



OPEN ACCESS

EDITED BY

Weihua Yang,
Southern Medical University, China

REVIEWED BY

Tianming Huo,
Wuhan University, China
Xiang Liu,
The First Affiliated Hospital of Nanchang
University, China
Jeewoo Yoon,
Raondata, Republic of Korea

*CORRESPONDENCE

Hui Qi,
✉ qihui@sysush.com
Xiang Tu,
✉ tuxiang@mail.sysu.edu.cn

RECEIVED 11 September 2025

REVISED 19 November 2025

ACCEPTED 19 November 2025

PUBLISHED 11 December 2025

CITATION

Duan M-M, Qi H and Tu X (2025)
Knowledge-enhanced AI drives diagnosis of
multiple retinal diseases in fundus fluorescein
angiography.
Front. Cell Dev. Biol. 13:1703606.
doi: 10.3389/fcell.2025.1703606

COPYRIGHT

© 2025 Duan, Qi and Tu. 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.

Knowledge-enhanced AI drives diagnosis of multiple retinal diseases in fundus fluorescein angiography

Ming-Ming Duan¹, Hui Qi^{2*} and Xiang Tu^{2*}

¹Department of ophthalmology, Jiujiang City Key Laboratory of Cell Therapy, JiuJiang No. 1 People's Hospital, JiuJiang, Jiangxi, China, ²Department of Otolaryngology, The Seventh Affiliated Hospital, Sun Yat-sen University, ShenZhen, Guangzhou, China

Purpose: This study aimed to develop and validate a deep learning model for the accurate multi-class classification of six retinal diseases using fundus fluorescein angiography (FFA) images.

Methods: We applied a knowledge-enhanced pre-training strategy (KeepFIT) using a ResNet-50 image encoder on two large FFA corpora: a curated atlas and a clinical report dataset. The resulting visual encoder was fine-tuned to classify six conditions, including diabetic retinopathy and macular degeneration. The model's performance and generalizability were assessed on two independent test sets, one of which was sourced from an external institution.

Results: Our proposed deep learning model, leveraging a knowledge-enhanced pre-training strategy, demonstrated robust performance in classifying six distinct retinal diseases using fundus fluorescein angiography images. The model achieved a strong and consistent micro-average area under the curve (AUC) of 0.92 across two independent test sets. Notably, it showed excellent classification performance for critical conditions such as venous occlusion (VO) and neovascular age-related macular degeneration (nAMD), with AUC values reaching 0.95 and 0.96, respectively.

Conclusion: The knowledge-enhanced pre-training strategy significantly improves the diagnostic accuracy and generalizability of deep learning models for FFA analysis. This approach provides a scalable and effective framework for automated retinal disease screening, holding significant potential for clinical decision support, especially in resource-limited settings.

KEYWORDS

artificial intelligence, fundus fluorescein angiography, KeepFIT, retinal diseases, deep learning

Introduction

With the aging of the global population, retinal diseases, including diabetic retinopathy and age-related macular degeneration, are increasing worldwide and are increasingly becoming a significant cause of eye health hazards, leading to visual

impairment and even blindness (Burton et al., 2021). Early screening and timely intervention can help prevent or minimize eye damage caused by retinal diseases (Vujosevic et al., 2020). However, large-scale manual screening is difficult to achieve due to limited medical resources. Therefore, accurate, cost-effective, and efficient screening methods for retinal diseases are essential. Fundus fluorescein angiography (FFA) is an important ophthalmic imaging technique. The technique uses sodium fluorescein as a contrast agent injected intravenously, and its fluorescent properties allow details of the retinal blood vessels to be visualized when the fundus is illuminated with specific wavelengths (Wang et al., 2017; Liu et al., 2025). Thus, FFA can assess microvascular structure and blood flow and is the gold standard for the diagnosis of fundus diseases such as diabetic retinopathy (Yu et al., 2025). However, the interpretation of FFA images is time-consuming, and the diagnosis is dependent on the ophthalmologist's expertise and subjective, which makes the efficiency and accuracy of the diagnosis, to some extent, compromised (Gao et al., 2023).

In recent years, the rapid development of artificial intelligence (AI) technology has brought new possibilities for the screening and diagnosis of fundus diseases. Deep learning-based diagnostic systems have been widely used in the detection of various fundus diseases, such as diabetic retinopathy and age-related macular degeneration (Long et al., 2025; Dai et al., 2021; Dai et al., 2024; Yim et al., 2020; Kazemzadeh, 2025), showing great potential for clinical decision support. It can quickly and accurately identify and analyze the characteristics of fundus lesions, which greatly improves the efficiency and accuracy of early screening and diagnosis of various fundus diseases. However, many previous studies have typically utilized separate datasets to train task-specific models, which leads to poor generalization across different scenarios and the need for large amounts of annotated training data (Yang et al., 2022). MM-Retinal is a multimodal dataset that includes graphic paired data from FFA, color fundus photography (CFP), and optical coherence tomography (OCT) images (Wu R. et al., 2024). The data were obtained from fundus chart books containing comprehensive ophthalmologic knowledge and accurate graphic descriptions provided by ophthalmologists. The knowledge-enhanced foundational pre-training model which incorporates fundus image-text expertise, known as KeepFIT, is a knowledge-enhanced base model that injects MM-Retinal's fundus expert knowledge into model training through an image similarity-guided text modification approach and a hybrid training strategy (Wu R. et al., 2024). It exhibits strong robustness and generalizability.

Therefore, this study uses the KeepFIT model introduced with MM-Retinal for pre-training, with ResNet-50 as the image encoder E_V and BioClinical-BERT as the text encoder E_T . The domain knowledge of MM-Retinal-FFA is injected into the labeled cue text of fluorescein fundus angiography image repository (FFA-IR) via an image similarity-based text correction module. The trained visual backbone E_V was then transferred to the classification task to classify the FFA dataset containing central serous chorioretinopathy (CSC), venous occlusion (VO), neovascular age-related macular degeneration (nAMD), non-proliferative diabetic retinopathy (NPDR), proliferative diabetic retinopathy (PDR), and healthy control. The aim of this study was to improve the efficiency and accuracy of diagnosis of fundus diseases.

Methods

Pre-training data

The foundation model is first exposed to two complementary vision-language corpora that both belong to the FFA modality. The first source is MM-Retinal-FFA, a subset of the MM-Retinal atlas that offers 1,947 high-resolution ($\geq 800 \times 800$) FFA photographs, each accompanied by a bilingual expert description. The second source is FFA-IR, a clinical collection composed of 1,048,584 FFA frames aligned with 10,790 bilingual diagnostic reports that follow a 46-label schema. Because FFA-IR is two orders of magnitude larger than MM-Retinal-FFA, every mini-batch is constructed with a one-to-one ratio of samples from the two corpora to maintain a balance between scale and expert coverage.

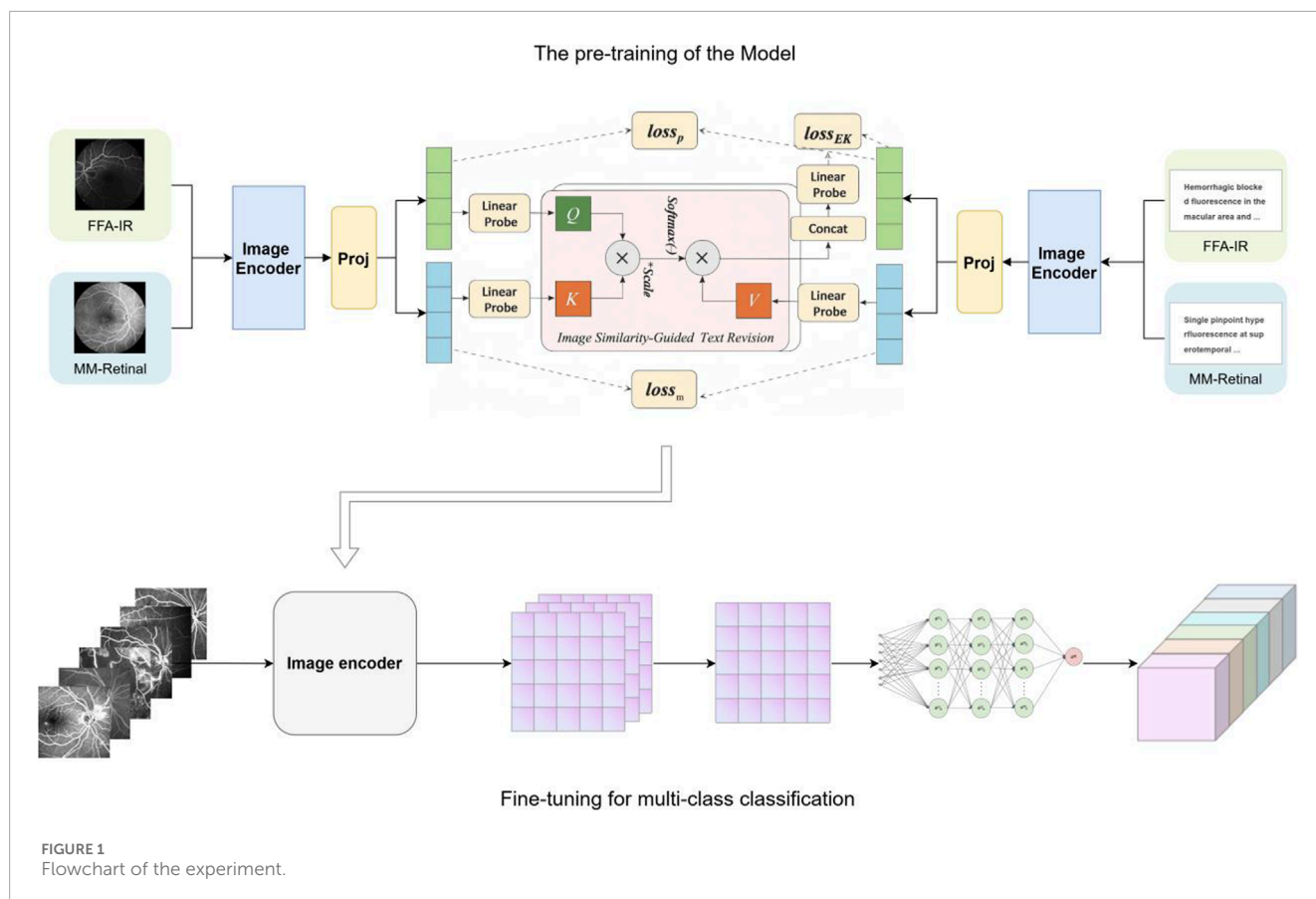
Downstream classification data

In this retrospective study, we sourced a primary dataset of FFA images from patients treated at Jiujiang No. 1 People's Hospital between January 2017 and December 2022. This dataset was randomly divided, with 80% of the images allocated for training and 20% for validation. Additionally, an independent internal test set (Test Set 1) was constructed using images acquired at the same hospital between January 2023 and January 2025. To further assess generalizability, an external test set (Test Set 2) was compiled using FFA images collected during the same period from Seventh Affiliated Hospital, Sun Yat-sen University. The research was conducted in accordance with the tenets of the Declaration of Helsinki and received approval from the institutional review boards of both hospitals. Given that the study relied exclusively on anonymized data devoid of personal identifiers, the ethics committees granted a waiver for patient-specific informed consent. The overall study framework is illustrated in Figure 1.

Knowledge-enhanced vision-language pre-training

Pre-training follows the KeepFIT strategy that was introduced together with MM-Retinal. A ResNet-50 model initialized on ImageNet-1 K serves as the image encoder E_V ; BioClinical-BERT acts as the text encoder E_T . In addition, the standard ResNet-50 model from the Torchvision library serves as the image encoder without any architectural modification; BioClinical-BERT acts as the text encoder. Another domain-specific language model based on BERT, pre-trained on large-scale clinical notes, is included, enabling it to effectively understand medical terminology and context (Lee et al., 2020). Their outputs are projected into a shared 512-dimensional space through linear layers with layer normalization. To inject domain knowledge from MM-Retinal-FFA into the label-prompt texts of FFA-IR, an image similarity-guided text-revision module is implemented as an eight-head cross-attention block.

Let B be a mini-batch that mixes the two datasets. Genuine image-caption pairs from MM-Retinal-FFA are trained with a



symmetric InfoNCE contrastive loss L_m . Prompt pairs from FFA-IR are optimized by a category-aware variant L_p that attracts images whose prompts share the same disease label. The revised text embedding E_K produced by the cross-attention module is encouraged to remain close to the original prompt embedding t_p via a mean-squared error term L_{EK} . The complete objective is given as

$$L = L_m + L_p + \alpha L_{EK}, \alpha = 100.$$

Optimization uses AdamW, with an initial learning rate of 1×10^{-4} and weight decay of 1×10^{-5} . Training runs for 20 epochs on four RTX 4090 GPUs, with a per-GPU batch size of 64.

Fine-tuning for multi-class classification

The pre-trained visual backbone E_V is transferred to the classification task. All layers are unfrozen. The final convolutional feature map is transformed by global average pooling into a 2048-dimensional vector that feeds a four-layer multilayer perceptron. Cross-entropy is minimized with Adam (a learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}). The learning rate decreases by a factor of 10 every 10 epochs; training terminates after 10 epochs. Data augmentation at training time includes random horizontal flip, $\pm 10^\circ$ rotation, and mild color jitter. At inference time, the softmax of the six logits yields posterior probabilities for each diagnosis.

Results

Characteristics of the datasets

Fine-tuning and evaluation are conducted on an independent FFA dataset that covers six clinically important categories: CSC, VO, nAMD, NPDR, PDR, and healthy control. The training set contains 5,270 images (CSC 790, VO 1,650, nAMD 1,300, NPDR 692, PDR 429, and normal 409). The validation set comprises 1,054 images (CSC 300, VO 300, nAMD 300, NPDR 75, PDR 40, and normal 39). Two disjoint test sets are used: Test Set 1 with 1,908 images (CSC 500, VO 500, nAMD 400, NPDR 215, PDR 166, and normal 127) and Test Set 2 with 896 images (CSC 300, VO 300, nAMD 120, NPDR 98, PDR 55, and normal 23). All frames are resized to 224×224 pixels, center-cropped if necessary, and normalized using the ImageNet mean and standard deviation (mean = [0.485, 0.456, and 0.406]; std = [0.229, 0.224, and 0.225]). The distribution of images of each retinal disease category in the training set, validation set, and test sets is detailed in Table 1. Representative examples of each disease category are illustrated in Figure 2.

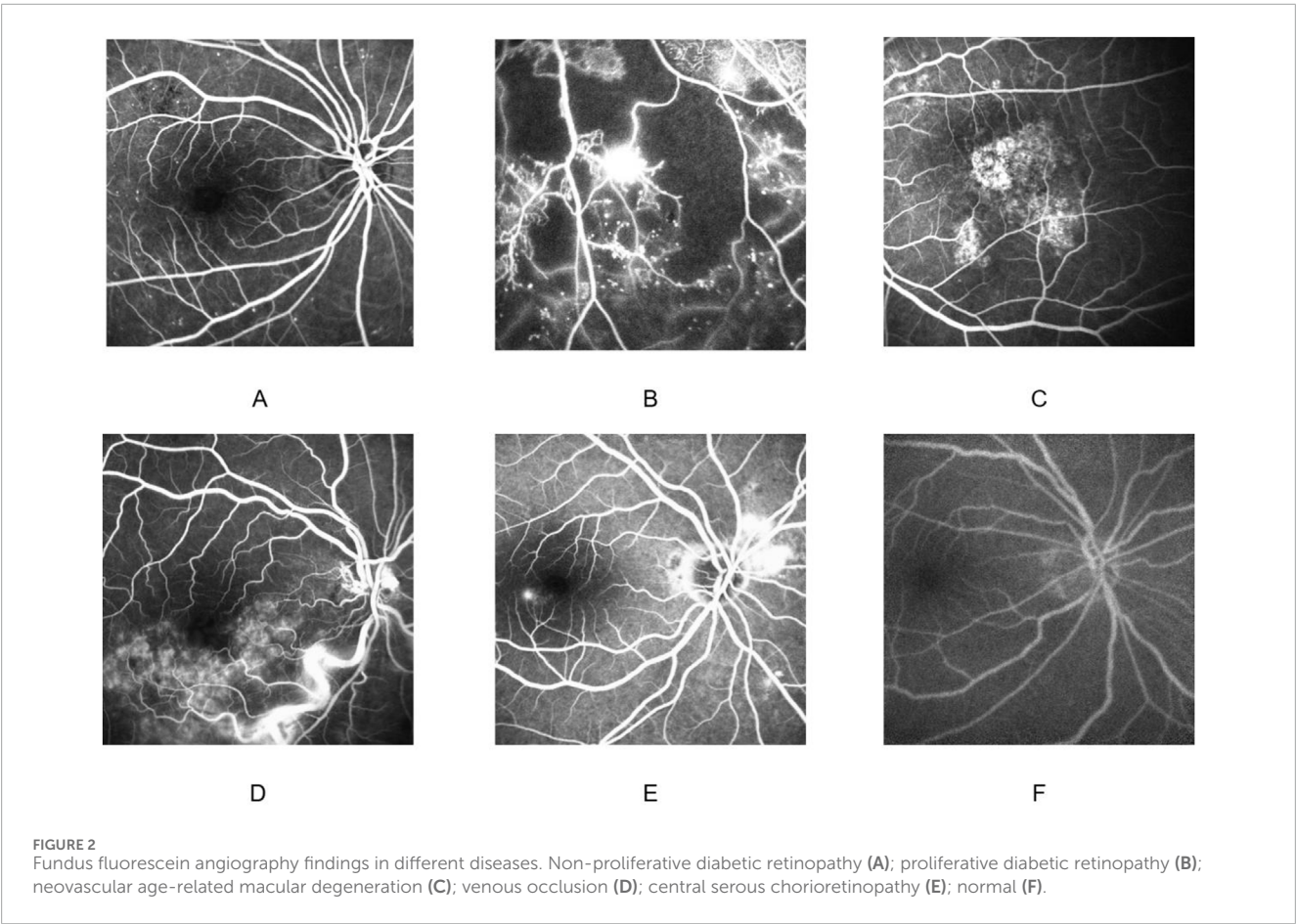
Evaluation of model diagnostic performance

The deep learning model developed in this study demonstrated favorable multi-class classification performance across six retinal

TABLE 1 Dataset characteristics.

Category	Training set	Validation set	Test set 1	Test set 2
CSC	790	300	500	300
VO	1,650	300	500	300
nAMD	1,300	300	400	120
NPDR	692	75	215	98
PDR	429	40	166	55
Normal	409	39	127	23

CSC, central serous chorioretinopathy; VO, venous occlusion; nAMD, neovascular age-related macular degeneration; NPDR, non-proliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.



disease categories in both internal and external test sets. The overall accuracy ranged from 0.868 to 0.967, and the micro-average AUC remained consistently high at 0.92 in both datasets. Nevertheless, classification performance varied across disease types, with notable differences observed in precision, sensitivity, and F1-scores.

The model achieved the most consistent and robust performance in identifying central serous chorioretinopathy and retinal vein occlusion. In Test Set 1, the F1-scores for these two categories

reached 0.746 and 0.795, respectively, and further improved to 0.788 and 0.818, respectively, in Test Set 2. Both precision and sensitivity remained high, indicating a well-balanced classification capability.

These findings were further supported by the receiver operating characteristic analysis. The AUCs for central serous chorioretinopathy were 0.91 and 0.92 in the two test sets, while those for retinal vein occlusion were 0.89 and 0.90, respectively. The micro-average AUC of 0.92 in both test sets reflects the model's overall stable classification performance.

TABLE 2 Model performance evaluation on the validation and test sets.

Task	Accuracy	Precision	Sensitivity	Specificity	F1-score
Test set 1					
CSC	0.871	0.772	0.720	0.925	0.746
PDR	0.883	0.481	0.512	0.930	0.497
Normal	0.940	0.537	0.630	0.962	0.580
NPDR	0.911	0.488	0.476	0.952	0.481
VO	0.888	0.765	0.826	0.910	0.795
nAMD	0.900	0.787	0.713	0.949	0.748
Test set 2					
CSC	0.868	0.858	0.727	0.940	0.788
PDR	0.890	0.495	0.490	0.939	0.492
Normal	0.967	0.400	0.609	0.976	0.483
NPDR	0.915	0.373	0.564	0.939	0.449
VO	0.878	0.819	0.817	0.909	0.818
nAMD	0.913	0.664	0.708	0.945	0.685

CSC, central serous chorioretinopathy; VO, venous occlusion; nAMD, neovascular age-related macular degeneration; NPDR, non-proliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.

Confusion matrix analysis also revealed that central serous chorioretinopathy and retinal vein occlusion had the highest numbers of correctly classified cases, with 360 and 218 true positives for central serous chorioretinopathy and 413 and 245 for retinal vein occlusion in the two test sets, respectively. Misclassification rates for these categories were comparatively low, underscoring the model's strong and generalizable discriminatory capacity. The detailed diagnostic performance metrics are summarized in Table 2, and the corresponding ROC curves and confusion matrices are presented in Figure 3.

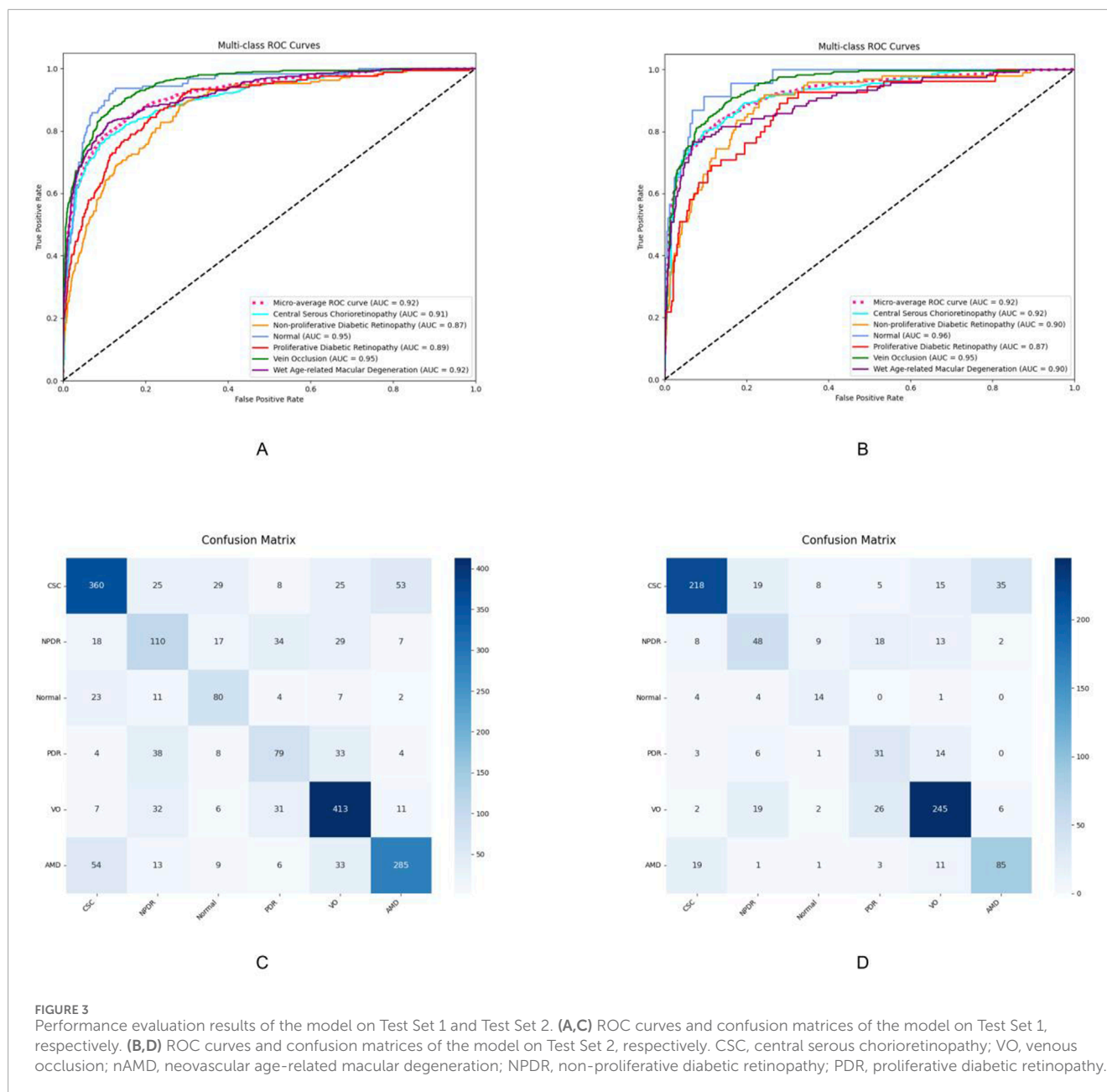
Discussion

In this study, we developed and validated a deep learning model for the classification of six common retinal diseases using FFA images. Our primary finding is that a visual model pre-trained with KeepFIT demonstrates both high accuracy and robust generalizability. The model achieved a consistent micro-average AUC of 0.92 across two independent test sets, one of which was sourced from an external institution, underscoring its potential for broader clinical application and affirming our initial hypothesis.

Our model's capacity for simultaneous differential diagnosis among six distinct retinal conditions marks a significant advancement over prior work, which has predominantly focused on binary classification tasks (Lee et al., 2021; Heydon et al., 2021; Alsaih et al., 2017; N et al., 2021). Although previous models

have achieved high performance in detecting single, specific lesions such as neovascular leakage in PDR (Yang et al., 2022), their utility is confined to narrow clinical questions. In contrast, our model's ability to differentiate between multiple pathologies, including CSC, VO, nAMD, and different stages of DR (Long et al., 2025; Gan et al., 2023), more closely mirrors the complex diagnostic challenges encountered in real-world clinical practice (Shao et al., 2025; Ueno et al., 2019). This multi-class capability significantly broadens its potential applicability as a decision support tool for ophthalmologists.

The key innovation of our study lies in the KeepFIT pre-training methodology, which distinguishes our approach from conventional supervised learning paradigms used in earlier multi-class FFA classifiers (Gao et al., 2023; Huang et al., 2024). Traditional models rely heavily on learning pixel-level patterns from labeled images, a process that is often constrained by the quantity and quality of available annotations, as labels can be highly subjective and exhibit large variance between experts (Wu J. et al., 2024; Chen, 2023; Hu et al., 2021). Our strategy fundamentally circumvents this limitation by leveraging two heterogeneous data sources: a curated atlas dataset providing precise, granular knowledge and a large-scale clinical report dataset capturing the rich, nuanced language of expert interpretation. This allows the model to move beyond superficial feature recognition and learn the deep semantic correlations between imaging findings and their underlying pathology. Consequently, the model acquires a foundational "understanding" of retinal



diseases, which enhances its generalization capabilities and reduces its dependency on massive, meticulously annotated datasets for fine-tuning.

The model's performance, while strong overall, varied across different disease categories. It demonstrated exceptional classification accuracy for nAMD and VO, with AUC values reaching as high as 0.96 and 0.95, respectively. This success can likely be attributed to the distinct and often dramatic morphological changes these conditions present on FFA imaging, such as neovascular leakage or clear vascular blockages, providing strong signals for the model to learn from. Conversely, the F1-scores for NPDR and PDR were comparatively lower. This disparity may stem from two factors. First, the training dataset contained fewer examples of PDR and NPDR compared to VO and nAMD, creating a data imbalance that could impede

learning, a common challenge in medical AI development (Gan et al., 2025a; Gan et al., 2025b; Li et al., 2021). Second, the pathological features of early diabetic retinopathy (Gan et al., 2025b) can be subtle and varied, posing a greater challenge for automated detection compared to the more overt features of other diseases.

However, we acknowledge several limitations. First, the aforementioned class imbalance in the fine-tuning dataset remains a significant constraint. Future efforts should prioritize the curation of larger and more balanced multi-class datasets to further refine model performance, particularly for underrepresented diseases. Second, although validated on data from two centers, establishing true global generalizability requires further testing on more geographically and ethnically diverse multinational cohorts as models trained on data from a single center or

limited population may lack generalizability (Gan et al., 2023; Gong et al., 2024). Third, the “black box” nature of the current model is a common barrier to clinical translation. Integrating established interpretability techniques, such as class activation maps (CAMs), is a critical next step (Wang et al., 2020). This will not only enhance transparency by providing visual evidence for the model’s predictions but also foster greater trust and adoption among clinicians who need to understand the basis of the AI’s diagnostic reasoning.

Conclusion

This study demonstrates that a knowledge-enhanced vision-language pre-training strategy significantly improves the diagnostic accuracy and generalizability of deep learning models for fundus fluorescein angiography image classification. By addressing the challenge of efficient and scalable retinal disease screening, the proposed approach meets the critical need for intelligent diagnostic tools in ophthalmology. These findings advance the integration of clinical expertise into medical AI and offer a promising framework for real-world application in resource-limited settings.

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 study received approval from the ethics committees of First People’s Hospital of Jiujiang City and the Seventh Affiliated Hospital. 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. 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

M-MD: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Software,

Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing. HQ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. XT: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review and editing.

Funding

The authors declare that financial support was received for the research and/or publication of this article. The study was supported by the Sanming Project of Medicine in Shenzhen (SZSM202111005).

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

- Alsaii, K., Lemaitre, G., Rastgoo, M., Massich, J., Sidibé, D., and Meriaudeau, F. (2017). Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomed. Eng. Online* 16 (1), 68. doi:10.1186/s12938-017-0352-9
- Burton, M. J., Ramke, J., Marques, A. P., Bourne, R. R. A., Congdon, N., Jones, I., et al. (2021). The lancet Global Health Commission on Global eye Health: vision beyond 2020. *Lancet Glob. Health* 9 (4), e489–e551. doi:10.1016/S2214-109X(20)30488-5
- Chen, X. (2023). A deep learning based automatic report generator for retinal optical coherence tomography images.
- Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., et al. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* 12 (1), 3242. doi:10.1038/s41467-021-23458-5
- Dai, L., Sheng, B., Chen, T., Wu, Q., Liu, R., Cai, C., et al. (2024). A deep learning system for predicting time to progression of diabetic retinopathy. *Nat. Med.* 30 (2), 584–594. doi:10.1038/s41591-023-02702-z
- Gan, F., Wu, F.-P., and Zhong, Y.-L. (2023). Artificial intelligence method based on multi-feature fusion for automatic macular edema (ME) classification on spectral-domain optical coherence tomography (SD-OCT) images. *Front. Neurosci.* 17, 1097291. doi:10.3389/fnins.2023.1097291
- Gan, F., Long, X., Wu, X., Luo, L., Ji, W., Fan, H., et al. (2025a). Deep learning-enabled transformation of anterior segment images to corneal fluorescein staining images

for enhanced corneal disease screening. *Comput. Struct. Biotechnol. J.* 28, 94–105. doi:10.1016/j.csbj.2025.02.039

Gan, F., Cao, J., Fan, H., Qin, W., Li, X., Wan, Q., et al. (2025b). Development and validation of the Artificial Intelligence-Proliferative Vitreoretinopathy (AI-PVR) Insight system for deep learning-based diagnosis and postoperative risk prediction in proliferative vitreoretinopathy using multimodal fundus imaging. *Quant. Imaging Med. Surg.* 15 (4), 2774–2788. doi:10.21037/qims-24-1644

Gao, Z., Pan, X., Shao, J., Jiang, X., Su, Z., Jin, K., et al. (2023). Automatic interpretation and clinical evaluation for fundus fluorescein angiography images of diabetic retinopathy patients by deep learning. *Br. J. Ophthalmol.* 107 (12), 1852–1858. doi:10.1136/bjo-2022-321472

Gong, D., Li, W.-T., Li, X.-M., Wan, C., Zhou, Y.-J., Wang, S. J., et al. (2024). Development and research status of intelligent ophthalmology in China. *Int. J. Ophthalmol.* 17 (12), 2308–2315. doi:10.18240/ijo.2024.12.20

Heydon, P., Egan, C., Bolter, L., Chambers, R., Anderson, J., Aldington, S., et al. (2021). Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br. J. Ophthalmol.* 105 (5), 723–728. doi:10.1136/bjophthalmol-2020-316594

Hu, Y., Xiao, Y., Quan, W., Zhang, B., Wu, Y., Wu, Q., et al. (2021). A multi-center study of prediction of macular hole status after vitrectomy and internal limiting membrane peeling by a deep learning model. *Ann. Transl. Med.* 9 (1), 51. doi:10.21037/atm-20-1789

Huang, S., Jin, K., Gao, Z., Yang, B., Shi, X., Zhou, J., et al. (2024). Automated interpretation of retinal vein occlusion based on fundus fluorescein angiography images using deep learning: a retrospective, multi-center study. *Heliyon* 10 (13), e33108. doi:10.1016/j.heliyon.2024.e33108

Kazemzadeh, K. (2025). Artificial intelligence in ophthalmology: opportunities, challenges, and ethical considerations. *Med. Hypothesis Discov. Innov. Ophthalmol. J.* 14 (1), 255–272. doi:10.51329/mehdiophthal1517

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinforma. Oxf. Engl.* 36 (4), 1234–1240. doi:10.1093/bioinformatics/btz682

Lee, A. Y., Yanagihara, R. T., Lee, C. S., Blazes, M., Jung, H. C., Chee, Y. E., et al. (2021). Multicenter, Head-to-Head, real-world validation Study of seven automated artificial intelligence diabetic Retinopathy screening systems. *Diabetes Care* 44 (5), 1168–1175. doi:10.2337/dc20-1877

Li, M., Cai, W., Liu, R., Weng, Y., Zhao, X., et al. (2021). FFA-IR: towards an explainable and reliable medical report generation benchmark.

Liu, X., Xie, J., Hou, J., Xu, X., Guo, Y., D-Get, (2025). D-GET: Group-Enhanced transformer for diabetic retinopathy severity classification in

fundus fluorescein angiography. *J. Med. Syst.* 49 (1), 34. doi:10.1007/s10916-025-02165-4

Long, X., Gan, F., Fan, H., Qin, W., Li, X., Ma, R., et al. (2025). EfficientNetB0-Based end-to-end diagnostic system for diabetic retinopathy grading and macular edema detection. *Diabetes Metab. Syndr. Obes. Targets Ther.* 18, 1311–1321. doi:10.2147/DMSO.S506494

N, R., So, G., and A, W. (2021). Automatic detection of pathological myopia using machine learning. *Sci. Rep.* 11 (1), 16570. doi:10.1038/s41598-021-95205-1

Shao, A., Liu, X., Shen, W., Li, Y., Wu, H., Pan, X., et al. (2025). Generative artificial intelligence for fundus fluorescein angiography interpretation and human expert evaluation. *NPJ Digit. Med.* 8 (1), 396. doi:10.1038/s41746-025-01759-z

Ueno, Y., Oda, M., Yamaguchi, T., Fukuoka, H., Nejima, R., et al. (2019). Deep learning model for extensive smartphone-based diagnosis and triage of cataracts and multiple corneal diseases.

Vujosevic, S., Aldington, S. J., Silva, P., Hernández, C., Scanlon, P., Peto, T., et al. (2020). Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol.* 8 (4), 337–347. doi:10.1016/S2213-8587(19)30411-5

Wang, S., Zuo, Y., Wang, N., and Tong, B. (2017). Fundus fluorescence Angiography in diagnosing diabetic retinopathy. *Pak. J. Med. Sci.* 33 (6), 1328–1332. doi:10.12669/pjms.336.13405

Wang, D., Haytham, A., Pottenburgh, J., Saeedi, O., and Tao, Y. (2020). Hard attention net for automatic retinal vessel segmentation. *IEEE J. Biomed. Health Inf.* 24 (12), 3384–3396. doi:10.1109/JBHI.2020.3002985

Wu, R., Zhang, C., Zhang, J., Zhou, Y., Zhou, T., et al. (2024a). MM-Retinal: knowledge-enhanced foundational pretraining with fundus image-text expertise. *arXiv:2405.11793*. doi:10.48550/arXiv.2405.11793

Wu, J., Fang, H., Zhu, J., Zhang, Y., Li, X., Liu, Y., et al. (2024b). Multi-rater prism: learning self-calibrated medical image segmentation from multiple raters. *Sci. Bull.* 69 (18), 2906–2919. doi:10.1016/j.scib.2024.06.037

Yang, J., Soltan, A. A. S., and Clifton, D. A. (2022). Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit. Med.* 5 (1), 69. doi:10.1038/s41746-022-00614-9

Yim, J., Chopra, R., Spitz, T., Winkens, J., Obika, A., Kelly, C., et al. (2020). Predicting conversion to wet age-related macular degeneration using deep learning. *Nat. Med.* 26 (6), 892–899. doi:10.1038/s41591-020-0867-7

Yu, T., Shao, A., Wu, H., Su, Z., Shen, W., Zhou, J., et al. (2025). A systematic review of advances in AI-Assisted analysis of Fundus Fluorescein Angiography (FFA) images: from detection to report generation. *Ophthalmol. Ther.* 14 (4), 599–619. doi:10.1007/s40123-025-01109-y