eng*. (2019) 76:56–64. doi: 10.1016/j.compeleceng.2019.03.004

13. Khan Z, Khan FG, Khan A, Rehman ZU, Shah S, Qummar S, et al. Diabetic retinopathy detection using VGG-NIN a deep learning architecture. *IEEE Access*. (2021) 9:61408–16. doi: 10.1109/ACCESS.2021.3074422

14. Kobat SG, Baygin N, Yusufoglu E, Baygin M, Barua PD, Dogan S, et al. Automated diabetic retinopathy detection using horizontal and vertical patch division-based pre-trained DenseNET with digital fundus images. *Diagnostics (Basel)*. (2022) 12:1975. doi: 10.3390/diagnostics12081975

15. Al-Moosawi NMA-MM, Khudeyer RS. ResNet-34/DR: a residual convolutional neural network for the diagnosis of diabetic retinopathy. *IJCAI*. (2021) 45:115–24. doi: 10.31449/inf.v45i7.3774

16. Prayas Pal Swagata Kundu Ashis Kumar Dhara. Detection of red lesions in retinal fundus images using YOLO V3. *Curr Indian Eye Res J Ophthalmic Res Group*. (2020) 7:49–53.

17. Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*. (2015) 149:708–17. doi: 10.1016/j.neucom.2014.07.059

18. Das S, Kharbanda K, Raman R. Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. *Biomed Sig Process Control*. (2021) 68:102600. doi: 10.1016/j.bspc.2021.102600

19. Mohamed N.A., Zulkifley M.A., Abdani S.R. "Spatial pyramid pooling with atrous convolutional for MobileNet". In: 2020 IEEE Student Conference on Research and Development (2020).

20. Santos C, Aguiar M, Welfer D, Belloni B. A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model. *Sensors (Basel)*. (2022) 22:6441. doi: 10.3390/s22176441

21. Xu C, Chen Z, Zhang X, Peng Y, Tan Z, Fan Y, et al. Accurate C/D ratio estimation with elliptical fitting for OCT image based on joint segmentation and detection network. *Comput Biol Med*. (2023) 160:106903. doi: 10.1016/j.compbiomed.2023.106903

22. Wang X, Zhang Y, Ma Y, Kwapong WR, Ying J, Jiayi L, et al. Automated evaluation of retinal hyperreflective foci changes in diabetic macular edema patients before and after intravitreal injection. *Front Med*. (2023) 10:1280714. doi: 10.3389/fmed.2023.1280714

23. Yao Chenpu, Zhu Weifang, Wang Meng, Zhu Liangjiu, Huang Haifan, Chen Haoyu, et al. "SANet: a self-adaptive network for hyperreflective foci segmentation in retinal OCT images." (2021). In: Proceedings SPIE 11596, Medical Imaging 2021: Image Processing. 115962Y.

24. Guo Chao, Weifang Zhu, Ting Wang, Tian Lin, Haoyu Chen, Xinjian Chen. "Retinal OCT image report generation based on visual and semantic topic attention model." (2022). In: Proceedings SPIE 12032, Medical Imaging 2022: Image. 120322C-3.

25. Wang J, Luo J, Liu B, Feng R, Lu L, Zou H. Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm. *IET Comput Vis*. (2020) 14:1–8. doi: 10.1049/iet-cvi.2018.5508

26. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:2481–95. doi: 10.1109/TPAMI.2016.2644615

27. Saha O, Sathish R, Sheet D. Fully convolutional neural network for semantic segmentation of anatomical structure and pathologies in colour fundus images associated with diabetic retinopathy. *arXiv* (2019). doi: 10.48550/arXiv.1902.03122

28. Ananda S, Kitahara D, Hirabayashi A, Reddy KUK. "Automatic fundus image segmentation for diabetic retinopathy diagnosis by multiple modified U-nets and SegNets." (2019). In: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA).

29. Hengshuang Z, Jianping S, Xiaojuan Q, Xiaogang W, Jiaya J. "Pyramid scene parsing network." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*. (2017).

30. Fang X, Shen Y, Zheng B, Zhu S, Wu M. "Optic disc segmentation based on phase-fusion PSPNet." (2021). In: Proceedings of the 2nd international symposium on artificial intelligence for medicine sciences (ISAIMS 2021).

31. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv (2014). Available at: https://arxiv.org/abs/1412.7062v4

32. Philipp K, Vladlen K. Efficient inference in fully connected CRFs with gaussian edge potentials. *Adv Neural Inf Proces Syst*. (2011) 24:1–9.

33. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustainability*. (2021) 13:1224. doi: 10.3390/su13031224

34. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv (2017). Available at: https://arxiv.org/abs/1706.05587v3

35. Liang-Chieh C, Yukun Z, George P, Florian S, Hartwig A. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In: *Proceedings of the European Conference on Computer vision (ECCV)*. (2018).

36. Liu Q, Wang S, Dai Y, Zhang J, Wang Y, Zhou R. "Improved PSP-net segmentation network for automatic detection of neovascularization in color fundus images." (2022). In: IEEE International Conference on Visual Communications and Image Processing.

37. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. arXiv (2015). Available at: https://arxiv.org/abs/1505.04597v1 [Accessed October 7, 2023].

38. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. arXiv (2016). Available at: https://arxiv.org/abs/1606.06650v1 [Accessed October 7, 2023].

39. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-net: learning where to look for the pancreas. arXiv (2018). Available at: https://arxiv.org/abs/1804.03999v3 [Accessed October 7, 2023].

40. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-net. *arXiv* (2018). doi: 10.1109/LGRS.2018.2802944

41. Md Z A, Mahmudul H, Chris Y, Tarek M T, Vijayan K A. Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation. arXiv (2018). Available at: https://arxiv.org/abs/1802.06955

42. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging*. (2019) 38:2281–92. doi: 10.1109/TMI.2019.2903562

43. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: a nested U-net architecture for medical image segmentation. arXiv (2018). Available at: https://arxiv.org/abs/1807.10165v1 [Accessed October 7, 2023].

44. Yang Y, Dasmahapatra S, Mahmoodi S. ADS_UNet: a nested UNet for histopathology image segmentation. arXiv (2023). Available at: https://arxiv.org/abs/2304.04567v1 [Accessed October 7, 2023].

45. Li Z, Zhang H, Li Z, Ren Z. Residual-attention UNet++: a nested residual-attention U-net for medical image segmentation. *Appl Sci*. (2022) 12:7149. doi: 10.3390/app12147149

46. Rajput V. Robustness of different loss functions and their impact on networks learning capability. arXiv (2021). Available at: https://arxiv.org/abs/2110.08322 [Accessed October 7, 2023].

47. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging*. (2004) 23:501–9. doi: 10.1109/TMI.2004.825627

48. Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. *Int J Biomed Imaging*. (2013) 2013:154860. doi: 10.1155/2013/154860

49. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng*. (2012) 59:2538–48. doi: 10.1109/TBME.2012.2205687

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fmed.2023.1309097

Check for updates

OPEN ACCESS

EDITED BY
Yalin Zheng,
University of Liverpool, United Kingdom

REVIEWED BY
Amr Elsawy,
National Library of Medicine (NIH), United States
Jiong Zhang,
University of Southern California, United States

*CORRESPONDENCE
Weihua Yang
✉ benben0606@139.com
Jiantao Wang
✉ wangjiantao65@126.com
Zhe Zhang
✉ whypotato@126.com

RECEIVED 07 October 2023
ACCEPTED 11 December 2023
PUBLISHED 04 January 2024

CITATION
Wan C, Mao Y, Xi W, Zhang Z, Wang J and Yang W (2024) DBPF-net: dual-branch structural feature extraction reinforcement network for ocular surface disease image classification.
Front. Med. 10:1309097.
doi: 10.3389/fmed.2023.1309097

COPYRIGHT
© 2024 Wan, Mao, Xi, Zhang, Wang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

DBPF-net: dual-branch structural feature extraction reinforcement network for ocular surface disease image classification

Cheng Wan[1], Yulong Mao[1], Wenqun Xi[2], Zhe Zhang[2]*, Jiantao Wang[2]* and Weihua Yang[2]*

[1]College of Electronic Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, [2]Shenzhen Eye Institute, Shenzhen Eye Hospital, Jinan University, Shenzhen, China

Pterygium and subconjunctival hemorrhage are two common types of ocular surface diseases that can cause distress and anxiety in patients. In this study, 2855 ocular surface images were collected in four categories: normal ocular surface, subconjunctival hemorrhage, pterygium to be observed, and pterygium requiring surgery. We propose a diagnostic classification model for ocular surface diseases, dual-branch network reinforced by PFM block (DBPF-Net), which adopts the conformer model with two-branch architectural properties as the backbone of a four-way classification model for ocular surface diseases. In addition, we propose a block composed of a patch merging layer and a FReLU layer (PFM block) for extracting spatial structure features to further strengthen the feature extraction capability of the model. In practice, only the ocular surface images need to be input into the model to discriminate automatically between the disease categories. We also trained the VGG16, ResNet50, EfficientNetB7, and Conformer models, and evaluated and analyzed the results of all models on the test set. The main evaluation indicators were sensitivity, specificity, F1-score, area under the receiver operating characteristics curve (AUC), kappa coefficient, and accuracy. The accuracy and kappa coefficient of the proposed diagnostic model in several experiments were averaged at 0.9789 and 0.9681, respectively. The sensitivity, specificity, F1-score, and AUC were, respectively, 0.9723, 0.9836, 0.9688, and 0.9869 for diagnosing pterygium to be observed, and, respectively, 0.9210, 0.9905, 0.9292, and 0.9776 for diagnosing pterygium requiring surgery. The proposed method has high clinical reference value for recognizing these four types of ocular surface images.

KEYWORDS

subconjunctival hemorrhage, pterygium, visual recognition, deep learning, computer aided diagnosis

1 Introduction

Pterygium is a common ocular surface disease caused by overgrowth of fibro vascularity in the subconjunctival tissue, resulting in invasion of the inner eyelid and outer cornea (1). It is most prevalent in areas with high ultraviolet light; in some areas, 9.5% of the pterygium patient population is associated with prolonged exposure to high ultraviolet

light (2). Clinically, pterygium can be categorized into active and fixed stages. In the fixed stage, the pterygium invades the cornea to a lesser extent, with thin fibro vascular tissue and a smooth, transparent cornea. In the active stage, the pterygium severely invades the cornea, resulting in a cloudy cornea, which if not properly controlled can obscure the pupil and cause irritation and astigmatism, with a more serious effect on vision and limited eye movement accompanied by pain (3). In the medical field, the width of the pterygium (WP) invading the cornea is commonly used as an indicator of whether to operate; the patient is in the stage to be observed when the width of invasion is less than 3 mm, and in the stage to be operated when the width of invasion is greater than 3 mm (4). Subconjunctival hemorrhage is also a common ocular surface disease characterized by painless, acute, obvious red, swollen hemorrhages in the absence of secretions under the conjunctiva, which may evolve from punctate to massive hemorrhages, rendering the underlying sclera invisible (5, 6). Subconjunctival hemorrhage can be defined histologically as bleeding between the conjunctiva and the outer layer of the sclera, and the blood component will be found in the lamina propria of the conjunctiva when the blood vessels under the conjunctiva rupture (7). In contrast to pterygium, subconjunctival hemorrhage is not vision-threatening and is predominantly found in hypertensive groups over 50 years of age (8). Pterygium and subconjunctival hemorrhage often cause uneasiness and anxiety in patients; however, most cases do not require much medical management in the early stages.

Traditional screening methods for ocular surface diseases rely primarily on capturing anterior segment images using a slit lamp for patient sampling, followed by clinical diagnosis by experienced ophthalmologists for early screening and analysis. However, a lack of ophthalmologists in remote areas with poor healthcare resources means that screening for ocular surface diseases still faces great difficulties.

Recently, the increasing application of artificial intelligence in ophthalmology has led to the rapid development of research on intelligent ophthalmic diagnosis. Many researchers have used deep learning algorithms to detect common fundus diseases on fundus images (9–13). In addition, researchers have used deep learning for the diagnosis of ocular surface diseases. In 2018, Zhang et al. implemented an interpretable and scalable deep learning automated diagnostic architecture for four ophthalmic diseases, including subconjunctival hemorrhage and pterygium (14). In 2020, a team from the U.S. improved VggNet16 and applied transfer learning to apply it to screening for pterygium (15). In 2022, Wan et al. improved the U-Net++ segmentation algorithm and proposed a system to diagnose and measure the progression of pterygium pathology (16). To provide high-quality diagnostic services for ocular surface diseases, we designed an automatic diagnostic model for ocular surface diseases using deep learning techniques. The proposed model simultaneously accomplishes the detection of multiple diseases from ocular surface images and achieves fast recognition with high accuracy. This capability is crucial for early screening of ocular surface diseases in remote areas where access to professional medical personnel and equipment is limited.

2 Dataset description

The dataset used in this study was provided by the Affiliated Eye Hospital of Nanjing Medical University, and contains color images of the ocular surface with good image quality captured by a professional ophthalmologist. To prevent the leakage of patients' personal information, the images do not contain patients' personal information, including but not limited to age, sex, and name.

In this study, 2855 ocular surface images were collected from patients of different age groups and sexes, including 1312 normal ocular surfaces, 251 ocular surface hemorrhages, 909 pterygiums to be observed, and 383 pterygiums requiring surgery. Examples of the four types of ocular surface images are shown in **Figure 1**. The camera used was a Canon DSLR, model Canon EOS 600D, with diffuse illumination from a slit lamp and an image resolution of 5184 × 3456. The quality of the images was verified by a professional ophthalmologist. We followed the guidelines proposed by Yang et al. (17).

3 Materials and methods

Currently, image classification algorithms based on deep learning are primarily composed of convolutional neural networks or visual transformer modules. Convolutional neural networks were first proposed by Lecun et al. (18), and several representative modeling algorithms have subsequently emerged. Among them, the residual network architecture proposed by He et al. (19) is an important milestone in the field of computer vision that solves the problem of network training difficulty owing to gradient vanishing and gradient explosion in convolutional neural networks. The vision transformer (20), proposed by researchers at Google Brain, is an image classification algorithm based on the transformer model that allows images to be viewed as sequences and uses a self-attention mechanism to extract features. Traditional convolutional neural networks perform excellently in the field of image processing; however, the convolutional kernel limits its receptive field and may ignore global information in the image. The transformer can consider all the pixels in the image simultaneously, thus capturing global information more reliably. We adopted the conformer model as the main body, which combines the convolutional neural network and transformer models by parallel fusion to fuse local and global features effectively (21). In addition, we propose a structural feature extraction block composed of a patch merging layer and a FReLU layer (PFM block), which improves the conformer to further differentiate between pterygium to be observed and pterygium to be operated. We propose this dual-branch network reinforced by PFM block (DBPF-Net).

3.1 Network structure

In computer vision, local and global features are an important pair of concepts that have been extensively studied in the

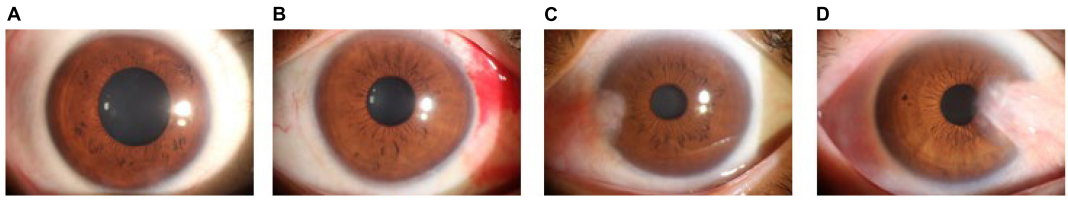

FIGURE 1
Examples of ocular surface samples. **(A)** Normal ocular surface; **(B)** subconjunctival hemorrhages; **(C)** pterygium to be observed; and **(D)** pterygium requiring surgery.

long history of visual feature description. Local features characterize local regions of images and are represented by compact vectors in local image domains (22); global features include contour representations, shape descriptors, and object representations at long distances (23). Local features provide information about the details in the image, whereas global features provide information about the image as a whole. Using both local and global features helps improve the model performance. In deep learning, a convolutional neural network collects local features in a hierarchical manner through convolutional operations and retains local cues as feature maps, and a vision transformer aggregates global representations in compressed plots by cascading self-attention modules. The conformer efficiently fuses local and global features through concatenation and bridging.

The overall architecture of the conformer is shown as the backbone in Figure 2A, which is mainly composed of ConvTrans blocks. The stem block, consisting of a 7 × 7 convolutional layer with stride 2 followed by a 3 × 3 max pooling layer with stride 2, is used to extract the initial local features, which are then fed into the two branches. The internal structure of the ConvTrans block is shown in Figure 2B. This consists of the convolutional neural network (CNN) branch on the left and the transformer branch on the right, with the feature interactions between them accomplished by the upsampling and downsampling branches. The local features extracted from the CNN branch are transformed into the form of patch embeddings for the transformer branch through the downsampling branch; the global features extracted from the transformer branch are transformed into the form of feature maps for the CNN branch through the upsampling branch. The downsampling operation of the downsampling branch is performed through the max pooling layer, whereas the upsampling operation of the upsampling branch is performed through bilinear interpolation. In every ConvTrans block except for the first one, there are upsampling and downsampling branches for feature exchange.

The core of the transformer branch is multi-head self-attention (24), as shown in Figure 2C. Multi-head self-attention is a technique that introduces multiple heads into the self-attention mechanism, which is used to process sequential data and assign a weight to each element in the sequence to better capture the relationships between them. In the traditional self-attention mechanism, only one head is used to compute the attention weights. In contrast, the multi-head self-attention mechanism introduces multiple heads, each of which has its own weight calculation system to learn different semantic information, thus improving the expressive power of the model.

The input sequence X is first subjected to three different linear transformations to obtain the representations of Q (query), K (key), and V (value). Subsequently, Q, K, and V are divided into multiple heads, denoted Q_i, K_i, V_i. Then, for each head, the attention weights are computed separately by computing the dot product of Q_i and K_i and then performing softmax normalization. Next, a weighted summation is performed on V_i using the attention weights to obtain the attention output for each head, which is concatenated and linearly transformed to obtain the final multi-head self-attention output. The calculation procedure is shown in Eqs 1–4, where $W_Q \in R^{d_{model} \times d_{model}}$, $W_K \in R^{d_{model} \times d_{model}}$, $W_V \in R^{d_{model} \times d_{model}}$, $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_v}$, *and* $W^0 \in R^{hd_v \times d_{model}}$

$$Q = XW_Q; K = XW_K; V = X \tag{1}$$

$$MultiHead\,(Q, K, V) = Concat\left(head_1, \ldots, head_h\right) W^o \tag{2}$$

$$head_i = Attention\,(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

$$Attention\,(Q_i, K_i, V_i) = softmax(\frac{Q_i K_i^T}{\sqrt{d_k}})\, V_i \tag{4}$$

The PFM block comprises two novel layer structures: the patch merging (25) and flexible rectified linear unit (FReLU) nonlinear activation layers (26). The operating principles for these are shown in Figures 2D, E, respectively. Patch merging acts as downsampling for resolution reduction, which is a similar operation to pooling; however, unlike pooling, patch merging does not lose feature information. FReLU is a context-conditional activation function that relies on the local information of the center pixel to obtain pixel-level constructive capabilities. It operates on a localization of the feature map through a parameter-learning convolution kernel, compares it with the center pixel point, and takes the maximum value. This provides each pixel with an option to view the contextual information, which enables spatial structure extraction of the feature map. Formally, the joint action of multiple FReLUs can provide a wider selection of information for each pixel, which helps focus on the structural features of the pterygium and differentiate effectively between the two subclasses of pterygium. The structure of the PFM block is shown in Figure 2F, where the downsampling operation is performed

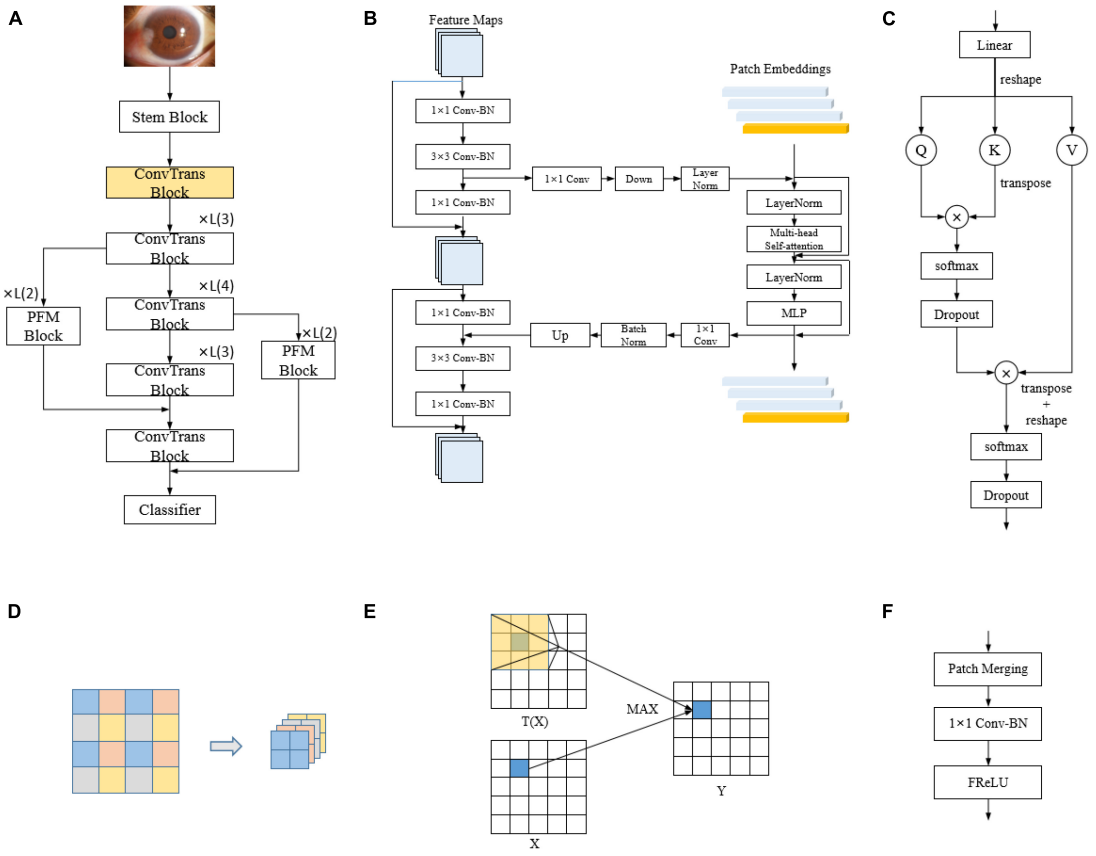

FIGURE 2
Model structure. **(A)** DBPF-net; **(B)** ConvTrans block; **(C)** multi-head self-attention; **(D)** schematic of patch merging; **(E)** schematic of the FReLU activation function, which can be expressed as $Y = MAX(X, T(x))$; and **(F)** PFM block.

by patch merging, followed by a 1×1 convolutional layer to change the number of channels. Finally, non-linear activation is performed by the FReLU activation function, which is in line with the design concept of traditional convolutional neural networks.

We designed the PFM block to further differentiate between the similar cases of pterygium to be observed and pterygium requiring surgery. The relationship between the spatial structure of the pterygium and the cornea is particularly important in the medical field, where the depth of pterygium invasion into the cornea is usually used as a discriminator. Relying on the basic lossless downsampling of patch merging and the spatial structure feature extraction of the FReLU activation function, the bottom-layer feature map output from the first 3×3 convolutional layer in the ConvTrans block is passed to the top-layer feature map through the two processes of the PFM block, which causes the network to focus on extracting the spatial structure features. The overall architecture of DBPF-Net is shown in Figure 2A.

3.2 Data division and pre-processing

The original dataset used in this study consisted of 2855 ocular surface images. Considering the reliability of the model's performance on the validation set and its generalization on the test set, we divided the dataset into training, validation, and test sets in a ratio of 7:1:2. In the original dataset, there is generally only one image for an eye, and images from the same eye only appear inside one dataset (i.e., training set, validation set, test set). The number of samples for each category in each subset is shown in Table 1. Owing to the different difficulties in obtaining samples for each image category, the number of samples for the four categories is not balanced, which may lead to the model focusing excessively on categories with a large number of training samples and lack of attention to categories with a small number of samples. Therefore, we used enhancement methods to reduce the impact of data imbalance, including image bilinear interpolation stretching, random horizontal flipping, random small-angle rotation, central region cropping, and normalization. The purpose of these steps was to minimize the influence of the upper and lower eyelids during training while preserving the conjunctiva. These augmentation techniques do not eliminate the pathological regions present in the original images, such as hemorrhages on the conjunctiva and pterygium invading the cornea.

3.3 Model training

We used the Adam optimization algorithm (27) during model training, with a weight decay of 0.0005. The training batch size was 4, the total number of training iterations was 90, and the initial value of the learning rate was 0.0001. Two cross-entropy loss

TABLE 1 Data division.

Category	Train	Validation	Test	Total
C0	919	131	262	1,312
C1	176	25	50	251
C2	637	91	181	909
C3	269	38	76	383
Total	2001	285	569	2,855

C0, C1, C2, and C3, respectively, represent normal ocular surface, subconjunctival hemorrhages, pterygium to be observed, and pterygium requiring surgery.

functions were used to supervise the classifiers of the two branches separately, and the importance of the loss function was identical for both. The learning rate was adjusted dynamically using the cosine annealing strategy (28), which helps prevent the model from falling into local optimal solutions during the training process, as well as to avoid the impact of sudden learning rate changes on the training process. In addition, we selected VGG16 (29), ResNet50 (19), EfficientNetB7 (30), and conformer models to compare the classification results, all of which used ImageNet pretrained model parameters as initial conditions.

The central processor used in our experiments was a 3.6 GHz Intel i7-7700, and the graphics processor was an NVIDIA RTX 2080Ti with 11 GB of RAM. The operating system was Windows 10, the programming language was Python 3.6, and the deep learning framework was Pytorch 1.7.

3.4 Model evaluation indicators

This study is a multi-categorization task, and we evaluate the effectiveness of the model from two perspectives. The first approach involves evaluating the overall performance of multi-class classification using the kappa coefficient, which demonstrates consistent agreement. The calculation of the kappa coefficient is based on the confusion matrix, and its value typically ranges from 0 to 1. A higher kappa coefficient indicates a higher level of agreement between the model's evaluation and the diagnostic assessment by experts. The formula for the kappa coefficient is as follows:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{5}$$

$$p_e = = \frac{a_1 \times b_1 + a_2 \times b_2 + \ldots + a_c \times b_c}{n \times n} \tag{6}$$

where p_o is the sum of all correctly classified samples divided by the total number of samples, a_i is the number of true samples in category i, and b_i is the number of predicted samples in category i.

Another approach is to convert a multi-classified problem into multiple independent binary classification problems. For example, to identify the normal ocular surface, the normal ocular surface is labeled as a positive sample, whereas the three categories of subconjunctival hemorrhage, pterygium to be observed, and pterygium requiring surgery are labeled as negative samples. To calculate the evaluation indicators for the binary classification problem, the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) samples were first obtained from the confusion matrix, and then the accuracy (ACC), sensitivity (SE), specificity (SP), and F1-score (F1) were calculated. Accuracy indicates the proportion of correctly diagnosed samples to the total number of samples; sensitivity indicates the proportion of samples predicted to be positive and actually positive to the proportion of all actual positive samples; specificity indicates the proportion of samples predicted to be negative and actually negative to the proportion of all actual negative samples; and F1-score is defined as the harmonic average of accuracy and sensitivity, which is meaningful for datasets with unbalanced samples.

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

$$SE = \frac{TP}{TP + FN} \tag{8}$$

$$SP = \frac{TN}{TN + FP} \tag{9}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{10}$$

Receiver operating characteristics (ROC) curves are commonly used to analyze the classification performance of different models, owing to their visualization features. The area under the ROC curve (AUC) was used to evaluate the classification accuracy. Generally speaking, an AUC value of 0.50–0.70 is regarded as a low diagnostic value, 0.70–0.85 is regarded as a general diagnostic value, and 0.85 and above is regarded as a good diagnostic value.

4 Results

In this study, 569 ocular surface images were randomly selected as a test set containing 262 images of a normal ocular surface, 50 images of subconjunctival hemorrhage, 181 images of pterygium to be observed, and 76 images of pterygium requiring surgery. The model with the best accuracy on the validation set was considered the optimal model for evaluating the performance of the models on the test set.

The best diagnostic results of each model on the test set are presented in Figure 3, in the form of confusion matrices.

The purpose of this study was to correctly diagnose four categories of ocular surface images: normal ocular surface, subconjunctival hemorrhage, pterygium to be observed and pterygium requiring surgery. To demonstrate the performance of the models clearly, we quantified their evaluation indicators, with results as listed in Table 2. These evaluation indicators are calculated according to Eqs 5–10.

In summary, the DBPF-Net model achieved high sensitivity and specificity values, indicating that it performs well in differentiating between positive and negative samples, which is valuable for clinical diagnoses that require accurate identification and differentiation of different disease categories. In addition, its high F1-score and kappa coefficient indicate that the model has excellent classification performance when the data are unbalanced, and high consistency with the evaluation of the expert diagnostic group. The ROC curves for each model with the best accuracy are shown in Figure 4. Moreover, we used Grad-CAM (31) to analyze the region of interest of the models for the ocular surface images, as shown in Figure 5.

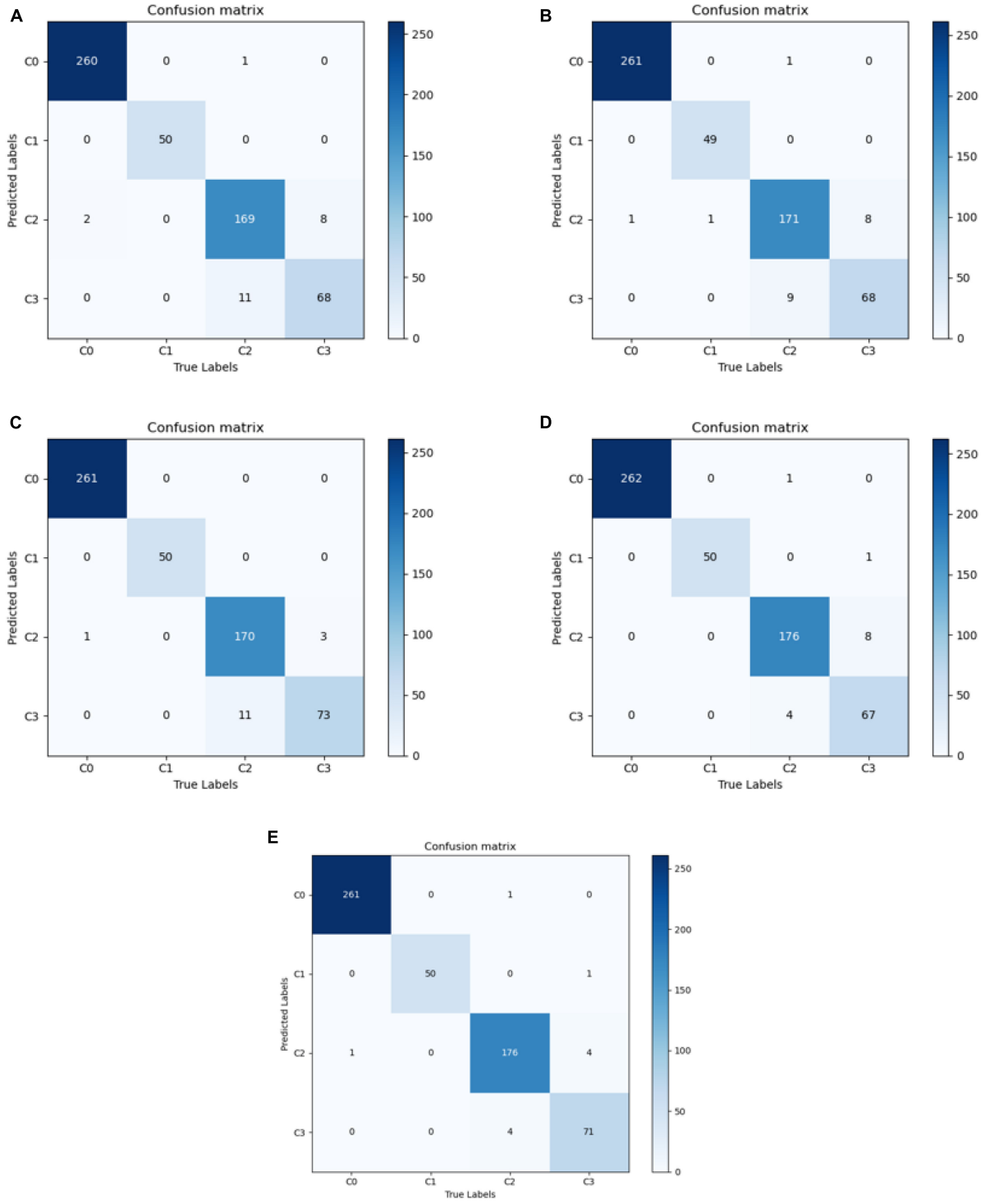

FIGURE 3
Confusion matrix for each model. **(A)** VGG16; **(B)** ResNet50; **(C)** EfficientNetB7; **(D)** Conformer; and **(E)** DBPF-Net.

5 Discussion

Ocular surface diseases have received worldwide attention as a common public health problem. The variety and complexity of these diseases are important factors that should not be ignored in their diagnosis. Therefore, diagnosis and treatment require doctors with rich experience and professional knowledge to be able to determine the condition accurately and take appropriate treatment measures. Currently, the lack of specialized ophthalmologists in areas with a high prevalence of ocular surface diseases leaves many patients without timely diagnosis and treatment. Therefore, it is important to develop an automatic diagnostic model for initial screening and diagnosis.

TABLE 2 Evaluation indicators for each model in each category.

Models	Evaluation indicators	C0	C1	C2	C3
VGG16	Sensitivity	0.9987 ± 0.0017	0.96 ± 0.0163	0.9152 ± 0.0213	0.8947
	Specificity	0.9945 ± 0.003	1.0	0.9759 ± 0.0012	0.9702 ± 0.0084
	F1-score	0.9962 ± 0.0015	0.9795 ± 0.0084	0.9306 ± 0.0103	0.8577 ± 0.0229
	AUC	1.0	1.0	0.9944 ± 0.0008	0.991 ± 0.0018
	Kappa	0.9319 ± 0.0105			
	Accuracy	0.9548 ± 0.007			
ResNet50	Sensitivity	0.9949 ± 0.0017	1.0	0.9281 ± 0.0078	0.8859 ± 0.0223
	Specificity	0.9967	1.0	0.9742 ± 0.0042	0.9756 ± 0.0028
	F1-score	0.9955 ± 0.0008	1.0	0.9359 ± 0.0021	0.8668 ± 0.0090
	AUC	1.0	1.0	0.9960 ± 0.0001	0.9931 ± 0.0012
	Kappa	0.9389 ± 0.0022			
	Accuracy	0.9595 ± 0.0014			
EfficientNetB7	Sensitivity	0.9962	1.0	0.9355 ± 0.0051	0.9517 ± 0.0123
	Specificity	1.0	1.0	0.9879 ± 0.0024	0.9763 ± 0.0019
	F1-score	0.9981	1.0	0.9539 ± 0.0026	0.9041 ± 0.006
	AUC	0.9999	1.0	0.9909 ± 0.0006	0.9860 ± 0.0009
	Kappa	0.9568 ± 0.0024			
	Accuracy	0.9712 ± 0.0016			
Conformer	Sensitivity	0.9962 ± 0.0031	0.9933 ± 0.0094	0.9668 ± 0.0078	0.8903 ± 0.0223
	Specificity	0.9967 ± 0.0026	0.9987 ± 0.0008	0.9776 ± 0.0044	0.9892 ± 0.0038
	F1-score	0.9962 ± 0.0015	0.9901	0.9597 ± 0.0033	0.9082 ± 0.0031
	AUC	0.9999	1.0	0.9958 ± 0.0002	0.9934 ± 0.0006
	Kappa	0.9583 ± 0.0033			
	Accuracy	0.9723 ± 0.0021			
DBPF-Net	Sensitivity	0.9962 ± 0.0031	1.0	0.9723 ± 0.009	0.9210 ± 0.0186
	Specificity	0.9989 ± 0.001	0.9987 ± .0008	0.9836 ± 0.0048	0.9905 ± 0.0034
	F1-score	0.9974 ± 0.0017	0.9934 ± 0.0046	0.9688 ± 0.0025	0.9292 ± 0.0079
	AUC	0.9989 ± 0.0006	1.0	0.9869 ± 0.0109	0.9776 ± 0.0155
	Kappa	0.9681 ± 0.0022			
	Accuracy	0.9789 ± 0.0014			

C0, C1, C2, and C3, respectively, represent normal ocular surface, subconjunctival hemorrhages, pterygium to be observed, and pterygium requiring surgery. The variable was expressed as the mean ± standard deviation.

The application of artificial intelligence to the field of medical image processing has been based on traditional convolutional neural networks and has achieved remarkable research results in recent years. The emergence of vision transformers has confirmed the advantages of global features in image recognition, and a variety of deformation models have been derived (25, 32, 33). The DBPF-Net model proposed in this study selects the conformer as the backbone of the four-way classification model for ocular surface diseases. Compared with other models, the conformer's ability to extract and fuse global and local features gives it better feature extraction capability. In addition, we propose a PFM block for enhancing the conformer's extraction of spatial structural features to differentiate further between the two pterygium categories.

Several research groups have investigated the classification and diagnosis of ocular surface diseases. Elsawy et al. (34) employed an improved VGG19 model to classify corneal diseases automatically, achieving an overall F1-score in excess of 86%. Zhang et al. (14) implemented an automated diagnostic architecture with deep learning interpretability and scalability, achieving over 95% accuracy for pterygium. Xu et al. (35) Proposed a computer-aided pterygium diagnosis system based on EfficientNetB6 with transfer learning, achieving a sensitivity of 90.06% for pterygium to be observed and 92.73% for pterygium requiring surgery. Huang et al. (36) developed a deep learning system for pterygium grading, using a classification algorithm to categorize pterygiums from non-pterygiums, and then a segmentation algorithm to segment pterygiums for grading, achieving sensitivities ranging from 80 to 91.67%. These studies exclusively employed CNN models without specific disease-targeted feature modules. Our study focused on the practical situation of whether or not patients with pterygium need

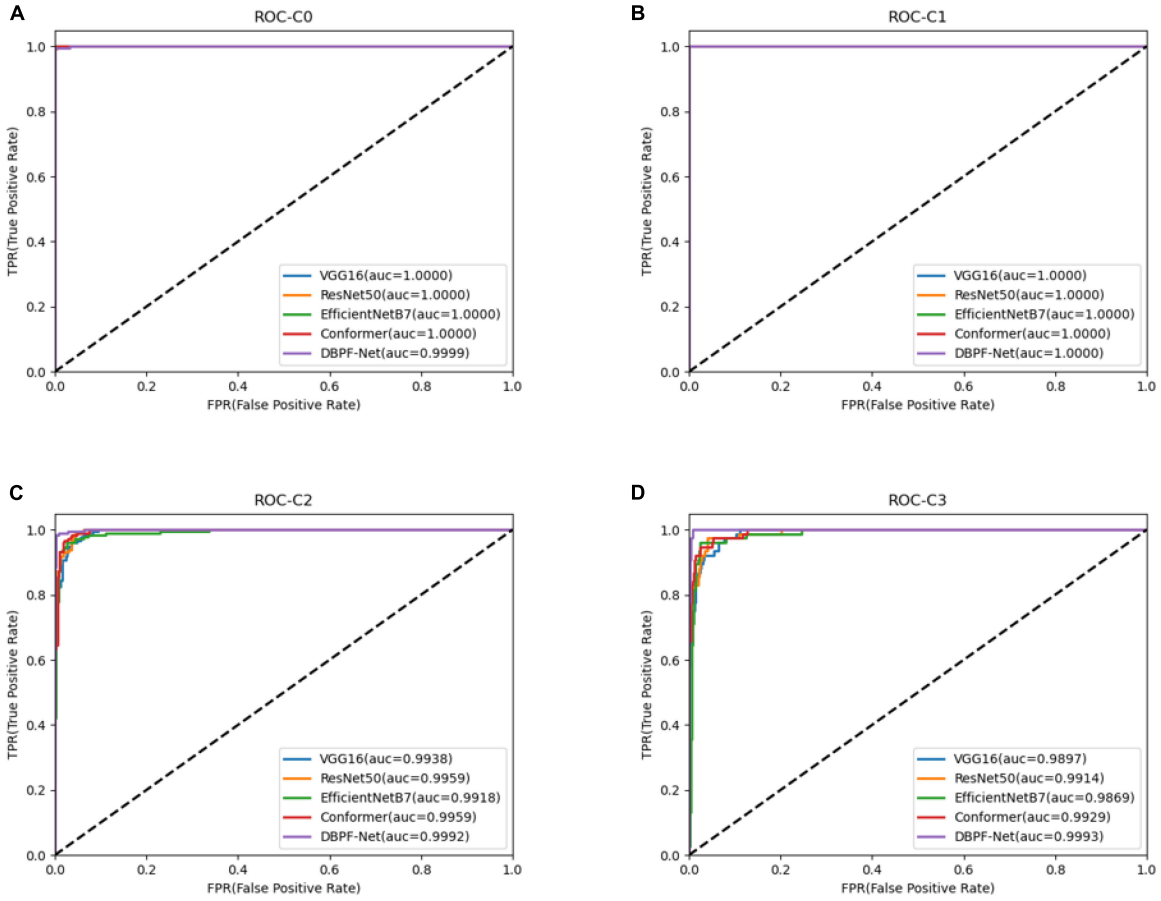

FIGURE 4
Receiver operating characteristics curves for each model. **(A)** Normal ocular surface; **(B)** subconjunctival hemorrhage; **(C)** pterygium to be observed; and **(D)** pterygium requiring surgery.

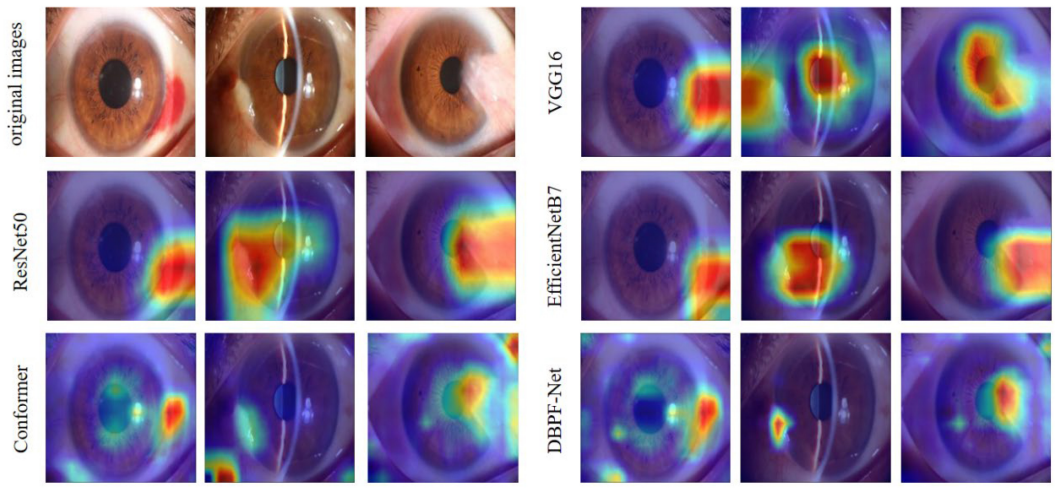

FIGURE 5
Heat maps of the models for subconjunctival hemorrhage, pterygium to be observed, and pterygium requiring surgery.

surgery. The proposed DBPF-Net achieved an accuracy of 97.89% for the four categories, demonstrating promising results. In our experiments, we compared it with three other representative CNN models and the original conformer model.

As shown in Figure 3 and Table 2, the overall evaluation indicators of DBPF-Net were generally higher than those of the other models. Among the test results for all models, evaluation indicators for the normal ocular surface and subconjunctival

hemorrhage categories reached a high level, mainly because of the sufficient number of samples in the normal ocular surface category and the significant characteristics of the subconjunctival hemorrhage category. The test results for the categories of pterygium to be observed and pterygium requiring surgery demonstrate that the conformer model exhibits superior discriminative ability compared to VGG16 and ResNet50, While in comparison with EfficientNetB7, each of the two was dominant. Compared to the conformer model, DBPF-Net showed an improvement of 0.55% in sensitivity, 0.6% in specificity, and 0.91% in F1-score for the category of pterygium to be observed. For the category of pterygium requiring surgery, DBPF-Net achieved an increase of 3.07% in sensitivity and 2.1% in F1-score. Overall, DBPF-Net showed further improvement in pterygium diagnosis compared with Conformer. Although the proposed method has a slightly lower AUC than Conformer, the proposed method outperforms in terms of the F1 Score. The heat map shown in Figure 5 demonstrates that DBPF-Net focuses on the area of hemorrhage in the category of subconjunctival hemorrhage, the area of pterygium tipping into the cornea in the category of pterygium to be observed, and the area of pterygium approaching the center of the cornea in the category of pterygium requiring surgery. The heat maps generated by VGG16, ResNet50, and EfficientNetB7 indicate that their attention on the lesion area is not adequately concentrated, as well as on the pupil area. In comparison, Conformer exhibits a similar focus area to DBPF-Net, the latter is more focused.

Our study has some limitations. First, the dataset used in this study has a limited number of samples and an uneven number of samples per category, which leads to poorer generalization and precision for categories with fewer samples. Second, the hardware configuration of the experimental platform in this study was ordinary, and the model performance was limited by the amount of GPU RAM. In the future we will continue to collect datasets, improve the model to increase its accuracy, and consider a method of semantic segmentation of images to assist in classification.

6 Conclusion

In this paper, we propose DBPF-Net, a model that achieves high classification performance on four categories of ocular surface images: normal ocular surface, subconjunctival hemorrhage, pterygium to be observed, and pterygium requiring surgery. This model is hopefully to achieve initial screening for ocular surface diseases in remote areas where access to professional medical personnel and equipment is limited. In addition, we hope to help reduce the workload of medical personnel in primary care facilities.

Data availability statement

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

Author contributions

CW: Data curation, Methodology, Software, Writing – original draft. YM: Data curation, Methodology, Software, Writing – original draft. WX: Data curation, Software, Validation, Writing – review and editing. ZZ: Data curation, Validation, Writing – review and editing. JW: Data curation, Formal analysis, Methodology, Project administration, Supervision, Writing – review and editing. WY: Data curation, Formal analysis, Methodology, Project administration, Supervision, Writing – review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by Shenzhen Fund for Guangdong Provincial High-level Clinical Key Specialties (SZGSP014), Sanming Project of Medicine in Shenzhen (SZSM202311012), Shenzhen Science and Technology Program (JCYJ20220530153604010), and Shenzhen Fundamental Research Program (JCYJ20220818103207015).

Conflict of interest

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

Publisher's note

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

References

1. Chen B, Fang X, Wu M, Zhu S, Zheng B, Liu B, et al. Artificial intelligence assisted pterygium diagnosis: current status and perspectives. *Int J Ophthalmol.* (2023) 16:1386–94. doi: 10.18240/ijo.2023.09.04

2. Asokan R, Venkatasubbu R, Velumuri L, Lingam V, George R. Prevalence and associated factors for pterygium and pinguecula in a South Indian population. *Ophthalmic Physiol Opt.* (2012) 32:39–44. doi: 10.1111/j.1475-1313.2011.00882.x

3. Tarlan B, Kiratli H. Subconjunctival hemorrhage: risk factors and potential indicators. *Clin Ophthalmol.* (2013) 7:1163–70. doi: 10.2147/OPTH.S35062

4. Tan D, Chee S, Dear K, Lim A. Effect of pterygium morphology on pterygium recurrence in a controlled trial comparing conjunctival autografting with bare sclera excision. *Arch Ophthalmol.* (1997) 115:1235–40. doi: 10.1001/archopht.1997.01100160405001

5. Leibowitz H. The red eye. *N Engl J Med.* (2000) 343:345–51. doi: 10.1056/NEJM200008033430507

6. Mimura T, Usui T, Yamagami S, Funatsu H, Noma H, Honda N, et al. Recent causes of subconjunctival hemorrhage. *Ophthalmologica.* (2010) 224:133–7. doi: 10.1159/000236038

7. Mimura T, Yamagami S, Usui T, Funatsu H, Noma H, Honda N, et al. Location and extent of subconjunctival hemorrhage. *Ophthalmologica.* (2010) 224:90–5. doi: 10.1159/000235798

8. Fukuyama J, Hayasaka S, Yamada K, Setogawa T. Causes of subconjunctival hemorrhage. *Ophthalmologica.* (1990) 200:63–7. doi: 10.1159/000310079

9. Gulshan V, Peng L, Coram M, Stumpe M, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

10. Li Z, He Y, Keel S, Meng W, Chang R, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology.* (2018) 125:1199–206. doi: 10.1016/j.ophtha.2018.01.023

11. Yim J, Chopra R, Spitz T, Winkens J, Obika A, Kelly C, et al. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat Med.* (2020) 26:892–9. doi: 10.1038/s41591-020-0867-7

12. Bhati A, Gour N, Khanna P, Ojha A. Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. *Comput Biol Med.* (2023) 153:106519. doi: 10.1016/j.compbiomed.2022.106519

13. Zhu S, Zhan H, Wu M, Zheng B, Liu B, Zhang S, et al. Research on classification method of high myopic maculopathy based on retinal fundus images and optimized ALFA-Mix active learning algorithm. *Int J Ophthalmol.* (2023) 16:995–1004. doi: 10.18240/ijo.2023.07.01

14. Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, et al. An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study. *J Med Internet Res.* (2018) 20:e11144. doi: 10.2196/11144

15. Zamani NS, Zaki WM, Huddin AB, Hussain A, Mutalib H, Ali A. Automated pterygium detection using deep neural network. *IEEE Access* (2020) 8:191659–72.

16. Wan C, Shao Y, Wang C, Jing J, Yang WA. Novel system for measuring pterygium's progress using deep learning. *Front Med.* (2022) 9:819971. doi: 10.3389/fmed.2022.819971

17. Yang W, Shao Y, Xu Y. Guidelines on clinical research evaluation of artificial intelligence in ophthalmology (2023). *Int J Ophthalmol.* (2023) 16:1361–72. doi: 10.18240/ijo.2023.09.02

18. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* (1998) 86:2278–324.

19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv* [preprint]. (2016) doi: 10.48550/arXiv.1512.03385

20. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv* [preprint]. (2010) doi: 10.48550/arXiv.2010.11929

21. Peng Z, Guo Z, Huang W, Wang Y, Xie L, Jiao J, et al. Conformer: local features coupling global representations for recognition and detection. *IEEE Trans Pattern Anal Mach Intell.* (2023) 45:9454–68. doi: 10.1109/TPAMI.2023.3243048

22. Niethammer M, Betelu S, Sapiro G, Tannenbaum A, Giblin P. Area-Based Medial Axis of Planar Curves. *Int J Comput Vis.* (2004) 60:203–24. doi: 10.1023/B:VISI.0000036835.28674.d0

23. Lisin DA, Mattar MA, Blaschko MB, Learned-Miller EG, Benfield MC. Combining local and global image features for object class recognition[C]. 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-Workshops. *IEEE* (2005) 2005:47–47.

24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *arXiv* [preprint]. (2017) doi: 10.48550/arXiv.1706.03762

25. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv* [preprint]. (2021) doi: 10.48550/arXiv.2103.14030

26. Ma N, Zhang X, Sun J. Funnel activation for visual recognition[C]. Computer Vision–ECCV 2020. *16th European Conference, Proceedings, Part XI 16.* Berlin (2020).

27. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv* [preprint]. (2014) doi: 10.48550/arXiv.1412.6980

28. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* [preprint]. (2016) doi: 10.48550/arXiv.1608.03983

29. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* [preprint]. (2014) doi: 10.48550/arXiv.1409.1556

30. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *Int Conf Mach Learn.* (2019) 2019:6105–14.

31. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *arXiv* [preprint]. (2017) doi: 10.48550/arXiv.1610.02391

32. Ding M, Xiao B, Codella N, Luo P, Wang J, Yuan L. Davit: Dual attention vision transformers. *arXiv* [preprint]. (2022) doi: 10.48550/arXiv.2204.03645

33. Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A, et al. Maxvit: Multi-axis vision transformer. European conference on computer vision. *arXiv* [preprint]. (2022) doi: 10.48550/arXiv.2204.01697

34. Elsawy A, Eleiwa T, Chase C, Ozcan E, Tolba M, Feuer W, et al. Multidisease deep learning neural network for the diagnosis of corneal diseases. *Am J Ophthalmol.* (2021) 226:252–61. doi: 10.1016/j.ajo.2021.01.018

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

36. Hung KH, Lin C, Roan J, Kuo CF, Hsiao CH, Tan HY, et al. Application of a deep learning system in pterygium grading and further prediction of recurrence with slit lamp photographs. *Diagnostics.* (2022) 12:888. doi: 10.3390/diagnostics12040888

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 February 2024
DOI 10.3389/fmed.2024.1326004

Check for updates

OPEN ACCESS

EDITED BY
Qiang Chen,
Nanjing University of Science and Technology, China

REVIEWED BY
Mahmud Omar,
Ziv Medical Center, Israel
Fei Shi,
Soochow University, China

*CORRESPONDENCE
Juan Ye
✉ yejuan@zju.edu.cn

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

RECEIVED 23 November 2023
ACCEPTED 19 January 2024
PUBLISHED 06 February 2024

CITATION
Li H, Cao J, You K, Zhang Y and Ye J (2024) Artificial intelligence-assisted management of retinal detachment from ultra-widefield fundus images based on weakly-supervised approach.
Front. Med. 11:1326004.
doi: 10.3389/fmed.2024.1326004

COPYRIGHT
© 2024 Li, Cao, You, Zhang and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence-assisted management of retinal detachment from ultra-widefield fundus images based on weakly-supervised approach

Huimin Li[1†], Jing Cao[1†], Kun You[2], Yuehua Zhang[2] and Juan Ye[1*]

[1]Eye Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, [2]Zhejiang Feitu Medical Imaging Co., Ltd, Hangzhou, Zhejiang, China

Background: Retinal detachment (RD) is a common sight-threatening condition in the emergency department. Early postural intervention based on detachment regions can improve visual prognosis.

Methods: We developed a weakly supervised model with 24,208 ultra-widefield fundus images to localize and coarsely outline the anatomical RD regions. The customized preoperative postural guidance was generated for patients accordingly. The localization performance was then compared with the baseline model and an ophthalmologist according to the reference standard established by the retina experts.

Results: In the 48-partition lesion detection, our proposed model reached an 86.42% (95% confidence interval (CI): 85.81–87.01%) precision and an 83.27% (95%CI: 82.62–83.90%) recall with an average precision (PA) of 0.9132. In contrast, the baseline model achieved a 92.67% (95%CI: 92.11–93.19%) precision and limited recall of 68.07% (95%CI: 67.25–68.88%). Our holistic lesion localization performance was comparable to the ophthalmologist's 89.16% (95%CI: 88.75–89.55%) precision and 83.38% (95%CI: 82.91–83.84%) recall. As to the performance of four-zone anatomical localization, compared with the ground truth, the un-weighted Cohen's κ coefficients were 0.710(95%CI: 0.659–0.761) and 0.753(95%CI: 0.702–0.804) for the weakly-supervised model and the general ophthalmologist, respectively.

Conclusion: The proposed weakly-supervised deep learning model showed outstanding performance comparable to that of the general ophthalmologist in localizing and outlining the RD regions. Hopefully, it would greatly facilitate managing RD patients, especially for medical referral and patient education.

KEYWORDS

weakly supervised, deep learning, localization, retinal detachment, ultra-widefield fundus images

1 Introduction

Retinal detachment (RD) is a sight-threatening condition that occurs when the neurosensory retina is separated from the retinal pigment epithelium (1). Several population-based epidemiological studies of RD find an annual incidence of around 1 in 10,000 (2). It has been estimated that the lifetime risk of RD is about 0.1% (3, 4). However, early intervention facilitates the prevention of disease progression and improves prognosis. Clinically, scleral buckle, vitrectomy, and pneumatic retinopexy are the most common surgical approaches to repairing RD (5–7). Before the surgery, patients should be instructed to lie in the appropriate position to minimize the detachment extending and improve visual outcomes (6, 8, 9). Postural guidance is consistent with the localization of the lesion throughout the management. However, corresponding patient education is not often adequate in busy clinical situations which may lead to poor patient compliance (10). Therefore, an efficient and reliable method for localizing and estimating the detached retinal regions is fundamental for detailed postural instruction and medical referrals, especially in remote areas with insufficient fundus specialists.

In recent years, artificial intelligence (AI) models for RD detection based on color fundus photography (CFP) and optical coherence tomography (OCT) have been gradually established (11–14). However, the emergence of the ultra-widefield fundus (UWF) imaging system promotes the intelligent diagnosis of fundus diseases to a new height. A panoramic image of the retina with 200° views allows for detailed rendering of the peripheral retina, which compensates for the deficiency of traditional fundus images (15). Ohsugi et al. (16) made a pioneering attempt to diagnose rhegmatogenous RD with a small sample of UWF images based on deep learning algorithms. Later, Li et al. (17) proposed a cascaded deep learning system using UWF images for various RD detection and macula status discerning. Despite promising advancements, their work mainly focused on the presence or absence of the target disease. However, the concrete localization of the RD lesions, a crucial need for therapeutic decision-making including the preoperative posture and surgical options, is not fully emphasized (18–21).

Generally, the extent of the retinal lesion is obtained using the supervised models which requires elaborate labeling for most existing algorithms. Whereas, the equivocal boundaries of lesion, as well as the lack of expert annotations considerably hinder the efficient development of related models. In this context, weakly supervised learning, where the learning model can be trained with incomplete and simplified annotations, has attracted great attention (22). It typically fits for training lesion localization and segmentation models in medical images. For instance, Ma et al. (23) resorted to classification-based Class Activation Maps (CAMs) to segment geographic atrophy in retinal OCT images. Monaro et al. (24) proposed an architectural setting that enabled the weakly-supervised coarse segmentation of age-related macular degeneration lesions in color fundus images. The incorporation of lesion-specific activation maps provides more meaningful information for diagnosis with great explainability. In medical imaging, Gradient-weighted CAM (Grad-CAM) (25) is one of the most commonly used techniques to generate coarse localization maps. However, most approaches derived from it only focus on the discriminative image regions but ignore much detailed information. To alleviate this issue, Qin et al. (26) proposed an activation modulation and recalibration (AMR) scheme. The combination architecture of a compensation branch and spotlight branch could achieve better performance on image-level weakly supervised segmentation tasks. Given our purpose of achieving lesion-specific holistic localization, working under coarse image-level annotation instead of bounding box annotation is highly desirable (22, 27–29). Moreover, incorporating the AMR scheme mentioned above with our approaches could generate high-quality activation maps to compensate for previous detail-loss issues.

Therefore, we proposed a weakly supervised learning model to generate localization maps that outline the RD lesions based on UWF images. Relying on the localization maps, the potential diagnostic evidence will be instantaneously transmitted to the clinicians for reference. Furthermore, individual postural guidance will be generated for healthcare reference to the patients.

2 Materials and methods

This study was conducted adhering to the tenets of the Declaration of Helsinki. It was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Zhejiang University, School of Medicine.

2.1 Data acquisition

A total of 30,446 UWF images were retrospectively obtained from visitors presenting for ophthalmic examinations between 1 May 2016 and 15 August 2022, at Eye Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University. Images insufficient for interpretation were excluded, including (1) Poor-view images, referring to images with significant deficiencies in focus or illumination, visibility of the optic disc, or over one-third of the field obscured by the eyelashes or eyelids. (2) Poor-position images, referring to images with significantly off-center optic disc and macula due to incorrect gazing in the image capture process. The UWF images were captured using an OPTOS nonmydriatic camera (OPTOS Daytona, Dunfermline, United Kingdom) with 200-degree fields of view. The subjects underwent the examinations without mydriasis. All of the UWF images were anonymized before being involved in this study.

2.2 Image labeling and the definition of RD regions

A professional image labeling team was recruited to generate the ground truth. The team consisted of two retinal specialists with more than 5 years of clinical experience and one senior specialized ophthalmologist with over 20 years of clinical experience.

At first, the included UWF images were annotated with image-level labels after quality filtration. Two specialists, respectively, classified all images into two types: RD and Non-RD. The ground truth was determined based on their consensus. Any divergences were finally arbitrated by the senior specialized ophthalmologist. Figure 1 illustrates the workflow of image classification.

Then, the uninvolved fovea of each RD image (Macula ON) was marked manually to further obtain the specific anatomical zone for

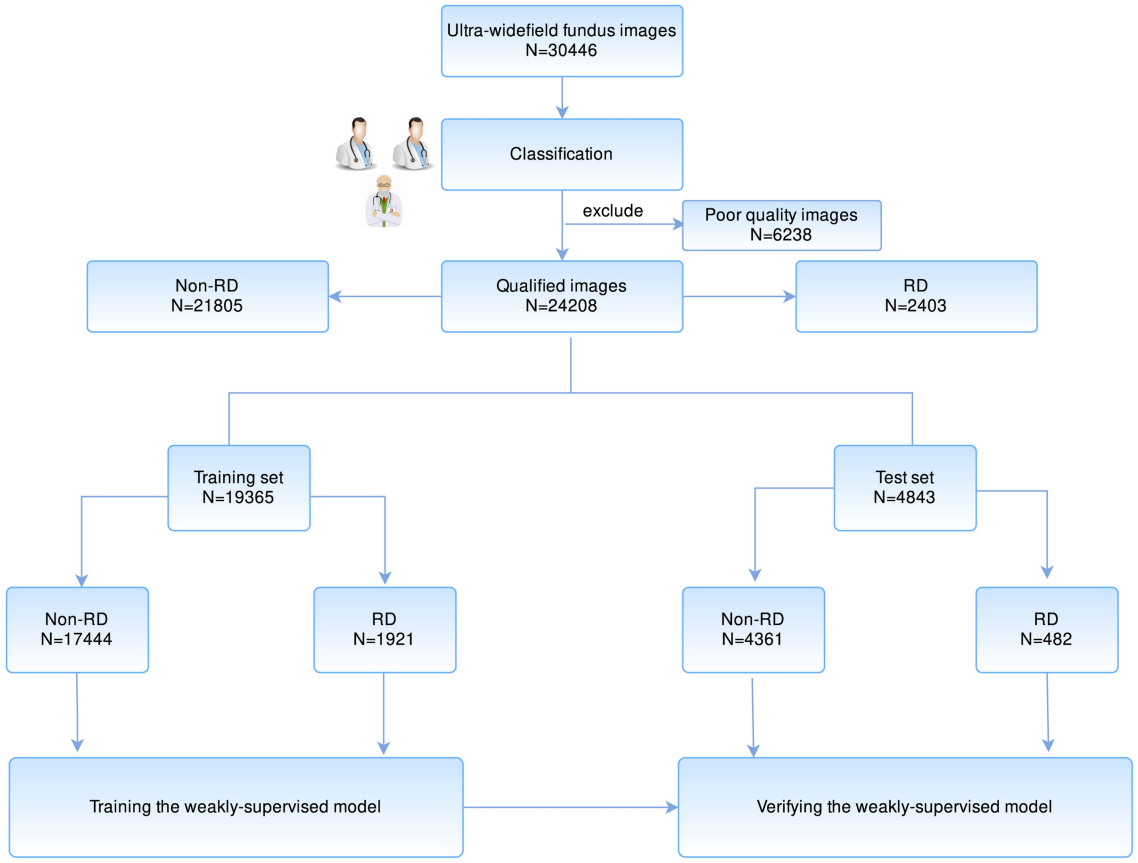

FIGURE 1
The workflow of developing a weakly supervised learning model for the localization of RD region based on UWF images. RD, retinal detachment; UWF, ultra-widefield fundus.

postural guidance. Besides, the RD regions of images in the test set were independently contoured by two specialists. The ground truth of the RD region was determined based on the intersection of their labeled areas. Any image with less than 0.9 intersection-over-union (IOU) of the labeled RD regions was confirmed by the senior specialized ophthalmologist.

2.3 Development of the weakly-supervised deep learning model to localize the RD regions

UWF images incorporate a vast of critical information about the profile and distribution of the lesions, which is essential for the healthcare of RD patients. Clinically, typical RD is recognized by an elevated and corrugated retinal appearance accompanied by retinal breaks, and such features can often be recognized by the deep learning algorithm. Based on this rationale, we propose a model that enables the localization of RD regions based on weakly supervised training. The design of the model consisted of two sections: localizing the RD lesions and generating postural guidance according to the anatomical zone of the lesion.

In the localization section, an attention modulation module (AMM) (26) was involved in our scheme to realize recalibration supervision and generate lesion-specific activation maps. In the first place, it was necessary to extract the fundus' region of interest (ROI). The four corners (left and right top, left and right bottom) in a UWF image were called irrelevant areas since there was no fundus information in these four regions. These irrelevant regions from different images were variable in texture but highly similar in extent. We manually crafted an ROI template to erase pixels in these irrelevant regions. Local contrast enhancement (CLAHE) was applied to image augmentation afterward.

A ResNet-101 (30) was pre-trained to identify RD cases with a learning rate of 0.01 and focal loss (alpha was set to 0.65, gamma was set to 1.15). Then, AMM was employed to emphasize region-essential features for the segmentation task between every two stages, as shown in Figure 2. Features from the discriminative regions were considered to be the most sensitive features, and the minor features referred to features that are important but easily ignored (31). The AMM can rearrange the distribution of the feature importance to highlight sensitive and minor activations, which is crucial to generating semantic segmentation masks. The ResNet-101 with AMMs was fine-tuned with a learning rate of 0.001. Probability maps were generated based on feature maps from stage 4 by Grad-CAM and resampled to the original size afterward.

In the guidance section, the coarse segmentation of the RD region with pseudo labels obtained from localization maps with a probability threshold of 0.5 was carried out. As shown in Figure 2, a

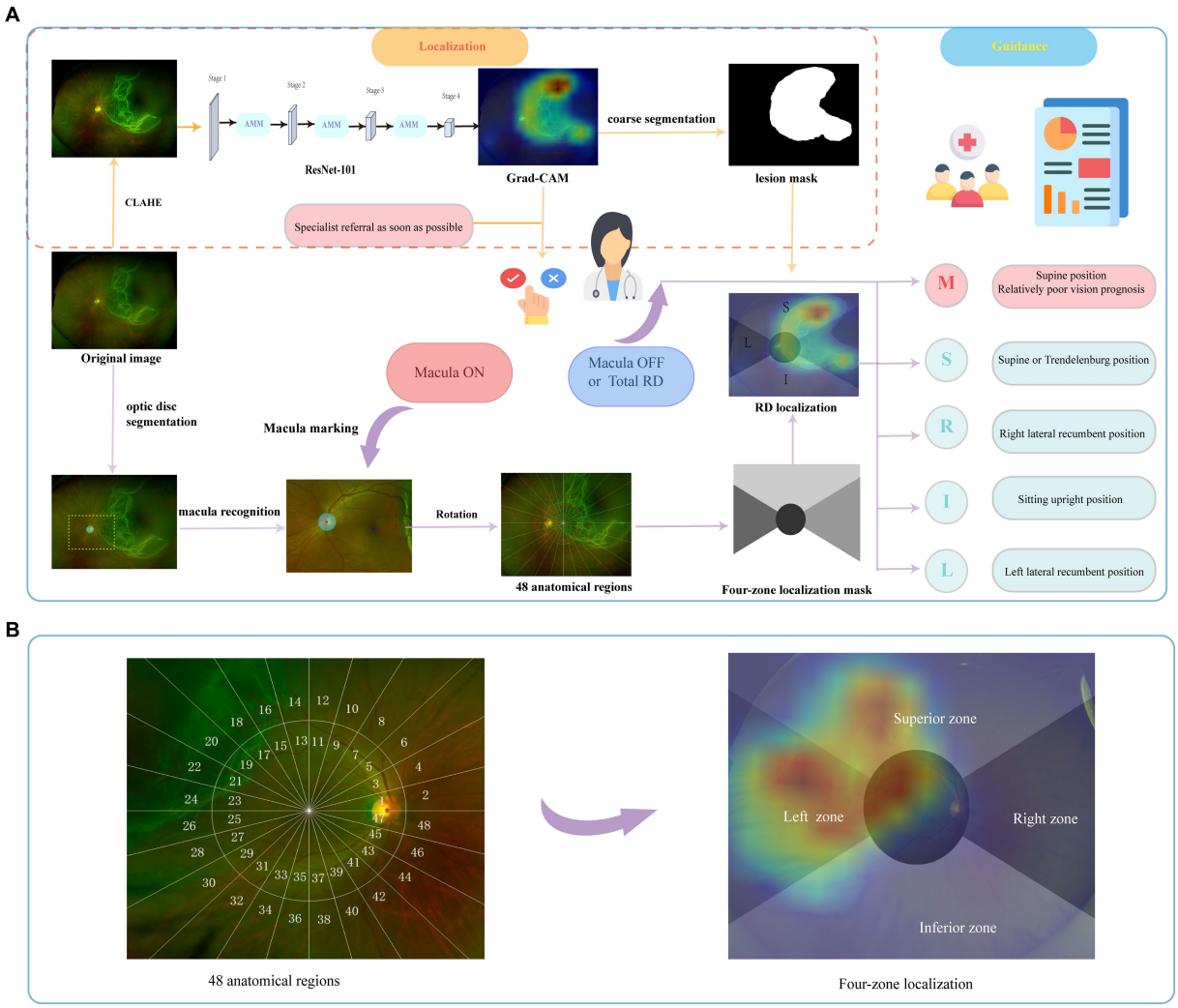

FIGURE 2
The schema of the overall study. The brief illustration of RD region localization and corresponding postural guidance **(A)**. The retina was divided into 48 anatomical regions to evaluate the holistic localization performance. The final four-zone overlaid image was generated for postural guidance **(B)**. RD, retinal detachment; M, macula zone; S, superior zone; R, right zone; I, inferior zone; L, left zone.

coordinate system was constructed based on the recognition of fovea marked manually and optical disc segmented by U-Net (32). Details on the established zoning principles are presented below. The primary zone label is assigned corresponding to the largest number of RD pixels. Then, the predicted label was output based on the coarse lesion segmentation results to generate the customized postural guidance.

2.4 Zoning principles

The panoramic retinal view is divided into four zones centered on the manually labeled macula fovea. It is calibrated with a horizontal line through the fovea and optic disc center. In a clockwise direction, we designate 2–4 o'clock as the right zone, 10–2 o'clock as the superior zone, 8–10 o'clock as the left zone, and 4–8 o'clock as the inferior zone. In addition, we pay more attention to the posterior pole, designated as the circle centered on the macula and including the optic disc (33), which is closely associated with the surgical option and visual prognosis (4, 6). To further evaluate the holistic localization performance, each zone was divided clockwise by 15° to obtain 48 anatomical regions defining the entire retina as shown in Figure 2. Each image has a 48-length vector label for 48-partition localization. The label is assigned as 1 when more than 50 RD pixels fall into this partition.

2.5 Sensitivity analysis

Given that difference in image resolution of input data may have impacts on the localization outcome. We implemented sensitivity analyses based on three common image resolutions including 256 × 256, 512 × 512, and 1,024 × 1,024 pixels. We evaluated the 48-partition localization performance of our weakly-supervised model in these contexts separately and selected the optimal resolution model for further evaluation.

TABLE 1 Baseline characteristics of the training and test datasets.

	Training set (80%)		Test set (20%)	
	(*n* = 19,365)		(*n* = 4,843)	
	RD	Non-RD	RD	Non-RD
Total no. of images	1,921	17,444	482	4,361
No. of OD images	1,019	9,163	256	2,325
No. of OS images	902	8,281	226	2,036

RD, retinal detachment; OD, right eye; OS, left eye.

TABLE 2 The holistic localization performance of our weakly-supervised model with different image resolutions.

Resolutions (pixels)	Precision (95%CI)[1]	Recall (95%CI)[1]	F1 score (95%CI)[1]
256 × 256	0.8718 (0.8653–0.8780)	0.7381 (0.7304–0.7457)	0.7994 (0.7931–0.8055)
512 × 512	0.8642 (0.8581–0.8701)	**0.8327 (0.8262–0.8390)**	**0.8481 (0.8426–0.8535)**
1,024 × 1,024	**0.8914 (0.8852–0.8973)**	0.7284 (0.7206–0.7361)	0.8017 (0.7955–0.8078)

[1]The localization performance of the weakly-supervised deep learning model was evaluated with a probability threshold of 0.5.
The bold values are the optimal indicator results of different resolutions.

2.6 Comparisons of the proposed model with the baseline model and general ophthalmologist

A comparison experiment with the proposed model was conducted using a baseline model without AMM to explore the performance enhancement that comes with the AMR scheme. Meanwhile, to evaluate our weakly-supervised deep learning model in the localization of the RD region, we recruited a general ophthalmologist with 3 years of clinical experience. It is challenging to clearly define the contours of the RD region, considering its equivocal borders, even for clinicians. Given that the final localization is the essential factor for postural instruction, we evaluated their performance of lesion body localization rather than the edge segmentation performance. According to the defined ground truth, we compared the localization performance of the proposed model with that of the baseline model and general ophthalmologist based on the test set, respectively.

2.7 Statistical analysis

The precision, recall, F1 score, sensitivity, specificity, and accuracy of the models and general ophthalmologist were calculated according to the reference standard. The F1 score is the harmonic mean of precision and recall, which is calculated as:

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

The precision-recall curve was generated to visualize the localization performance of the deep learning models. The Cohen's Kappa value of the model and general ophthalmologist compared with the reference standard for the four-zone localization was calculated to evaluate the consistency. All statistical analyses for the study were conducted using SPSS 26.0 (Chicago, IL, United States) and Python 3.7 (Wilmington, DE, United States).

3 Results

3.1 Data characteristics

In total, 30,446 images were obtained for preliminary model development. After filtering 6,238 poor-quality images that are insufficient for interpretation, 24,208 eligible images were annotated. Two thousand four hundred and three were classified as RD, while the remaining 21,805 images were classified as non-RD. The dataset was randomly split in 80:20 ratios according to the Pareto principle, with 19,365 (80%) images as a training set and 4,843 (20%) as a test set. The baseline characteristics of collected images are summarized in Table 1.

3.2 Evaluation of the weakly-supervised deep learning model to localize the RD regions

In the test set, the associated lesions of 480 RD images are successfully localized with activation maps. Only two cases have been missed due to the inconspicuous shallow detachment. In 467 Macula-ON RD images, the entire retina is divided into 48 anatomical regions based on the location of the optic disc and macula fovea, as illustrated in Figure 2, to evaluate the holistic localization of the RD region in the test set. The following anatomical localization evaluation will be specific to these 467 RD images.

Table 2 exhibited the holistic localization performance of our weakly supervised model with three image resolutions for sensitivity analysis. The results showed that the image resolution of 1,024 × 1,024 pixels had the highest precision of 89.14% (95%CI: 88.52–89.73%). However, the image resolution of 512 × 512 pixels achieved the highest recall of 83.38% (95%CI: 82.91–83.84%) and acceptable precision with an optimal F1 score of 84.81% (95%CI: 84.26–85.35%). As a result, the following localization evaluation adopted the image resolution of 512 × 512 uniformly.

The performance of the baseline model, the proposed model, and general ophthalmologist to identify whether the posterior pole is

TABLE 3 The localization of RD in the posterior pole area by the weakly-supervised deep learning model and the general ophthalmologist compared with the ground truth in the test set.

Index	Sensitivity (95%CI)	Specificity (95%CI)	Accuracy (95%CI)
Baseline model (without AMM)[1]	0.7453 (0.7339–0.7564)	0.9189 (0.9113–0.9259)	0.8295 (0.8224–0.8363)
Weakly-supervised model[1]	0.8249 (0.8150–0.8344)	0.9116 (0.9034–0.9191)	0.8651 (0.8586–0.8713)
General ophthalmologist	0.8649 (0.8585–0.8710)	0.8630 (0.8563–0.8693)	0.8639 (0.8593–0.8683)

[1]The localization performance of the weakly-supervised deep learning model was evaluated with a probability threshold of 0.5. The image resolution of the input data is 512 × 512 pixels. AMM, attention modulation module.

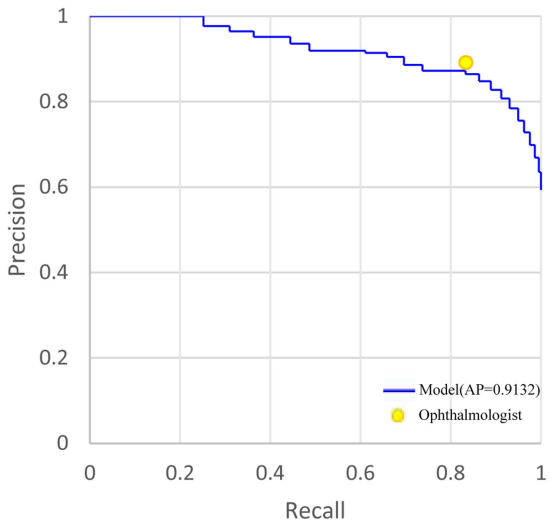

FIGURE 3
The precision-recall curve of holistic localization performance of RD region based on the weakly-supervised model and general ophthalmologist, compared with the ground truth in the test set. AP, average precision. RD, retinal detachment. AMM, attention modulation module.

TABLE 4 The performance of localizing RD lesions in 48 anatomical regions by the baseline model, weakly-supervised model, and the general ophthalmologist, compared with the ground truth in the test set.

Index	Precision (95%CI)	Recall (95%CI)	F1 score (95%CI)
Baseline model (without AMM)[1]	0.9267 (0.9211–0.9319)	0.6807 (0.6725–0.6888)	0.7849 (0.7785–0.7912)
Weakly-supervised model[1]	0.8642 (0.8581–0.8701)	0.8327 (0.8262–0.8390)	0.8481 (0.8426–0.8535)
General ophthalmologist	0.8916 (0.8875–0.8955)	0.8338 (0.8291–0.8384)	0.8617 (0.8564–0.8668)

[1]The localization performance of the weakly-supervised deep learning model was evaluated with a probability threshold of 0.5. The image resolution of the input data is 512 × 512 pixels. AMM, attention modulation module.

involved or not is shown in Table 3. The general ophthalmologist had an 86.49% (95%CI: 85.85–87.10%) sensitivity and an 86.30% (95%CI: 85.63–86.93%) specificity, whereas the model had an 82.49% (95%CI: 81.50–83.44%) sensitivity and a 91.16% (95%CI: 90.34–91.91%) specificity with a probability threshold of 0.5. Despite a high specificity of 91.89% (95%CI: 91.13–92.59%) achieved, the baseline model showed limited sensitivity of 74.53% (95%CI: 73.39–75.64%) for early identification.

As for localizing RD lesions in 48 anatomical regions, the general ophthalmologist had an 89.16% (95%CI: 88.75–89.55%) precision and 83.38% (95%CI: 82.91–83.84%) recall. In contrast, our model had an 86.42% (95%CI: 85.81–87.01%) precision and an 83.27% (95%CI: 82.62–83.90%) recall with an average precision (AP) of 0.9132. Though the baseline model achieved a 92.67% (95%CI: 92.11–93.19%) precision which could be attributed to the most discriminative response region, it showed limited recall of 68.07% (95%CI: 67.25–68.88%). For visualizing the model performance when different probability thresholds are applied, the precision-recall curve of the model is shown in Figure 3. The performance of localizing RD lesions in all 48 anatomical regions by the proposed model and general ophthalmologist is shown in Table 4.

Compared with the ground truth, the unweighted Cohen's κ coefficients were 0.710 (95%CI: 0.659–0.761) and 0.753 (95%CI: 0.702–0.804) for the weakly-supervised model and the general ophthalmologist, respectively. The four-zone location accuracy of our model is 0.8051 (95%CI: 0.7656–0.8395), which is slightly inferior to the general ophthalmologist's accuracy of 0.8437 (95%CI: 0.8068–0.8748). The confusion matrixes are shown in Figure 4.

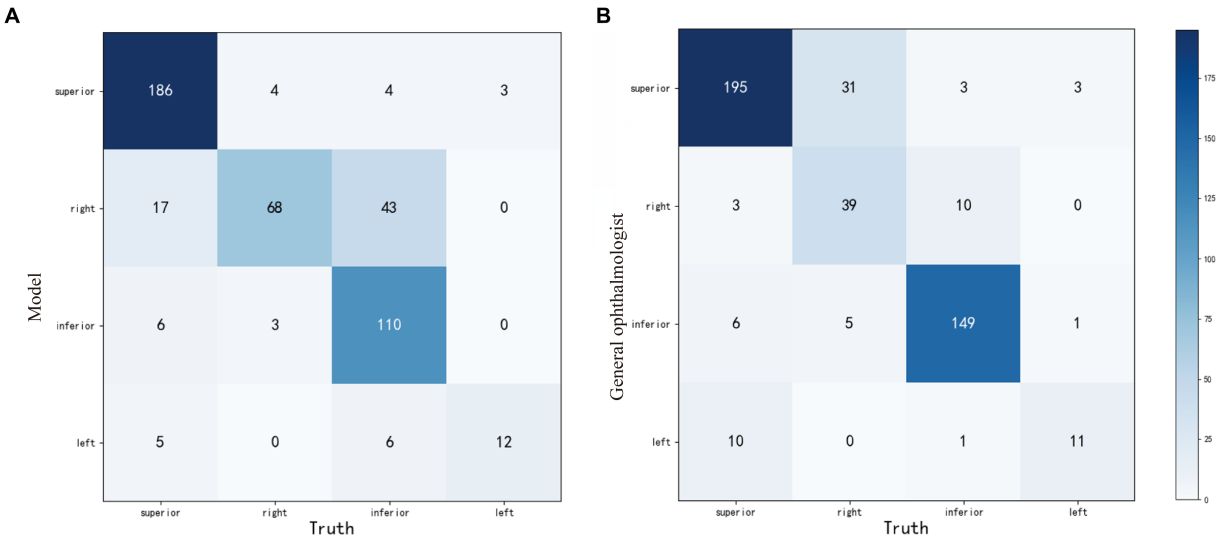

FIGURE 4
The confusion matrixes of four-zone RD localization performance based on the weakly-supervised model **(A)** and the general ophthalmologist **(B)**, compared with the ground truth in the test set. RD, retinal detachment.

4 Discussion

RD is a typical ophthalmic emergency. Early medical interventions based on the precise localization of lesions could increase the success rate of surgical repair and avoid permanent visual impairment (6, 34). Here, we established a standardized procedure for RD localization from UWF images using a weakly-supervised approach. It could provide a corresponding medical reference to both clinicians and RD patients throughout the early-stage management. Compared with the baseline model which only focused on the most discriminative regions with limited recall, our weakly supervised model incorporated the AMR scheme. For this reason, the generated localization maps yielded a comprehensive presentation of RD lesion-related information. The four-zone anatomical localization performance of our model, which was highly related to posture regimens (6, 35, 36), showed substantial consistency with the specialists according to the unweighted Cohen's kappa coefficients of 0.710(95%CI: 0.659–0.761). The human-model comparisons also demonstrated its localization performance with high precision and recall, almost equaled to a general ophthalmologist's judging ability. In general, our model exhibits acceptable performance for the holistic localization of the RD regions. To the best of our knowledge, this is the first attempt to precisely localize the RD regions.

Previously, several deep learning systems in identifying RD in fundus images presented favorable performance (16, 17, 37, 38). Similarly, our model also showed a perfect capacity of discernment for RD from UWF images. Nevertheless, previous deep learning models were mainly proposed for classification tasks, and CAMs were employed for post-hoc interpretability. Since such heatmaps were classification-oriented, they tended to resort to some discriminative regions instead of the holistic bound of the whole object. Even though Li et al. (17) attempted to visualize the decisive regions with saliency maps and embedded an arrow according to the hot regions for head positioning guidance, the most decisive regions in the heatmaps may not be the primary location of RD lesions. The classifiers may only focus on a small part of the target lesions (26, 39). Moreover, the limited localization results from true-positive samples had yet to be thoroughly evaluated for general feasibility. In contrast to simply utilizing classification-oriented heatmaps, our model presents the edge of providing lesion-specific holistic activation maps to localize RD regions. For digging out the regions that are essential but easily ignored for lesion segmentation by the weakly supervised algorithm, we introduce AMM to our scheme to provide recalibration supervision and task-specific concepts. The lesion information of clinical interests provided by this interpretable method complies with cognitive law, which could indicate the diagnostic reference to the clinicians and could be verified easily. Moreover, in the coordinates established above, the model could elaborate on the anatomical zones of the RD lesions. According to the most affected zone, a supine preoperative position is advised for RD in the superior zones and a sitting position for RD in the inferior zones (9). Patients with RD lesions in the right or left zones were positioned flat on the right or left side of the affected eye, respectively (40). The involvement of the posterior pole is almost suggestive of a relatively poor vision prognosis if emergency repair surgery is not available before the macula is involved (4, 7, 41). Patients should maintain a supine position during this time and take an urgent referral.

In our research, most cases can realize holistic localization of RD lesions with great satisfaction. As shown in Figure 5, the corrugated retinal appearance of RD lesions makes them more distinguishable, whereas the shallow RDs are easily missed due to their atypical appearance. In addition, interference from irrelevant factors can also be misleading for automatic localization. The OPTOS camera pads and artifacts with RD-similar edges may result in mistaken highlights in localization maps. In future work, these problems could be improved by further training based on large-scale images with corresponding issues.

This study has several limitations. First, blurred border and texture feature differences within the RD regions made it difficult for

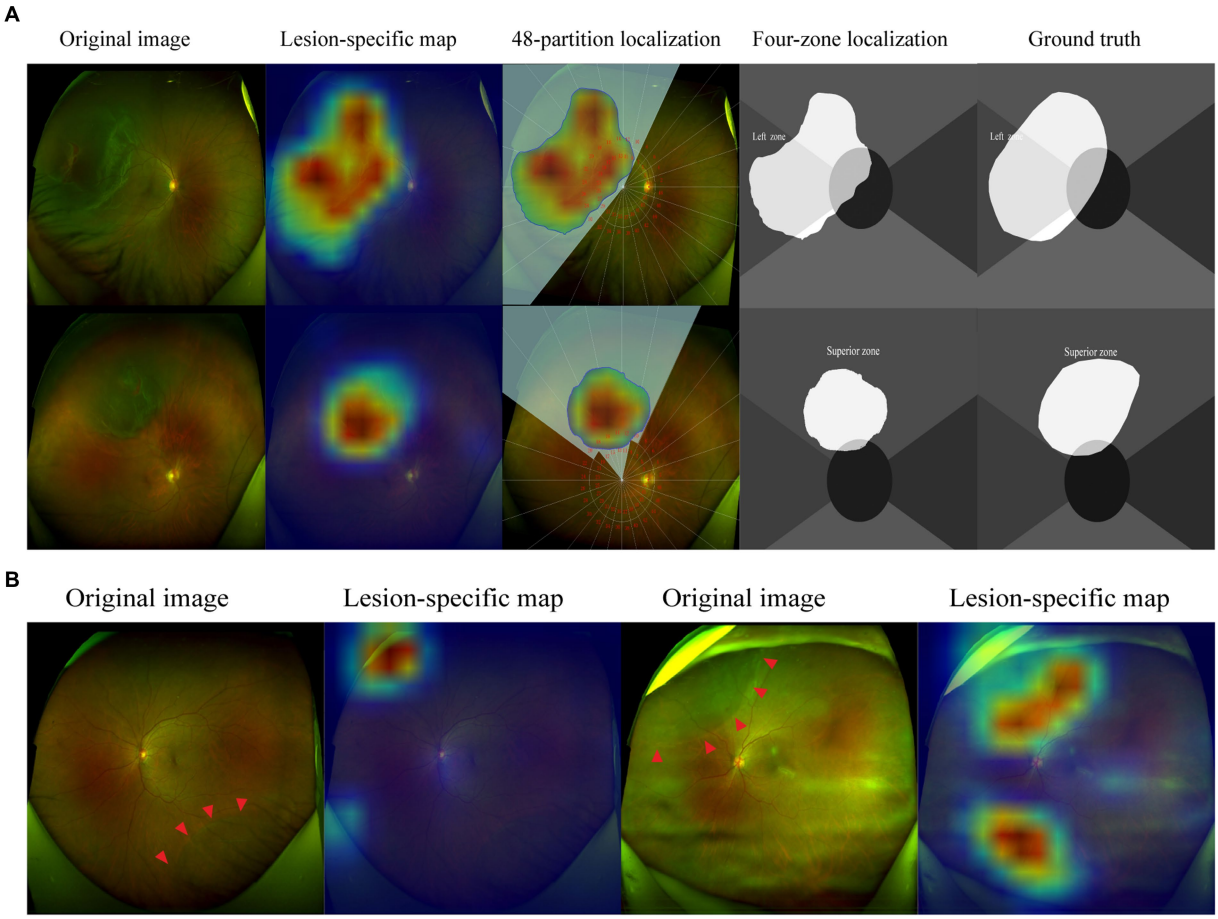

FIGURE 5
Visualization of representative cases. The corrugated retina and the edge of breaks are highlighted in lesion-specific maps, the detached regions were demonstrated in 48-partition localization maps and four-zone localization maps **(A)**. The shallow retinal detachments are not detected in the inferior quadrant, while OPTOS camera pads are highlighted. Artifacts caused by opaque refractive media are highlighted in localization maps **(B)**. The red arrowheads indicate the borders of the RD region.

the activation maps to highlight the whole area of the lesion. The regions with inconspicuous texture features were easily missed even though the advanced AMM module had been incorporated, which may result in some inconsistency in anatomical localization. Strictly, the localization of breaks of the rhegmatogenous RD had more significance for posture instructions. However, the small breaks in the retina were not always visible, especially in the peripheral regions. Given most of the breaks are within the detached retina, the localization of the RD region could extend its clinical applications considerably. In addition, the determination of whether the posterior pole was involved may not represent the status of the macula, especially when the macula was located near the borderline of the RD regions. Hence, further work is warranted to accurately discern the status of the macula for determining operation time and predicting visual prognosis. Furthermore, automatic postural guidance had a relatively limited application range due to the high-quality images required for anatomical localization. The anatomical localization of RD was highly dependent on the clear presentation of the retina. Those fundus images with significant opaque refractive media, inappropriate illumination, and invisible optic disc were not eligible for inclusion in this study. Finally, our model was developed based on single-center retrospective datasets with limited generalization. The evaluation of localization accuracy was conducted on a single-disease dataset and was not strictly validated in the cases of fundus comorbidities. In the future, we expect to explore more advanced methods to aid the full-stage management of RD, incorporating the medical history and other imaging data. Meanwhile, we will expand the training samples of fundus comorbidity images and facilitate the evaluation based on the large-scale test scenario.

5 Conclusion

In this study, we developed a weakly-supervised deep learning model to localize RD regions based on UWF images. The lesion-specific localization maps could be incorporated into the diagnostic process and personalized postural guidance of RD patients for reference. Moreover, the implementation of this task considerably surmounted the current "label-hunger" difficulty. It would greatly facilitate managing RD patients when insufficient specialists are available, especially for medical referral and postural guidance. The application of this model could significantly equilibrate medical resources and improve healthcare efficiency.

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 humans were approved by Institutional Review Board of the Second Affiliated Hospital of Zhejiang University, School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. 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

HL: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft. JC: Conceptualization, Data curation, Validation, Writing – review & editing, Writing – original draft. KY: Data curation, Formal analysis, Methodology, Software, Writing – original draft. YZ: Resources, Writing – review & editing. JY: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by National Natural Science Foundation Regional Innovation and Development Joint Fund, grant number U20A20386, National key research and development program of China, grant number 2019YFC0118400, and Key research and development program of Zhejiang Province, grant number 2019C03020.

Conflict of interest

KY and YZ were employed by Zhejiang Feitu Medical Imaging Co., Ltd.

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

Publisher's note

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

References

1. Ghazi NG, Green WR. Pathology and pathogenesis of retinal detachment. *Eye*. (2002) 16:411–21. doi: 10.1038/sj.eye.6700197

2. Mitry D, Charteris DG, Fleck BW, Campbell H, Singh J. The epidemiology of rhegmatogenous retinal detachment: geographical variation and clinical associations. *Br J Ophthalmol*. (2010) 94:678–84. doi: 10.1136/bjo.2009.157727

3. Gariano RF, Kim CH. Evaluation and management of suspected retinal detachment. *Am Fam Physician*. (2004) 69:1691–8.

4. Kwok JM, Yu CW, Christakis PG. Retinal detachment. *CMAJ*. (2020) 192:E312. doi: 10.1503/cmaj.191337

5. Felfeli T, Teja B, Miranda RN, Simbulan F, Sridhar J, Sander B, et al. Cost-utility of Rhegmatogenous retinal detachment repair with pars Plana vitrectomy, scleral buckle, and pneumatic Retinopexy: a microsimulation model. *Am J Ophthalmol*. (2023) 255:141–54. doi: 10.1016/j.ajo.2023.06.002

6. Kang HK, Luff AJ. Management of retinal detachment: a guide for non-ophthalmologists. *BMJ*. (2008) 336:1235–40. doi: 10.1136/bmj.39581.525532.47

7. Steel D. Retinal detachment. *BMJ Clin Evid*. (2014) 2014:0710.

8. de Jong JH, de Koning K, den Ouden T, van Meurs JC, Vermeer KA. The effect of compliance with preoperative Posturing advice and head movements on the progression of macula-on retinal detachment. *Transl Vis Sci Technol*. (2019) 8:4. doi: 10.1167/tvst.8.2.4

9. de Jong JH, Vigueras-Guillén JP, Simon TC, Timman R, Peto T, Vermeer KA, et al. Preoperative Posturing of patients with macula-on retinal detachment reduces progression toward the fovea. *Ophthalmology*. (2017) 124:1510–22. doi: 10.1016/j.ophtha.2017.04.004

10. Li Y, Li J, Shao Y, Feng R, Li J, Duan Y. Factors influencing compliance in RRD patients with the face-down position via grounded theory approach. *Sci Rep*. (2022) 12:20320. doi: 10.1038/s41598-022-24121-9

11. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: current status and future perspectives. *Adv Ophthalmol Practice Res*. (2022) 2:100078. doi: 10.1016/j.aopr.2022.100078

12. Lake SR, Bottema MJ, Williams KA, Lange T, Reynolds KJ. Retinal shape-based classification of retinal detachment and posterior vitreous detachment eyes. *Ophthalmol Ther*. (2023) 12:155–65. doi: 10.1007/s40123-022-00597-6

13. Li B, Chen H, Zhang B, Yuan M, Jin X, Lei B, et al. Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography. *Br J Ophthalmol*. (2022) 106:316290. doi: 10.1136/bjophthalmol-2020-316290

14. Yadav S, Das S, Murugan R, Dutta Roy S, Agrawal M, Goel T, et al. Performance analysis of deep neural networks through transfer learning in retinal detachment diagnosis using fundus images. *Sādhanā*. (2022) 47:49. doi: 10.1007/s12046-022-01822-5

15. Nagiel A, Lalane RA, Sadda SR, Schwartz SD. ULTRA-WIDEFIELD Fundus Imaging: a review of clinical applications and future trends. *Retina*. (2016) 36:660–78. doi: 10.1097/IAE.0000000000000937

16. Ohsugi H, Tabuchi H, Enno H, Ishitobi N. Accuracy of deep learning, a machine-learning technology, using ultra–wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. *Sci Rep*. (2017) 7:9425. doi: 10.1038/s41598-017-09891-x

17. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. *Commun Biol*. (2020) 3:15. doi: 10.1038/s42003-019-0730-x

18. Alexander P, Ang A, Poulson A, Snead MP. Scleral buckling combined with vitrectomy for the management of rhegmatogenous retinal detachment associated with inferior retinal breaks. *Eye*. (2008) 22:200–3. doi: 10.1038/sj.eye.6702555

19. Chronopoulos A, Hattenbach LO, Schutz JS. Pneumatic retinopexy: A critical reappraisal. *Surv Ophthalmol*. (2021) 66:585–93. doi: 10.1016/j.survophthal.2020.12.007

20. Hillier RJ, Felfeli T, Berger AR, Wong DT, Altomare F, Dai D, et al. The pneumatic Retinopexy versus vitrectomy for the Management of Primary Rhegmatogenous Retinal Detachment Outcomes Randomized Trial (PIVOT). *Ophthalmology*. (2019) 126:531–9. doi: 10.1016/j.ophtha.2018.11.014

21. Warren A, Wang DW, Lim JI. Rhegmatogenous retinal detachment surgery: a review. *Clin Experiment Ophthalmol*. (2023) 51:271–9. doi: 10.1111/ceo.14205

22. Wang J, Li W, Chen Y, Fang W, Kong W, He Y, et al. Weakly supervised anomaly segmentation in retinal OCT images using an adversarial learning approach. *Biomed Opt Express*. (2021) 12:4713–4729. doi: 10.1364/BOE.426803

23. Ma X., Ji Z., Niu S., Leng T., Rubin DL, Chen Q. MS-CAM: Multi-Scale Class Activation Maps for Weakly-Supervised Segmentation of Geographic Atrophy Lesions in SD-OCT Images. *IEEE Journal of Biomedical and Health Informatics*, (2020) 24:3443–3455. doi: 10.1109/JBHI.2020.2999588

24. Morano J, Hervella ÁS, Rouco J, Novo J, Fernández-Vigo JI, Ortega M. Weakly-supervised detection of AMD-related lesions in color fundus images using explainable deep learning. *Comput Methods Prog Biomed*. (2023) 229:107296. doi: 10.1016/j.cmpb.2022.107296

25. Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. in *2017 IEEE international Conference on Computer vision (ICCV)*, 618–626. (2017).

26. Qin J., Wu J., Xiao X., Li L., Wang X. Activation modulation and recalibration scheme for weakly supervised semantic segmentation. (2021). *Proceedings of the AAAI Conference on Artificial Intelligence*, 36. (Palo Alto, CA: AAAI Press), 2117–2125.

27. Cinbis RG, Verbeek J, Schmid C. Weakly supervised object localization with multi-fold multiple instance learning. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:189–203. doi: 10.1109/TPAMI.2016.2535231

28. Zhang D., Guo G., Zeng W., Li L., Han J. Generalized weakly supervised object localization. in: *IEEE Transactions on Neural Networks and Learning Systems*, (2022a). PP. 1, 12.

29. Zhang D, Han J, Cheng G, Yang MH. Weakly supervised object localization and detection: a survey. *IEEE Trans Pattern Anal Mach Intell*. (2022b) 44:1–5885. doi: 10.1109/TPAMI.2021.3074313

30. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. in 2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR) IEEE Conference on Computer Vision and Pattern Recognition. (2016). (New York: IEEE), 770–778.

31. Jiang P., Hou Q., Cao Y., Cheng M., Wei Y., Xiong H. Integral object mining via online attention accumulation. in *2019 IEEE/CVF international Conference on Computer vision (ICCV)* (Seoul, Korea (South): IEEE), (2019). 2070–2079.

32. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation In: N Navab, J Hornegger, WM Wells and AF Frangi, editors. *Medical image computing and Computer-assisted intervention—MICCAI 2015 lecture notes in Computer science*. Cham: Springer International Publishing (2015). 234–41.

33. Quinn N, Csincsik L, Flynn E, Curcio CA, Kiss S, Sadda SR, et al. The clinical relevance of visualising the peripheral retina. *Prog Retin Eye Res*. (2019) 68:83–109. doi: 10.1016/j.preteyeres.2018.10.001

34. Wickham L, Bunce C, Wong D, Charteris DG. Retinal detachment repair by vitrectomy: simplified formulae to estimate the risk of failure. *Br J Ophthalmol*. (2011) 95:1239–44. doi: 10.1136/bjo.2010.190314

35. Reeves MG, Pershing S, Afshar AR. Choice of primary Rhegmatogenous retinal detachment repair method in US commercially insured and Medicare advantage patients, 2003-2016. *Am J Ophthalmol*. (2018) 196:82–90. doi: 10.1016/j.ajo.2018.08.024

36. Sverdlichenko I, Lim M, Popovic MM, Pimentel MC, Kertes PJ, Muni RH. Postoperative positioning regimens in adults who undergo retinal detachment repair: a systematic review. *Surv Ophthalmol*. (2023) 68:113–25. doi: 10.1016/j.survophthal.2022.09.002

37. Sun G, Wang X, Xu L, Li C, Wang W, Yi Z, et al. Deep learning for the detection of multiple fundus diseases using ultra-widefield images. *Ophthalmol Ther*. (2023) 12:895–907. doi: 10.1007/s40123-022-00627-3

38. Zhang C, He F, Li B, Wang H, He X, Li X, et al. Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study. *Graefes Arch Clin Exp Ophthalmol*. (2021) 259:2225–34. doi: 10.1007/s00417-021-05105-3

39. Meng Q, Liao L, Satoh S. Weakly-supervised learning with complementary Heatmap for retinal disease detection. *IEEE Trans Med Imaging*. (2022) 41:2067–78. doi: 10.1109/TMI.2022.3155154

40. Johannigmann-Malek N, Stephen BK, Badawood S, Maier M, Baumann C. Influence of preoperative POSTURING on SUBFOVEAL fluid height in macula-off retinal detachments. *Retina*. (2023) 43:1738–44. doi: 10.1097/IAE.0000000000003864

41. Diederen RMH, La Heij EC, Kessels AGH, Goezinne F, Liem ATA, Hendrikse F. Scleral buckling surgery after macula-off retinal detachment: worse visual outcome after more than 6 days. *Ophthalmology*. (2007) 114:705–9. doi: 10.1016/j.ophtha.2006.09.004

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/fmed.2024.1377479

Check for updates

OPEN ACCESS

EDITED BY
Yanda Meng,
University of Exeter, United Kingdom

REVIEWED BY
Xu Chen,
University of Cambridge, United Kingdom
Peng Xue,
Shandong University, China

*CORRESPONDENCE
Zhangrong Chen
✉ chenzhangrong71@163.com

RECEIVED 27 January 2024
ACCEPTED 09 May 2024
PUBLISHED 22 May 2024

CITATION
Liu X, Tan H, Wang W and Chen Z (2024) Deep learning based retinal vessel segmentation and hypertensive retinopathy quantification using heterogeneous features cross-attention neural network. *Front. Med.* 11:1377479. doi: 10.3389/fmed.2024.1377479

COPYRIGHT
© 2024 Liu, Tan, Wang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning based retinal vessel segmentation and hypertensive retinopathy quantification using heterogeneous features cross-attention neural network

Xinghui Liu[1,2], Hongwen Tan[2], Wu Wang[3] and Zhangrong Chen[1,4]*

[1]School of Clinical Medicine, Guizhou Medical University, Guiyang, China, [2]Department of Cardiovascular Medicine, Guizhou Provincial People's Hospital, Guiyang, China, [3]Electrical Engineering College, Guizhou University, Guiyang, China, [4]Department of Cardiovascular Medicine, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

Retinal vessels play a pivotal role as biomarkers in the detection of retinal diseases, including hypertensive retinopathy. The manual identification of these retinal vessels is both resource-intensive and time-consuming. The fidelity of vessel segmentation in automated methods directly depends on the fundus images' quality. In instances of sub-optimal image quality, applying deep learning-based methodologies emerges as a more effective approach for precise segmentation. We propose a heterogeneous neural network combining the benefit of local semantic information extraction of convolutional neural network and long-range spatial features mining of transformer network structures. Such cross-attention network structure boosts the model's ability to tackle vessel structures in the retinal images. Experiments on four publicly available datasets demonstrate our model's superior performance on vessel segmentation and the big potential of hypertensive retinopathy quantification.

KEYWORDS

retinal vessel segmentation, hypertensive retinopathy quantification, deep learning, cross-attention network, color fundus images

1 Introduction

Hypertension (HT) is a chronic ailment posing a profound menace to human wellbeing, manifesting in vascular alterations (1). Its substantial contribution to the global prevalence and fatality rates of cardiovascular diseases (CVD) cannot be overstated. The escalated incidence and mortality rates are not solely attributable to HT's correlation with CVD but also to the ramifications of hypertension-mediated organ damage (HMOD). This encompasses structural and functional modifications across pivotal organs, including arteries, heart, brain, kidneys, vessels, and the retina, signifying preclinical or asymptomatic CVD (2, 3). HT management's principal aim remains to deter CVD incidence and mortality rates. Achieving this goal mandates meticulous adherence to HT guidelines, emphasizing precise blood pressure monitoring and evaluating target organ damage (4). Consequently, the early identification of HT-mediated organ damage emerges as a pivotal concern.

The retinal vascular system shares commonalities in structural, functional, and embryological aspects with the vascular systems of the heart, brain, and kidneys (5–9). Compared to other microvascular territories, the distinctive attributes of the retinal microcirculation enable relatively straightforward detection of localized HMOD (5, 9). Its capacity to offer a non-invasive and uncomplicated diagnostic tool positions retinal visualization as the simplest means of elucidating the microcirculatory system. In hypertensive patients, retinal microvasculature gives insight into the wellbeing of the heart, kidneys, and brain (5, 10, 11). Early detection of HT-mediated retinal changes indirectly mirrors the vascular status of these organs, facilitating refined evaluation of cardiovascular risk stratification, timely interventions, and improved prognostication, thereby holding substantial clinical significance. Traditional clinical methodologies for diagnosing HT-mediated retinal alterations, while reliant on the proficiency of ophthalmic professionals, often demand considerable time and specialized expertise (12). Figure 1 presents a sample fundus image, demonstrating the complexity of the retinal vasculature and image intensity variation. However, integrating AI-based models in ophthalmology holds promising prospects for revolutionizing this paradigm. Leveraging machine learning algorithms and deep neural networks, AI-enabled diagnostic tools have demonstrated the potential to expedite and enhance the assessment of HT-related retinal vessel changes (13–17). These AI models learn from extensive datasets of annotated medical images, swiftly recognizing subtle retinal anomalies that might elude human detection. By automating the analysis and interpretation of retinal images, AI-based systems offer the prospect of reducing diagnostic timeframes, improving accuracy, and potentially mitigating the need for extensive human oversight. In this work, we proposed a heterogeneous features cross-attention neural network to tackle the retinal vessel segmentation task with color fundus images.

2 Related work

Segmenting blood vessels in retinal color fundus images plays a pivotal role in the diagnostic process of hypertensive retinopathy. Over the years, researchers have explored computer-assisted methodologies to tackle this task. For instance, Annunziata and Trucco (18) introduced a novel curvature segmentation technique leveraging an accelerating filter bank implemented via a speed-up convolutional sparse coding filter learning approach. Their method employs a warm initialization strategy, kickstarted by meticulously crafted filters. These filters are adept at capturing the visual characteristics of curvilinear structures, subsequently fine-tuned through convolutional sparse coding. Similarly, Marín et al. (19) delved into the realm of hand-crafted feature learning methods, harnessing gray-level and moment invariant-based features for vessel segmentation. However, despite the efficacy of such techniques, the manual crafting of filters is inherently time-intensive and prone to biases, necessitating a shift toward more automated and data-driven approaches in this domain.

Deep learning techniques based on data analysis have demonstrated superior performance to conventional retinal vessel segmentation approaches (18–20). For instance, Maninis et al. (21) developed a method wherein feature maps derived from a side output layer contributed to vessel and optic disc segmentation. Along a similar line, Oliveira et al. (22) combined the benefits of stationary wavelet transform's multi-scale analysis with a multi-scale full convolutional neural network, resulting in a technique adept at accommodating variations in the width and orientation of retinal vessel structures. In terms of exploiting the advance of the Unit structure, there are previous methods that achieved promising performance. For example, Yan et al. (23) implemented a joint loss function in U-Net, comprising two components responsible for pixel-wise and segment-level losses, aiming to enhance the model's ability to balance segmentation between thicker and thinner vessels. Mou et al. (24) embedded dense dilated convolutional blocks between encoder and decoder cells at corresponding levels of a U-shaped network, employing a regularized walk algorithm for post-processing model predictions. Similarly, Wang et al. (25) proposed a Dual U-Net with two encoders: one focused on spatial information extraction and the other on context information. They introduced a novel module to merge information from both paths.

Despite the proficiency of existing deep learning methodologies in segmenting thicker vessels, there remains a challenge in combining heterogeneous features from different stages of the deep learning models via Transformers and CNN models. Generally, improving deep learning-based techniques for vessel segmentation can be approached from various angles, including multi-stage feature fusion and optimization of loss functions. This work proposes a heterogeneous feature cross-attention neural network to address the above challenge.

3 Materials and methods

3.1 Heterogeneous features cross-attention neural network

A detailed model structure overview is shown in Figure 2. In detail, two brunches of feature extraction modules are proposed to extract heterogeneous features from different stages of the backbone network. In detail, there is CNN-based (Conv-Block) and transformer-based (Trans-Block) brunch, which focus on local semantic and long-range spatial information. Those two features' information are both important for the vessel segmentation task.

The interaction between the two branches is used as a cross-attention module to emphasize the essential heterogeneous (semantic and spatial) features. It is used as the main structure to facilitate the interaction and integration of local and long-range global features. Drawing inspiration from the work by Peng et al. (26), the intersecting network architecture within our model ensures that both Conv-Block and Trans-Block can concurrently learn features derived from the preceding Conv-Block and Trans-Block, respectively.

3.1.1 CNN blocks

In the structure depicted in Figure 2, the CNN branch adopts a hierarchical structure, leading to a reduction in the resolution of feature maps as the network depth increases and the channel count expands. Each phase of this structure

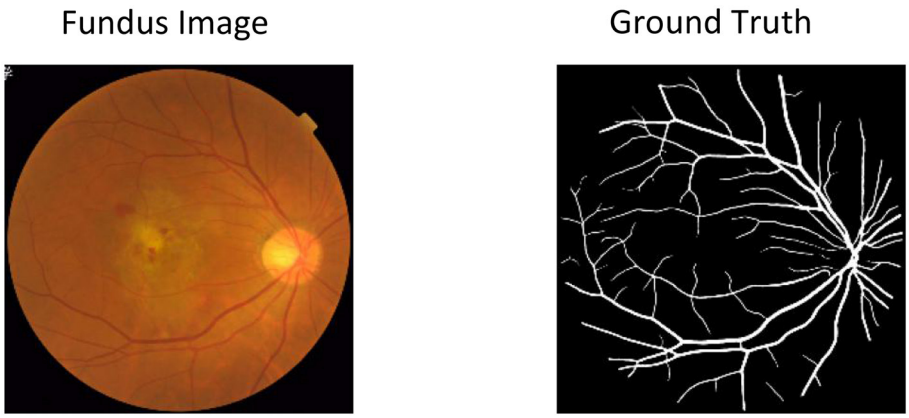

FIGURE 1
Sample retinal fundus image for vessel segmentation and hypertensive retinopathy quantification. The yellow areas in Ground Truth represent the retinal vessel area that needs to be segmented for disease analysis.

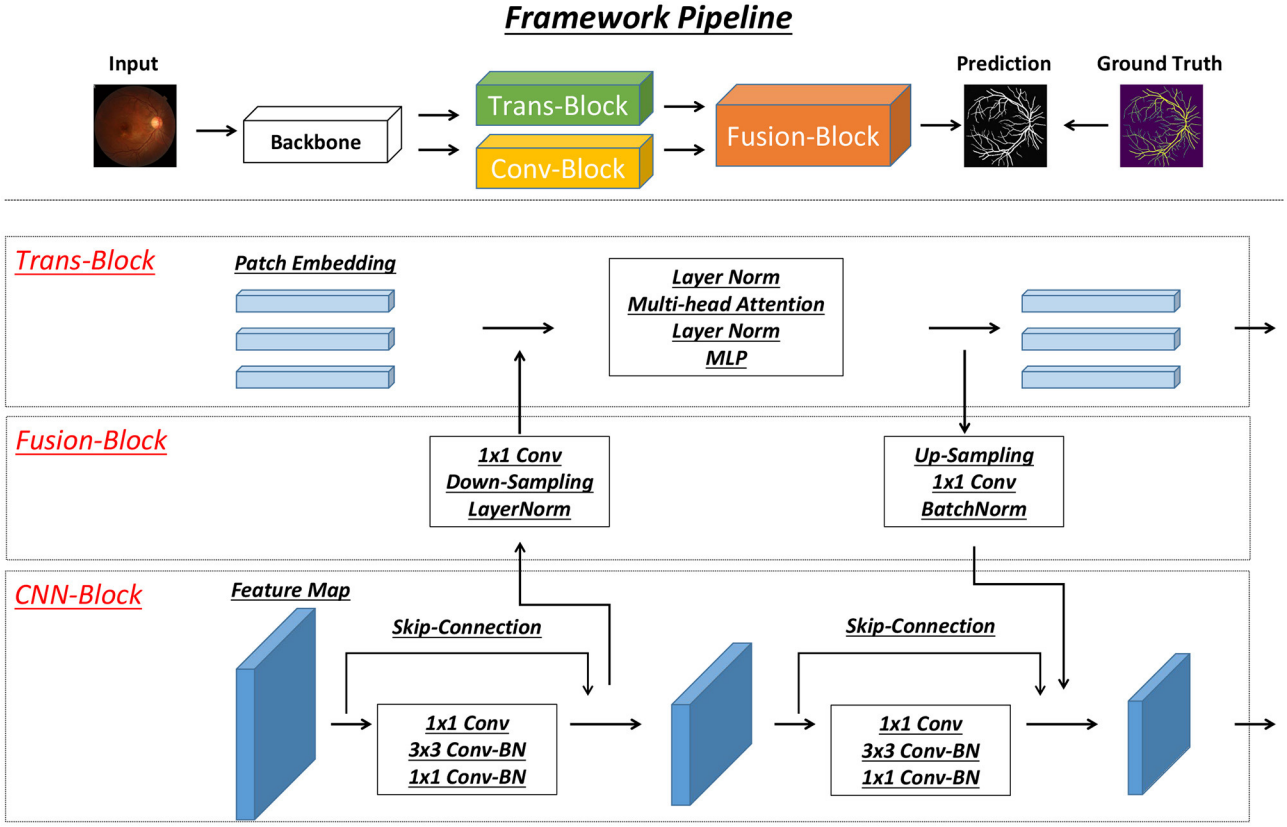

FIGURE 2
Figure of our proposed model structure. Our model contains three modules, including Trans-Block, CNN-Block and Fusion-Block. The detailed structure of each module is shown in the figure.

consists of several convolution blocks, each housing multiple bottlenecks. These bottlenecks, in accordance with the ResNet framework (27), comprise a sequence involving down-projection, spatial convolution, up-projection, and a residual connection to maintain information flow within the block. Distinctly, visual transformers (28) condense an image patch into a vector in one step, which unfortunately leads to the loss of localized details. Conversely, in CNNs, the convolutional kernels operate on feature maps, overlapping to retain intricate local features. Consequently, the CNN branch ensures a sequential provision of localized feature intricacies to benefit the transformer branch.

3.1.2 Transformer blocks

In line with the approach introduced in ViT (28), this segment consists of N sequential transformer blocks, as showcased in Figure 2. Each transformer block combines a multi-head self-attention module with an MLP block, encompassing an up-projection fully connected layer and a down-projection fully connected layer. Throughout this structure, LayerNorms (29) are applied before each layer, and residual connections are integrated into both the self-attention layer and the MLP block. For tokenization purposes, the feature maps generated by the backbone module are compressed into 16 × 16 patch embeddings without overlap. This compression is achieved using a linear projection layer, implemented via a 3 × 3 convolution with a stride of 1. Notably, considering that the CNN branch (3 × 3 convolution) encodes both local features and spatial location information, the necessity for positional embeddings diminishes. This strategic adaptation results in an improved image resolution, advantageous for subsequent tasks related to vision.

3.1.3 Feature fusion blocks

Aligning the feature maps derived from the CNN branch with the patch embeddings within the transformer branch poses a significant challenge. To tackle this, we introduce the feature fusion block, aiming to continuously and interactively integrate local features with global representations. The substantial difference in dimensionalities between the CNN and transformer features is noteworthy. While CNN feature maps are characterized by dimensions $C \times H \times W$ (representing channels, height, and width, respectively), patch embeddings assume a shape of $(L + 1) \times J$, where L, 1, and J denote the count of image patches, class token, and embedding dimensions, respectively. To reconcile these disparities, feature maps transmitted to the transformer branch undergo an initial 1 × 1 convolution to align their channel numbers with the patch embeddings. Subsequently, a down-sampling module (depicted in Figure 2) aligns spatial dimensions, following which the feature maps are amalgamated with patch embeddings, as portrayed in Figure 2. Upon feedback from the transformer to the CNN branch, the patch embeddings necessitate up-sampling (as illustrated in Figure 2) to match the spatial scale. Following this, aligning the channel dimension with that of the CNN feature maps through a 1×1 convolution is performed, integrating these adjusted embeddings into the feature maps. Furthermore, LayerNorm and BatchNorm modules are employed to regularize the features. Moreover, a significant semantic disparity arises between feature maps and patch embeddings. While feature maps stem from local convolutional operators, patch embeddings arise from global self-attention mechanisms. Consequently, the feature fusion block is incorporated into each block (excluding the initial one) to bridge this semantic gap progressively.

3.2 Experiments

3.2.1 Datasets

Four public datasets, *DRIVE* (30), *CHASEDB1* (31), *STARE* (32), and *HRF* (33), were used in our experiments. The images of these datasets were captured by different devices and with different image sizes. A detailed description of each dataset is elaborated below:

1). *DRIVE* dataset: the dataset known as *DRIVE* comprises 40 pairs of fundus images accompanied by their respective labels for vessel segmentation. Each image within this dataset measures 565 × 584 pixels. Furthermore, the dataset has been partitioned into distinct training and test sets, encompassing 20 pairs of images and corresponding labels within each set. Notably, in the test set, every image has undergone labeling by two medical professionals. Typically, the initial label is considered the reference standard (ground truth), while the second label serves as a human observation used to assess accuracy.

2). *CHASEDB1* dataset: the CHASEDB1 dataset encompasses a collection of 28 images, comprising samples from both the left and right eyes, with each image possessing dimensions of 999 × 960 pixels. Past investigations have specifically delineated the dataset's utilization, designating a distinct partition for training and testing purposes. According to prior scholarly research (31), a selection strategy has been employed, with the final eight images demarcated for evaluation as testing samples, while the preceding images have been earmarked for utilization as training samples. This segmentation strategy in the dataset facilitates a structured approach for model training and evaluation, enabling a systematic analysis of algorithm performance on separate subsets of images to ensure robustness and generalizability in vessel segmentation tasks.

3). *STARE* dataset: each image within the *STARE* dataset measures 700 × 605 pixels. This dataset comprises 20 color fundus images without a predefined division into training and test sets. Previous studies have employed two common schemes for test set allocation to assess method performance. One approach involves assigning 10 images to the training set and the remaining 10 to the test set. Alternatively, the Leave-One-Out method has been utilized, wherein each image successively serves as the test set while the remaining images form the training set for evaluation purposes in different iterations.

4). *HRF* dataset: the HRF dataset comprises 45 fundus images with a resolution of 3,504 × 2,336 pixels. From this dataset, 15 images from are allocated to the training set, while the remaining 30 images constitute the test set. To mitigate computational expenses, both the images and their corresponding labels are downsampled twice, as noted in (34).

3.2.2 Loss functions

Commonly utilized region-based losses, like Dice loss (35), often result in highly precise segmentation. However, they tend to disregard the intricate vessel shapes due to a multitude of pixels outside the target area, overshadowing the significance of those delineating the vessel (36–40). This oversight may contribute to relatively imprecise retinal vessel segmentation and, consequently, inaccurate quantification of hypertensive retinopathy. In response,

we incorporated the TopK loss (Equation 1) (41, 42) to emphasize the retinal vessels during the training process specifically. When objects exhibit sizes that are not notably smaller in comparison to the convolutional neural network's (CNN) receptive field, the vessel emerges as the most variable component within the prediction, displaying the least certainty; thus, the loss within the vessel region tends to be the highest among the predictions (43). Building upon these observations and rationale, the TopK loss is formulated as follows:

$$L_{TopK} = -\frac{1}{N}\sum_{i\in K} g_i \log s_i \tag{1}$$

where g_i is the ground truth of pixel i, s_i is the corresponding predicted probability, and K is the set of the $k\%$ pixels with the lowest prediction accuracy. While sole vessel-focused loss often causes training instability (44), region-based loss, such as Dice loss (Equation 2) (35), is needed at the early stage of the training. We represent Dice loss as follows:

$$L_{Dice} = 1 - \frac{2|V_s \cap V_g|}{|V_s| + |V_g|} \tag{2}$$

where V_g is the ground truth label and V_s is the prediction result of segmentation. We coupled TopK with region-based Dice loss as our final loss function (Equation 3) for the retinal vessel segmentation.

$$L = L_{TopK} + L_{Dice} \tag{3}$$

3.2.3 Experimental setting

To enrich the dataset, we introduce random rotations on the fly to the input images in the training dataset, applied to both segmentation tasks. Specifically, these rotations span from −20 to 20 degrees. Additionally, 10% of the training dataset is randomly chosen to serve as the validation dataset. The proposed network was implemented utilizing the PyTorch Library and executed on the Nvidia GeForce TITAN Xp GPU. Throughout the training phase, we employed the AdamW optimizer to fine-tune the deep model. To ensure effective training, a gradually decreasing learning rate was adopted, commencing at 0.0001, alongside a momentum parameter set at 0.9. For each iteration, a random patch of size 118 × 118 from the image was selected for training purposes, with a specified batch size of 16. A backbone of ResNet50 (27) is used in this work.

3.2.4 Evaluation metrics

The model's output is represented as a probability map, assigning to each pixel the probability of being associated with the vessel class. Throughout the experiments, a probability threshold of 0.5 was employed to yield the results. To comprehensively assess the efficacy of our proposed framework during the testing phase, the subsequent metrics will be computed:

- Acc (accuracy) = (TP + TN) / (TP + TN + FP + FN),
- SE (sensitivity) = TP / (TP + FN),
- SP (specificity) = TN / (TN + FP)
- F1 (F1 score) = (2 × TP) / (2 × TP + FP + FN)
- AUROC = area under the receiver operating characteristic curve.

In this context, the correct classification of a vessel pixel is categorized as a true positive (TP), while misclassification is identified as a false positive (FP). Correspondingly, accurate classification of a non-vessel pixel is considered a true negative (TN), whereas misclassification is denoted as a false negative (FN).

3.3 Compared methods

We compared our approach to other classic and state-of-the-art models that have achieved promising performance on different medical image segmentation tasks. All of the experiments are conducted under the same experimental setting. The compared methods are briefly introduced below:

- Unet (45): Unet is a CNN architecture used for image segmentation tasks. Its U-shaped design includes an encoder (contracting path) for feature extraction and a symmetric decoder (expansive path) for generating segmented outputs. The network uses skip connections to preserve fine details and context, making it effective for tasks like biomedical image segmentation.
- Unet++ (46): Unet++ is an advanced version of the U-Net architecture designed for image segmentation tasks. It improves upon U-Net by introducing nested skip connections and aggregation pathways, allowing better multi-scale feature integration and context aggregation. This enhancement leads to more accurate and precise segmentation results compared to the original U-Net model.
- Swin-Transformer (47): Swin-Transformer is a hierarchical vision transformer (28) structure. It uses shifted windows to process image patches hierarchically, allowing for improved global context understanding. This architecture has demonstrated competitive segmentation performance with efficient computation.
- AttenUnet (48): The AttenUnet enhances the traditional U-Net architecture that integrates attention mechanisms. These mechanisms enable the network to focus on important image features during segmentation tasks. It improves accuracy by refining object delineation and suppressing irrelevant information. This variant is particularly effective in tasks like medical image segmentation, where precise localization of structures is essential.
- TransUnet (49): TransUNet is a proposed architecture to improve medical image segmentation, addressing limitations seen in the widely used U-Net model. It combines the strengths of Transformers' global self-attention with U-Net's precise localization abilities. The Transformer part encodes image patches from a CNN feature map to capture global context, while the decoder integrates this with high-resolution feature maps for accurate localization.

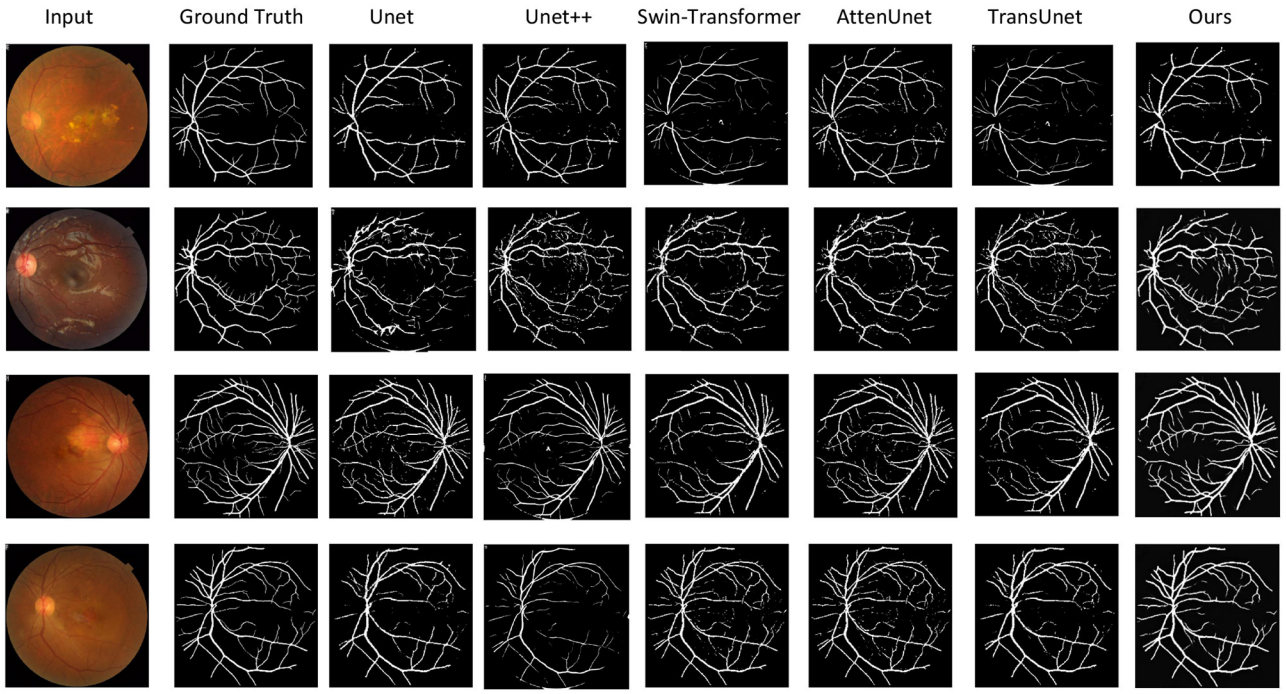

FIGURE 3
Qualitative results of the vessel segmentation. We compare our model with Unet (45), Unet++ (46), Swin-Transformer (47), AttenUnet (48), TransUnet (49). Our method can produce more accurate segmentation results than the other methods compared with the ground truth.

TABLE 1 **Quantitative results comparison between our methods and other compared state-of-the-art methods on *DRIVE* dataset.**

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
Unet	90.1 (89.1, 90.8)	76.5 (74.2, 78.1)	97.7 (95.8, 99.1)	80.3 (78.3, 82.3)	97.2 (95.0, 98.0)
Unet++	91.3 (90.4, 92.7)	79.2 (78.0, 80.6)	97.9 (95.2, 99.0)	81.0 (79.2, 82.5)	97.1 (95.8, 99.0)
Swin-Transformer	92.3 (91.5, 92.9)	79.0 (77.9, 80.6)	98.1 (96.4, 99.2)	82.0 (81.0, 84.0)	97.6 (96.1, 98.3)
AttenUnet	92.1 (91.3, 93.2)	80.0 (78.3, 82.0)	98.3 (96.1, 99.5)	80.4 (78.5, 82.1)	97.4 (96.2, 98.6)
TransUnet	91.8 (91.2, 93.0)	80.3 (79.1, 81.3)	98.3 (97.2, 99.6)	80.1 (78.8, 80.9)	97.3 (96.4, 99.0)
Ours	**93.8** (92.9, 94.8)	**81.0** (80.2, 82.6)	**98.5** (96.7, 99.1)	**83.3** (78.8, 82.1)	**97.9** (96.2, 98.8)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

4 Results

4.1 Vessel segmentation performance

Figure 3 illustrates qualitative comparison with other compared methods on the test dataset. Tables 1–4 shows the quantitative performance of *Ours* and other methods on four different datasets, respectively.

Our proposed method can outperform other compared methods on *DRIVE*, *CHASEDB1*, *STARE*, and *HRF* datasets, respectively. In detail, *Ours* achieved 83.3% *F1* on *DRIVE* dataset, which outperformed *Unet* (45) by 3.6%, outperformed *Swin-Transformer* (47) by 1.6% and outperformed *TransUnet* (49) by 4.0%. *Ours* achieved 81.6% *F1* on *CHASEDB1* dataset, which outperformed *Unet++* (46) by 1.9%, outperformed *AttenUnet* (48) by 2.1% and outperformed *TransUnet* (49) by 1.5%. *Ours* achieved 86.6% *F1* on *STARE* dataset, which outperformed *Unet* (45) by 2.7%, outperformed *AttenUnet* (48) by 2.4% and outperformed *TransUnet* (49) by 1.6%. *Ours* achieved 79.9% *F1* on *HRF* dataset, which outperformed *Unet++* (46) by 0.8%, outperformed *Swin-Transformer* (47) by 0.5% and outperformed *TransUnet* (49) by 1.3%. Notably, *Swin-Transformer* (47) and *TransUnet* (49) belong to the transformer-based model structure, which demonstrates a superior performance on many tasks. However, in this work, the limited data size is one of the leading reasons for the relatively low performance of those datasets. Another reason could be the task's own nature of vessel segmentation, where more local information is needed rather than the long-range relationship between pixels. Thus, given two brunches with transformer and CNN structures and fusion modules, our proposed model can simultaneously tackle both the local semantic information and long-range spatial information for the segmentation task.

Figure 3 shows the qualitative comparison between ours and other compared methods. It demonstrated that our proposed

TABLE 2 Quantitative results comparison between our methods and other compared state-of-the-art methods on *CHASEDB1* dataset.

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
Unet	91.2 (89.8, 92.3)	60.3 (58.2, 61.4)	97.1 (96.4, 97.9)	79.7 (76.9, 81.0)	97.7 (96.6, 98.2)
Unet++	91.6 (89.8, 93.2)	63.0 (61.2, 65.0)	97.3 (95.5, 98.3)	80.1 (78.5, 82.1)	97.7 (96.2, 98.3)
Swin-Transformer	92.3(91.0, 94.1)	62.9 (61.4, 64.0)	97.8 (96.2, 98.5)	80.3 (78.7, 81.7)	97.9 (96.2, 98.8)
AttenUnet	92.4 (91.0, 94.2)	67.7 (65.5, 68.3)	97.7 (96.2, 98.4)	79.9 (77.4, 80.6)	97.8 (97.0, 98.5)
TransUnet	92.6 (90.2, 94.4)	66.1 (64.6, 67.7)	98.0 (96.7, 99.0)	80.4 (78.9, 82.1)	98.2 (96.3, 99.9)
Ours	**93.7** (91.7, 95.2)	**69.0** (67.4, 70.5)	**98.9** (97.2, 99.3)	**81.6** (81.0, 93.0)	**98.9** (98.1, 99.3)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

TABLE 3 Quantitative results comparison between our methods and other compared state-of-the-art methods on *STARE* dataset.

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
Unet	93.3 (91.7, 95.2)	80.8 (78.7, 81.8)	98.1 (97.1, 99.0)	84.3 (82.2, 86.3)	98.1 (97.0, 99.0)
Unet++	94.2 (92.5, 96.0)	82.6 (81.6, 83.1)	98.0 (96.4, 99.0)	84.5 (83.7, 85.2)	98.3 (97.1, 99.2)
Swin-Transformer	93.9 (92.8, 94.7)	83.0 (82.0, 84.2)	98.2 (96.9, 99.1)	84.1 (82.5, 86.2)	98.5 (97.4, 99.3)
AttenUnet	93.6 (92.7, 94.7)	82.9 (81.7, 84.2)	98.6 (96.2, 99.3)	84.6 (82.9, 86.3)	98.6 (96.7, 99.5)
TransUnet	93.4 (91.9, 94.7)	83.2 (81.6, 85.0)	98.7 (96.6, 99.4)	85.2 (83.7, 86.9)	98.1 (97.2, 99.1)
Ours	**94.8** (92.9, 95.6)	**84.2** (82.6, 86.1)	**99.2** (97.7, 99.4)	**86.6** (85.9, 87.4)	**99.3** (98.4, 99.7)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

methods can segment the vessels more accurately. This is important for vessel segmentation tasks and hypertensive retinopathy quantification with more accurate vessel area calculation.

4.2 Ablation study

4.2.1 Ablation study on loss functions

We did ablation study experiments on loss functions. We maintain the same model structure and only change the loss functions. In detail, we remove Dice loss and TopK loss, respectively, to evaluate their respective contribution to the performance of the proposed models. Furthermore, we replace TopK loss with a cross-entropy loss to validate the effectiveness of TopK loss in the segmentation task. Table 5 demonstrates that Dice Loss can lead to a 6.2% *F1* and *TopK* loss can lead to a 2.9% *F1* performance. On the other hand, Dice loss can lead to 15.5% *SE* performance, and *TopK* loss can lead to a 2.8% *SE* performance on *Drive* dataset. Additionally, compared with cross-entropy loss, the TopK loss could lead to a 1.5% *F1* improvement and 2.3% *SE* improvement. Each loss function can boost the model's performance in different evaluation metrics. This demonstrated that the adopted loss function can both contribute to the learning process and benefit the vessel segmentation performance.

4.2.2 Ablation study on the models' components

We did ablation study experiments on the model's components. In detail, we maintain the same model structure and only change the models' structure by removing different modules, including *Trans-Block*, *CNN-Block* and *Fusion-Block*, respectively. In detail, we remove each of those three modules, respectively, to evaluate their respective contribution to the performance of the proposed models. Table 6 demonstrates that *Trans-Block* can lead to a 10% *F1*, *CNN-Block* can lead to a 10.3% *F1* performance and *Fusion-Block* can lead to a 7.9% *F1* performance boost. On the other hand, *Trans-Block* can lead to a 3.3% *SE* performance, *CNN-Block* can lead to a 2.3% *SE* performance, and *Fusion-Block* can lead to an 0.9% *SE* performance on *Drive* dataset. Each module can boost the model's performance in different evaluation metrics. This demonstrated that the proposed modules can all contribute to the learning process and benefit the vessel segmentation performance.

5 Hypertensive retinopathy quantification

The proposed method has demonstrated a promising retinal vessel segmentation performance on different datasets and benchmarks. Additionally, precise segmentation of retinal vessels plays a vital role in hypertensive retinopathy detection, whereas manual segmentation tends to be cumbersome and time-consuming (50). The model proposed can generate a binary mask distinguishing vessel pixels as one and background pixels as zero. This mask effectively quantifies the total count of vessel pixels within each mask. The ratio (R_{vessel}) between the count of vessel pixels and non-vessel pixels is defined as follows:

$$R_{vessel} = \frac{N_v}{N_{non} - N_v}, \tag{4}$$

where N_v represents the count of vessel pixels, and N_{non} denotes the count of non-vessel pixels. The ratio R_{vessel} (Equation 4) serves as a valuable metric in identifying hypertensive retinopathy within fundus images. Hypertensive retinopathy leads to vascular

TABLE 4 Quantitative results comparison between our methods and other compared state-of-the-art methods on *HRF* dataset.

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
Unet	94.4 (92.3, 96.0)	77.7 (75.8, 79.0)	95.1 (93.8, 96.7)	78.6 (76.9, 79.1)	97.2 (96.0, 98.0)
Unet++	94.8 (92.8, 96.2)	78.9 (78.0, 79.6)	95.1 (93.8, 96.4)	79.3 (78.7, 80.5)	97.3 (96.1, 98.3)
Swin-Transformer	94.6 (92.9, 96.0)	79.1 (77.9, 80.5)	94.4 (92.7, 96.0)	79.5 (77.7, 80.6)	97.8 (96.2, 98.6)
AttenUnet	95.8 (93.9, 96.9)	77.6 (75.8, 79.1)	94.6 (93.9, 95.4)	78.8 (76.9, 79.5)	98.2 (97.0, 99.0)
TransUnet	95.3 (94.2, 96.3)	78.6 (77.4, 79.8)	94.7 (92.9, 96.3)	78.9 (77.0, 79.9)	98.3 (97.2, 99.1)
Ours	**96.2** (95.0, 97.1)	**79.9** (78.0, 81.0)	**94.9** (92.8, 96.0)	**79.9** (77.9, 81.2)	**98.8** (97.9, 99.3)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

TABLE 5 Quantitative ablation study results of the loss function on DRIVE dataset.

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
w/o Dice loss	86.4 (85.0, 88.0)	70.1 (68.2, 72.5)	94.4 (92.3, 96.0)	75.6 (74.1, 76.2)	94.5 (92.8, 95.6)
w/o TopK loss	88.9 (87.3, 89.6)	78.8 (76.9, 80.3)	96.0 (94.2, 97.2)	78.0 (77.0, 79.2)	96.3 (94.8, 97.7)
w/ Cross-entropy loss	90.3 (89.6, 91.0)	79.2 (78.5, 80.0)	96.9 (95.8, 97.4)	79.1 (78.0, 80.2)	96.9 (95.8, 97.5)
Ours	**93.8** (92.9, 94.8)	**81.0** (80.2, 82.6)	**98.5** (96.7, 99.1)	**80.3** (78.8, 82.1)	**97.9** (96.2, 98.8)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

TABLE 6 Quantitative ablation study results of the model's components on DRIVE dataset.

Methods	*Acc*	*SE*	*SP*	*F1*	*AUROC*
w/o Trans-Block	88.9 (87.6, 89.5)	78.4 (76.8, 79.3)	92.1 (91.2, 92.9)	73.0 (71.5, 74.6)	95.2 (93.7, 96.6)
w/o CNN-Block	89.1 (87.9, 90.8)	79.2 (78.2, 80.6)	92.3 (91.4, 92.9)	72.8 (71.6, 73.5)	95.3 (93.8, 96.6)
w/o Fusion-Block	91.2 (89.9, 92.3)	80.3 (78.8, 81.6)	93.1 (92.1, 94.4)	74.4 (72.6, 76.6)	96.3 (95.8, 96.7)
Ours	**93.8** (92.9, 94.8)	**81.0** (80.2, 82.6)	**98.5** (96.7, 99.1)	**80.3** (78.8, 82.1)	**97.9** (96.2, 98.8)

Performance is reported with *Acc*, *SE*, *SP*, *F1* and *AUROC*. 95% confidence interval is presented in the bracket. The best performance is highlighted in bold.

constriction (51, 52), resulting in a decrease in the count of vessel pixels (R_{vessel}).

Detection of hypertensive retinopathy, characterized by vascular constriction, involves assessing changes in R_{vessel} across sequential examinations. Increases or decreases in R_{vessel} indicate the occurrence or progression of hypertensive retinopathy, respectively. Hence, our proposed methods offer a straightforward approach for detecting hypertensive retinopathy.

In the future, with increased datasets comprising fundus images from hypertensive and healthy patients, we can further analyze vessel changes within these images. In real-world clinical practice, comparing the R_{vessel} obtained from consecutive visits can serve as a diagnostic tool. Additionally, the detection of newly formed vessels can be achieved by subtracting images from successive visits post-segmentation. This approach enables the identification and tracking of changes in vasculature over time, offering potential insights for clinical assessment and monitoring.

6 Limitation and future works

While our deep learning method has shown promising results in the challenging tasks of retinal vessel segmentation and hypertensive retinopathy quantification, it's important to acknowledge the nuanced landscape of limitations accompanying such endeavors. One notable factor is the inherent variability present in medical imaging datasets. Our model's performance could be influenced by factors such as variations in image quality and disease severity across different datasets. Moreover, despite achieving commendable results overall, there are instances where the model might struggle to accurately delineate intricate vascular structures or detect subtle manifestations of hypertensive retinopathy. This suggests the need for further exploration and refinement of our approach.

In future research, attention could be directed toward enhancing the model's robustness and adaptability to diverse imaging conditions and patient populations. Techniques such as advanced data augmentation and domain adaptation strategies could prove instrumental in achieving this goal. Additionally, integrating complementary sources of information, such as clinical metadata or genetic markers, holds promise for enriching the predictive capabilities of our model and enhancing its clinical relevance. Furthermore, the pursuit of interpretability and explainability remains paramount. Providing clinicians with insights into how the model arrives at its predictions can foster trust and facilitate its integration into real-world clinical workflows. However, this pursuit must be balanced with ethical considerations, particularly concerning patient privacy, algorithmic bias, and the potential consequences of automated decision-making in healthcare settings. By addressing these multifaceted challenges, we

can pave the way for more effective and responsible deployment of deep learning technologies in ophthalmology and beyond.

7 Conclusion

We have proposed a novel and comprehensive framework for retinal vessel segmentation and hypertensive retinopathy quantification. It takes advantage of heterogeneous feature cross-attention with the help of local emphasis CNN and long-range emphasis transformer structure with a fusion module to aggregate the information. Our experiments on four large-scale datasets have demonstrated that our framework can simultaneously conduct accurate segmentation and potential hypertensive retinopathy quantification performance.

Data availability statement

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

Author contributions

XL: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. HT: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Formal analysis, Conceptualization. WW: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. ZC: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Clinical special of Science and Technology Department of Guizhou Province (No. Qiankehechengguo-LC[2021]023) and the Youth Foundation of Guizhou Provincial People's Hospital (No. GZSYQN[2019]06). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

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

Publisher's note

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

References

1. Houben AJ, Martens RJ, Stehouwer CD. Assessing microvascular function in humans from a chronic disease perspective. *J Am Soc Nephrol*. (2017) 28:3461. doi: 10.1681/ASN.2017020157

2. Rizzoni D, Agabiti-Rosei C, De Ciuceis C, Boari GEM. Subclinical hypertension-mediated organ damage (HMOD) in hypertension: atherosclerotic cardiovascular disease (ASCVD) and calcium score. *High Blood Press Cardiovasc Prev*. (2023) 30:17–27. doi: 10.1007/s40292-022-00551-4

3. Meng Y, Bridge J, Addison C, Wang M, Merritt C, Franks S, et al. Bilateral adaptive graph convolutional network on CT based Covid-19 diagnosis with uncertainty-aware consensus-assisted multiple instance learning. *Med Image Anal*. (2023) 84:102722. doi: 10.1016/j.media.2022.102722

4. Mancia G, De Backer G, Dominiczak A, Cifkova R, Fagard R, Germano G, et al. 2007 Guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur Heart J*. (2007) 28:1462–536. doi: 10.1093/eurheartj/ehm236

5. Flammer J, Konieczka K, Bruno RM, Virdis A, Flammer AJ, Taddei S. The eye and the heart. *Eur Heart J*. (2013) 34:1270–8. doi: 10.1093/eurheartj/eht023

6. Wong TY, Mitchell P. Hypertensive retinopathy. *N Engl J Med*. (2004) 351:2310–7. doi: 10.1056/NEJMra032865

7. Bidani AK, Griffin KA. Pathophysiology of hypertensive renal damage: implications for therapy. *Hypertension*. (2004) 44:595–601. doi: 10.1161/01.HYP.0000145180.38707.84

8. Del Pinto R. Mulè G, Vadalà M, Carollo C, Cottone S, Agabiti Rosei C, et al. Arterial hypertension and the hidden disease of the eye: diagnostic tools and therapeutic strategies. *Nutrients*. (2022) 14:2200. doi: 10.3390/nu14112200

9. Rizzoni D, Agabiti Rosei C, De Ciuceis C, Semeraro F, Rizzoni M, Docchio F. New methods to study the microcirculation. *Am J Hypertens*. (2018) 31:265–73. doi: 10.1093/ajh/hpx211

10. Peng SY, Lee YC. Wu IWn, Lee CC, Sun CC, Ding JJ, et al. Impact of blood pressure control on retinal microvasculature in patients with chronic kidney disease. *Sci Rep*. (2020) 10:14275. doi: 10.1038/s41598-020-71251-z

11. Rizzoni D, De Ciuceis C, Porteri E, Paiardi S, Boari GE, Mortini P, et al. Altered structure of small cerebral arteries in patients with essential hypertension. *J Hypertens*. (2009) 27:838–45. doi: 10.1097/HJH.0b013e32832401ea

12. Arsalan M, Haider A, Lee YW, Park KR. Detecting retinal vasculature as a key biomarker for deep learning-based intelligent screening and analysis of diabetic and hypertensive retinopathy. *Expert Syst Appl*. (2022) 200:117009. doi: 10.1016/j.eswa.2022.117009

13. Wu H, Wang W, Zhong J, Lei B, Wen Z, Qin J. Scs-net: a scale and context sensitive network for retinal vessel segmentation. *Med Image Anal*. (2021) 70:102025. doi: 10.1016/j.media.2021.102025

14. Lin J, Huang X, Zhou H, Wang Y, Zhang Q. Stimulus-guided adaptive transformer network for retinal blood vessel segmentation in fundus images. *Med Image Anal*. (2023) 89:102929. doi: 10.1016/j.media.2023.102929

15. Wei J, Zhu G, Fan Z, Liu J, Rong Y, Mo J, et al. Genetic U-Net: automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *IEEE Trans Med Imaging*. (2021) 41:292–307. doi: 10.1109/TMI.2021.3111679

16. Tan Y, Yang KF, Zhao SX, Li YJ. Retinal vessel segmentation with skeletal prior and contrastive loss. *IEEE Trans Med Imaging*. (2022) 41:2238–51. doi: 10.1109/TMI.2022.3161681

17. Li Y, Zhang Y, Cui W, Lei B, Kuang X, Zhang T. Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation. *IEEE Trans Med Imaging*. (2022) 41:1975–89. doi: 10.1109/TMI.2022.3151666

18. Annunziata R, Trucco E. Accelerating convolutional sparse coding for curvilinear structures segmentation by refining SCIRD-TS filter banks. *IEEE Trans Med Imaging*. (2016) 35:2381–92. doi: 10.1109/TMI.2016.2570123

19. Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imaging*. (2010) 30:146–8. doi: 10.1109/TMI.2010.2064333

20. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging*. (2006) 25:1214–22. doi: 10.1109/TMI.2006.879967

21. Maninis KK, Pont-Tuset J, Arbeláez P, Van Gool L. Deep retinal image understanding. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Cham: Springer (2016), p. 140–8. doi: 10.1007/978-3-319-46723-8_17

22. Oliveira A, Pereira S, Silva CA. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Syst Appl*. (2018) 112:229–42. doi: 10.1016/j.eswa.2018.06.034

23. Yan Z, Yang X, Cheng KT. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng*. (2018) 65:1912–23. doi: 10.1109/TBME.2018.2828137

24. Mou L, Chen L, Cheng J, Gu Z, Zhao Y, Liu J. Dense dilated network with probability regularized walk for vessel detection. *IEEE Trans Med Imaging*. (2019) 39:1392–403. doi: 10.1109/TMI.2019.2950051

25. Wang B, Qiu S, He H. Dual encoding u-net for retinal vessel segmentation. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part I 22*. Cham: Springer (2019), p. 84–92. doi: 10.1007/978-3-030-32239-7_10

26. Peng Z, Huang W, Gu S, Xie L, Wang Y, Jiao J, et al. Conformer: local features coupling global representations for visual recognition. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, QC: IEEE (2021), p. 367–76. doi: 10.1109/ICCV48922.2021.00042

27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV: IEEEE (2016), p. 770–8. doi: 10.1109/CVPR.2016.90

28. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv*. (2020) [Preprint]. arXiv:2010.11929. doi: 10.48550/arXiv:2010.11929

29. Ba JL, Kiros JR, Hinton GE. Layer normalization. *arXiv*. (2016) [Preprint]. arXiv:1607.06450. doi: 10.48550/arXiv:1607.06450

30. Staal J. Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging*. (2004) 23:501–9. doi: 10.1109/TMI.2004.825627

31. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng*. (2012) 59:2538–48. doi: 10.1109/TBME.2012.2205687

32. Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging*. (2000) 19:203–10. doi: 10.1109/42.845178

33. Odstrcilik J, Kolar R, Budai A, Hornegger J, Jan J, Gazarek J, et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Process*. (2013) 7:373–83. doi: 10.1049/iet-ipr.2012.0455

34. Cherukuri V, Bg VK, Bala R, Monga V. Deep retinal image segmentation with regularization under geometric priors. *IEEE Trans Image Process*. (2019) 29:2552–67. doi: 10.1109/TIP.2019.2946078

35. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. In: *Deep Learning and Data Labeling for Medical Applications*. Cham: Springer (2016), p. 179–87. doi: 10.1007/978-3-319-46976-8_19

36. Meng Y, Zhang H, Zhao Y, Gao D, Hamill B, Patri G, et al. Dual consistency enabled weakly and semi-supervised optic disc and cup segmentation with dual adaptive graph convolutional networks. *IEEE Trans Med Imaging*. (2022) 42:416–29. doi: 10.1109/TMI.2022.3203318

37. Meng Y, Chen X, Zhang H, Zhao Y, Gao D, Hamill B, et al. Shape-aware weakly/semi-supervised optic disc and cup segmentation with regional/marginal consistency. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2022), p. 524–34. doi: 10.1007/978-3-031-16440-8_50

38. Meng Y, Wei M, Gao D, Zhao Y, Yang X, Huang X, et al. CNN-GCN aggregation enabled boundary regression for biomedical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part IV 23*. Cham: Springer (2020), p. 352–62. doi: 10.1007/978-3-030-59719-1_35

39. Meng Y, Meng W, Gao D, Zhao Y, Yang X, Huang X, et al. Regression of instance boundary by aggregated CNN and GCN. In: *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VIII 16*. Cham: Springer (2020), p. 190–207. doi: 10.1007/978-3-030-58598-3_12

40. Meng Y, Zhang Y, Xie J, Duan J, Joddrell M, Madhusudhan S, et al. Multi-granularity learning of explicit geometric constraint and contrast for label-efficient medical image segmentation and differentiable clinical function assessment. *Med Image Anal*. (2024) 95:103183. doi: 10.1016/j.media.2024.103183

41. Wu Z, Shen C, Hengel AD. Bridging category-level and instance-level semantic image segmentation. *arXiv*. (2016) [Preprint]. arXiv:1605.06885. doi: 10.48550/arXiv.1605.06885

42. Zhang Y, Meng Y, Zheng Y. Automatically segment the left atrium and scars from LGE-MRIs using a boundary-focused nnU-Net. In: *Challenge on Left Atrial and Scar Quantification and Segmentation*. Cham: Springer (2022), p. 49–59. doi: 10.1007/978-3-031-31778-1_5

43. Yang X, Wang N, Wang Y, Wang X, Nezafat R, Ni D, et al. Combating uncertainty with novel losses for automatic left atrium segmentation. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018), p. 246–54. doi: 10.1007/978-3-030-12029-0_27

44. Meng Y, Zhang H, Zhao Y, Yang X, Qiao Y, MacCormick IJ, et al. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE Trans Med Imaging*. (2021) 41:690–701. doi: 10.1109/TMI.2021.3123567

45. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2015), p. 234–41. doi: 10.1007/978-3-319-24574-4_28

46. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: a nested U-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer (2018), p. 3–11. doi: 10.1007/978-3-030-00889-5_1

47. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, QC: IEEE (2021), p. 10012–22. doi: 10.1109/ICCV48922.2021.00986

48. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-net: learning where to look for the pancreas. *arXiv [Preprint]*. arXiv:1804.03999 (2018).

49. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: transformers make strong encoders for medical image segmentation. *arXiv [Preprint]*. arXiv:2102.04306 (2021). doi: 10.48550/arXiv:2102.04306

50. Laibacher T, Weyde T, Jalali S. M2u-net: effective and efficient retinal vessel segmentation for real-world applications. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach, CA: IEEE (2019). doi: 10.1109/CVPRW.2019.00020

51. Hua D, Xu Y, Zeng X, Yang N, Jiang M, Zhang X, et al. Use of optical coherence tomography angiography for assessment of microvascular changes in the macula and optic nerve head in hypertensive patients without hypertensive retinopathy. *Microvasc Res*. (2020) 129:103969. doi: 10.1016/j.mvr.2019.103969

52. Irshad S, Akram MU. Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. In: *2014 Cairo International Biomedical Engineering Conference (CIBEC)*. Giza: IEEE (2014), p. 133–6. doi: 10.1109/CIBEC.2014.7020937

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 19 June 2024
DOI 10.3389/fmed.2024.1372091

Check for updates

OPEN ACCESS

EDITED BY
Haoyu Chen,
The Chinese University of Hong Kong, China

REVIEWED BY
Fei Shi,
Soochow University, China
Songtao Yuan,
Nanjing Medical University, China

*CORRESPONDENCE
Biao Yan
✉ biao.yan@fdeent.org
Zhenhua Wang
✉ zh-wang@shou.edu.cn
Qin Jiang
✉ jiangqin710@126.com

†These authors have contributed equally to this work

RECEIVED 17 January 2024
ACCEPTED 21 May 2024
PUBLISHED 19 June 2024

CITATION
Li J, Ma Q, Yao M, Jiang Q, Wang Z and Yan B (2024) Segmentation of retinal microaneurysms in fluorescein fundus angiography images by a novel three-step model. *Front. Med.* 11:1372091. doi: 10.3389/fmed.2024.1372091

COPYRIGHT
© 2024 Li, Ma, Yao, Jiang, Wang and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Segmentation of retinal microaneurysms in fluorescein fundus angiography images by a novel three-step model

Jing Li[1,2†], Qian Ma[3†], Mudi Yao[4,5], Qin Jiang[4*], Zhenhua Wang[2*] and Biao Yan[1,5*]

[1]Eye Institute and Department of Ophthalmology, Eye and ENT Hospital, State Key Laboratory of Medical Neurobiology, Fudan University, Shanghai, China, [2]College of Information Science, Shanghai Ocean University, Shanghai, China, [3]Department of Ophthalmology, General Hospital of Ningxia Medical University, Ningxia, China, [4]Department of Ophthalmology and Optometry, The Affiliated Eye Hospital, Nanjing Medical University, Nanjing, China, [5]Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Introduction: Microaneurysms serve as early signs of diabetic retinopathy, and their accurate detection is critical for effective treatment. Due to their low contrast and similarity to retinal vessels, distinguishing microaneurysms from background noise and retinal vessels in fluorescein fundus angiography (FFA) images poses a significant challenge.

Methods: We present a model for automatic detection of microaneurysms. FFA images were pre-processed using Top-hat transformation, Gray-stretching, and Gaussian filter techniques to eliminate noise. The candidate microaneurysms were coarsely segmented using an improved matched filter algorithm. Real microaneurysms were segmented by a morphological strategy. To evaluate the segmentation performance, our proposed model was compared against other models, including Otsu's method, Region Growing, Global Threshold, Matched Filter, Fuzzy c-means, and K-means, using both self-constructed and publicly available datasets. Performance metrics such as accuracy, sensitivity, specificity, positive predictive value, and intersection-over-union were calculated.

Results: The proposed model outperforms other models in terms of accuracy, sensitivity, specificity, positive predictive value, and intersection-over-union. The segmentation results obtained with our model closely align with benchmark standard. Our model demonstrates significant advantages for microaneurysm segmentation in FFA images and holds promise for clinical application in the diagnosis of diabetic retinopathy.

Conclusion: The proposed model offers a robust and accurate approach to microaneurysm detection, outperforming existing methods and demonstrating potential for clinical application in the effective treatment of diabetic retinopathy.

KEYWORDS

diabetic retinopathy, segmentation model, microaneurysms, fluorescein fundus angiography, computer-aided diagnosis

1 Introduction

Diabetic retinopathy (DR) is known as a blinding eye disease in the working population. Most of the patients with type 1 diabetes mellitus and nearly 60% of the patients with type 2 diabetes mellitus will develop retinopathy following a long duration of diabetes (≥20 years). However, it is difficult to detect DR until it develops into the advanced vision-threatening stage (1). DR is often divided into two stages: non-proliferative DR (NPDR) and proliferative DR (PDR). In the NPDR stage, hyperglycemia can cause

serious injuries to retinal capillaries, which can weaken the capillary walls and lead to the occurrence of microaneurysms (MAs). MAs are the small outpouchings of retinal capillaries and the early signs of NPDR, as well as the indicators for DR progression (2, 3). MAs appear as small, reddish, and circular shapes in color fundus images. They can be clinically identified by ophthalmoscopy as the deep-red dots varying from 10 to 100 μm in diameter (4, 5). Thus, automatic detection of MAs is important for DR diagnosis, which can help in controlling and retarding visual loss.

Previous studies have reported that several imaging modalities have been developed for MA detection, including color fundus images (6), optical coherence tomography angiography (OCTA) (7), and fluorescein fundus angiography (FFA) (8). Colored fundus photography has often been used due to its low cost compared with Optical coherence tomography machines. Walter et al. proposed a method for the automatic detection of MAs based on diameter closure and kernel density estimation (6). Melo et al. proposed a method for MA detection using the sliding band filter algorithm in color fundus images (9). MAs are situated on retinal capillaries and are not often visible, which makes them difficult to distinguish from the noises and pigmentation variations in color fundus images. An OCTA can provide detailed visualization of vascular perfusions. However, optical coherence tomography (OCT) machines are very expensive, and the interpretation of OCTA data is still challenging due to the complicated image artifacts and elusive algorithmic details of OCTA data (10, 11). FFA can be used for the detection of small changes in retinal vessels. The small and leaky MAs are easily ignored without the aid of FFA. FFA is highly effective in detecting MAs, especially when MAs are close to the vessels or too small to distinguish (12, 13). However, objective segmentation of MAs in FFA images is still challenging because MA segmentation requires laborious manual segmentation by experienced graders. Therefore, it is necessary to develop a model for automatic detection of MAs in FFA images for DR diagnosis.

Computer-assisted MA detection is important for DR diagnosis. Baudoin et al. used a mathematical morphology method to remove vessels and applied a top-hat transformation with the linear structuring elements to detect MAs (14). Spencer et al. proposed an image correction procedure for MA segmentation by calculating the true- and false-positive rates (15). Mendonca et al. further improved this method by altering the pre-filtering and classification procedures. However, shade corrections may produce false positives caused by the darkening of regions close to the bright patterns (16). Walter applied mathematical morphology to segment the vascular trees of retinal angiograms. This algorithm can extract patterns if vein width is constant, but it cannot extract them from narrower/wider veins (17). Zhang et al. proposed a model based on the dynamic thresholding and correlation coefficients of a multi-scale Gaussian template (18). Antal and Hajdu proposed an ensemble-based method for MA detection by selecting an optimal combination of pre-processing methods and candidate extractors (19). Saleh et al. developed a DR detection system based on the Gaussian filter, a multi-layered dark object filtering method, and a singular spectrum analysis (20). Despite their clinical significance, MAs pose challenges for accurate detection due to their low-contrast and close resemblance to blood vessels. Thus, further study is necessary to refine MA detection algorithms and enhance accuracy, particularly in FFA images. In this study, we present a novel model for the automatic detection of MA lesions in FFA images. Our proposed model comprises pre-processing of FFA images, followed by coarse segmentation of candidate MA regions and fine segmentation of MA regions. Subsequently, comparative studies were conducted to assess the MA detection performance of the proposed model.

2 Materials and methods

2.1 The proposed model for MA detection

The flowchart of the proposed MA detection model is shown in Figure 1, including pre-processing of FFA images, coarse segmentation of candidate MA regions by the matched filter (MF) algorithm, and fine segmentation of MA regions by the morphological strategy.

2.2 Pre-processing of FFA images

High-noise and low-contrast can pose great difficulties for the identification of MAs in FFA images. In the pre-processing step, the FFA images underwent decomposition into individual channels to alleviate computational demands, given that the pixel values across each channel were identical. Subsequently, each single channel underwent processing, employing top-hat transformation and gray-stretching (21) to enhance the contrasts between MAs and the background. Following this processing, the processed result underwent further refinement via a Gaussian filter to reduce noise.

The top-hat transformation was defined according to Equation (1) (22):

$$I_{th}(x,y) = I(x,y) - I(x,y) \circ B(u,v) \tag{1}$$

where $I(x,y)$ refers to the grayscale image, $B(u,v)$ refers to the structural element constructed as a circle with a radius of 45 pixels, and $\circ$ refers to the open operation. Opening of $I(x,y)$ by $B(u,v)$ was defined according to Equation (2):

$$I(x,y) \circ B(u,v) = (I(x,y) \ominus B(u,v)) \oplus B(u,v) \tag{2}$$

where $\ominus$ and $\oplus$ refer to the erosion and dilation operations, respectively. The erosion and dilation of $I(x,y)$ by $B(u,v)$ were defined according to Equations (3) and (4):

$$I(x,y) \ominus B(u,v) = \min_{u,v} (I(x+u, y+v) - B(u,v)) \tag{3}$$

$$I(x,y) \oplus B(u,v) = \min_{u,v} (I(x-u, y-v) + B(u,v)) \tag{4}$$

Gray-stretching was defined according to Equation (5) (23):

$$I_{new} = \left(\frac{G_{\max} - G_{\min}}{I_{\max} - I_{\min}}\right)(I - I_{\min}) + G_{\min} \tag{5}$$

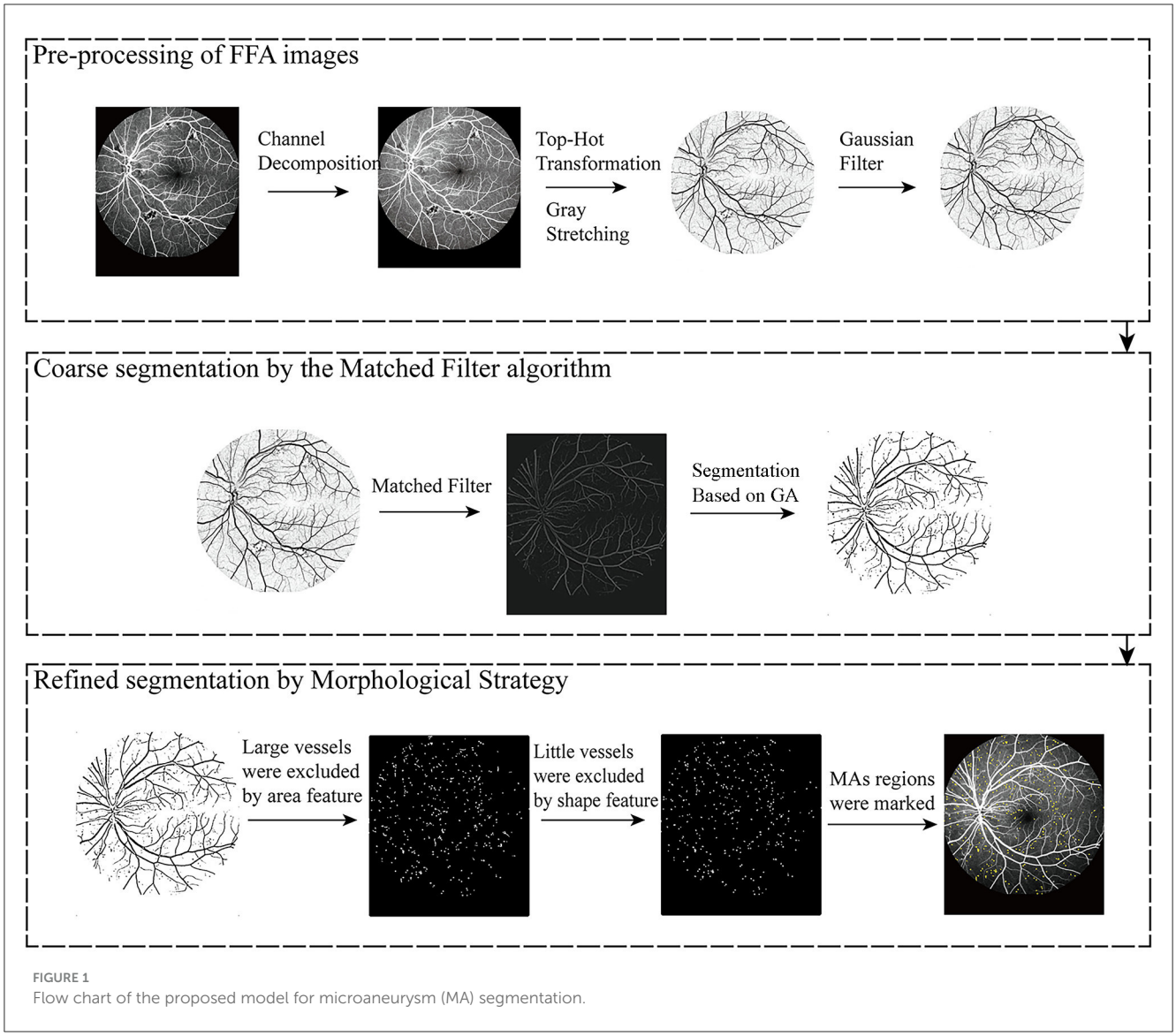

FIGURE 1
Flow chart of the proposed model for microaneurysm (MA) segmentation.

where I_{max} and I_{min} refer to the largest and smallest gray values in the original images, respectively. G_{max} and G_{min} refer to the largest and smallest gray values in the transformed images.

The Gaussian filter was defined according to Equation (6) (24):

$$G(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{6}$$

where σ^2 refers to the variance of the Gaussian filter.

In the pre-processing step, top-hat transformation, gray-stretching, and a Gaussian filter were employed for MA extraction by strengthening, enhancing, and denoising. A top-hat transformation was used to highlight the object edges and remove distracting information such as background noises. Gray-stretching mapped the grayscale ranges of FFA images. The Gaussian filter smoothed FFA images and removed irregular details such as noise points and burrs in the FFA images.

2.3 Coarse segmentation of MAs by the MF algorithm

The candidate MA regions in the FFA images were detected using the MF algorithm. MF was initially proposed by Chaudhuri et al. (25) for blood vessel extraction. Analogous to the matching filter concept in signal processing, a blood vessel image can be interpreted as a signal. Blood vessels exhibit characteristics such as a narrow range of width variation and parallel inner walls. Based on the prior knowledge, MF can construct a template to match the cross-sectional structure of blood vessels. Consequently, when the blood vessel component is input, a higher value is yielded, whereas a lower value is produced for the background, facilitating the separation of blood vessels. Hence, MF effectively enhances blood vessels and suppresses background noises.

MF was defined according to Equation (7) (26):

$$f(x,y) = \frac{1}{\sqrt{2\pi s^2}} e^{\frac{-x^2}{2s^2}} - m, \ |x| \le t \times s, |y| \le \frac{L}{2} \tag{7}$$

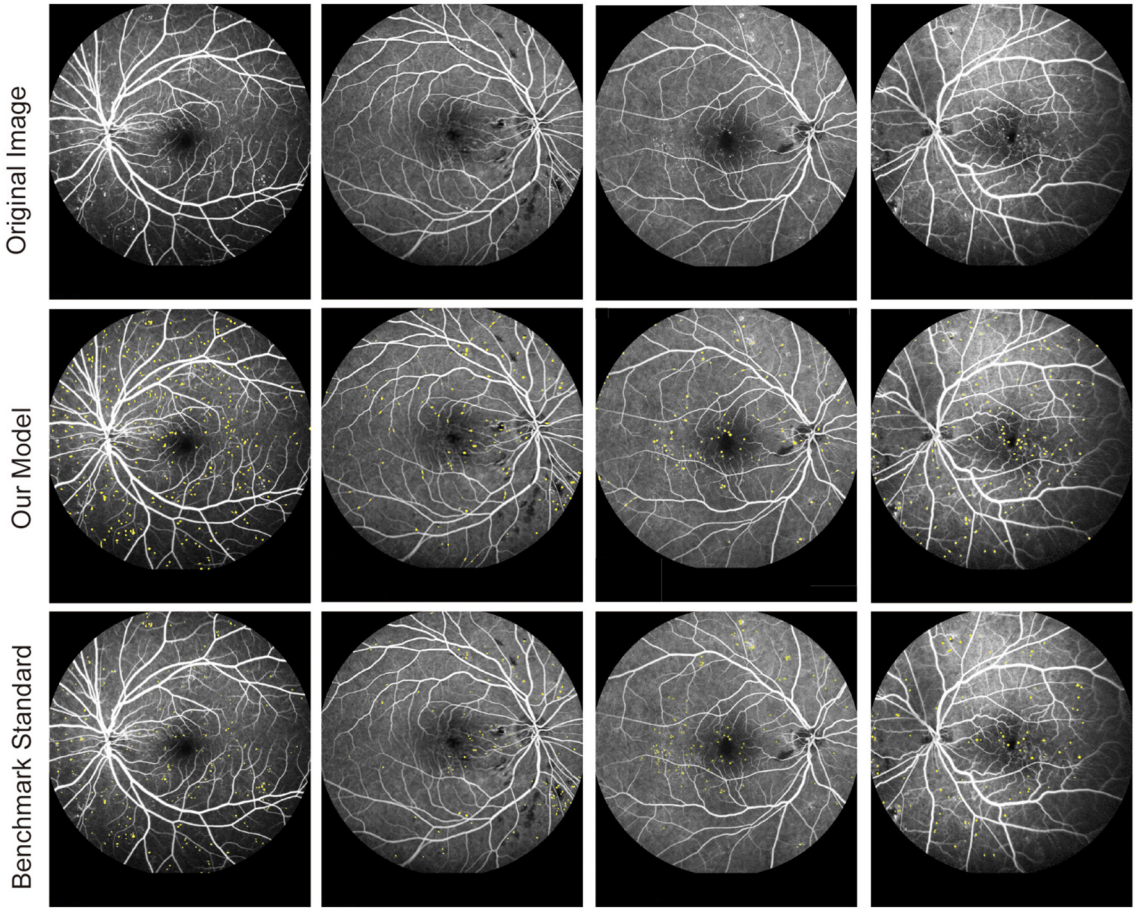

FIGURE 2
Original fluorescein fundus angiography (FFA) images and segmentation results of MAs by the proposed model and retinal clinicians.

where s refers to the filter scale, m is used for normalizing the mean value of the filter to 0, which is defined as Equation (8), L refers to the neighborhood length along the y-axis and is used to smooth the noises. L was deduced by s. When s was small, L was set relatively small, and vice versa. The criterion t is a constant and was set to 3 (27).

$$m = \frac{\int_{-ts}^{ts} \frac{1}{\sqrt{2\pi s^2}} e^{\frac{-x^2}{2s^2}} dx}{2ts} \tag{8}$$

The performance of the MF algorithm is heavily reliant on the design of the template. Poorly designed templates or significant deviations from the actual blood vessel structure can result in inaccurate extraction or an abundance of noise. Genetic algorithms (GA), an optimization technique introduced by John Holland, offer a solution to this challenge. GA mimics natural selection and genetic mechanisms to search for optimal solutions within the solution space. By using GA, one can efficiently explore and identify template configurations that yield improved accuracy and robustness in vessel extraction.

Hence, GA can be utilized to automatically adjust the threshold value of MF to accommodate the morphological features of blood vessels in various images. The GA process comprises five key steps: population initialization, fitness assessment, selection, crossover, and mutation. In the population initialization step, chromosome length was set to 8 and population size was set to 10. In the fitness assessment step, the efficacy of a solution was determined using a fitness function, where solutions with higher fitness were deemed superior. In our study, the fitness function of the GA is defined as in Equation (9) (28). In the selection step, the elitism strategy was adopted. In the crossover step, the crossover probability was set to 0.7. In the mutation step, the mutation probability was set to 0.4. In the later stages of the genetic algorithm's evolution, adjustments were made to both the crossover and mutation probabilities, setting them to 0.3 each.

Through iterative optimization via GA, the MF template that most accurately aligns with blood vessels can be gradually identified, enabling the identification of all candidate MAs.

$$f = p_1 \times p_2 \times (\mu_1 - \mu_2)^2 \tag{9}$$

where p_1 and p_2 refer to the number of the target pixels and background pixels, respectively, μ_1 and μ_2 refer to the average gray values of the target pixels and background pixels, respectively. f is the fitness value.

2.4 Fine segmentation of MAs by the morphological strategy

Real MA regions were determined by the morphological strategy, including removing vessels, hemorrhages, and exudates from the candidate MA regions based on area features and shape features, respectively. Previous studies have developed multiple image processing and machine learning algorithms for the automatic detection of MAs and recognized that the area size of MAs was typically between 5 and 100 pixels. In addition, real MAs were often localized next to the capillaries, appearing as dotted or rounded structures (29–31). The vessels, hemorrhages, and exudates were removed from the candidate MA regions according to Equation (10). Hemorrhages and exudates caused by the injured vessels were removed from the candidate MA regions according to Equation (11) and the threshold for roundness was set to 0.51.

$$I(x,y) = \begin{cases} 0, & S > 100 \\ 1, & 5 \le S \le 100 \\ 0, & S < 5 \end{cases} \tag{10}$$

$$Roundess = \frac{4\pi S}{C^2} \tag{11}$$

where S refers to the pixels of the candidate MA regions and C refers to the circumference of the contour.

2.5 Dataset

The FFA dataset comprises 1,010 FFA images, each with dimensions of 768 × 868 pixels, obtained from 65 eyes of 60 DR patients aged between 31 and 81 years. These patients underwent FFA examinations at the Eye Hospital affiliated with Nanjing Medical University between 2015 and 2019. The FFA images were captured using Heidelberg Retina Angiography (Heidelberg Engineering, Germany) by experienced clinicians. Notably, the FFA dataset did not include blurry or overexposed images. For labeling MAs in FFA images, three retinal clinicians with over 10 years of experience independently annotated MAs, serving as the benchmark standard. Patients with FFAs indicating mild or moderate DR were eligible for inclusion. The following exclusion criteria were used: (1) presence of other ocular diseases unrelated to diabetes, such as retinal arteriovenous obstruction, age-related macular degeneration, glaucoma, and uveitis; (2) any condition causing poor image quality or inability to visualize the optic disc and vessels, such as dense cataracts or corneal opacity; and (3) history of previous ophthalmological interventions, such as laser photocoagulation, vitrectomy, or anti-vascular endothelial growth factor injection. To ensure the reliability and validity of segmentation results, FFA images were independently divided into three sets: 830 images for training, 90 images for testing, and 90 images for validation. Figure 2 shows the original FFA images and MA detection results by the proposed model and benchmark standard.

Another publicly available dataset was utilized to assess the performance of MA detection. This dataset consisted of FFA images obtained from diabetic patients. The images were captured as part of a study conducted at the Persian Eye Clinic (Feiz Hospital), affiliated with the Isfahan University of Medical Sciences. The dataset comprised retinal images from 70 patients, with 30 samples categorized as normal and 40 samples representing various stages of DR.

2.6 Evaluation metrics

Five different metrics, including accuracy (Acc) (30), sensitivity (Se) (30), specificity (Sp) (30), positive predictive value (PPV) (31), and intersection-over-union (IOU) (32), were employed to evaluate the detection performance of MAs according to Equations (12–16):

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$Se = \frac{TP}{TP + FN} \tag{13}$$

$$Sp = \frac{TN}{TN + FP} \tag{14}$$

$$PPV = \frac{TP}{TP + FP} \tag{15}$$

$$IOU = \frac{TP}{TP + FP + FN} \tag{16}$$

where TP denotes the region that was predicted as MAs and was real MAs; FP denotes the region that was predicted as MAs but was background; TN denotes the region that was predicted as background and was real background; and FN denotes the region that was predicted as background but was MAs. Accuracy (Acc) is defined as the measure providing the ratio of total well-segmented pixels based on the gold standard for hand-labeled detection. Sensitivity (Se) and specificity (Sp) measures the ability of the model to detect well-segmented MAs and background pixels, respectively. PPV represents the correct proportion of the sample with a positive prediction. The IOU reflects the degree of coincidence between the MA detection result of the proposed model and the benchmark standard.

2.7 Implementation

All experiments were conducted on a PC with an Intel Core processor running at 2.50 GHz and equipped with 8 GB of RAM, using the MATLAB 2013a software.

3 Results

Our proposed model encompassed the pre-processing of FFA images, followed by coarse segmentation of candidate MA

TABLE 1 Performance comparison between our proposed model and the microaneurysms-matched filter (MAs-MF) model.

Model	Evaluation metrics				
	Acc (%)	Se (%)	Sp (%)	PPV (%)	IOU (%)
Clinician	99.94 ± 0.04	96.65 ± 0.08	99.96 ± 0.02	92.91 ± 0.09	90.02 ± 0.08
MAs-MF	99.43 ± 0.06	90.95 ± 0.46	99.46 ± 0.05	42.42 ± 0.97	40.64 ± 1.07
Our model	99.80 ± 0.05	92.10 ± 0.20	99.85 ± 0.04	75.07 ± 0.44	70.57 ± 0.55

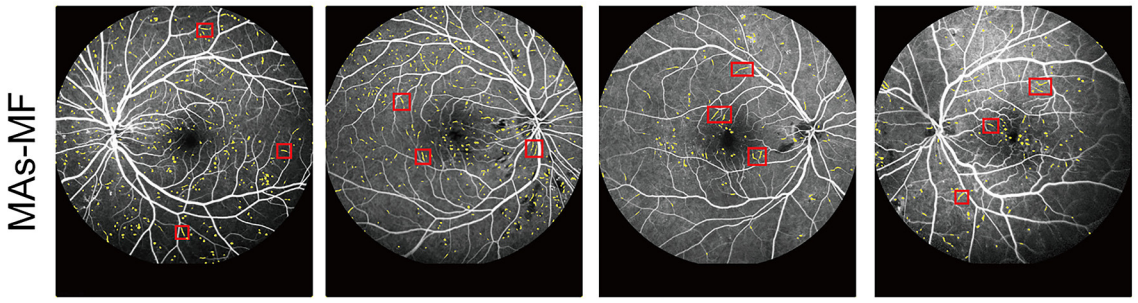

FIGURE 3
Detection results of MAs by microaneurysms-matched filter (MAs-MF).

regions and fine segmentation of MA regions. To assess the MA detection performance of our proposed model, two distinct experiments were conducted. In Experiment 1, our proposed model was juxtaposed against the MF model optimized by the GA algorithm (referred to as MAs-MF). In Experiment 2, our proposed model was compared against previous MA detection models. To maintain the integrity of our experiments, the outcomes presented for the clinician in Table 1 were segmented by a skilled clinician who did not participate in the dataset labeling process.

3.1 The ablation experiment suggests that our proposed model improves MA detection performance

We compared our proposed model against the MAs-MF model to evaluate MA detection performance. The results of MA detection are shown in Figure 3. The metrics of MA detection are shown in Table 1.

From Figure 3 and Table 1, we can observe that there were several label errors of small blood vessels for MA detection results in the MAs-MF model, as shown in the red squares in Figure 3. Compared with the MAs-MF model, the MA detection performance of the proposed model was close to the MA detection results of the clinicians. Compared with the MAs-MF model, the proposed model had greater values of accuracy (Acc), sensitivity (Se), specificity (Sp), PPV, and IOU, which were 99.80 (0.37↑), 92.10 (1.15↑), 99.85 (0.39↑), 75.07 (32.65↑), and 75.57 (29.93↑), respectively.

3.2 The comparison experiment suggests that the proposed model has an obvious MA detection advantage over previous MA detection models

We further compared our proposed model against other MA detection models, such as Otsu's method (33), Region Growing (34), MF (25), Global Threshold (35), K-means, (36) and Fuzzy c-means, (37) to evaluate MA detection performance. The results of MA detection are shown in Figure 4, and the metrics of MA evaluation are shown in Table 2.

As shown in Figure 3 and Table 2, Otsu's method, Region Growing, MF, Global Threshold, K-means, and Fuzzy c-means models could not accurately detect the boundaries of MA regions and normal regions. Additionally, there were some omissions and false detections, which are marked by red squares in Figure 4. The proposed model had greater values of PPV and IOU than other models. Furthermore, our proposed model demonstrated performance in MA detection that closely aligned with the benchmark standard, surpassing the performance of other MA detection models.

To further evaluate MA detection performance, we used a publicly available dataset, which was obtained during a study conducted at the Persian Eye Clinic (Feiz Hospital) in Isfahan University of Medical Sciences (32), including retinal images from 70 patients, with 30 samples classified as normal and 40 samples representing different stages of DR. As shown in Table 3, MA detection using our proposed model had an average accuracy of 99.42%, a sensitivity of 90.21%, a specificity of 98.86%, a PPV of 71.93%, and an IOU of 64.89%, showing an obvious advantage over other MA detection models.

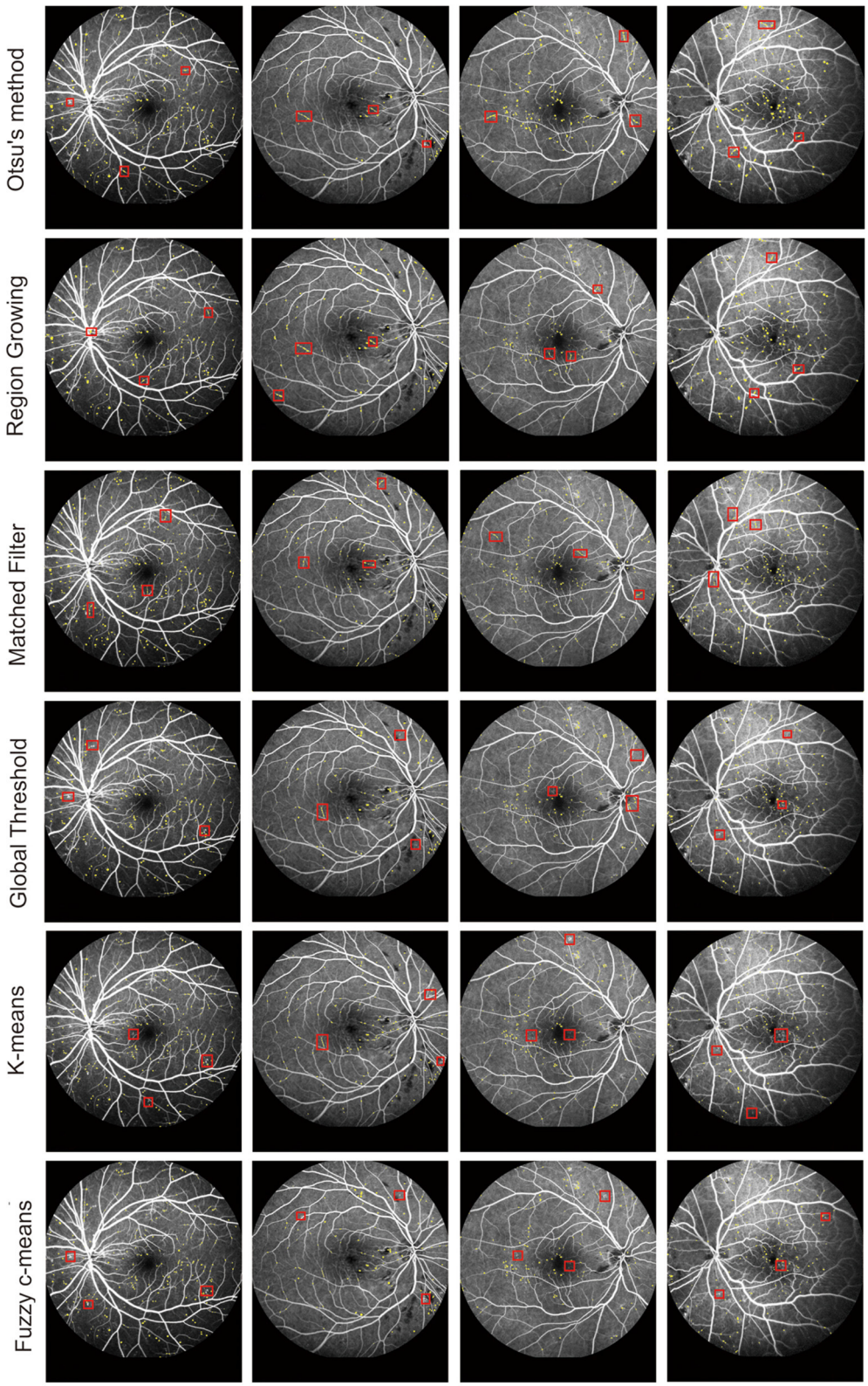

FIGURE 4
MA segmentation results from different models.

TABLE 2 Comparison of microaneurysm (MA) segmentation performance between our proposed model and other previously reported models.

Model	Evaluation metrics				
	Acc (%)	Se (%)	Sp (%)	PPV (%)	IOU (%)
Otsu's method	99.71 ± 0.08	79.55 ± 0.97	99.80 ± 0.08	63.85 ± 0.85	54.43 ± 1.11
Region growing	99.73 ± 0.07	81.80 ± 0.89	99.80 ± 0.07	63.92 ± 0.82	55.42 ± 1.07
Matched filter	99.73 ± 0.07	84.60 ± 0.57	99.79 ± 0.06	63.57 ± 0.80	57.40 ± 0.99
Global threshold	99.73 ± 0.08	78.06 ± 0.75	99.82 ± 0.06	62.88 ± 0.88	53.43 ± 1.12
K-means	99.76 ± 0.07	80.39 ± 0.68	99.83 ± 0.05	60.27 ± 0.75	52.36 ± 0.98
Fuzzy c-means	99.74 ± 0.06	76.52 ± 0.56	99.84 ± 0.05	63.23 ± 0.57	52.71 ± 0.92
Our model	99.80 ± 0.05	92.10 ± 0.20	99.85 ± 0.04	75.07 ± 0.44	70.57 ± 0.55

TABLE 3 Comparison of MA detection performance between our proposed model and previous detection models using the publicly available dataset.

Model	Evaluation metrics				
	Acc (%)	Se (%)	Sp (%)	PPV (%)	IOU (%)
Otsu's method	95.37 ± 0.12	79.76 ± 0.97	97.43 ± 0.16	64.43 ± 0.95	55.83 ± 1.05
Region growing	98.78 ± 0.11	82.22 ± 0.29	98.75 ± 0.21	67.21 ± 0.65	56.87 ± 1.21
Matched filter	98.54 ± 0.15	83.76 ± 0.29	98.46 ± 0.11	66.45 ± 0.86	56.87 ± 0.76
Global threshold	98.76 ± 0.09	79.12 ± 0.69	99.29 ± 0.16	64.64 ± 0.72	55.65 ± 1.08
K-means	98.55 ± 0.12	81.43 ± 0.87	99.12 ± 0.21	63.32 ± 0.86	54.65 ± 0.91
Fuzzy c-means	97.87 ± 0.11	79.47 ± 0.72	98.54 ± 0.13	64.54 ± 0.75	53.81 ± 0.87
Our model	99.42 ± 0.35	90.21 ± 0.54	98.86 ± 0.12	71.93 ± 0.41	64.89 ± 0.35

4 Discussion

MA detection is highly important for the diagnosis of DR (5). FFA is a technique used for the evaluation of retinal and choroidal circulation. MAs are immediately visible following the arterial phase of FFA (33). In this study, we propose a three-step model for MA detection in FFA images. Initially, FFA image pre-processing is conducted to enhance the contrasts of FFA images. Subsequently, candidate MA regions are coarsely segmented using an improved MF algorithm. Finally, real MA regions are identified through a morphological strategy. This proposed model aims to enhance the accuracy and efficiency of MA detection in FFA images, thus aiding in the early diagnosis and management of DR.

Automatic segmentation of MAs is still a tricky problem due to their tiny sizes, low contrasts, and high similarities to retinal vessels. The high-noise and low-contrast of FFA images can also affect the quality of FFA images and reduce the accuracy of MA detection (33). The goal of image enhancement is to decrease image noise and enhance the contrasts of the targets and backgrounds. In this study, top-hat transformation, gray-stretching, and a Gaussian filter were used for the improvement of FFA image quality. Top-hat transformation and gray-stretching can efficiently solve the problem of uneven illumination, while a Gaussian filter can efficiently reduce the potential impacts of retinal noises on FFA images.

We also evaluated the MA detection performance of the proposed model by comparing it with other MA detection methods. Compared with Otsu's method, Region Growing, MF, K-means, Global Threshold, and Fuzzy c-means (3, 25, 34–37), the proposed model has the greatest accuracy and efficiency for MA detection in FFA images. The evaluation metrics of the proposed model, including accuracy, sensitivity, specificity, PPV, and IOU, have the highest value. Moreover, the proposed model has a similar MA detection performance as the clinicians.

Recently, deep learning-based algorithms have gained popularity for medical image analysis. However, these algorithms typically demand high-performance computing resources, such as central processing units (CPUs) and graphics processing units (GPUs), as well as a substantial amount of labeled data for training. Unfortunately, many hospitals lack access to such resources and specialized personnel (38). Given this context, there is a pressing need for simpler methods for analyzing FFA images. In contrast to deep learning-based approaches, the proposed model does not necessitate a large number of labeled images or high-performance computing resources. Moreover, it offers comparable accuracy to manual labeling by clinicians but with faster detection speed. This feature makes it a practical and efficient solution for MA detection in clinical settings where resources and expertise may be limited.

5 Conclusion

This study provides a new model for the detection of MAs in FFA images, which consists of three steps. First, the quality of FFA images was improved by the image enhancement methods, including top-hat transformation, gray-stretching, and

the Gaussian filter. Then, the candidate MAs were coarsely segmented by the MF algorithm. Finally, real MA regions were determined by the morphological strategy. Compared with manual MA labeling or other existing MA detection algorithms, the proposed model shows promising performance for the early diagnosis of DR by detecting MA lesions. This model is expected to assist ophthalmologists in efficiently detecting MA lesions, thereby enhancing the overall efficiency of DR diagnosis.

6 Limitations of this study

The number of MAs tends to increase as the severity of DR worsens. While the proposed model effectively detects the presence of MA lesions in FFA images, there are limitations to its clinical application. Indeed, MA formation is associated with various pathological changes such as basement membrane thickening, pericyte degeneration, and endothelial injury, which can lead to retinal vessel leakage, edema, and even hemorrhage. Given that vessel leakage, edema, and hemorrhage are closely linked to the size and volume of MAs, accurately detecting these parameters can provide additional valuable information for DR screening and monitoring. To achieve broader clinical applicability, the proposed model should be integrated with algorithms for detecting the size and volume of MAs, as well as for identifying edema and hemorrhages. This enhanced model would significantly improve the accuracy of assessing DR severity and estimating DR risk. Due to the high variability in pathological features and the quality of FFA images, deep learning techniques could play a crucial role in detecting and quantifying these features more accurately and efficiently. Therefore, in the future, we plan to incorporate deep learning approaches to further enhance the efficiency of MA detection in FFA images and improve the overall diagnostic capabilities for DR.

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the (patients/participants OR patients/participants legal guardian/next of kin) was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

QJ: Data curation, Writing – original draft, Writing – review & editing. JL: Conceptualization, Data curation, Writing – original draft. QM: Data curation, Formal analysis, Writing – original draft. MY: Conceptualization, Formal analysis, Writing – original draft. ZW: Conceptualization, Software, Writing – original draft. BY: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the grants from the National Natural Science Foundation of China (Grant No. 82171074).

Conflict of interest

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

The reviewer SY declared a shared parent affiliation with the author QJ to the handling editor at the time of review.

Publisher's note

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

References

1. Jin K, Pan X, You K, Wu J, Liu Z, Cao J, et al. Automatic detection of non-perfusion areas in diabetic macular edema from fundus fluorescein angiography for decision making using deep learning. *Sci Rep.* (2020) 10:1–7. doi: 10.1038/s41598-020-71622-6

2. Saleh MD, Eswaran C. An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection. *Comput Methods Progr Biomed.* (2012) 108:186–96. doi: 10.1016/j.cmpb.2012.03.004

3. Chudzik P, Majumdar S, Calivá F, Al-Diri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. *Comput. Methods Progr. Biomed.* (2018) 158:185–92. doi: 10.1016/j.cmpb.2018.02.016

4. Sehirli E, Turan MK, Dietzel A. Automatic detection of microaneurysms in rgb retinal fundus images. *Int J Sci Technol Res.* (2015) 1:1–7.

5. Ganjee R, Azmi R, Ebrahimi Moghadam M. A novel microaneurysms detection method based on local applying of Markov random field. *J Med Syst.* (2016) 40:1–9. doi: 10.1007/s10916-016-0434-4

6. Walter T, Massin P, Erginay A, Ordonez R, Jeulin C, Klein JC. Automatic detection of microaneurysms in color fundus images. *Med Image Anal.* (2007) 11:555–66. doi: 10.1016/j.media.2007.05.001

7. Schreur V, Domanian A, Liefers B, Venhuizen FG, Klevering BJ, Hoyng CB, et al. Morphological and topographical appearance of microaneurysms on

optical coherence tomography angiography. *Br J Ophthalmol.* (2019) 103:630–5. doi: 10.1136/bjophthalmol-2018-312258

8. Tavakoli M, Shahri RP, Pourreza H, Mehdizadeh A, Banaee T, Toosi MHB. A complementary method for automated detection of microaneurysms in fluorescein angiography fundus images to assess diabetic retinopathy. *Pattern Recognit.* (2013) 46:2740–53. doi: 10.1016/j.patcog.2013.03.011

9. Melo T, Mendonça AM, Campilho A. Microaneurysm detection in color eye fundus images for diabetic retinopathy screening. *Comput Biol Med.* (2020) 126:103995. doi: 10.1016/j.compbiomed.2020.103995

10. Xu J, Song S, Li Y, Wang RK. Complex-based OCT angiography algorithm recovers microvascular information better than amplitude-or phase-based algorithms in phase-stable systems. *Phys Med Biol.* (2017) 63:015023. doi: 10.1088/1361-6560/aa94bc

11. De Carlo TE, Romano A, Waheed NK, Duker JS. A review of optical coherence tomography angiography (OCTA). *Int J Retina Vitreous.* (2015) 1:1–15. doi: 10.1186/s40942-015-0005-8

12. Cheung CMG, Yanagi Y, Mohla A, Lee SY, Mathur R, Chan CM, et al. Characterization and differentiation of polypoidal choroidal vasculopathy using swept source optical coherence tomography angiography. *Retina.* (2017) 37:1464–74. doi: 10.1097/IAE.0000000000001391

13. Kwan CC, Fawzi AA. Imaging and biomarkers in diabetic macular edema and diabetic retinopathy. *Curr Diabetes Rep.* (2019) 19:1–10. doi: 10.1007/s11892-019-1226-2

14. Baudoin CE, Lay BJ, Klein JC. Klein, Automatic detection of microaneurysms in diabetic fluorescein angiography. *Rev Epidemiol Sante Publique.* (1984) 32:254–61.

15. Spencer T, Olson JA, McHardy KC, Sharp PF, Forrester JV. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput Biomed Res.* (1996) 29:284–302. doi: 10.1006/cbmr.1996.0021

16. Mendonca AM, Campilho A, Nunes J. Automatic segmentation of microaneurysms in retinal angiograms of diabetic patients. In: *Proceedings 10th International Conference on Image Analysis and Processing*. Venice: IEEE (1999). p. 728–33.

17. Walter T. Automatic segmentation and registration of retinal fluorescein angiographies-application to diabetic retinopathy. In: *First International Workshop on Computer Assisted Fundus Image Analysis*. Copenhagen (2000). p. 15–20.

18. Zhang B, Wu X, You J, Li Q, Karray F. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recognit.* (2010) 43:2237–48. doi: 10.1016/j.patcog.2009.12.017

19. Antal B, Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans Biomed Eng.* (2012) 59:1720–6. doi: 10.1109/TBME.2012.2193126

20. Saleh GM, Wawrzynski J, Caputo S, Peto T, Al Turk LI, Wang S, et al. An automated detection system for microaneurysms that is effective across different racial groups. *J Ophthalmol.* (2016) 2016:4176547. doi: 10.1155/2016/4176547

21. Li K, Qi X, Luo Y, Yao Z, Zhou X, Sun M. Accurate retinal vessel segmentation in color fundus images via fully attention-based networks. *IEEE J Biomed Health Inform.* (2020) 25:2071–81. doi: 10.1109/JBHI.2020.3028180

22. Román JCM, Escobar R, Martínez F, Noguera JLV, Legal-Ayala H, Pinto-Roa DP. Medical image enhancement with brightness and detail preserving using multiscale top-hat transform by reconstruction. *Electron Notes Theor Comput.* (2020) 349:69–80. doi: 10.1016/j.entcs.2020.02.013

23. Liu L, Yang N, Lan J, Li J. Image segmentation based on gray stretch and threshold algorithm. *Optik.* (2015) 126:626–9. doi: 10.1016/j.ijleo.2015.01.033

24. Nasor M, Obaid W. Segmentation of osteosarcoma in MRI images by K-means clustering, Chan-Vese segmentation, and iterative Gaussian filtering. *IET Image Proc.* (2021) 15:1310–8. doi: 10.1049/ipr2.12106

25. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans Med Imag.* (1989) 8:263–9. doi: 10.1109/42.34715

26. Saroj SK, Kumar R, Singh NP. Frechet PDF based matched filter approach for retinal blood vessels segmentation. *Comput Methods Progr Biomed.* (2020) 194:105490. doi: 10.1016/j.cmpb.2020.105490

27. Zhang B, Zhang L, Zhang L, Karray F. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Comput Biol Med.* (2010) 40:438–45. doi: 10.1016/j.compbiomed.2010.02.008

28. Wang H, Zhang L, Yao L. Application of genetic algorithm based support vector machine in selection of new EEG rhythms for drowsiness detection. *Expert Syst Appl.* (2021) 171:114634. doi: 10.1016/j.eswa.2021.114634

29. Li X, Xie J, Zhang L, Cui Y, Zhang G, Wang J, et al. Differential distribution of manifest lesions in diabetic retinopathy by fundus fluorescein angiography and fundus photography. *BMC Ophthalmol.* (2020) 20:1–8. doi: 10.1186/s12886-020-01740-2

30. Wu B, Zhu W, Shi F, Zhu S, Chen X. Automatic detection of microaneurysms in retinal fundus images. *Comput Med Imag Graph.* (2017) 55:106–12. doi: 10.1016/j.compmedimag.2016.08.001

31. Dashtbozorg B, Zhang J, Huang F, ter Haar Romeny BM. Retinal microaneurysms detection using local convergence index features. *IEEE Trans Image Process.* (2018) 27:3300–15. doi: 10.1109/TIP.2018.2815345

32. Hajeb Mohammad Alipour S, Rabbani H, Akhlaghi M. A new combined method based on curvelet transform and morphological operators for automatic detection of foveal avascular zone. *SIViP.* (2014) 8:205–22. doi: 10.1007/s11760-013-0530-6

33. Otsu N. A threshold selection method from gray-level histograms. *Autom.* (1975) 11:23–7.

34. Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, et al. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Med.* (2002) 19:105–12. doi: 10.1046/j.1464-5491.2002.00613.x

35. Jang JW, Lee S, Hwang HJ, Baek KR. Global thresholding algorithm based on boundary selection. In: *Proceedings 13th International Conference on Control, Automation and Systems*. Gwangju: IEEE (2013). p. 704–6.

36. Zheng X, Lei Q, Yao R, Gong Y, Yin Q. Image segmentation based on adaptive K-means algorithm. *EURASIP J Imag Video Process.* (2018) 2018:1–10. doi: 10.1186/s13640-018-0309-3

37. Ghosh S, Dubey SK. Comparative analysis of k-means and fuzzy c-means algorithms. *Int J Adv Comput Sci Appl.* (2013) 4:35–39. doi: 10.14569/IJACSA.2013.040406

38. Wang Z, Zhang W, Sun Y, Yao M, Yan B. Detection of diabetic macular edema in Optical Coherence Tomography image using an improved level set algorithm. *Biomed Res Int.* (2020) 2020:6974215. doi: 10.1155/2020/6974215

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 16 July 2024
DOI 10.3389/fmed.2024.1421439

Check for updates

OPEN ACCESS

EDITED BY
Yitian Zhao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Dongxu Gao,
University of Portsmouth, United Kingdom
Xujiong Ye,
University of Lincoln, United Kingdom
Gilbert Yong San Lim,
SingHealth, Singapore

*CORRESPONDENCE
Yalin Zheng
✉ yalin.zheng@liverpool.ac.uk

RECEIVED 22 April 2024
ACCEPTED 25 June 2024
PUBLISHED 16 July 2024

CITATION
Chen X, Fan X, Meng Y and Zheng Y (2024) AI-driven generalized polynomial transformation models for unsupervised fundus image registration.
Front. Med. 11:1421439.
doi: 10.3389/fmed.2024.1421439

COPYRIGHT
© 2024 Chen, Fan, Meng and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

AI-driven generalized polynomial transformation models for unsupervised fundus image registration

Xu Chen[1], Xiaochen Fan[2], Yanda Meng[3] and Yalin Zheng[4,5]*

[1]Department of Medicine, University of Cambridge, Cambridge, United Kingdom, [2]Institute of Ophthalmology, University College London, London, United Kingdom, [3]Department of Computer Science, University of Exeter, Exeter, United Kingdom, [4]Department of Eye and Vision Sciences, University of Liverpool, Liverpool, United Kingdom, [5]Liverpool Centre for Cardiovascular Science, Liverpool, United Kingdom

We introduce a novel AI-driven approach to unsupervised fundus image registration utilizing our Generalized Polynomial Transformation (GPT) model. Through the GPT, we establish a foundational model capable of simulating diverse polynomial transformations, trained on a large synthetic dataset to encompass a broad range of transformation scenarios. Additionally, our hybrid pre-processing strategy aims to streamline the learning process by offering model-focused input. We evaluated our model's effectiveness on the publicly available AREDS dataset by using standard metrics such as image-level and parameter-level analyzes. Linear regression analysis reveals an average Pearson correlation coefficient (R) of 0.9876 across all quadratic transformation parameters. Image-level evaluation, comprising qualitative and quantitative analyzes, showcases significant improvements in Structural Similarity Index (SSIM) and Normalized Cross Correlation (NCC) scores, indicating its robust performance. Notably, precise matching of the optic disc and vessel locations with minimal global distortion are observed. These findings underscore the potential of GPT-based approaches in image registration methodologies, promising advancements in diagnosis, treatment planning, and disease monitoring in ophthalmology and beyond.

KEYWORDS

image registration, unsupervised learning, polynomial transformation, foundational model, color fundus photography

1 Introduction

Image registration is an essential process in vision applications where multiple images obtained from different viewpoints or spaces, are aligned. In medical imaging, this technique holds significant importance, enabling the comparison and analysis of images to gain insights into structural changes, disease progression, and treatment efficacy. The primary objective of image registration is to align two images, denoted as a fixed image (target) F and a moving image (source) M, by establishing spatial correspondence within a shared coordinate system. In a simpler term, assuming x and y represent the column and row indices, image registration involves mapping a position (x, y) from M to a new warped/aligned image W at position $(u(x, y), v(x, y))$, where u and v denote different types of transformation functions. Image registration encompasses linear and non-linear transformations. Linear transformations involve global geometric adjustment of the

moving image, while non-linear transformations allow for local or regional deformations to the moving image. Linear transformations often serve as the prerequisite step for non-linear registration techniques by addressing global distortions from differing viewpoints, making them an essential component in the image registration pipeline. The most basic linear transformation type for image registration is translation, wherein u and v can be expressed in Equation (1):

$$u = x + t_x \quad \text{and} \quad v = y + t_y \tag{1}$$

Here, t_x and t_y represent the translation lengths along the respective axes. Affine transformation is a common linear technique employed in image registration to address distortions arising from non-ideal camera angles. Typically, an affine transformation encompasses four fundamental operations: rotation, translation, scaling, and shearing. The expressions for u and v in the context of affine transformation are given (see Equation 2):

$$u = a_{00}x + a_{01}y + t_x \quad \text{and} \quad v = a_{10}x + a_{11}y + t_y \tag{2}$$

where, a_{00}, a_{01}, a_{10}, a_{11}, t_x and t_y are the transformation parameters. The planetary of surfaces, parallelism and angles between lines are all preserved in affine transformation. Furthermore, projective transformation is a type of geometric transformation that maps points in one plane to another plane using a projective matrix. It involves transforming points in a two-dimensional space, such as an image, to another two-dimensional space, allowing for changes in perspective, rotation, skewing, and other distortions. The expressions for u and v in the context of projective transformation is given in Equation 3:

$$u = \frac{b_{00}x + b_{01}y + b_{02}}{b_{03}x + b_{04}y + c}, \quad v = \frac{b_{10}x + b_{11}y + b_{12}}{b_{13}x + b_{14}y + c} \tag{3}$$

where b_{00}-b_{14} are the projective transformation parameters; c represents the coefficient associated with the z-coordinate in homogeneous coordinates. It is commonly referred to as the projective invariant and is used to represent the translation component of the transformation. Projective transformations are frequently employed in retinal image registration and geometric correction (1, 2). Retinal image registration is crucial in the diagnosis of eye diseases as it enables the accurate assessment of disease-related features and progression. Fundus imaging, including color fundus photography, optical coherence tomography (OCT), fluorescein angiography and other advanced imaging modalities, provides essential visual information for the diagnosis and management of retinal diseases and systemic diseases (3–5). The registration of fundus images allows for the alignment and comparison of images over time, facilitating the identification of changes in related features such as drusen, geographic atrophy (GA), and choroidal neovascularization (CNV) (6, 7). Fundus image registration is particularly important in the context of multi-modal imaging, where the integration of different imaging modalities such as OCT and fluorescein angiography enhances the comprehensive assessment (4, 5). By registering fundus images with other imaging modalities, clinicians can obtain a more comprehensive understanding of the structural and functional changes, leading to improved diagnostic accuracy and prognostic evaluation (8, 9). Moreover, the application of advanced technologies such as deep learning has shown promise in leveraging fundus image registration for the differential diagnosis, as well as for the automated segmentation of related lesions such as GA (10–12). These technological advancements enable the precise analysis of fundus images, contributing to the development of prognostic biomarkers and the prediction of disease progression (13). Deep learning-based image registration has emerged as a promising approach, offering solutions for linear transformations using convolutional neural networks (CNNs). The Spatial Transformer Network (STN) was among the pioneering CNN-based methods, focusing on learning two-dimensional affine transformations for distorted MNIST digit classification through supervised learning. Miao et al. (14) introduced a supervised CNN approach to regress three-dimensional transformation matrices for affine registration of X-ray images, utilizing synthesized transformation parameters as ground truth. However, the reliance on labeled ground truth for supervised methods can be limiting, prompting the development of unsupervised models that do not require transformation ground truth. De Vos et al. (15) proposed an unsupervised Deep Learning Image Registration (DLIR) framework, enabling joint affine and nonlinear registration without the need for labeled ground truth. The affine transformation framework within DLIR employs a multi-stage approach tailored for multi-temporal image registration. Additionally, Chen et al. (16) proposed an unsupervised CNN approach focused on explicitly learning specific geometric transformation parameters such as translations, rotations, scaling, and shearing. Unlike traditional methods that regress affine transformation matrices, this approach targets individual transformation parameters, offering a tailored solution for affine registration tasks in multi-modality image registration scenarios.

Current limitations in deep learning-based models for image registration are: (1) While much attention has been devoted to affine transformation for linear registration in deep learning-based models, real-world scenarios often involve more complex distortions that may not be adequately addressed by affine transformation alone. Powerful and complex linear registration techniques, such as projective transformation or polynomial transformation, offer additional flexibility in capturing the intricacies of image distortions. Affine transformations, while effective for linear registration tasks, have limitations in capturing non-linear distortions or irregular deformations present in many medical imaging applications. By incorporating projective or polynomial transformations, which allow for non-linear and higher-order transformations, these techniques can better model the intricate variations and deformations encountered in medical images. This enhanced flexibility enables more accurate alignment and registration of images, leading to improved diagnostic and analytical outcomes. However, the exploration of these techniques in the context of deep learning-based image registration remains limited. (2) Lack of generalized models for image transformation: One significant limitation in the realm of deep learning-based image registration lies in the absence of generalized models capable of learning image transformations universally. Many existing models are meticulously designed for specific images and modalities, hindering their adaptability to a broader range of scenarios. This limitation restricts the scalability of these models,

making them less effective in scenarios where a diverse set of images or modalities is encountered. Consequently, the field faces challenges in achieving a more comprehensive and generalized approach to image transformation learning and expansion.

2 Methods

Addressing the constraints observed in existing deep learning-based fundus image registration models, we proposed a generalized model that introduces an unsupervised approach tailored specifically for quadratic transformations, the second degree of polynomial transformation. Polynomial transformation is a process in which the input features are transformed by using a polynomial function of a certain degree. The goal of polynomial transformation is to capture more complex relationships between the features and the target variable than a simple linear model would. It can be useful when the relationship between variables is curvilinear rather than linear. However, higher-degree polynomials can also lead to over-fitting, so the degree of the polynomial should be chosen carefully based on the characteristics of the data. Mathematically, the u and v for polynomial transformation can be defined in Equation 4:

$$u = \sum_{d=0}^{p} \sum_{d=0}^{p-i} a_d x^d y^d \quad \text{and} \quad v = \sum_{d=0}^{p} \sum_{d=0}^{p-i} b_d x^d y^d \tag{4}$$

where p is the degree of polynomial and a_d, b_d, are the transformation parameters. These transformations include linear ($p = 1$), quadratic ($p = 2$), cubic ($p = 3$), bi-quadratic ($p = 4$) and quintic ($p = 5$) ones as special cases. For this work, quadratic ($p = 2$) transformation is used and can be expressed as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{Q} \begin{bmatrix} x^2 & y^2 & xy & x & y & 1 \end{bmatrix}^T = \begin{bmatrix} q_{00} & q_{01} & q_{02} & q_{03} & q_{04} & q_{05} \\ q_{10} & q_{11} & q_{12} & q_{13} & q_{14} & q_{15} \end{bmatrix} \begin{bmatrix} x^2 & y^2 & xy & x & y & 1 \end{bmatrix}^T \tag{5}$$

where $\mathbf{Q}$ is the quadratic transformation matrix. For image registration tasks, quadratic transformation can be formulated as an energy minimization problem (see Equation 6):

$$\mathbf{Q}^* = \arg\max_{\mathbf{Q}} \left\{ \mathbf{Q} \mid \mathcal{S}(F, \mathbf{Q}M) \right\} \tag{6}$$

where $\mathcal{S}$ is the metrics to measure the similarity between a fixed image F and the warped image $\mathbf{Q}M$. Our model aims to optimize each individual transformation parameter q_{00} - q_{15}, instead of directly optimizing the transformation matrix $\mathbf{Q}$. In the following sections, we provide more details of our framework, highlighting its two distinct features: the Generalized Polynomial Transformation (GPT) model and an unsupervised GPT-based transformation model specifically tailored for fundus image registration. The overview of our proposed model is represented in Figure 1.

Firstly, we propose the GPT model, serving as a foundational model to emulate diverse polynomial transformations. To construct a synthetic dataset to acquire knowledge of the quadratic transformation, we randomly selected each q parameter from the raw $\mathbf{Q}$ and then generated a synthetically wrapped image by the new $\mathbf{Q}_s$ matrix according to Equation (5). More specifically, each $q \in Q_s$ is derived from the distribution of the corresponding $q \in Q$. For example, q_15 ranges from 651.0 to –278.0 across the non-hold out testing set, so the new q_15 is assigned a random value within this range. Given unlimited combination, we developed an "on-the-fly" synthetic dataset generation approach during training steps, continuously generating synthetic data until the model achieved full convergence. To achieve full convergence, the "on-the-fly" synthetic data generator continuously produces random parameters for each epoch. This process involves generating a diverse set of synthetic image pairs by applying various quadratic transformations. The synthetic data generator operates iteratively, introducing new transformation parameters in each epoch to ensure that the model is exposed to a wide range of transformation scenarios. The training continues until the evaluation accuracy on the validation dataset stabilizes, indicating that the model has effectively learned the transformation characteristics and can generalize unseen data. This dynamic approach helps prevent over-fitting and ensures robust performance by leveraging an ever-expanding dataset that reflects the complex nature of real-world transformations. This step offers the advantage of automatically generating ground truth data without the need for manual annotations. It enables GPT to investigate a broad spectrum of polynomial transformation scenarios, encompassing nearly all possible transformation combinations. This strategy allows its convolutional neurons to be activated appropriately when capturing relevant features for geometric transformation. Without employing the "on-the-fly" synthetic dataset generation approach, the convolutional neurons in the model might be influenced by potential biases arising from a limited number of training samples. This could lead to sub-optimal learning outcomes and reduced model generalization ability, as the network may not adequately capture the full variability and complexity of the transformation space in $\mathbf{Q}$. By continuously generating synthetic data on the fly, the model receives a diverse and extensive training dataset, mitigating the risk of over-fitting and enhancing its ability to learn robust representations of quadratic transformations.

In this study, the GPT model is trained using binary masks extracted from fundus images, where non-black areas are encoded as 1, and black areas are assigned a value of 0. This strategic approach enables the model to focus on capturing the global features of transformation between images while filtering out irrelevant local features such as vessels. By prioritizing the essential structural elements of the images, the GPT model can effectively learn and reproduce accurate geometric transformations, leading to improved image registration performance. A well-tuned GPT model can be extended as a generalized model across various imaging modalities where polynomial transformations are required, providing a versatile solution for image registration tasks.

The development of our GPT model is based on the EfficientNetV2 architecture (17), which is chosen for its well-established balance between model complexity and computational efficiency, rendering it ideal for training on a large synthetic dataset. The global max pooling layer was introduced in GPT because it can enhance the GPT model's ability to focus on essential features contributing to overall image transformation. The output layer is a linear activation function, facilitating the generation of

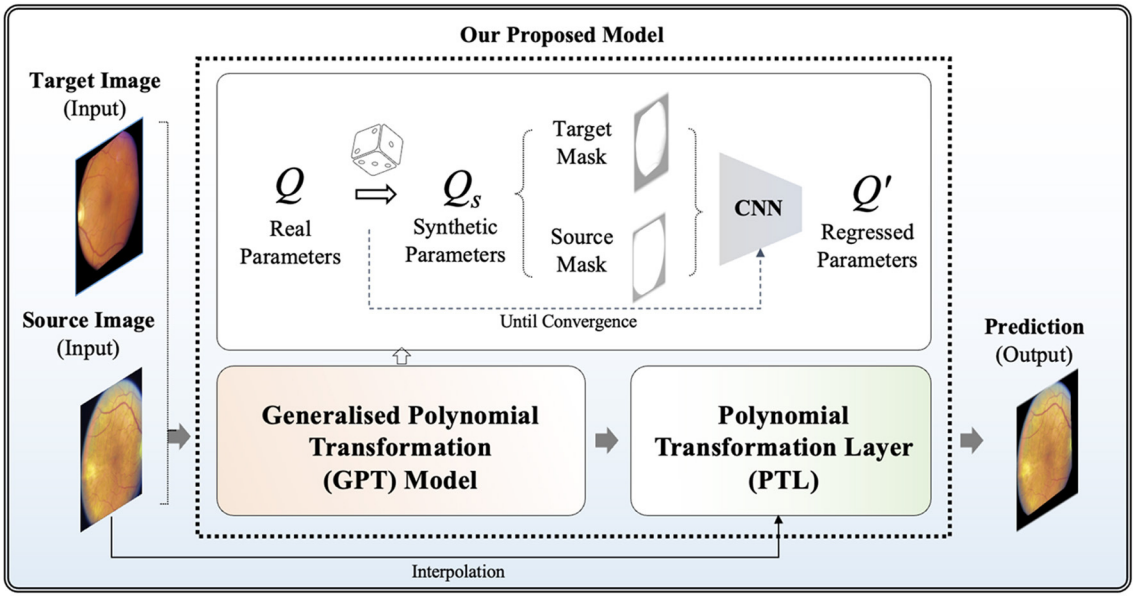

FIGURE 1
Overview of our proposed model for unsupervised polynomial image registration.

regressed parameters for the randomly polynomial-transformed image. To guide the training process effectively, we propose a hybrid loss function in Equation 7, denoted as $\mathcal{L}_{hybrid}$, which combines Mean Squared Logarithmic Error (MSLE) (Equation 8) and Cosine Similarity (CoS) (Equation 9). In which, ω represents the weighting factor for balancing between two terms. Specifically, in our implementation, we set ω to 0.5 to ensure equal contribution from both terms.

$$\mathcal{L}_{hybird}(\mathbf{Q}, \mathbf{Q}') = \omega \mathbf{MSLE} + (1-\omega)(1-\mathbf{CoS}) \tag{7}$$

where,

$$\mathbf{CoS}(\mathbf{Q}, \mathbf{Q}') = \frac{\mathbf{Q}\cdot\mathbf{Q}'}{\|\mathbf{Q}\| \times \|\mathbf{Q}'\|} = \frac{\sum(\mathbf{Q}\times\mathbf{Q}')}{\sqrt{\sum \mathbf{Q}^2} \times \sqrt{\sum \mathbf{Q}'^2}} \tag{8}$$

$$\mathbf{MSLE}(\mathbf{Q}, \mathbf{Q}') = \frac{1}{N}\sum_{i=0}^{N}[\log(\mathbf{Q}_i + 1) - \log(\mathbf{Q}'_i + 1)]^2 \tag{9}$$

Given the diverse ranges of parameters ($N = 16$) within $\mathbf{Q}$, MSLE serves as a robust loss measure. By utilizing a logarithmic scale, MSLE effectively addresses large outliers, treating them comparably to smaller deviations. This feature is particularly advantageous for ensuring model balance, especially when striving for uniform percentage errors across $\mathbf{Q}$. To address negative values of parameters within $\mathbf{Q}$, CoS evaluates the directional consistency between vectors of $\mathbf{Q}$ and $\mathbf{Q}'$, offering significant utility when handling transformations that incorporate negative values. Our $\mathcal{L}_{hybrid}$ loss functions fortify the GPT model, empowering it to adeptly capture and mimic diverse polynomial transformations with resilience and efficacy.

Note that the pre-trained GPT model cannot be directly applied to real fundus image pairs because it was tuned using binary masks and is not trained with any local features such as vessels and the optic disc. In the methodology of our model for unsupervised fundus image registration, the pre-trained GPT model is severed as the foundation, namely pre-trained weights, leveraging its capabilities in capturing diverse polynomial transformations to train a new tailored model for fundus image registration. In which, we proposed a new Polynomial Transformation Layer (PTL) to warp M by the regressed transformations $\mathbf{Q}'$. In PTL, interpolations of $\mathbf{Q}'M$ can be formulated in Equation (10) according to Equation (5):

$$\begin{aligned} u &= q_{00}x^2 + q_{01}y^2 + q_{02}xy + q_{03}x + q_{04}y + q_{05}, \\ v &= q_{10}x^2 + q_{11}y^2 + q_{12}xy + q_{13}x + q_{14}y + q_{15} \end{aligned} \tag{10}$$

The objective is to maximize the similarity between the transformed image $\mathbf{Q}'M$ and the target image F, facilitating unsupervised image registration as the model encounters real transformed images. The loss function $\mathcal{L}_{unsupervised}$ (see Equation 11) is based on Normalized Cross Correlation (NCC) by measuring the correlation between corresponding pixel values.

$$\begin{aligned} &\mathcal{L}_{unsupervised}(F, \mathbf{Q}'M) \\ &= 1 - \frac{\sum_{x,y}(F(x,y)-\bar{F})(\mathbf{Q}'M(x,y)-\bar{\mathbf{Q}'M})}{\sqrt{\sum_{x,y}(F(x,y)-\bar{F})^2 \sum_{x,y}(\mathbf{Q}'M(x,y)-\bar{M})^2}} \end{aligned} \tag{11}$$

3 Experiments

In this section, we detail experiments conducted to validate our GPT-based model for unsupervised fundus image registration. Through a series of experiments and analyzes, we aim to assess the model's ability to accurately align fundus images without the need for ground truth transformation parameters. By detailing

the experimental methodology, dataset characteristics, evaluation metrics, and results, we provide insights into the robustness and reliability of our proposed approach in the context of ophthalmic imaging and clinical practice.

3.1 Dataset

Our methodology is applied to a longitudinal dataset comprising color fundus images from the AREDS study (18), captured using the Zeiss FF-series 30-degree fundus camera at baseline, 2-year, and subsequently annually (19). Extracting longitudinal color fundus images from 4,903 eyes (involving 2,702 participants) sourced from the AREDS study, each patient underwent a minimum of three follow-up visits after the baseline examination. Categorizing the fundus images into non-advanced (early/intermediate stage) and advanced (late stage) AMD, with advanced AMD characterized by the presence of drusen or geographic atrophy, the dataset is publicly available upon request from the database of Genotypes and Phenotypes (dbGaP; accession: phs000001.v3.p1). All analyzes adhere to the approved research use statement.

3.2 Pre-processing

In our approach to unsupervised fundus image registration, we recognize the significance of targeted pre-processing to enhance the model's focus on crucial features. We introduced a hybrid pre-processing approach incorporating both Contrast Limited Adaptive Histogram Equalization (CLAHE) (20) and bilateral filter (21).

CLAHE is a preprocessing technique particularly beneficial for enhancing contrast and improving image quality in fundus images. By locally adapting the contrast enhancement process, CLAHE ensures that the contrast improvements are tailored to the specific characteristics of different regions within the image. This helps in bringing out subtle details and structures in fundus images, such as blood vessels and pathological features. Additionally, CLAHE helps in reducing the impact of uneven illumination and varying brightness levels often present in fundus images, thereby aiding in standardizing the image appearance and facilitating more reliable analysis algorithms. However, as observed in Figure 2C, CLAHE may inadvertently over-enhance unnecessary features in fundus images. Consequently, to address this issue and further denoize the images, a bilateral filter was introduced as a subsequent step in the preprocessing pipeline.

A bilateral filter is employed to systematically eliminate irrelevant and non-diagnostic elements from fundus images. The bilateral filter acts as a selective tool, smoothing the images while preserving essential features such as the optic disc and blood vessels. By doing so, we effectively reduce noise and unwanted details, creating a cleaner input for the subsequent learning stages. This refined dataset allows our model to concentrate on the pertinent anatomical structures, namely the optic disc and vessels, optimizing its ability to learn and predict image transformations accurately. The impact of the bilateral filter can be observed in Figures 2B, D, where a bilateral filter is applied to raw images (Figure 2A) and post-CLAHE images (Figure 2C), respectively.

With our hybrid pre-processing strategy, the objective is to optimize the learning process by offering the model a concentrated and pertinent input (see Figure 2D). This approach enhances the model's interpretability and fosters a more efficient understanding of fundus images.

3.3 Training

The dataset was partitioned at the patient level, with 60% allocated for training, 20% for validation, and 20% for hold-out testing purposes. Training employed the Adam optimizer, a widely embraced algorithm in deep learning, with an initial learning rate of 0.001, facilitating effective model convergence. Input images were resized to 256x256 pixels, and for normalization, we adopted a scale spanning from −1 to 1 instead of the conventional 0–1 range. This deliberate choice prevents the suppression of convolutional neuron activation in black areas, which often contain relevant features for geometric transformation. To enhance model generalization, we applied data augmentation techniques, including flipping, rotation, random brightness, and random contrast.

Our GPT-based model, constructed upon the pre-trained GPT architecture, served as the foundational framework for image registration. Fine-tuning the training set showcased the model's adaptability in capturing diverse Polynomial Transformations, proving advantageous for aligning fundus images. Transparency in our methodology is maintained by providing access to the code, models, and data employed in this experiment, implemented using TensorFlow (version 2.10). During the training phase, where the regression work involves various parameter ranges, we took into account the potential inefficiency of the last linear layer. To address this, our model regressed on normalized values within the [0, 1] range. This strategic approach facilitated a more effective learning process. Once the model converged, we implemented a scaling process to transform each parameter back to its actual range. This scaling step is particularly crucial for subsequent interpolation work, ensuring that the model's learned parameters align accurately with the original data characteristics. By incorporating this normalization and scaling strategy, our methodology enhances the model's adaptability to diverse parameter ranges and contributes to the precision of the final predictions.

3.4 Evaluation metrics

To assess the effectiveness of our model, we employed standard evaluation metrics for image registration at both the image level and parameter level. At the parameter-level, Bland-Altman plots and Pearson correlation coefficients were utilized to evaluate the agreement between predicted and ground truth parameters. Bland-Altman plots visually display the agreement between two quantitative measurements by plotting the difference between the paired measurements against their mean. Additionally, correlation coefficients provide a numerical measure of the strength and direction of the linear relationship between two

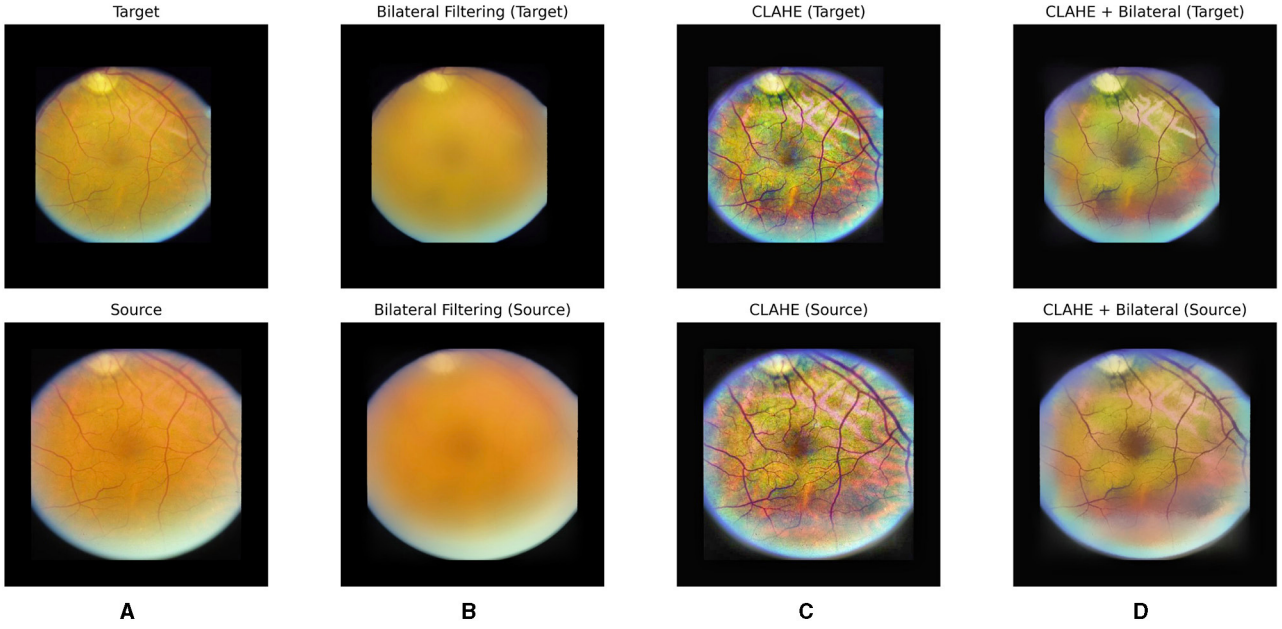

FIGURE 2
Our hybrid pre-processing strategy incorporates both Contrast Limited Adaptive Histogram Equalization (CLAHE) and bilateral filter. From left to right, raw images **(A)**, post-bilateral filtering **(B)**, post-CLAHE **(C)** and images after hybrid pre-processing for a pair of fundus images are displayed, respectively.

variables, indicating the degree of agreement between predicted and ground truth parameters. Meanwhile, at the image level, Structural Similarity Index (SSIM) and Normalized Cross Correlation (NCC) were employed. These metrics provided a comprehensive assessment of the overall quality of image alignment by measuring both structural and pixel-wise similarity between the predicted and target images (see Equations 11 and 12).

$$\mathbf{SSIM}(I_t, I_w) = \frac{(2\mu_{I_t}\mu_{I_w} + C_1)(2\sigma_{I_t}\sigma_{I_w} + C_2)}{(\mu_{I_t}^2 + \mu_{I_w}^2 + C_1)((\sigma_{I_t}^2 + \sigma_{I_w}^2 + C_2)} \quad (12)$$

4 Results

In the results section, we extensively assess the performance of our GPT-based model for unsupervised fundus image registration using the AREDS dataset.

Firstly, We normalized the predicted and target transformation parameters to a range of 0 to 1 to ensure consistency and comparability between the values. This normalization allows for a standardized scale across all parameters, facilitating easier interpretation and analysis of the data. Additionally, by scaling the parameters to a common range, we mitigate the effects of varying magnitudes and ensure that each parameter contributes proportionally to the overall transformation. Upon comparison of their distributions (see Figure 3), it allows us to visually assess the similarity between the predicted and target parameter distributions, providing insights into the model's performance in capturing the transformation characteristics accurately.

Then, we analyzed the non-zero parameters individually, examining their respective distributions and correlations with the target parameters in terms of the correlation coefficient R. For each non-zero parameter in $\mathbf{Q}$, the correlation coefficient R ranges from 0.895 to 0.990, with associated p-values <0.00001, indicating a strong linear relationship between the predicted and target values. These correlation coefficients signify the degree of agreement between the predicted and target parameters. Figure 4 illustrates the corresponding Bland-Altman plots, showcasing the mean difference and upper/lower limits, providing visual insights into the agreement and potential biases between the predicted and target parameters.

While individual parameters show promising results, the overall mean performance of GPT lacked evaluation. To address this, linear regression analysis was conducted across all quadratic transformation parameters, yielding an average correlation coefficient R of 0.9876. Figure 5 illustrates the regression results and corresponding Bland-Altman plot. This high level of correlation underscores the GPT model's ability to accurately predict transformation parameters, demonstrating its efficacy in aligning fundus images without the need for ground truth transformation data.

At the image level evaluation, we conducted both qualitative and quantitative analyzes to comprehensively assess the performance of our model. For the quantitative analysis, we initially evaluated the SSIM and NCC scores before alignment to establish a baseline measurement of similarity between the fixed and moving images prior to any transformations. This baseline provides insights into the initial degree of correspondence before considering the contributions of our models. The SSIM and NCC scores before alignment are 0.6096 and 0.524, respectively. According to Equation (9), the warped images were generated using the transformation parameters (model outputs) based on the corresponding moving images.

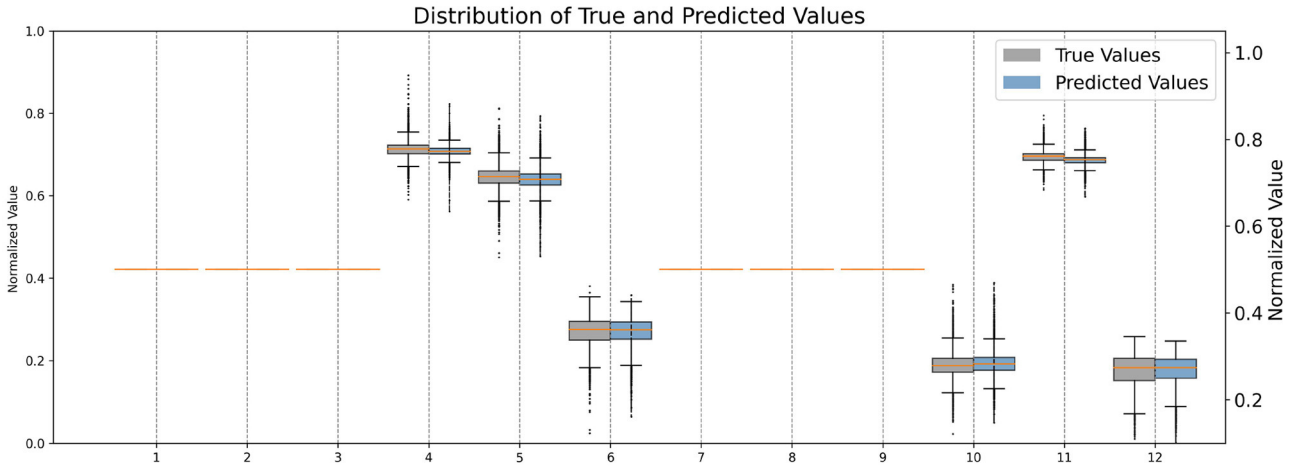

FIGURE 3
Comparison of normalized distributions of each quadratic transformation parameter.

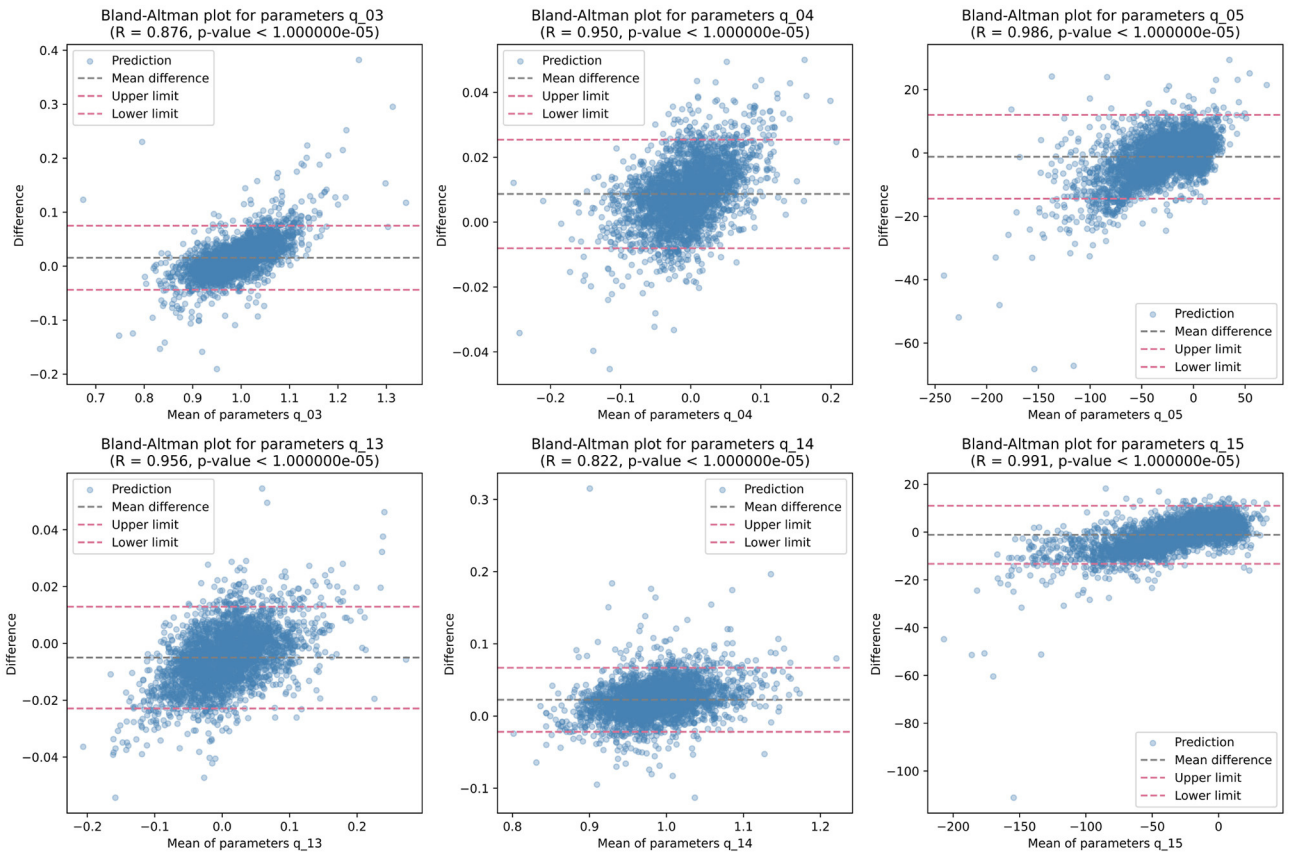

FIGURE 4
Bland-Altman plots illustrating the distribution of each non-zero quadratic transformation parameter, accompanied by correlation coefficient *R*, *p*-value, mean difference, and upper/lower limits.

The SSIM and NCC scores after alignment by our model are 0.8075 and 0.6765, respectively, demonstrating a huge improvement over the baseline. Additionally, the contribution of the pre-processing steps is significant. When these steps are omitted, the SSIM and NCC scores decrease to 0.7649 and 0.6305, respectively.

For qualitative analysis, overlapping and heat maps are employed to visualize differences between images. Figure 6

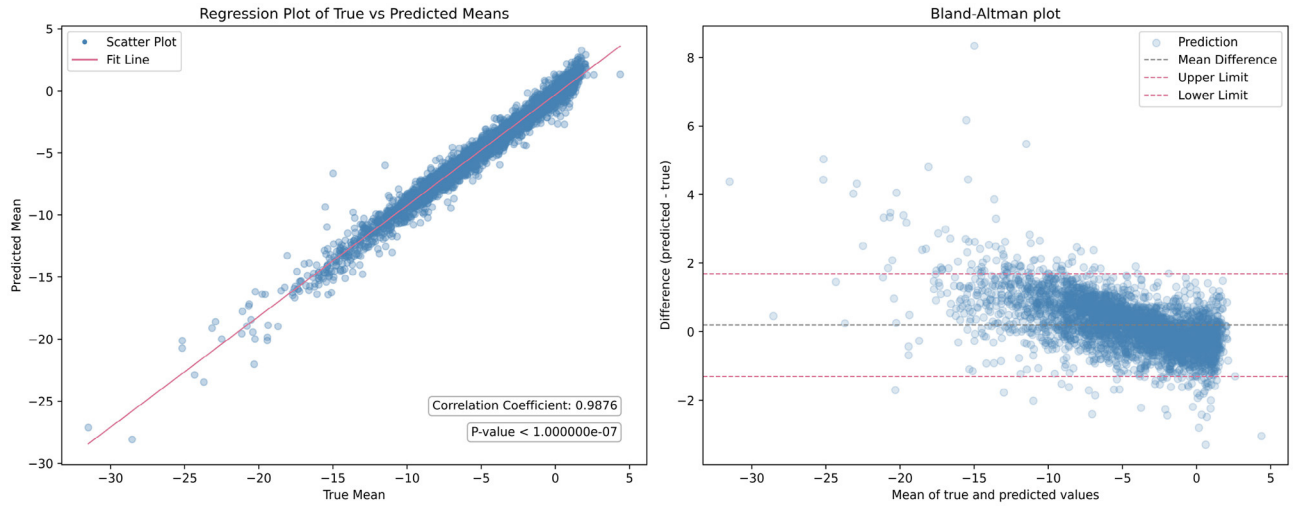

FIGURE 5
Comparison of the overall mean performance of the generalized polynomial transformation (GPT) model. The **(left figure)** displays the results of linear regression analysis, while the **(right figure)** presents the Bland-Altman plot.

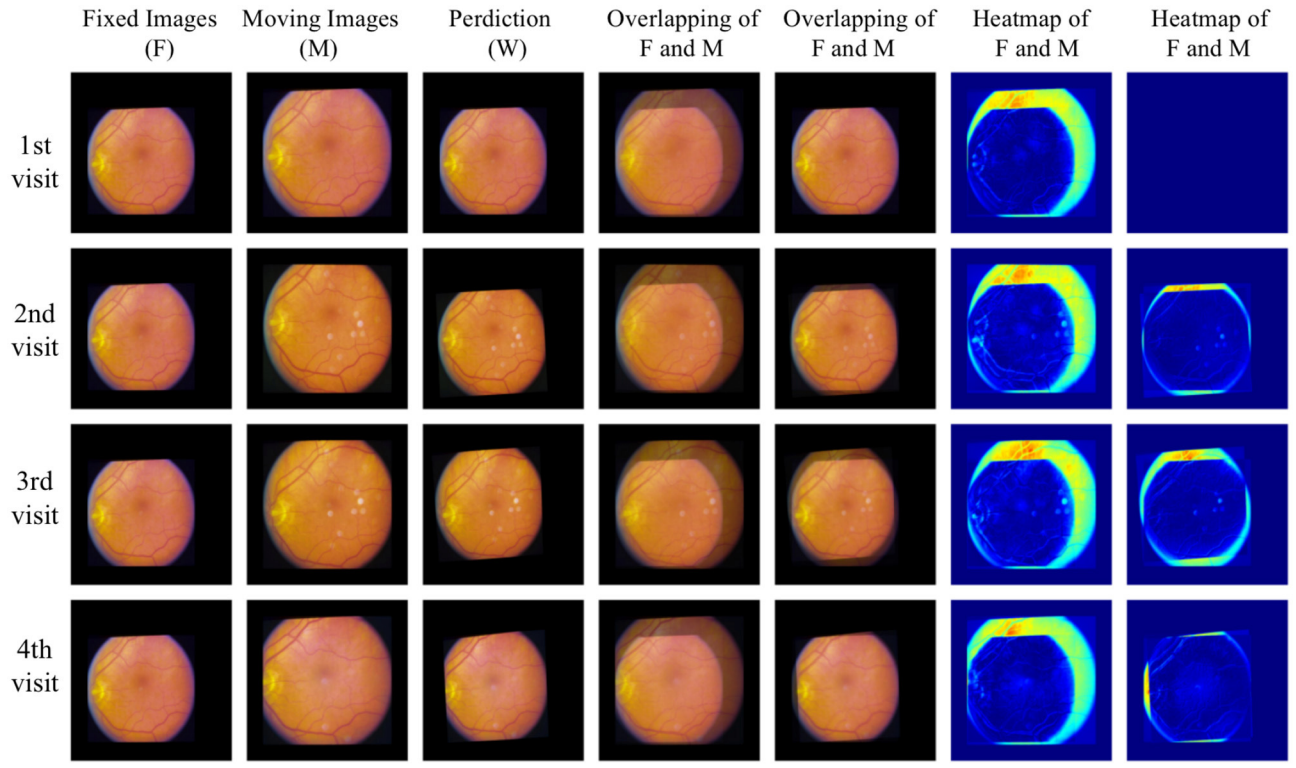

FIGURE 6
Qualitative analysis with overlapping and heat maps reveals four different visits of one eye.

illustrates the fixed images, moving images, and our warped images in the first three columns across four different visits of one eye. Subsequent columns display the overlap between the fixed images and the moving/warped images, followed by heat maps showcasing differences. Significant differences are observed between fixed images and moving images in the overlapping and heat maps, attributed to variations in image acquisition such as differing camera angles or positions. However, comparing fixed images with our warped images reveals a reduction in differences. Particularly, global distortion is minimized, and the locations of

the optic disc and vessels are matched precisely. To the best of our knowledge, our work represents the first unsupervised registration method specifically targeting polynomial transformations. This novel approach sets it apart from the majority of existing models, which predominantly focus on nonlinear or affine transformations. The unique nature of our method introduces challenges in direct comparisons with other models, as their underlying objectives differ significantly.

A limitation of our work is that image intensity-based metrics like SSIM and NCC may not be sufficient for evaluating performance. Variations in illumination or field of view between image pairs can lead to an underestimation of our model's capabilities. It would be beneficial to use ground truth segmentation of the optic disc or vessels for evaluation, as this approach can eliminate irrelevant features from images taken from different viewpoints, providing a more accurate assessment of performance. Moreover, exploring alternative evaluation metrics without segmentation labels that account for these challenges, such as domain-specific similarity measures or perceptual metrics, could provide a more comprehensive assessment of performance in real-world scenarios. In addition to the limitations mentioned, variations in image quality across different datasets or imaging devices may also pose challenges for our model. Addressing these factors and developing robust techniques to handle artifacts could further enhance the reliability and applicability of our approach.

Overall, our GPT model showcases its efficacy in aligning fundus images, presenting a notable advancement in the field of medical image registration. By harnessing the power of deep learning and unsupervised learning techniques, our model achieves remarkable results without relying on ground truth transformation data. This not only streamlines the registration process but also mitigates the need for labor-intensive manual annotation, making the approach more scalable and applicable to large-scale datasets. Furthermore, the versatility of the GPT model allows it to adapt to diverse transformation scenarios, offering a robust solution for aligning fundus images acquired from different sources and modalities.

5 Conclusion

Our work presents a novel approach to unsupervised fundus image registration using the GPT model. Through GPT, we introduced a foundational model capable of emulating diverse polynomial transformations, trained on a large synthetic dataset to cover a wide spectrum of transformation scenarios. Additionally, our hybrid pre-processing strategy aims to optimize the learning process by providing the model with focused input. To assess our model's effectiveness, we employed standard evaluation metrics on the publicly available AREDS dataset, including image-level and parameter-level analyzes. Linear regression analysis yielded an average correlation coefficient R of 0.9876 across all quadratic transformation parameters. In image-level evaluation, both qualitative and quantitative analyzes were conducted, revealing significant improvements in SSIM (20%) and NCC (15%) scores, indicating robust performance. Particularly noteworthy is the precise matching of optic disc and vessel locations and the minimization of global distortion. Our findings highlight the potential of GPT-based approaches in image registration methodologies, and promising advancements in diagnosis, treatment planning, and disease monitoring in ophthalmology.

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 humans were approved by Age-Related Eye Disease Study Research Group (1999). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

XC: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. XF: Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. YM: Resources, Writing – review & editing. YZ: Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partly supported by the EPSRC (Engineering and Physical Sciences Research Council) (grant ref: EP/R014094/1).

Conflict of interest

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

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

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

References

1. Lee S, Reinhardt JM, Cattin PC, Abràmoff MD. Objective and expert-independent validation of retinal image registration algorithms by a projective imaging distortion model. *Med Image Anal*. (2010) 14:539–49. doi: 10.1016/j.media.2010.04.001

2. Lee S, Abramoff MD, Reinhardt JM. Validation of retinal image registration algorithms by a projective imaging distortion model. In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE (2007). p. 6471–4.

3. Olcay K, Cakir A, Sönmez M, Duzgun E, Yildirim Y. Analysing the progression rates of macular lesions with autofluorescence imaging modes in dry age-related macular degeneration. *Turk J Ophthalmol*. (2015) 2015:tjo.93276. doi: 10.4274/tjo.93276

4. Bhuiyan A, Xiao D, Kanagasingam Y. A review of disease grading and remote diagnosis for sight threatening eye condition: age related macular degeneration. *J Comput Sci Syst Biol*. (2014) 2014:139. doi: 10.4172/jcsb.1000139

5. Kanagasingam Y, Bhuiyan A, Abràmoff MD, Rjh S, Goldschmidt L, Wong TY. Progress on retinal image analysis for age related macular degeneration. *Progr Retin Eye Res*. (2014). 38:20–42. doi: 10.1016/j.preteyeres.2013.10.002

6. Schmidt-Erfurth U, Klimscha S, Waldstein SM, Bogunović H. A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration. *Eye*. (2016) 31:26–44. doi: 10.1038/eye.2016.227

7. Maruyama-Inoue M, Kitajima Y, Mohamed S, Inoue T, Sato S, Ito A, et al. Sensitivity and specificity of high-resolution wide field fundus imaging for detecting neovascular age-related macular degeneration. *PLoS ONE*. (2020) 15:e0238072. doi: 10.1371/journal.pone.0238072

8. Ly A, Nivison-Smith L, Zangerl B, Assaad N, Kalloniatis M. Advanced imaging for the diagnosis of age—related macular degeneration: a case vignettes study. *Clin Exp Optomet*. (2018) 101:243–54. doi: 10.1111/cxo.12607

9. Heo TY, Kim KM, Min HK, Gu SM, Kim JH, Yun J, et al. Development of a deep-learning-based artificial intelligence tool for differential diagnosis between dry and neovascular age-related macular degeneration. *Diagnostics*. (2020) 10:261. doi: 10.3390/diagnostics10050261

10. Feeny A, Tadarati M, Freund D, Bressler NM, Burlina P. Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Comput Biol Med*. (2015) 65:124–36. doi: 10.1016/j.compbiomed.2015.06.018

11. Gess AJ, Fung AE, Rodríguez J. Imaging in neovascular age-related macular degeneration. *Semin Ophthalmol*. (2011) 26:225–33. doi: 10.3109/08820538.2011.582533

12. Balyen L, Peto T. Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. *Asia-Pacific J Ophthalmol*. (2019) 8:264–72. doi: 10.22608/APO.2018479

13. Ly A, Yapp M, Nivison-Smith L, Assaad N, Hennessy M, Kalloniatis M. Developing prognostic biomarkers in intermediate age—related macular degeneration: their clinical use in predicting progression. *Clin Exp Opt*. (2018) 101:172–81. doi: 10.1111/cxo.12624

14. Miao S, Wang ZJ, Liao R. A CNN regression approach for real-time 2D/3D registration. *IEEE Trans Med Imag*. (2016) 35:1352–63. doi: 10.1109/TMI.2016.2521800

15. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*. (2019) 52:128–43. doi: 10.48550/arXiv.1809.06130

16. Chen X, Meng Y, Zhao Y, Williams R, Vallabhaneni SR, Zheng Y. Learning unsupervised parameter-specific affine transformation for medical images registration. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*. Berlin: Springer (2021). p. 24–34.

17. Tan M, Le Q. Efficientnetv2: smaller models and faster training. In: *International Conference on Machine Learning*. PMLR (2021). p. 10096–106.

18. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1. *Contr Clin Trials*. (1999) 20:573–600.

19. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. *Am J Ophthalmol*. (2001) 132:668–81. doi: 10.1016/s0002-9394(01)01218-1

20. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, et al. Adaptive histogram equalization and its variations. *Comput Vis Graph Image Process*. (1987) 39:355–68.

21. Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998). p. 839–46.

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

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