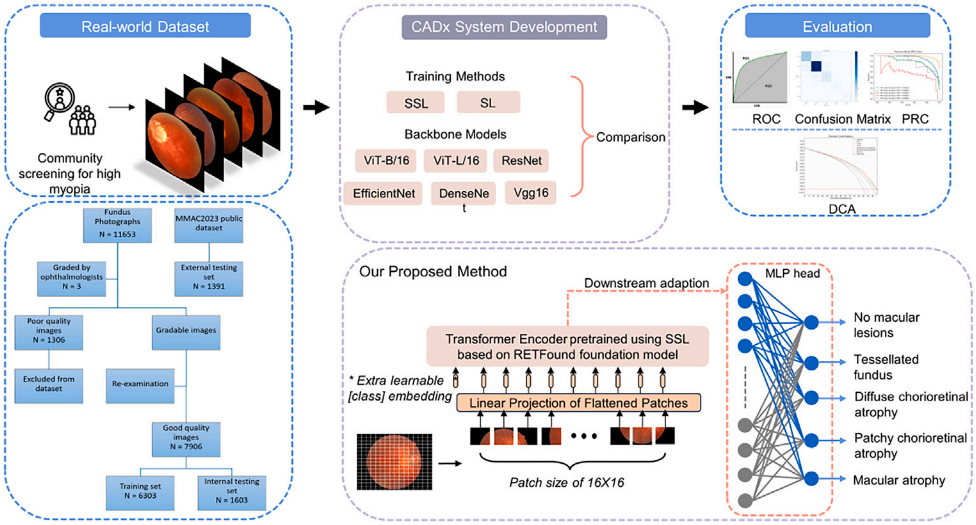
iScience

CellPress
OPEN ACCESS

Article

Self-supervised learning-enhanced deep learning method for identifying myopic maculopathy in high myopia patients


Real-world Dataset
Community screening for high myopia
Fundus Photographs N = 11653
MMAC2023 public dataset
Graded by ophthalmologists N = 3
External testing set N = 1391
Poor quality images N = 1306
Gradable images
Excluded from dataset
Re-examination
Good quality images N = 7906
Training set N = 6303
Internal testing set N = 1603
CADx System Development
Training Methods
SSL
SL
Backbone Models
ViT-B/16
ViT-L/16
ResNet
EfficientNet
DenseNet
Vgg16
Comparison
Evaluation
ROC
Confusion Matrix
PRC
DCA
Our Proposed Method
Downstream adaption
Transformer Encoder pretrained using SSL based on RETFound foundation model
* Extra learnable [class] embedding
Linear Projection of Flattened Patches
Patch size of 16X16
MLP head
No macular lesions
Tessellated fundus
Diffuse chorioretinal atrophy
Patchy chorioretinal atrophy
Macular atrophy

Juzhao Zhang, Fan Xiao, Haidong Zou, Rui Feng, Jiangnan He

fengrui@fudan.edu.cn (R.F.)
hejiangnan85@126.com (J.H.)

Highlights

The accuracy of automatic identification of myopic maculopathy (MM) is suboptimal

We propose a method for MM diagnosis based on self-supervised deep learning

Our method achieved higher accuracy and better net benefit in validation

Zhang et al., iScience *27*, 110566
August 16, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.110566

iScience

CellPress
OPEN ACCESS

Article

Self-supervised learning-enhanced deep learning method for identifying myopic maculopathy in high myopia patients

Juzhao Zhang,[1,2,3,4,7] Fan Xiao,[5,6,7] Haidong Zou,[1,2,3,4] Rui Feng,[4,5,6,*] and Jiangnan He[1,4,8,*]

SUMMARY

Accurate detection and timely care for patients with high myopia present significant challenges. We developed a deep learning (DL) system enhanced by a self-supervised learning (SSL) approach to improve the automatic diagnosis of myopic maculopathy (MM). Using a dataset of 7,906 images from the Shanghai High Myopia Screening Project and a public validation set of 1,391 images from MMAC2023, our method significantly outperformed conventional techniques. Internally, it achieved 96.8% accuracy, 83.1% sensitivity, and 95.6% specificity, with AUC values of 0.982 and 0.999. Externally, it maintained 89.0% accuracy, 71.7% sensitivity, and 87.8% specificity, with AUC values of 0.978 and 0.973. The model's Cohen's kappa values exceeded 0.8, indicating substantial agreement with retinal experts. Our SSL-enhanced DL approach offers high accuracy and potential to enhance large-scale myopia screenings, demonstrating broader significance in improving early detection and treatment of MM.

INTRODUCTION

Myopia is a severe global health concern.[1–3] It is estimated that by 2050, the global prevalence of myopia will reach 49.8% (4.758 billion), with 9.8% (938 million) suffering from high myopia.[4] As myopia advances, a subset of patients experience significant elongation of the eye axis and show various ocular tissue abnormalities, known as pathologic myopia (PM).[5] In cases of PM, the macular region often presents multiple retinal and choroidal lesions, referred to as myopic maculopathy (MM).[6] If current intervention strategies remain unchanged, the number of individuals with vision impairment due to MM will increase to 55.7 million by 2050, making it one of the leading causes of vision impairment and blindness worldwide.[7–9] Effective detection and timely intervention for this population is a significant public health challenge.[10] However, the vastness of the population in need of screening and the scarcity of ophthalmic resources render the task of large-scale fundus image acquisition and analysis a formidable challenge.[11] Moreover, the heterogeneity in image-based MM grading systems, coupled with the ambiguous nature of lesion morphological features, results in inconsistent interpretations among clinicians.[12–14] Therefore, there is an urgent need to integrate automated systems into the retinal imaging workflow, which will improve both efficiency and accuracy of diagnosis.

In recent years, deep learning techniques have reached a level of diagnostic precision equivalent to human experts in certain clinical tasks related to ophthalmology. This includes diseases, such as diabetic retinopathy (DR),[15] cataract,[16] age-related macular degeneration (AMD),[17] and glaucoma.[18] While deep learning presents a promising approach for analyzing fundus images, its success to date predominantly depends on supervised learning (SL) frameworks. These frameworks require extensive annotated datasets for optimal performance. Additionally, many models demonstrate limited generalization when extended to other institutions or different tasks.[19] These limitations likely stem from the SL training methodology, which promotes the creation of "specialist models" focused on label-specific features rather than on features that broadly represent data distribution. Therefore, developing strategies for training medical artificial intelligence (AI) models becomes critically important.

Self-supervised learning (SSL) is a training methodology that leverages unlabeled data to generate meaningful representations. Unlike SL, SSL facilitates the creation of "generalist models" capable of adapting to a variety of downstream tasks, thereby reducing dependence on extensive annotated datasets. Initially achieving significant success in the field of natural language processing (NLP), recent studies have demonstrated the advantages of SSL in the area of image processing.[20] On September 13, 2023, Zhou et al. published a research report wherein they employed SSL techniques on 1.6 million unlabeled fundus images to train a vision transformer (ViT).[21] This effort resulted in

[1] Shanghai Eye Disease Prevention & Treatment Center/Shanghai Eye Hospital, School of Medicine, Tongji University, Shanghai, China
[2] National Clinical Research Center for Eye Diseases, Shanghai, China
[3] Department of Ophthalmology, Shanghai General Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[4] Shanghai Engineering Center for Precise Diagnosis and Treatment of Eye Diseases, Shanghai, China
[5] School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China
[6] Academy for Engineering and Technology, Fudan University, Shanghai, China
[7] These authors contributed equally
[8] Lead contact
*Correspondence: fengrui@fudan.edu.cn (R.F.), hejiangnan85@126.com (J.H.)
https://doi.org/10.1016/j.isci.2024.110566

the creation of RETFound, the first foundational model in ophthalmology. Within their investigation, RETFound could be rapidly deployed to downstream tasks through transfer learning, and its accuracy surpassed traditional models trained via SL on ImageNet. This marks a fundamental departure from traditional technological pathways. Thus, to what extent can the accuracy of MM screenings be enhanced by employing deep learning models enhanced by SSL? What value could RETFound framework potentially bring to MM screenings? The key questions concerning the extended application of the RETFound framework are yet to be clarified.

The aim of this research is to assess the possible benefits of integrating SSL with DL techniques for the precise diagnosis of MM. We specifically investigate the capacity of SSL-augmented DL models to boost MM diagnostic and screening accuracy, along with their applicability in clinical practice. Additionally, as a pioneering effort in this field, this study will employ decision curve analysis (DCA) to assess the net benefit and impact of these models when implemented in large-scale high myopia screening.

RESULTS

Performance evaluation and comparison

Evaluation metrics of our proposed method

The test outcomes on our internal testing set (Figure 1; Table 1), show that our method achieved a general classification accuracy of 96.8%, with a sensitivity of 83.1%, specificity of 95.6%, and an F1 score of 0.842 in the five-category classification task. It achieved area under the receiver operating characteristic curve (AUROC) of 0.982 for normal versus other fundus types and 0.999 for macular atrophy versus others. On the external testing set, our approach demonstrated strong generalization capability, achieving an overall classification accuracy of 89.0%, sensitivity of 71.7%, specificity of 87.8%, and an F1 score of 0.683. The AUROCs for separating normal fundus from other types and macular atrophy from others were 0.978 and 0.973, respectively. Additionally, Cohen's kappa values exceeding 0.8 on both datasets indicate a high degree of concordance with expert-labeled results.

Comparison between different models

In the internal testing set, the classification accuracies across a variety of models in the five-category classification task demonstrated a range from 0.946 to 0.969. Sensitivity varied from 0.599 to 0.831, while specificity was recorded between 0.929 and 0.968, and AUROC ranged from 0.961 to 0.986, as delineated in Table 2. Models that were trained utilizing SL methods exhibited performances that were on par with those trained through SSL. The ViT-B/16 model achieved the better accuracy and AUROC compared to those of the RETFound-enhanced model. This marginal disparity in performance might be attributed to the larger parameter space of the latter, which potentially remained undertrained (Figures S2 and S3).

In the external testing set, different models demonstrated diverse generalization capabilities as summarized in Table 3. Within the scope of a five-category classification task, accuracy spanned from 0.831 to 0.890, sensitivity from 0.575 to 0.717, specificity from 0.832 to 0.897, and AUROC from 0.888 to 0.951. Models trained using SSL methods generally outperformed those trained through SL (Figures S4 and S5). Notably, our RETFound-enhanced model emerged as the leader, underscoring its satisfactory generalizability. Our methodology achieved an approximate 4% enhancement in accuracy compared to CNN-based models and a 2% improvement over models that utilized SL training methods and transformer architectures. It is pertinent to note that while some models, such as EfficientNet-B3 and Vgg16, achieved high specificity, this was at the potential expense of reduced sensitivity.

Detailed prediction results

To further analyze the performance of our models, we meticulously recorded the predictive outcomes for each fundus image in both testing datasets, as illustrated in Figure 2. For the predominantly high myopia categories C0 and C1, our approach reached sensitivities of 0.85 and 0.94 on the internal testing set. However, the external testing showed a notable misclassification of C0 as C1, significantly contributing to the performance decline in this dataset. This indicates a propensity of the model for more cautious predictions. Additionally, the model exhibited sensitivity below 0.6 in identifying C3, with 0.56 sensitivity in the internal testing set and 0.5 in the external testing set. This pattern of confusing C3 with C2 and C4 corresponds to the clinical findings observed by ophthalmologists.

Visualization of the prediction process

Using an illustrative true positive example from the testing dataset (Figure 3A), it was observed that most models effectively identified and highlighted the atrophic areas. Notably, our proposed method produced heatmaps (Figure 3B) that displayed deeper reds in lesion regions, indicating a more precise localization and concentrated model attention on these critical areas. However, the EfficientNet-B3 model appeared to distribute attention also to areas not directly associated with the lesions (Figure 3F), which could suggest a tendency toward overfitting, warranting further investigation.

Net benefit of our proposed method and other models

Decision curves for different strategies on the external testing set are depicted in Figure 4. For risk-averse doctors or patients (e.g., P_t below 20%), the most effective clinical strategy for balancing the number of unnecessary follow-up treatments against the detection of MM would be to forgo fundus examination results and immediately initiate further interventions, including more frequent follow-ups, medication, or surgery. The decision curves that exclude TestHarm, as presented in Figure 4B, may be more applicable to real-world scenarios. This is because

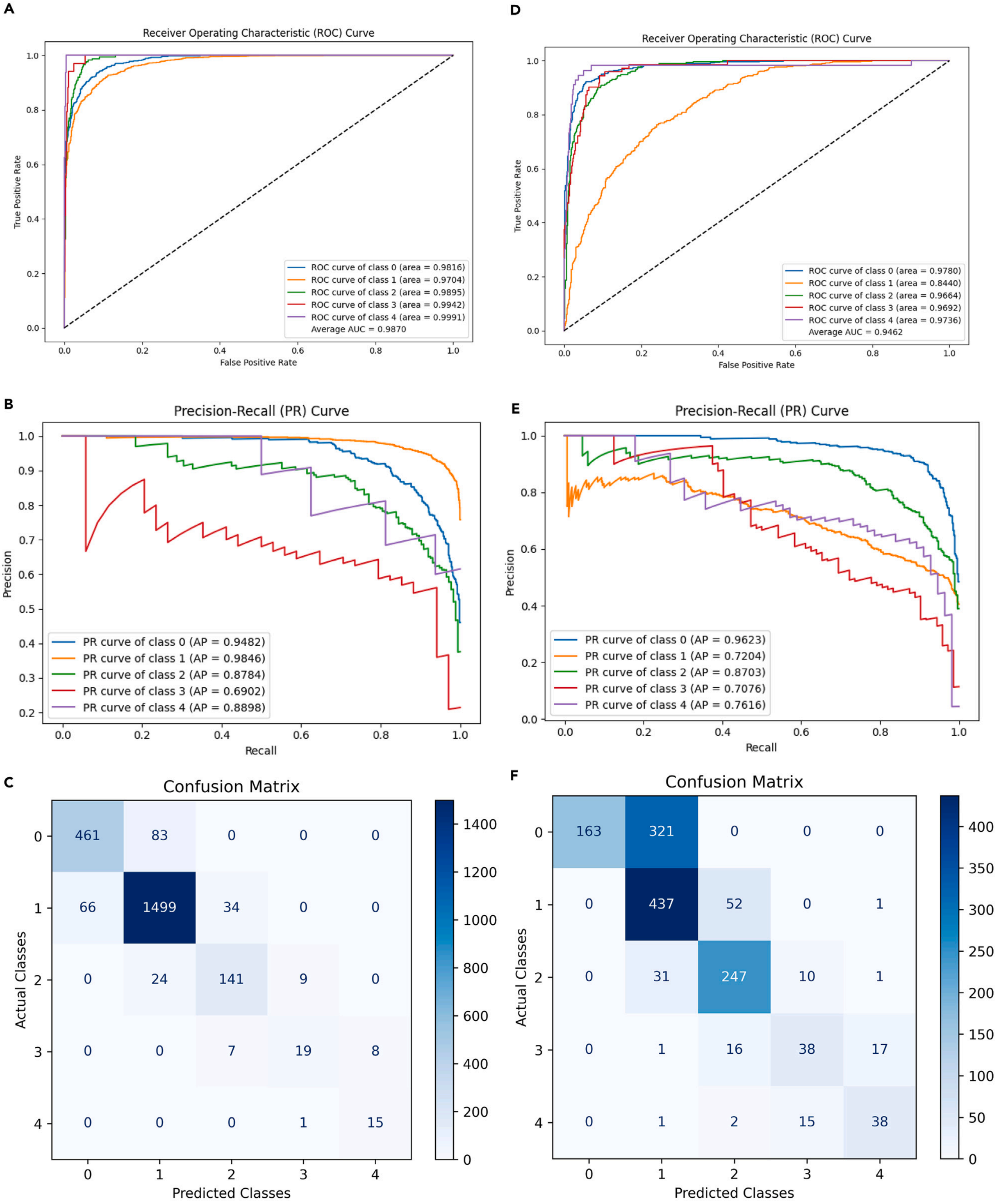


Figure 1. Performance evaluation of our proposed method using confusion matrix, ROC and PRC

(A) and (D) show the ROC curves and their AUROCs for binary classifications on internal and external datasets, respectively.

(B) and (E) depict the PRCs and corresponding AUPRCs for these tasks.

(C) and (F) present the confusion matrices for internal and external datasets, respectively. ROC, recipient operation curve; PRC, precision-recall curve; AUROC, area under recipient operation curve.

Table 1. Classification results of our proposed method

Dataset	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUROC	Kappa
Internal testing set	0.968(0.9664, 0.9733)	0.831(0.7945, 0.8901)	0.956(0.9547, 0.9649)	0.854(0.8018, 0.8916)	0.842(0.7979, 0.8849)	0.983(0.9843, 0.9889)	0.968
External testing set	0.890(0.8823, 0.8977)	0.717(0.6791, 0.7565)	0.878(0.8708, 0.8848)	0.691(0.6525, 0.7257)	0.683(0.6436, 0.7208)	0.951(0.9448, 0.9561)	0.819

the implementation of a CADx system for auxiliary reading in clinical practice primarily incurs economic costs. The results show that the net benefits of using SL-pretrained DenseNet121 and MMSSL-trained ResNet18 models are consistently lower, indicating these are suboptimal choices. Other models have similar and higher net benefits than treat-all, especially within a certain probability threshold range (45%–65%), where our model consistently offers the highest net benefit. This implies that when clinicians consider a risk level between 45% and 65% for high myopia as a threshold for further intervention, the use of our model in high myopia screening—taking into consideration the costs, inconvenience, side effects, and other adverse aspects of screening tests—offers the greatest net benefit.

DISCUSSION

In this study, we have for the first-time applied SSL pre-training method and ViT to the automated diagnosis and grading of MM. This research also represents one of the first evaluations of the RETFound foundation model in real-world screening scenarios. Our results suggest that our model achieves satisfactory performance in classification tasks according to META-PM criteria and remains stable in external testing. The implementation of our CADx system could improve both the accuracy and cost-efficiency of high myopia screenings in community hospitals.

Our developed model demonstrated satisfactory performance on an external testing set previously unseen by it. This can partly be attributable to our development dataset, which comprises 7,906 high-quality, well-annotated color fundus photographs sourced from community eye disease screenings in Shanghai (Table 4). These real-world data encompass heterogeneous image quality and diverse population sources, enhancing their representativeness. In contrast, previous studies have typically employed idealized datasets, either publicly accessible or sourced from clinical environments, for training and testing. However, these datasets frequently fail to encapsulate the complexities of

Table 2. Classification results of different models on internal testing set

Training Method	Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUROC
SSL							
MAE	RETFound (proposed method)	0.968(0.9664, 0.9733)	0.831(0.7945, 0.8901)	0.956(0.9547, 0.9649)	0.854(0.8018, 0.8916)	0.842(0.7979, 0.8849)	0.983(0.9843, 0.9889)
Uni4eye	ViT-L/16	0.946(0.9405, 0.9497)	0.677(0.6167, 0.7483)	0.929(0.9191, 0.9328)	0.756(0.6788, 0.8141)	0.713(0.6436, 0.7647)	0.961(0.9573, 0.9661)
Lesion-based CL	ResNet50	0.959(0.9562, 0.9640)	0.628(0.5782, 0.6980)	0.963(0.9599, 0.9696)	0.683(0.6051, 0.7530)	0.648(0.5823, 0.7043)	0.980(0.9777, 0.9835)
Multi-modal SSL	ResNet18	0.960(0.9557, 0.9638)	0.607(0.5382, 0.6682)	0.958(0.9554, 0.9656)	0.705(0.5853, 0.7610)	0.619(0.5479, 0.6752)	0.978(0.9751, 0.9813)
Rotation-Oriented	ResNet18	0.959(0.9570, 0.9647)	0.599(0.5371, 0.6692)	0.959(0.9576, 0.9676)	0.655(0.6070, 0.7541)	0.624(0.5711, 0.6974)	0.980(0.9779, 0.9836)
SL							
ImageNet	ViT-B/16	0.969(0.9651, 0.9722)	0.800(0.7410, 0.8592)	0.958(0.9527, 0.9633)	0.877(0.8475, 0.9086)	0.830(0.7767, 0.8769)	0.986(0.9835, 0.9883)
ImageNet	ViT-L/16	0.964(0.9601, 0.9677)	0.787(0.7293, 0.8389)	0.968(0.9627, 0.9723)	0.822(0.7612, 0.8725)	0.800(0.7423, 0.8460)	0.984(0.9809, 0.9862)
ImageNet	DenseNet-121	0.967(0.9635, 0.9708)	0.719(0.6564, 0.7862)	0.953(0.9477, 0.9589)	0.815(0.7457, 0.8685)	0.747(0.6734, 0.8107)	0.982(0.9782, 0.9849)
ImageNet	EfficientNet-B3	0.967(0.9639, 0.971)	0.778(0.7151, 0.8415)	0.957(0.9516, 0.9625)	0.863(0.8303, 0.8966)	0.809(0.7463, 0.8634)	0.985(0.9820, 0.987)
ImageNet	Vgg16	0.969(0.9649, 0.9723)	0.829(0.7704, 0.8790)	0.965(0.9597, 0.9699)	0.847(0.7875, 0.8960)	0.838(0.7783, 0.8817)	0.985(0.9819, 0.9870)
ImageNet	ResNet50	0.971(0.9670, 0.9739)	0.854(0.7970, 0.9034)	0.966(0.9612, 0.9712)	0.899(0.8701, 0.9261)	0.873(0.8279, 0.9107)	0.986(0.9841, 0.9887)

Table 3. Classification results of different models on external testing set

Training Method	Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUROC
SSL							
MAE	RETFound (proposed method)	0.890(0.8823, 0.8977)	0.717 (0.6791, 0.7565	0.878(0.8708, 0.8848)	0.691(0.6525, 0.7257)	0.683(0.6436, 0.7208)	0.951(0.9448, 0.9561)
Uni4eye	ViT-L/16	0.885(0.8767, 0.8934)	0.595(0.5575, 0.6346)	0.893(0.8863, 0.9000)	0.594(0.5497, 0.6401)	0.582(0.5418, 0.6228)	0.923(0.9151, 0.9330)
Lesion-based CL	ResNet50	0.866(0.8575, 0.8737)	0.611(0.5773, 0.6460)	0.862(0.8553, 0.8685)	0.543(0.5055, 0.5790)	0.543(0.5246, 0.5916)	0.921(0.9123, 0.9289)
Multi-modal SSL	ResNet18	0.863(0.8548, 0.8710)	0.575(0.5393, 0.6135)	0.851(0.8448, 0.8578)	0.577(0.5261, 0.6247)	0.546(0.5059, 0.5848)	0.909(0.8983, 0.9182)
Rotation-Oriented	ResNet18	0.863(0.8582, 0.8744)	0.575(0.5221, 0.5952)	0.851(0.8515, 0.8652)	0.577(0.5583, 0.6643)	0.546(0.4858, 0.5700)	0.913(0.9013, 0.9233)
SL							
ImageNet	ViT-B/16	0.831(0.8221, 0.8393)	0.608(0.5729, 0.6469)	0.832(0.8251, 0.8388)	0.504(0.4653, 0.5419)	0.529(0.4935, 0.5655)	0.888(0.8774, 0.899)
ImageNet	ViT-L/16	0.874(0.8656, 0.8825)	0.651(0.6139, 0.6900)	0.880(0.8723, 0.8868)	0.589(0.5471, 0.6311)	0.611(0.5725, 0.6489)	0.927(0.9198, 0.9340)
ImageNet	DenseNet-121	0.845(0.8361, 0.8533)	0.652(0.6114, 0.6936)	0.834(0.8277, 0.8401)	0.612(0.5744, 0.6504)	0.574(0.5331, 0.6102)	0.899(0.8889, 0.9085)
ImageNet	EfficientNet-B3	0.876(0.8674, 0.8835)	0.662(0.6261, 0.6972)	0.865(0.8581, 0.8721)	0.670(0.6289, 0.7077)	0.612(0.5691, 0.6522)	0.936(0.9287, 0.9430)
ImageNet	Vgg16	0.877(0.8684, 0.885)	0.679(0.6386, 0.7191)	0.871(0.8636, 0.8777)	0.615(0.5754, 0.6543)	0.627(0.5867, 0.6654)	0.923(0.9147, 0.9299)
ImageNet	ResNet50	0.888(0.8794, 0.8959)	0.649(0.6130, 0.6916)	0.897(0.8902, 0.9039)	0.579(0.5420, 0.6160)	0.605(0.5693, 0.6410)	0.929(0.9212, 0.9369)

real-world conditions, often resulting in significant performance degradation during external testing or practical application.[22–24] Furthermore, while some studies have reported favorable outcomes,[25,26] their classification criteria do not strictly adhere to the META-PM standards (Table 5). A significant factor is the low prevalence of high-grade lesions, resulting in a scarcity of corresponding images, which hampers adequate model training. By aggregating multiple high-grade lesions into a single category, performance can be effectively enhanced; however, this approach compromises the clinical utility of the model.

On the other hand, this success can also be attributed to the more advanced methods used in this study. It is well-known that CNNs have been the standard for automated medical image diagnosis in the last decade.[27] Transformers, notably the ViT, have started to gain prominence recently.[28–30] Unlike CNNs, which utilize local connections and shared parameters, Transformers can learn the relationships between various image segments and handling long-distance dependencies. This is particularly advantageous for fundus lesions, which typically exhibit alterations in multiple regions or structures. By exploiting the pathological links between these lesions, transformers have already demonstrated excellent results in DR tasks[31] and have been used in retinal blood vessel segmentation,[32] detection of retinal injuries,[33] prediction of visual field changes,[34] and analysis of choroidal structures.[33] MM typically presents as diverse alterations across the fundus, including stretching of retinal and choroidal regions along with vascular changes. Prior studies by Lu et al.[35] and Daniel et al.[36] have effectively utilized ResNet, a CNN-based architecture, and achieved notable results. Additionally, Du et al.[37] employed the more recent EfficientNet model, achieving an accuracy of 92.08% in a binary classification task for detecting pathologic myopia. In contrast, our research not only confirms the feasibility of utilizing transformer-based deep learning models for this task but also underscores the transformer's aptitude for analyzing MM fundus images. This effectiveness is largely attributable to its superior capability in modeling global image dependencies, which is critical for detecting the complex patterns inherent in MM.

As mentioned previously, the remarkable performance of our model can also be attributed to the application of advanced SSL pre-training methods. Traditional AI model development relies on pre-training with extensively annotated data, often limited in dataset size.[38] In contrast, the RETFound foundation model used in our study underwent self-supervised training on 1.6 million unlabeled fundus images. By leveraging generalized representations learned from these unlabeled retinal images, our method could unearth potential information across a broader range of data. This superior performance of the SSL model can be ascribed to its ability to learn without the limitations imposed by restricted and potentially biased annotated datasets, effectively harnessing the Transformer's self-attention mechanism for a more thorough capture of crucial image features. Such an attribute is particularly vital in processing images of MM, which are characterized by their high complexity and rich detail. Practically, in ophthalmic clinical practice filled with opportunities for automation, it is almost impossible to gather well-annotated

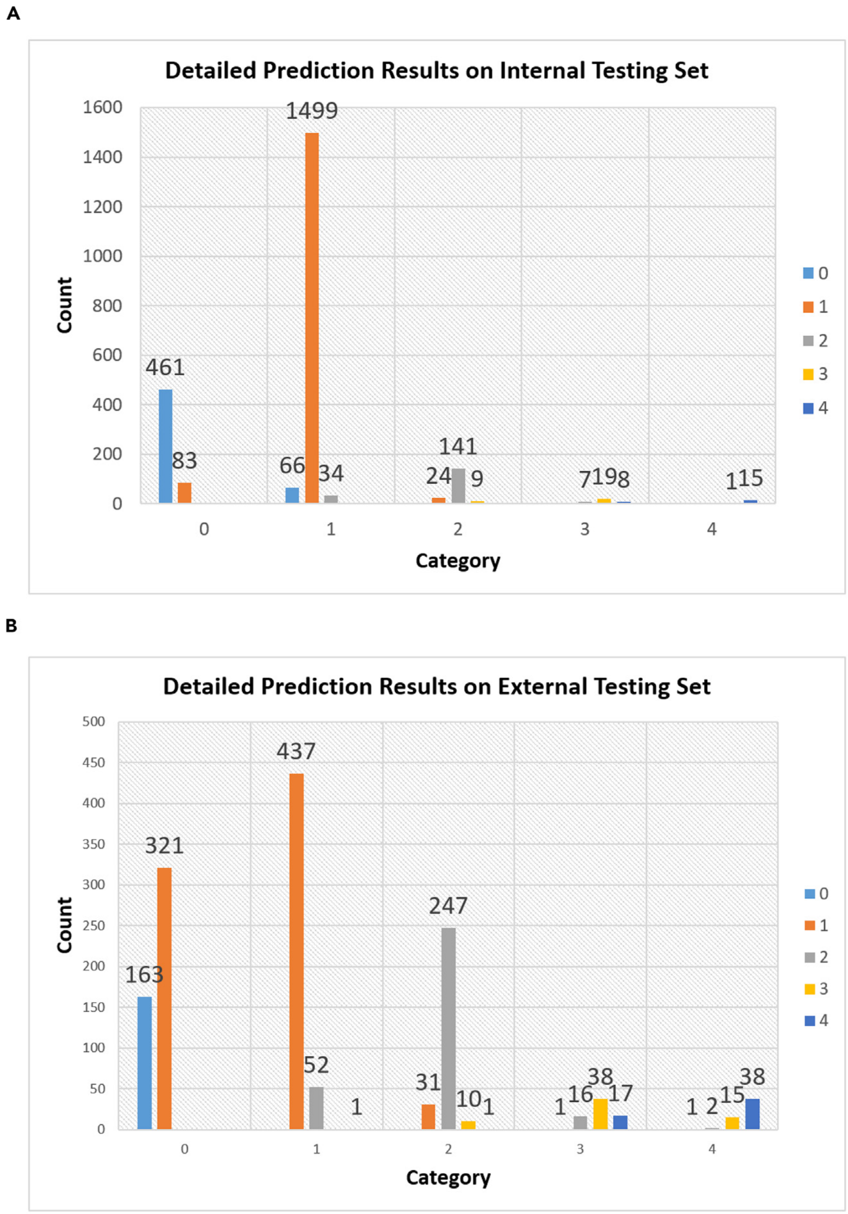


Figure 2. Analysis of prediction results using our proposed method

(A) and (B) show the distribution of various categories within the actual datasets on internal and external datasets, respectively. The x axis represents the true categories of the samples, and each bar represents the category into which that class of samples has been predicted.

datasets for all tasks to train supervised models. Thus, the approach adopted in our study, involving an initial phase of SSL pre-training followed by targeted fine-tuning for a specific downstream task, appears to be a more viable strategy. For ophthalmologists, this approach is less challenging to implement and offers higher label efficiency. In this context, our study extends beyond the development of a CADx system for the automatic diagnosis of MM. It also showcases the exceptional performance that foundational models like RETFound can achieve in specific downstream tasks. Following the methodology of our study, employing a foundational model to develop a specialized model for any automated task involving color fundus images or optical coherence tomography (OCT) fundus images becomes a streamlined process.

This is the first study that used DCA to evaluate DL models for automatic myopia screening, especially for MM screening. DCA has distinctive advantages in providing practical insights for clinical decision processes, especially when evaluating the potential benefits and risks of diagnostic tests. From the DCA results of our study, it was observed that our method had a higher net benefit in MM screening than other models, with the accuracy of screening being crucial. Our earlier research pointed out that in developing countries, AI-based community eye disease screening may not always be more cost-effective than tele-screening technologies due to

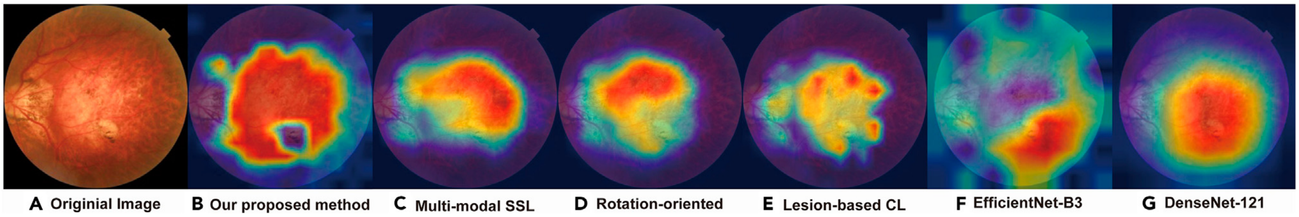


Figure 3. The typical "true positive" example, and the corresponding heatmaps of different models

(A) is the original picture, which does not involve any model.

(B–G) correspond to the following models: our proposed SSL-enhanced model, multi-modal SSL model, Rotation-oriented model, Lesion-based CL model, EfficientNet-B3 model, DenseNet-121 model. The color gradients utilized in the saliency maps serve to illustrate the magnitude of significance attributed to the corresponding regions within the fundus image in relation to the prediction. Areas exhibiting warmer colors denote heightened importance, whereas cooler colors signify diminished importance.

lower human resource costs for medical personnel.[39] However, this study demonstrates that the use of the free, open-source RETFound framework in community high myopia screening not only enhances accuracy but also increases net benefits. Future research should assess the health economic impact of the RETFound framework in real-world scenarios to further support and validate our results. Nonetheless, our study did not account for patient refusal rates, a common and inevitable factor in screening processes, due to the lack of available estimation data. Moreover, TestHarm is a generalized estimate based on clinicians' subjective assessments, which may vary across different contexts and populations. This variability implies that if TestHarm or refusal rates change significantly, conclusions might differ. For example, in communities more open to automated screening systems, the use of advanced CADx systems might be more welcomed. Conversely, a certain level of expert physician involvement could gain greater acceptance. Addressing this estimation challenge could involve conducting sensitivity analyses. However, current methodologies do not provide for such analyses using DCA, indicating a need for future research.

Limitations of the study

Despite the outstanding performance of our proposed model, there are still some limitations. First, while significant efforts have been made to expand our dataset, the real-world data are primarily sourced from Shanghai, China and does not describe features, such as posterior scleral staphyloma or "Plus" lesions (Fuchs spots, lacquer cracks, and choroidal neovascularization). More images of high myopia patients will be collected in future studies. Second, our approach did not utilize multimodal data. The growing use of OCT and ultra-wide-field fundus photography in examining high myopia patients underscores the importance of exploring the integration of diverse examination results and even textual medical records. Third, the presence of diseases other than myopic macular lesions, which is typical in practical use, was not accounted for. There is existing research showcasing the possibility of detecting multiple ocular diseases from fundus photographs, which is an avenue for future exploration.

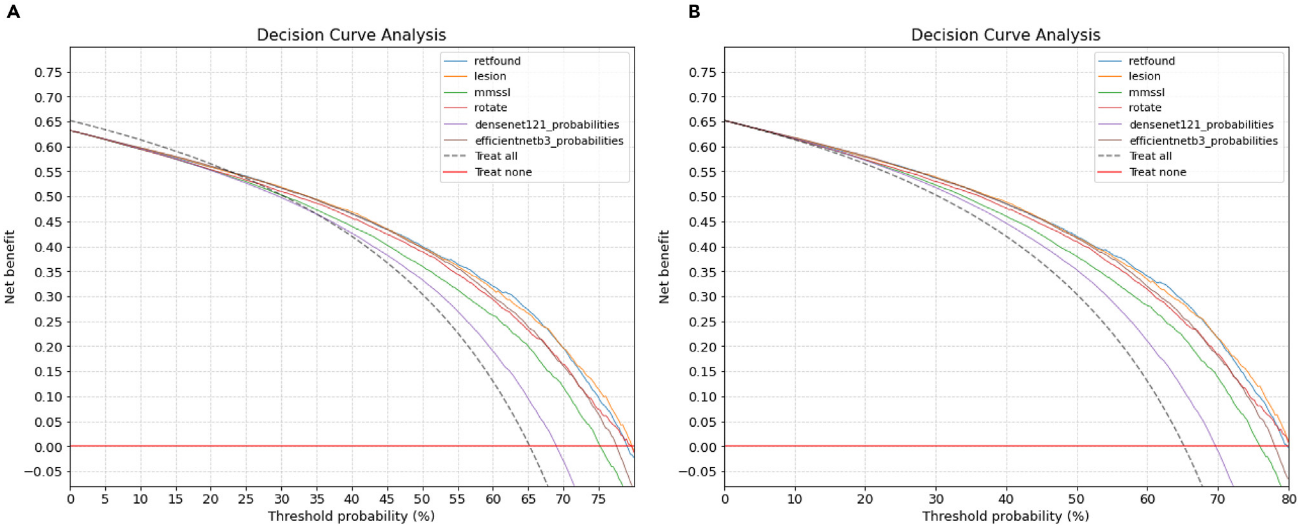


Figure 4. Comparison of using our proposed method and other models on the external testing set

(A) with TestHarm set as 0.02 (B) regardless of TestHarm.

Table 4. Detailed definition of MM classification criteria

Category	Definition
C0	No macular lesions
C1	Tessellated fundus
C2	Diffuse chorioretinal atrophy
C3	Patchy chorioretinal atrophy
C4	Macular atrophy

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Study design
  - Study approval
  - Data acquisition and quality control
  - Development of the CADx system
  - Evaluation of the CADx system
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Decision curve analysis
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110566.

ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China (project no. 2021YFC2702100), Shanghai engineering research center of precise diagnosis and treatment of eye diseases, Shanghai, China ((project no. 19DZ2250100) and Three-year Action Program of Shanghai Municipality for Strengthening the Construction of Public Health System (GWVI-11.1-22).

Table 5. Characteristics of the total dataset and external testing dataset

Characteristics	Training set	Internal testing set	External testing set
Dataset	SHMSP	SHMSP	MMAC2023
Number of fundus images	6303	1603	1391
Collection device	TOPCON DRI Triton	TOPCON DRI Triton	Not mentioned
Number of patients	5061	818	Not mentioned
Age (years), mean ± SD	56.8 ± 19.8	56.3 ± 19.1	53.4 ± 10.8
Gender (female, %)	60.8%	61.3%	Not mentioned
Images in each category			
0	1437	380	404
1	4265	1070	412
2	465	117	224
3	91	25	60
4	45	11	43

AUTHOR CONTRIBUTIONS

J.Z., F.X., H.Z., and J.H. conceptualized the study. J.Z. and J.H. collected the data. F.X. and J.Z. developed the system, conducted formal analysis, and prepared the first draft of the manuscript. H.Z., J.H., and R.F. critically reviewed drafts of the manuscript. All authors approved the final version to be submitted.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 8, 2024
Revised: April 28, 2024
Accepted: July 18, 2024
Published: July 23, 2024

REFERENCES

1. Wong, Y.L., and Saw, S.M. (2016). Epidemiology of Pathologic Myopia in Asia and Worldwide. Asia. Pac. J. Ophthalmol. *5*, 394–402. https://doi.org/10.1097/apo.0000000000000234.
2. Vitale, S., Sperduto, R.D., and Ferris, F.L. (2009). Increased prevalence of myopia in the United States between 1971-1972 and 1999-2004. Arch. Ophthalmol. *127*, 1632–1639. https://doi.org/10.1001/archophthalmol.2009.303.
3. Bikbov, M.M., Gilmanshin, T.R., Kazakbaeva, G.M., Zainullin, R.M., Rakhimova, E.M., Rusakova, I.A., Bolshakova, N.I., Safiullina, K.R., Zaynetdinov, A.F., Zinatullin, A.A., et al. (2020). Prevalence of Myopic Maculopathy Among Adults in a Russian Population. JAMA Netw. Open *3*, e200567. https://doi.org/10.1001/jamanetworkopen.2020.0567.
4. Holden, B.A., Fricke, T.R., Wilson, D.A., Jong, M., Naidoo, K.S., Sankaridurg, P., Wong, T.Y., Naduvilath, T.J., and Resnikoff, S. (2016). Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. Ophthalmology *123*, 1036–1042. https://doi.org/10.1016/j.ophtha.2016.01.006.
5. Ohno-Matsui, K., Shimada, N., Yasuzumi, K., Hayashi, K., Yoshida, T., Kojima, A., Moriyama, M., and Tokoro, T. (2011). Long-term development of significant visual field defects in highly myopic eyes. Am. J. Ophthalmol. *152*, 256–265.e1. https://doi.org/10.1016/j.ajo.2011.01.052.
6. Hopf, S., Korb, C., Nickels, S., Schulz, A., Münzel, T., Wild, P.S., Michal, M., Schmidtmann, I., Lackner, K.J., Pfeiffer, N., and Schuster, A.K. (2020). Prevalence of myopic maculopathy in the German population: results from the Gutenberg health study. Br. J. Ophthalmol. *104*, 1254–1259. https://doi.org/10.1136/bjophthalmol-2019-315255.
7. Xu, L., Wang, Y., Li, Y., Wang, Y., Cui, T., Li, J., and Jonas, J.B. (2006). Causes of blindness and visual impairment in urban and rural areas in Beijing: the Beijing Eye Study. Ophthalmology *113*, 1134.e1-11. https://doi.org/10.1016/j.ophtha.2006.01.035.
8. Wong, T.Y., Ferreira, A., Hughes, R., Carter, G., and Mitchell, P. (2014). Epidemiology and disease burden of pathologic myopia and myopic choroidal neovascularization: an evidence-based systematic review. Am. J. Ophthalmol. *157*, 9–25.e12. https://doi.org/10.1016/j.ajo.2013.08.010.
9. Fricke, T.R., Jong, M., Naidoo, K.S., Sankaridurg, P., Naduvilath, T.J., Ho, S.M., Wong, T.Y., and Resnikoff, S. (2018). Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: systematic review, meta-analysis and modelling. Br. J. Ophthalmol. *102*, 855–862. https://doi.org/10.1136/bjophthalmol-2017-311266.
10. Hayashi, K., Ohno-Matsui, K., Shimada, N., Moriyama, M., Kojima, A., Hayashi, W., Yasuzumi, K., Nagaoka, N., Saka, N., Yoshida, T., et al. (2010). Long-term pattern of progression of myopic maculopathy: a natural history study. Ophthalmology *117*, 1595-611–1611.e1-4. https://doi.org/10.1016/j.ophtha.2009.11.003.
11. Resnikoff, S., Lansingh, V.C., Washburn, L., Felch, W., Gauthier, T.M., Taylor, H.R., Eckert, K., Parke, D., and Wiedemann, P. (2020). Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? Br. J. Ophthalmol. *104*, 588–592. https://doi.org/10.1136/bjophthalmol-2019-314336.
12. Fang, Y., Du, R., Nagaoka, N., Yokoi, T., Shinohara, K., Xu, X., Takahashi, H., Onishi, Y., Yoshida, T., and Ohno-Matsui, K. (2019). OCT-Based Diagnostic Criteria for Different Stages of Myopic Maculopathy. Ophthalmology *126*, 1018–1032. https://doi.org/10.1016/j.ophtha.2019.01.012.
13. Ruiz-Medrano, J., Montero, J.A., Flores-Moreno, I., Arias, L., García-Layana, A., and Ruiz-Moreno, J.M. (2019). Myopic maculopathy: Current status and proposal for a new classification and grading system (ATN). Prog. Retin. Eye Res. *69*, 80–115. https://doi.org/10.1016/j.preteyeres.2018.10.005.
14. Avila, M.P., Weiter, J.J., Jalkh, A.E., Trempe, C.L., Pruett, R.C., and Schepens, C.L. (1984). Natural history of choroidal neovascularization in degenerative myopia. Ophthalmology *91*, 1573–1581. https://doi.org/10.1016/s0161-6420(84)34116-1.
15. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA *316*, 2402–2410. https://doi.org/10.1001/jama.2016.17216.
16. Gutierrez, L., Lim, J.S., Foo, L.L., Ng, W.Y., Yip, M., Lim, G.Y.S., Wong, M.H.Y., Fong, A., Rosman, M., Mehta, J.S., et al. (2022). Application of artificial intelligence in cataract management: current and future directions. Eye Vis. *9*, 3. https://doi.org/10.1186/s40662-021-00273-z.
17. Peng, Y., Dharssi, S., Chen, Q., Keenan, T.D., Agrón, E., Wong, W.T., Chew, E.Y., and Lu, Z. (2019). DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs. Ophthalmology *126*, 565–575. https://doi.org/10.1016/j.ophtha.2018.11.015.
18. Li, Z., He, Y., Keel, S., Meng, W., Chang, R.T., and He, M. (2018). Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology *125*, 1199–1206. https://doi.org/10.1016/j.ophtha.2018.01.023.
19. Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J., and Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. *15*, e1002683. https://doi.org/10.1371/journal.pmed.1002683.
20. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G.A. (2020). Simple Framework for Contrastive Learning of Visual Representations. In International conference on machine learning (PMLR), pp. 1597–1607.
21. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., 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.
22. Hemelings, R., Elen, B., Blaschko, M.B., Jacob, J., Stalmans, I., and De Boever, P. (2021). Pathological myopia classification with simultaneous lesion segmentation using deep learning. Comput. Methods Programs Biomed. *199*, 105920. https://doi.org/10.1016/j.cmpb.2020.105920.
23. Devda, J., and Eswari, R. (2019). Pathological Myopia Image Analysis Using Deep Learning. Procedia Comput. Sci. *165*, 239–244. https://doi.org/10.1016/j.procs.2020.01.084.
24. Li, M., Liu, S., Wang, Z., Li, X., Yan, Z., Zhu, R., and Wan, Z. (2023). MyopiaDETR: End-to-end pathological myopia detection based on transformer using 2D fundus images. Front.

Neurosci. *17*, 1130609. https://doi.org/10.3389/fnins.2023.1130609.

25. Wang, R., He, J., Chen, Q., Ye, L., Sun, D., Yin, L., Zhou, H., Zhao, L., Zhu, J., Zou, H., et al. (2023). Efficacy of a Deep Learning System for Screening Myopic Maculopathy Based on Color Fundus Photographs. Ophthalmol. Ther. *12*, 469–484. https://doi.org/10.1007/s40123-022-00621-9.
26. Li, J., Wang, L., Gao, Y., Liang, Q., Chen, L., Sun, X., Yang, H., Zhao, Z., Meng, L., Xue, S., et al. (2022). Automated detection of myopic maculopathy from color fundus photographs using deep convolutional neural networks. Eye Vis. *9*, 13. https://doi.org/10.1186/s40662-022-00285-3.
27. Zhang, J., and Zou, H. (2023). Artificial intelligence technology for myopia challenges: A review. Front. Cell Dev. Biol. *11*, 1124005. https://doi.org/10.3389/fcell.2023.1124005.
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention Is All You Need. Adv. Neural Inf. Process. Syst. 30.
29. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at arxiv. https://doi.org/10.48550/arXiv.2010.11929.
30. Khan, A., Rauf, Z., Sohail, A., Rehman, A., Asif, H., Asif, A., and Farooq, U. (2023). A Survey of the Vision Transformers and its CNN-Transformer Based Variants. Preprint at arxiv. https://doi.org/10.1007/s10462-023-10595-0.
31. Huang, S., Li, J., Xiao, Y., Shen, N., and Xu, T. (2022). RTNet: Relation Transformer Network for Diabetic Retinopathy Multi-Lesion Segmentation. IEEE Trans. Med. Imaging *41*, 1596–1607. https://doi.org/10.1109/tmi.2022.3143833.
32. Shen, X., Xu, J., Jia, H., Fan, P., Dong, F., Yu, B., and Ren, S. (2022). Self-attentional microvessel segmentation via squeeze-excitation transformer Unet. Comput. Med. Imaging Graph. *97*, 102055. https://doi.org/10.1016/j.compmedimag.2022.102055.
33. Philippi, D., Rothaus, K., and Castelli, M. (2023). A vision transformer architecture for the automated segmentation of retinal lesions in spectral domain optical coherence tomography images. Sci. Rep. *13*, 517. https://doi.org/10.1038/s41598-023-27616-1.
34. Hou, K., Bradley, C., Herbert, P., Johnson, C., Wall, M., Ramulu, P.Y., Unberath, M., and Yohannan, J. (2023). Predicting Visual Field Worsening with Longitudinal OCT Data Using a Gated Transformer Network. Ophthalmology *130*, 854–862. https://doi.org/10.1016/j.ophtha.2023.03.019.
35. Lu, L., Ren, P., Tang, X., Yang, M., Yuan, M., Yu, W., Huang, J., Zhou, E., Lu, L., He, Q., et al. (2021). AI-Model for Identifying Pathologic Myopia Based on Deep Learning Algorithms of Myopic Maculopathy Classification and "Plus" Lesion Detection in Fundus Images. Front. Cell Dev. Biol. *9*, 719262. https://doi.org/10.3389/fcell.2021.719262.
36. Tan, T.E., Anees, A., Chen, C., Li, S., Xu, X., Li, Z., Xiao, Z., Yang, Y., Lei, X., Ang, M., et al. (2021). Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. Lancet. Digit. Health *3*, e317–e329. https://doi.org/10.1016/s2589-7500(21)00055-8.
37. Du, R., Xie, S., Fang, Y., Igarashi-Yokoi, T., Moriyama, M., Ogata, S., Tsunoda, T., Kamatani, T., Yamamoto, S., Cheng, C.Y., et al. (2021). Deep Learning Approach for Automated Detection of Myopic Maculopathy and Pathologic Myopia in Fundus Images. Ophthalmol. Retina *5*, 1235–1244. https://doi.org/10.1016/j.oret.2021.02.006.
38. Zhang, J., and Zou, H. (2024). Insights into artificial intelligence in myopia management: from a data perspective. Graefes Arch. Clin. Exp. Ophthalmol. *262*, 3–17. https://doi.org/10.1007/s00417-023-06101-5.
39. Lin, S., Ma, Y., Xu, Y., Lu, L., He, J., Zhu, J., Peng, Y., Yu, T., Congdon, N., and Zou, H. (2023). Artificial Intelligence in Community-Based Diabetic Retinopathy Telemedicine Screening in Urban China: Cost-effectiveness and Cost-Utility Analyses With Real-world Data. JMIR Public Health Surveill. *9*, e41624. https://doi.org/10.2196/41624.
40. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int. J. Comput. Vis. *128*, 336–359. https://doi.org/10.1007/s11263-019-01228-7.
41. Ohno-Matsui, K., Kawasaki, R., Jonas, J.B., Cheung, C.M.G., Saw, S.M., Verhoeven, V.J.M., Klaver, C.C.W., Moriyama, M., Shinohara, K., Kawasaki, Y., et al. (2015). International photographic classification and grading system for myopic maculopathy. Am. J. Ophthalmol. *159*, 877–883.e7. https://doi.org/10.1016/j.ajo.2015.01.022.
42. Lu, L., Zhou, E., Yu, W., Chen, B., Ren, P., Lu, Q., Qin, D., Lu, L., He, Q., Tang, X., et al. (2021). Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images. Commun. Biol. *4*, 1225. https://doi.org/10.1038/s42003-021-02758-y.
43. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15979–15988.
44. Huang, S.-C., Pareek, A., Jensen, M., Lungren, M.P., Yeung, S., and Chaudhari, A.S. (2023). Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digit. Med. *6*, 74. https://doi.org/10.1038/s41746-023-00811-0.
45. Huang, Y., Lin, L., Cheng, P., Lyu, J., and Tang, X. (2021). Lesion-Based Contrastive Learning for Diabetic Retinopathy Grading Fro M Fundus Images (Springer International Publishing). https://doi.org/10.1007/978-3-030-87196-3_11.
46. Cai, Z., Lin, L., He, H., and Tang, X. (2022). Uni4Eye: Unified 2D and 3D Self-Supervised Pre-training via Masked Image Modeling Transformer for Ophthalmic Image Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention, L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, and S. Li, eds. (Springer Nature Switzerland), pp. 88–98.
47. Wei, W., Huang, C., Xia, L., and Zhang, C. (2023). Multi-Modal Self-Supervised Learning for Recommendation. In Proceedings of the ACM Web Conference 2023, pp. 790–800.
48. Li, X., Hu, X., Qi, X., Yu, L., Zhao, W., Heng, P.A., and Xing, L. (2021). Rotation-Oriented Collaborative Self-Supervised Learning for Retinal Disease Diagnosis. IEEE Trans. Med. Imaging *40*, 2284–2294. https://doi.org/10.1109/TMI.2021.3075244.
49. Vickers, A.J., Van Calster, B., and Steyerberg, E.W. (2016). Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ *352*, i6. https://doi.org/10.1136/bmj.i6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
RETFound foundation model	Zhou et al.[21]	https://github.com/rmaphoh/RETFound_MAE
Gradient-weighted Class Activation Mapping (Grad-CAM)	Selvaraju et al.[40]	https://github.com/ramprs/grad-cam/
PyTorch	Version 1.8.1	https://pytorch.org/blog/pytorch-1.8-released/
Matplotlib	Version 3.6.3	https://matplotlib.org/3.6.3/
Scikit-learn	Version 0.24.2	https://pypi.org/project/scikit-learn/0.24.2/
SAS	Version 9.4	https://www.sas.com/zh_cn/home.html
Python	Version 3.9.17	https://www.python.org/downloads/release/python-3917/
Our SSL-DL method for automatic identification of MM	The code for fine tuning and validation of our model.	https://github.com/Akemimadokami/ssl-enhanced-DL-for-MM-in-High-Myopia
Deposited data		
MMAC2023	The 26th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)	https://codalab.lisn.upsaclay.fr/competitions/12441

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the lead contact, Dr. Jiangnan He (hejiangnan85@126.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- SHMSP dataset contain human-related data, which is governed by the Ministry of Science and Technology of China (MOST) in accordance with the Regulations of the People's Republic of China on Administration of Human Genetic Resources (State Council No.717). Request for the non-profit use of the fundus images in the SHMSP should be sent to the lead contact.
- All original code can be found in GitHub as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This research incorporated 9,297 color fundus photographs (Figure S1). The development dataset was sourced from the Shanghai High Myopia Screening Project, while the external testing set was derived from the MMAC2023 public dataset (https://codalab.lisn.upsaclay.fr/competitions/12441). Both datasets were manually reviewed by professional ophthalmologists using uniform classification criteria, thereby ensuring consistency of the results. Detailed characteristics of the data are presented in Table 5.

The development dataset comprised color fundus photographs collected between 2016 and 2018 from the aforementioned screening program. Images were captured using the TOPCON DRI Triton device, centering on the macula. All images were independently classified and annotated by three ophthalmologists, with discrepancies resolved through group discussions involving the ophthalmologists and a senior retinal specialist. All images were re-evaluated based on specific criteria, excluding those that did not meet the following standards: incomplete visibility of the fovea or more than 50% obscuration, blurriness, significant artifacts, low contrast, uneven illumination, and excessive reflection. Ultimately, the development dataset consisted of 7,906 images from 5,879 patients (Figure S1). We randomly divided the development dataset into a training set (80%) and an internal testing set (20%), ensuring that the proportions of different categories were maintained consistently across both subsets and that there was no image-level overlap.

To evaluate the generalizability of our model, we further conducted external testing with the MMAC2023 (Myopic Maculopathy Analysis Challenge 2023) public dataset. The MMAC2023, a competition for myopic maculopathy image analysis, was released at the 26th

International Conference on Medical Image Computing and Computer Assisted Intervention. The dataset comprises 1,391 color fundus photographs annotated by experts. Each image conforms to the META-PM criteria for labeling, ensuring standardization across the dataset.

METHOD DETAILS

Study design

This study is a cross-sectional analysis using data from the Shanghai High Myopia Screening Project (SHMSP), carried out by the Shanghai Eye Disease Prevention and Control Center. As a leading ophthalmology specialty hospital in Shanghai, the center is in charge of several community-based eye disease screening programs and plays a crucial role in the prevention and treatment of eye diseases in Shanghai, China. The study's methodology encompasses three main stages. Initially, we collected 45° fundus images, which were subsequently reviewed and graded by experienced ophthalmologists. Following standardized preprocessing procedures, these images were saved in jpg format, with the diagnoses provided by the specialists used as labels. In the second step, we developed a computer-aided diagnosis (CADx) system based on SSL and the Transformer framework. The final step of our research involved a comparative analysis of our model's effectiveness in MM screening with existing approaches, focusing on two aspects: 1) evaluating common model performance metrics based on confusion matrices and receiver operating characteristic curves in both internal and external testing sets; 2) employing DCA to evaluate the impact of the widespread use of related models on the net benefits of implementing high myopia screening. Additionally, we generated heatmaps to visually represent the model's predictive process.

Study approval

Our research protocol adhered to the principles of the Declaration of Helsinki and was approved by the Ethics Committee of the First People's Hospital affiliated with Shanghai Jiao Tong University School of Medicine (Approval No. 2015KY156). Informed consent was obtained from all participants. All images were subjected to irreversible anonymization. This study exclusively utilized retrospective data and did not involve any active participation from the patients. No commercial interests were implicated in the design and conduct of this study.

Data acquisition and quality control

In the absence of a standardized classification system for MM, we adopted the META-PM criteria for grading.[41] The META-PM is a globally recognized pathologic myopia classification system based on color fundus photographs.[25,35,37,42] As outlined in Table 4, it categorizes myopic maculopathy into five groups: no maculopathy (Category 0), tessellated fundus (Category 1), diffuse choroidal atrophy (Category 2), patchy choroidal atrophy (Category 3), and macular atrophy (Category 4).

Development of the CADx system

We developed a CADx system for the automatic grading of MM from color fundus photographs. The approach adopted involves an initial phase of SSL pre-training, succeeded by fine-tuning for the specific downstream task. The pre-training was implemented based on the RETFound framework, utilizing the Mask Autoencoder (MAE)[43] to learn from 1.6 million unlabeled fundus images, allowing the model to generalize representations of fundus tissue structures. Following this, the model underwent fine-tuning on our training dataset for the specific downstream task, aiming to produce labels that correspond to the image labels. All images in the training dataset were resized to 256×256 pixels. Data augmentation techniques similar to those used in pre-training were applied, including random cropping (20% to 100% of the image), resizing cropped images to 224×224, random horizontal flipping, and image normalization. Training was conducted on 4 NVIDIA GeForce RTX 2080 Ti GPUs, using CUDA version 11.1, an Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz, on an Ubuntu 18.04 system with 86GB memory. The primary hyperparameters utilized in our study included a batch size of 16 and a total training duration of 50 epochs. We employed a cosine annealing algorithm to adjust the learning rate, initiating with a warm-up phase during the first 10 epochs (gradually increasing the learning rate from 0 to 5×10^-4), followed by a cosine annealing schedule that progressively reduced the learning rate from 5× 10^-4 to 1×10^-6 over the subsequent 40 epochs. After each epoch, the model was evaluated on the internal validation dataset, and the model weights with the highest AUROC were saved as checkpoints for subsequent evaluation and DCA analysis.

Evaluation of the CADx system

For a comprehensive evaluation of our system, several prevalent deep learning models were implemented. SSL typically encompasses four strategies: Contrastive, Self-prediction, Generative, and Innate relationship.[44] Accordingly, we employed Lesion-based CL for Contrastive SSL,[45] Uni4eye for Self-prediction SSL,[46] MMSSL for Generative SSL,[47] and Rotation-oriented for Innate relationship SSL[48] to train Transformer models (ViT-L/16) and CNN models (ResNet). Additionally, we developed two Transformer models (ViT-B/16 and ViT-L/16) and three CNN models (DenseNet-121, EfficientNet-b3, and Vgg16) using SL methods. Each model was trained using the same methodology and tested on identical datasets to ensure fairness and consistency in comparisons.

QUANTIFICATION AND STATISTICAL ANALYSIS

We compiled a confusion matrix and calculated key metrics such as sensitivity, precision, specificity, AUROC, and AUPRC. Considering the imbalance in the number of samples across different categories in our dataset, we employed Cohen''s Kappa (κ) as a statistical measure to gauge the level of agreement in classification or grading data among observers. The κ value ranges from 0 to 1, with values greater than 0.8 generally indicating almost perfect agreement. Additionally, we utilized the F1 score to assess the model's effectiveness in differentiating between positive and negative cases. The F1 score, ranging from 0 to 1, indicates that higher scores reflect the model's ability to correctly identify positive instances while minimizing false positives. The relationships between these parameters and the metrics are defined by the following equations.

$$Sensitivity = \frac{TP}{TP+FN} \tag{Equation 1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{Equation 2}$$

$$Precision = \frac{TP}{TP+FP} \tag{Equation 3}$$

$$F1\ score = \frac{2 * TP}{2 * TP+FP+FN} \tag{Equation 4}$$

$$Cohen's\ kappa = \frac{p_0 - p_e}{1 - p_e} \tag{Equation 5}$$

In the formula for Cohen's kappa calculation, p_0 is the observed agreement (the proportion of samples correctly classified), and p_e is the expected agreement under random chance. These metrics were calculated using Python 3.9.17.

To gain a deeper understanding of the predictive process of deep learning models, we conducted a significance analysis using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique.[40] This method generates colored heatmaps by employing a specific feature weight calculation, highlighting the contribution of different regions of the image to the detection of MM. This analysis enabled us to identify regions that significantly contribute to the decision-making process of the system.

Decision curve analysis

Unlike traditional evaluation metrics that fail to consider clinical consequences, DCA is a statistical technique focused on evaluating the 'net benefit' of predictive models or diagnostic tests.[49] This method accounts for critical clinical considerations, such as the consequences of a missed diagnosis (false negative) compared to unnecessary treatment (false positive). DCA evaluates a model's net benefit by comparing it to fundamental strategies like treating all or no patients. The formula of net benefits is as follows:

$$NetBenefit = \frac{TruePositiveCount}{N} - \frac{P_t}{1 - P_t} \times \frac{FalsePositiveCount}{N} - TestHarm \tag{Equation 6}$$

The probability threshold P_t denotes the equilibrium between the benefits of treatment or further testing and those of avoidance, reflecting clinicians' risk-benefit assessments. For example, a 10% risk of severe conditions might prompt some doctors to intervene, while more cautious ones may opt for intervention at a 20% risk. A model outperforms others if it yields a higher net benefit at a specified P_t. *TestHarm* quantifies a test's negative aspects, such as cost and side effects, relative to a true positive's value. As an example, if a clinician believes that overlooking a patient needing further treatment for high myopia is 50 times worse than an extra review using CADx system, then *TestHarm* would be valued at 0.02.

ADDITIONAL RESOURCES

This study did not generate additional data.