A novel clinical artificial intelligence model for disease detection via retinal imaging

Yidian Fu,[1,2,5] Liang Ma,[1,2,5] Sheng Wan,[3,4,5] Shengfang Ge,[1,2,*] and Zhi Yang[1,2,*]

[1]Department of Ophthalmology, Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai 200011, China
[2]Shanghai Key Laboratory of Orbital Diseases and Ocular Oncology, Shanghai 200011, China
[3]PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing 210000, China
[4]Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210000, China
[5]These authors contributed equally
*Correspondence: geshengfang@sjtu.edu.cn (S.G.); yangzhiscience@163.com (Z.Y.)

The Innovation

Artificial intelligence (AI) is an emerging field in which computerized systems are used to carry out complex tasks in place of humans. Medical AI algorithms have been developed for disease diagnosis and prediction and treatment recommendation across various clinical data types, e.g., chest X-rays, electrocardiograms, and other radiological images.[1] In ophthalmology, particularly, great progress has been made in AI systems over the past decade. Color fundus photography (CFP) and optical coherence tomography (OCT), which are readily available in routine clinical practice, are both mainstream and useful retinal imaging modalities in ophthalmology. In September 2023, the 2023 Lasker-Debakey Clinical Medical Research Award was awarded to three scientists for their work on OCT for accurate retinal disease detection.

However, the majority of AI models in the medical imaging field are calibrated with a large number of images with high-quality annotations, demanding intensive labor from specialists. Consequently, the large-scale production of medical datasets with advanced clinical labels is deemed impossible. Additionally, existing models usually have limited generalizability to distinct clinical tasks since they are useful for specific applications. In light of the above limitations, Y. Zhou et al. constructed a self-supervised learning (SSL)-based foundation model for retinal images (RETFound).[2] This model was trained on large volumes of unannotated data at scale to complete a variety of downstream tasks (Figure 1). By deriving supervisory signals directly from data, SSL performs well in alleviating data inefficiency and avoiding the use of expert knowledge to label images in the "pretrain-then-fine-tune" workflow. The SSL approach, used in RETFound for learning from unannotated retinal images, involves a predictive task. In this task, some parts of the image are masked out. The model is subsequently trained to predict these masked parts and enforced to understand and encode the underlying structure of retinal features. After learning from massive datasets, a rich representation is built that can be fine-tuned to recognize various patterns with minimal labeled data. Therefore, the model can achieve generalizable disease detection from retinal images.[2]

Two separate RETFound models were developed using CFP and OCT images from a natural image dataset (ImageNet-1k), the Moorfields diabetic image dataset (MEH-MIDAS), and public data (a total of 904,170 CFPs and 736,442 OCTs). The RETFound tool was adapted for complicated detection and prediction tasks, including ocular disease diagnosis (e.g., diabetic retinopathy and glaucoma), prognosis (e.g., wet-age-related macular degeneration), and predicting systemic diseases from oculomic challenges (e.g., cardiovascular diseases and neurodegenerative diseases).

As a result, the RETFound model performs well in terms of stability and label accuracy in contrast to those of three published comparison models, SL-ImageNet, SSL-ImageNet, and SSL-Retinal. For example, when used for identifying and further classifying diabetic retinopathy, RETFound achieved area under the receiver operating characteristic curves (AUROCs) of 0.943, 0.822, and 0.884 on the Kaggle APTOS-2019, IDRiD, and MESSIDOR-2 datasets, respectively, indicating that RETFound achieved significantly higher data quality than did SL-ImageNet (all $p < 0.001$). For predicting the 1 year prognosis of fellow eyes proceeding wet-age-related macular degeneration, RETFound exhibited satisfactory performance (AUROC = 0.862) compared to those of the three comparison groups ($p < 0.001$). Myocardial infarction, a cardiovascular disease, was predicted from CFP images using the RETFound model, and the AUROC was 0.737. Moreover, qualitative results and variable-controlling experiments were

used to evaluate the ability of RETFound to detect disease, revealing that disease-related areas were identified and inferred by this model. Surprisingly, although aging and disease progression for aging-associated systemic diseases result in clinical anatomical structure alterations, RETF identified such alterations well and exhibited stable performance for predicting systemic diseases, even when the age difference decreased.

The adapted RETFound model achieved good performance and generalizability in the diagnosis and prognosis of common ocular diseases, as well as in the prediction of complicated systemic disorders with fewer labeled images, leading to a wider range of clinical AI applications from retinal imaging. This achievement also highlights the strength of the correlation between systemic diseases and the information contained in various imaging modalities.

Nevertheless, RETFound showed a significant decrease in performance when tested against new cohorts that differ in demographic profile and imaging device. The current study cohorts were based in the UK and may not represent all populations worldwide. Therefore, introducing a larger dataset with retinal images worldwide is recommended to enhance the generalizability of this model. Since clinical information such as demographic and visual acuity data may influence ocular and oculomic studies, the relevant characteristics and multimodal information fusion between CFP and OCT should be further considered in the RETFound model. In addition, utilizing domain adaptation techniques to augment the training data enables diverse image types to be obtained. Therefore, training the model on such datasets promotes more generalizable features and enhances the model performance on unseen data.

Apart from SSL models, conventional supervised deep learning models are still worthy of consideration in the future. All the training images are labeled via supervised learning, and the model is directly optimized using image-label pairs. For instance, supervised contrastive learning was adopted in a vision transformer to decode elements of intraoperative surgical activity from videos, which might provide surgeons with feedback on their operating skills.[3] The EfficientNet-b2 network was employed to detect Alzheimer's disease dementia from retinal photographs. By integrating features from these photographs, researchers developed supervised deep learning models and equipped the network with unsupervised domain adaptation techniques to address dataset discrepancies among different studies.[4] HyperDenseNet was constructed for brain tissue segmentation in multimodal magnetic resonance (MR) images to address challenging medical image segmentation problems involving multimodal volumetric data.[5]

Analogous to most existing foundation models, RETFound is pretrained on a large corpus of unlabeled data and is designed to be adapted to a wide range of downstream tasks, whereas conventional supervised models are typically trained on labeled datasets and are usually optimized for specific tasks. Moreover, by adopting an SSL strategy, RETFound performs well in understanding the underlying structure and patterns within data, which is essential for generalizing across various medical imaging tasks. However, conventional supervised models, which depend on explicit signals from labeled data, may fail to capture subtle or complex patterns.

Great efforts still need to be made before applying AI systems from bench to bedside. Since AI studies are mainly based on fixed datasets and stable environments in the short term, their performance may be dominated and restricted by the development time background. In reality, however, the evolutionary world raises the requirement of a strong ability for AI models to learn over a lifetime and
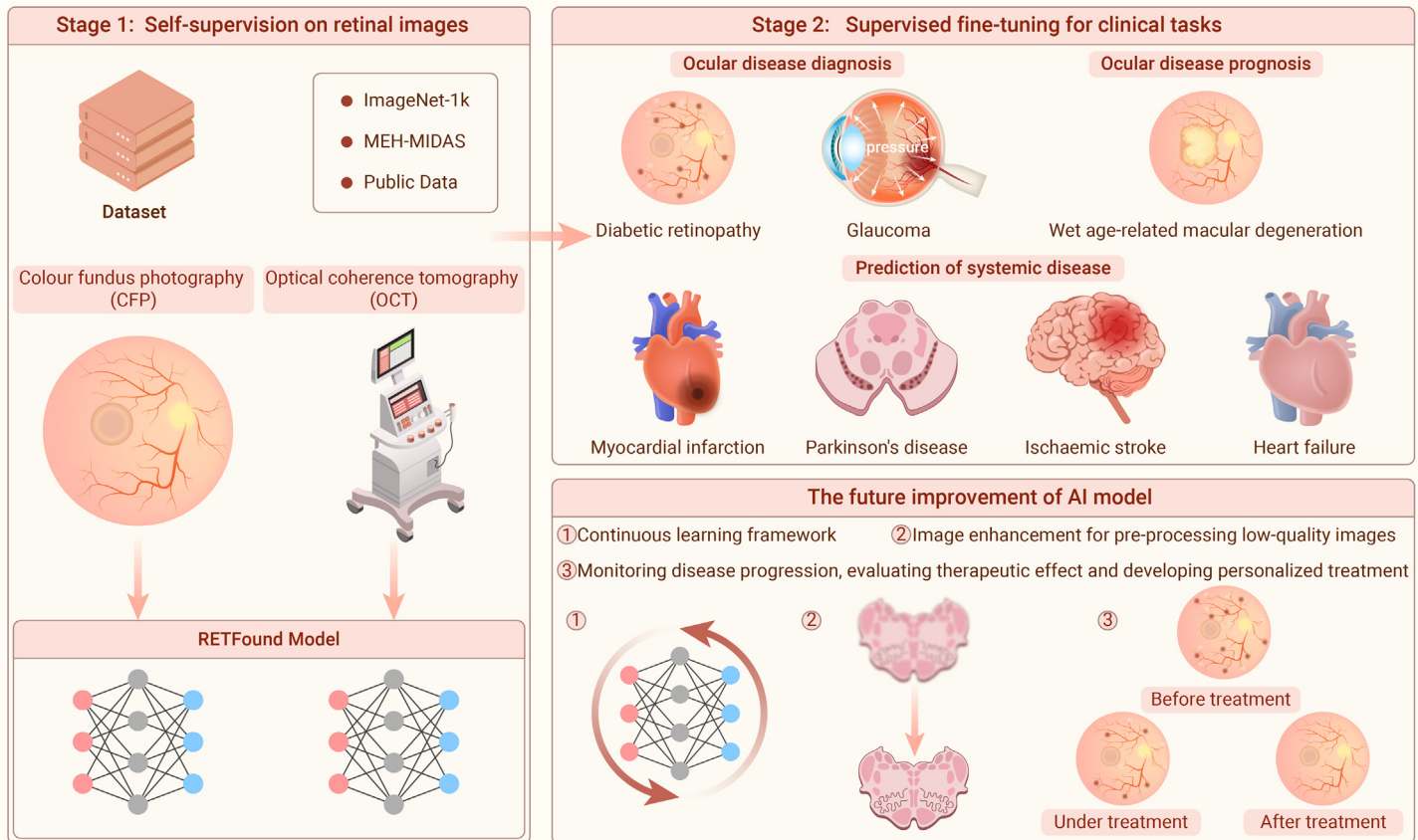
Figure 1. Schematic of RETFound model and future improvements Stage one constructs RETFound model using CFP and OCT images from three datasets by means of self-supervised technique. Stage two adapts RETFound to downstream tasks including ocular disease diagnosis and prognosis as well as prediction of systemic disease. Future improvements of AI models include continuous learning framework and image enhancement for preprocessing low-quality images as well as monitoring disease progression, evaluating therapeutic effect, and developing personalized treatment.

steadily evolve to thrive in dynamic learning settings. Therefore, to solve the problem of model outdating, a continuous learning framework should be implemented. This framework involves periodic model retraining on newly acquired medical data, enabling the system to adapt to evolving clinical knowledge and imaging techniques. Furthermore, a mechanism for automatic version control and model management should be utilized to ensure that the AI system is able to reflect the latest advancements in medical science and AI.

Additionally, the performance of AI systems is strongly dependent on high-quality images, whereas doctors make clinical decisions that are insusceptible to low-quality images, which are inevitable in real-world practice. Therefore, developing novel approaches to enhancing the stability and robustness of AI systems in low-quality images is also essential in clinical applications. A robust model for filtering out outliers should be integrated into the algorithm. In addition, preprocessing techniques, such as image enhancement, can be applied to improve the quality of input data prior to its introduction to an AI system. Furthermore, active learning strategies can also be employed to enhance the performance of AI models with a minimal set of precise human annotations.

Apart from current diagnostic and predictive tasks, AI methods may also be conducive to monitoring disease progression, evaluating therapeutic effects, and developing personalized treatments. For instance, a series of retinal images from a patient with diabetic retinopathy can be processed by AI models and may assist ophthalmologists in accurately evaluating whether the treatment is efficacious for this patient during follow-up visits. If retinopathy is mitigated, then the treatment is considered effective. Otherwise, the therapeutic schedule may need to be adjusted. Research exploring the value of AI models in monitoring disease progression during treatment is encouraged. Future improvements for AI models are illustrated in Figure 1.

Although the field of AI is still in its infancy, we hope that AI systems will have a profound impact on making healthcare more accurate, more efficient, and easier to access, especially in regions lacking clinicians and experts.

REFERENCES

1. Rajpurkar, P., Chen, E., Banerjee, O., et al. (2022). AI in health and medicine. Nat Med **28**: 31−38. https://doi.org/10.1038/s41591-021-01614-0.
2. Zhou, Y., Chia, M.A., Wagner, S.K., et al. (2023). A foundation model for generalizable disease detection from retinal images. Nature **622**: 156−163. https://doi.org/10.1038/s41586-023-06555-x.
3. Kiyasseh, D., Ma, R., Haque, T.F., et al. (2023). A vision transformer for decoding surgeon activity from surgical videos. Nat. Biomed. Eng. **7**: 780−796. https://doi.org/10.1038/s41551-023-01010-8.
4. Cheung, C.Y., Ran, A.R., Wang, S., et al. (2022). A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. Lancet Digit Health **4**: e806−e815. https://doi.org/10.1016/S2589-7500(22)00169-8.
5. Dolz, J., Gopinath, K., Yuan, J., et al. (2019). HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. IEEE Trans Med Imaging **38**: 1116−1126. https://doi.org/10.1109/TMI.2018.2878669.

DECLARATION OF INTERESTS

The authors declare no competing interests.