



OPEN ACCESS

EDITED BY

Yalin Zheng,
University of Liverpool, United Kingdom

REVIEWED BY

Jiong Zhang,
University of Southern California,
United States

Tae Keun Yoo,
Hangil Eye Hospital, Republic of Korea
Nan Chen,
Eye Hospital of Nanjing Medical University,
China

*CORRESPONDENCE

Lu Chen

✉ chenludoc@outlook.com

Wangting Li

✉ liwangting@hotmail.com

Yantao Wei

✉ weiyantao75@126.com

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

RECEIVED 08 September 2024

ACCEPTED 28 October 2024

PUBLISHED 13 November 2024

CITATION

Zhang Z, Gao Q, Fang D, Mijit A, Chen L, Li W and Wei Y (2024) Effective automatic classification methods via deep learning for myopic maculopathy.

Front. Med. 11:1492808.

doi: 10.3389/fmed.2024.1492808

COPYRIGHT

© 2024 Zhang, Gao, Fang, Mijit, Chen, Li and Wei. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Effective automatic classification methods via deep learning for myopic maculopathy

Zheming Zhang^{1†}, Qi Gao^{2,3†}, Dong Fang⁴, Alfira Mijit¹,
Lu Chen^{4*}, Wangting Li^{4*} and Yantao Wei^{1*}

¹State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China, ²School of Future Technology, South China University of Technology, Guangzhou, China, ³Pazhou Lab, Guangzhou, China, ⁴Shenzhen Eye Hospital, Jinan University, Shenzhen, China

Background: Pathologic myopia (PM) associated with myopic maculopathy (MM) is a significant cause of visual impairment, especially in East Asia, where its prevalence has surged. Early detection and accurate classification of myopia-related fundus lesions are critical for managing PM. Traditional clinical analysis of fundus images is time-consuming and dependent on specialist expertise, driving the need for automated, accurate diagnostic tools.

Methods: This study developed a deep learning-based system for classifying five types of MM using color fundus photographs. Five architectures—ResNet50, EfficientNet-B0, Vision Transformer (ViT), Contrastive Language-Image Pre-Training (CLIP), and RETFound—were utilized. An ensemble learning approach with weighted voting was employed to enhance model performance. The models were trained on a dataset of 2,159 annotated images from Shenzhen Eye Hospital, with performance evaluated using accuracy, sensitivity, specificity, F1-Score, Cohen's Kappa, and area under the receiver operating characteristic curve (AUC).

Results: The ensemble model achieved superior performance across all metrics, with an accuracy of 95.4% (95% CI: 93.0–97.0%), sensitivity of 95.4% (95% CI: 86.8–97.5%), specificity of 98.9% (95% CI: 97.1–99.5%), F1-Score of 95.3% (95% CI: 93.2–97.2%), Kappa value of 0.976 (95% CI: 0.957–0.989), and AUC of 0.995 (95% CI: 0.992–0.998). The voting ensemble method demonstrated robustness and high generalization ability in classifying complex lesions, outperforming individual models.

Conclusion: The ensemble deep learning system significantly enhances the accuracy and reliability of MM classification. This system holds potential for assisting ophthalmologists in early detection and precise diagnosis, thereby improving patient outcomes. Future work could focus on expanding the dataset, incorporating image quality assessment, and optimizing the ensemble algorithm for better efficiency and broader applicability.

KEYWORDS

myopic maculopathy, ensemble learning, deep learning, artificial intelligence, fundus image

1 Introduction

Pathologic myopia (PM) is one of the leading causes of visual impairment and blindness worldwide (1, 2). Over the past half-century, the prevalence of myopia has increased significantly, particularly in East Asia, where the proportion of high myopia cases has also risen. In these regions, up to 80% of 18-year-old high school graduates are myopic, with 20% of these cases classified as high myopia (3). The higher the degree of myopia, the greater the risk of developing PM. The growing incidence of PM, along with its associated severe ocular complications, underscores the critical need for effective screening and management strategies in global public health.

According to the meta-analysis for pathologic myopia (META-PM) classification system proposed by Ohno-Matsui et al., PM is defined as the presence of severe ocular lesions in fundus photographs that are equivalent to or exceed diffuse chorioretinal atrophy, or features such as lacquer cracks, myopic choroidal neovascularization (CNV), and Fuchs' spots (4). Due to the irreversible pathological changes in the shape and structure of the myopic eye, effective treatment options for PM remain limited, and the prognosis for PM-related complications is generally poor. Additionally, the slow progression of PM often leads patients to overlook symptoms, attributing them instead to issues with their corrective lenses, thus delaying diagnosis (5). Early diagnosis allows timely intervention and follow-up screenings, helping patients understand their condition and take a proactive role in managing their health. This is key to preventing further deterioration and improving outcomes. Therefore, regular screening of myopic individuals to detect PM early and prevent its progression is of paramount importance.

Fundus imaging has become a vital tool in ophthalmic diagnostics for common eye diseases due to its non-invasive, accessible, and easily processed nature (6). However, traditional clinical image analysis heavily relies on doctors' expertise and experience and is time-consuming (7). This has driven the development of efficient, automated, and accurate fundus image analysis systems, which are critical strategies for the future of preventing and treating eye diseases.

In recent years, artificial intelligence (AI) and deep learning technologies have advanced rapidly in the field of medical image processing (8–10), leading to the emergence of new techniques for analyzing fundus images related to high myopia (11, 12). AI can utilize structural changes in the eye, particularly those linked to high myopia, to predict specific conditions. High myopia is typically associated with the elongation of the eyeball and alterations in the retina, which can lead to various retinal complications. The correlation between ocular structure and conditions like high myopia highlights the significance of analyzing fundus images for predictive diagnostics. For instance, a recent study demonstrated that fundus photography can estimate corneal curvature, a crucial factor in refractive errors, showcasing AI's ability to extract valuable insights from these images (13). By recognizing these structural variations, AI models can greatly enhance their predictive accuracy and improve patient outcomes in myopic disease contexts.

These technologies hold significant potential in assisting ophthalmologists by enhancing diagnostic efficiency and accuracy. For instance, Cen et al. developed a deep learning platform capable of detecting 39 different fundus diseases and conditions, demonstrating excellent performance in multi-label classification tasks (14). Similarly, Li et al. proposed the MyopiaDTER model, which introduced a novel attention feature pyramid networks (FPN) architecture and generated multi-scale feature maps for the traditional detection transformer

(DETR), enabling the detection of normal myopia, high myopia, and pathologic myopia regions in fundus photographs, achieving three-class classification (15).

Moreover, due to the often limited availability of medical data, self-supervised learning is expected to gain significant traction in the field (16). In this regard, Zhou et al. introduced the RETFound model, trained on 1.6 million unlabeled retinal images using self-supervised learning and later adapted for disease monitoring tasks with labeled data (17). The model demonstrated superior performance compared to several baselines in diagnosing and predicting sight-threatening eye diseases, establishing itself as a foundational tool for ophthalmic image analysis.

Ensemble learning has demonstrated significant potential in various medical applications. For instance, Namamula and Chaytor proposed an ensemble learning approach that combined the results of the Edge Detection Instance Preference (EDIP) algorithm with Extreme Gradient Boosting (XGBoost), leading to enhanced accuracy in analyzing large-scale medical datasets. Their method achieved impressive success rates in diagnosing conditions such as blood cancer and diabetes (18).

In this study, we explore an effective automatic recognition system for pathologic myopia-related fundus lesions. We utilize five deep learning architectures to train models capable of recognizing five types of myopic maculopathy (MM) using color fundus photographs. The five architectures include ResNet50 (19) and EfficientNet-B0 (20), both of which have been proven effective in medical classification tasks; Vision Transformer (ViT) (21), which utilizes advanced transformer units for image feature extraction and analysis; Contrastive Language-Image Pre-Training (CLIP) (22) model, which enhances image understanding through language-vision alignment; and RETFound (17), which has been pre-trained on a large dataset of fundus images.

To enhance the system's accuracy and reliability, we employ an ensemble learning approach, integrating the outputs of these models through a weighted voting strategy (23). This research not only aims to reduce the workload of clinicians and address the shortage of medical resources but also enables rapid screening of pathologic myopia-related fundus lesions, serving a wide population and providing significant clinical and social value.

2 Materials and methods

2.1 Data

This study collects a total of 2,159 original color retinal fundus images from Shenzhen Eye Hospital, with analysis commencing in June 2024. The images are captured using a desktop non-mydratric retinal camera. All images are with a field of view 45 degrees, centered on the macula or on the connecting center of optic disc and macula. There are no quality issues in the images collected, such as images with obscured macular areas due to severe artifacts, defocus blurring or inadequate lighting, and with incorrect field position.

According to the META-PM classification system (4), MM is categorized into five grades (shown in Figure 1): no myopic retinopathy (C0), tessellated fundus (C1), diffuse chorioretinal atrophy (C2), patchy chorioretinal atrophy (C3), and macular atrophy (C4). Additionally, lacquer cracks, CNV, and Fuchs' spots are defined as "Plus" lesions. Grade C1 is characterized by distinct choroidal vessels visible around the fovea and arcade vessels. Grade C2 presents with a yellowish-white appearance of the posterior pole, with atrophy assessed relative to the

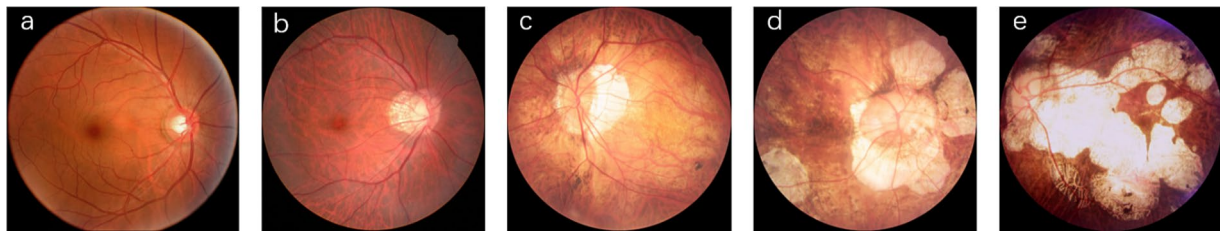


FIGURE 1

Representative images of five myopic macular degeneration categories. (a) no myopic retinopathy, (b) tessellated fundus, (c) diffuse chorioretinal atrophy, (d) patchy chorioretinal atrophy, (e) macular atrophy.

optic disc area. Grade C3 is marked by well-defined gray-white lesions in the macular region or around the optic disc. Grade C4 features well-defined, gray-white or white, round atrophic lesions in the foveal region. Grades C0 and C1 indicate low-risk high myopia, while Grades C2–C4 represent high-risk high myopia, also known as PM. In PM fundus images, “Plus” lesion features may be observed, which are not specific to any particular grade but can develop from or occur within any grade. This study primarily focuses on the five-class classification task among Grades C0–C4.

The color fundus photographs are annotated by two professional ophthalmologists according to the aforementioned classification system, with the distribution of the dataset detailed in Table 1.

2.2 Image preprocessing

To minimize the interference of black regions in fundus images on feature extraction, redundant black areas in the images are cropped. First, the images are loaded using the OpenCV library and converted to grayscale as follows:

$$I_{gray} = 0.299 \times I_R(x,y) + 0.587 \times I_G(x,y) + 0.114 \times I_B(x,y)$$

where $I_R(x,y)$, $I_G(x,y)$, and $I_B(x,y)$ respectively represent the value of pixel (x,y) in the red, green and blue channels. Subsequently, a binary mask $M(x,y)$ is generated using thresholding:

$$M(x,y) = \begin{cases} 255 & I_{gray}(x,y) \geq 20 \\ 0 & otherwise \end{cases}$$

By detecting the largest contour in the mask, the coordinates of its minimum bounding box (x_m, y_m, w, h) are calculated, and the region within this bounding box is cropped:

$$ROI = I[y_m : y_m + h, x_m : x_m + w]$$

This process results in a fundus image with black regions removed, preserving the relevant retinal information. Prior to developing the deep learning system, image normalization is performed. All fundus images are normalized to pixel values within the range of 0–1 and resized to a resolution of 224×224 pixels.

TABLE 1 Distribution of the dataset.

Category	Number
C0	510
C1	678
C2	401
C3	408
C4	162
Total	2,159

2.3 Development of the deep learning system

During the system development, the test set is constructed using a stratified random sampling method, where 20% of data from each category is randomly selected to form the test set. The remaining data is used for model training and validation through five-fold cross-validation. Specifically, the remaining data is randomly divided into five equally sized folds, ensuring that each image appeared in only one fold. The training process is conducted in two steps: first, four folds are selected for algorithm training and hyperparameter optimization, while the remaining fold is used for validation. This process is repeated five times, ensuring that each fold served as a validation set. This approach aims to ensure balanced data distribution across folds and effectively evaluate the model's generalization ability.

The model training involves five different architectures: ResNet50 (19), EfficientNet-B0 (20), ViT (24), CLIP (22), and RETFound (17). These architectures are chosen due to their superior performance in visual tasks and their success in similar research. Specifically, ResNet50 and EfficientNet-B0 excel in feature extraction and computational efficiency, Vision Transformer is effective in handling global image information, CLIP performs well in image-text matching and image understanding, and RETFound demonstrates outstanding performance in retinal image analysis. To initialize weights and leverage existing knowledge, the first four models are pre-trained on the ImageNet Large Scale Visual Recognition Challenge (25), a comprehensive database with 1.28 million images classified into 1,000 categories. The RETFound model is pre-trained through self-supervised learning on 1.6 million retinal images and validated across various disease detection tasks.

For the CLIP model, fine-tuning is performed by aligning image and text features through contrastive learning. Class labels are used as textual descriptions during training, enabling the model to adapt to our specific

task even with limited data. Similarly, other models are fine-tuned by initializing from pretrained weights and applying transfer learning, with five-fold cross-validation used to adapt them to our classification problem.

Our training platform utilizes the PyTorch framework, with all deep learning algorithms run on a NVIDIA 4090 graphics processing unit (GPU) (26). The batch size is set to 32, and model parameters are updated based on the mean of the samples. The training process employs the AdamW (27) optimizer with weight decay, with a learning rate set at 0.001. Each model is trained for 50 epochs, and performance is monitored at the end of each epoch using metrics including loss, accuracy, sensitivity, specificity, F1-Score, Kappa, and AUC on the validation dataset. The best model parameters that show the highest AUC on the validation set are saved.

Finally, we employ an ensemble learning approach using the voting strategy to combine the predictions of ResNet50 (19), EfficientNet-B0 (20), ViT (24), CLIP (22), and RETFound (17). The predictions are combined through weighted voting, where each model's voting weight is adjusted according to its average AUC score on the five-fold validation sets. The final classification result is determined by the ensembled prediction probability $prob_{final}$.

$$w_m = \frac{1}{5} \sum_{v=1}^5 AUC_{m,v} \quad m \in \{\text{ResNet50, EfficientNet-B0, ViT, CLIP, RETFound}\}$$

$$prob_{final} = \sum_m \sum_v w_m \cdot prob_{m,v}$$

where w_m represents the weight of the method m . v denotes the v th fold. $AUC_{m,v}$ represents the AUC on the verification set for the model under the v th fold among the method m . $prob_{m,v}$ is the probability of the model under the v th fold among the method m . This method is expected to enhance the overall accuracy and robustness of the system and improve the model's generalization ability on the test set.

2.4 Evaluation of the AI system

To comprehensively evaluate the classification performance of the models, this study employs a range of metrics including accuracy, sensitivity, specificity, F1-Score, weighted Cohen's Kappa, and AUC, with 95% confidence intervals (CI) calculated for all metrics. All metrics are derived from the results of five-fold cross-validation, and the average values are computed across the folds.

In multi-category classification task, sensitivity and specificity are calculated using a one-vs-rest strategy. The 95% CI for accuracy, sensitivity, and specificity are estimated using the Wilson Score method implemented in the Statsmodels package (version 0.13.5). For the F1-score, weighted Cohen's Kappa, and AUC, the 95% CI are calculated using the empirical Bootstrap method (28), with 1,000 resamples performed to ensure the robustness of the results.

In addition to numerical metrics, model performance is assessed through visualization techniques. The receiver operating characteristic (ROC) curve illustrates the model's performance across different threshold values, with an AUC value closer to 1.0 indicating better classification capability. The confusion matrix compares the true labels

with the predicted labels, clearly displaying the number of correct and incorrect classifications for each category. The ROC curves and confusion matrices are plotted using Matplotlib (version 3.8.3) and Scikit-learn (version 1.4.1) libraries.

2.5 Interpretability of AI system

To better understand the impact of different regions of fundus images on classification results, identify the causes of misclassification, and enhance the interpretability of the model, we employ visualization techniques to analyze the convolutional network model used in our experiments. Class Activation Mapping (CAM) (29) is a visualization technique that aggregates feature maps weighted by network parameters to generate heatmaps, which highlight the importance of each pixel in the image classification process. In these heatmaps, more important regions are indicated with warmer colors. However, CAM requires modifications to the network architecture and retraining of the model. To simplify implementation, this study utilizes Grad-CAM++ (30), which does not require any changes to the network structure. Grad-CAM++ provides a clear visualization of the features learned by the model while maintaining classification accuracy, making the model more transparent and interpretable.

3 Results

3.1 Evaluation of deep learning models

This study evaluates the performance of five deep learning models (ResNet50, EfficientNet-B0, ViT, CLIP, and RETFound) and their ensemble results for classifying five types of MM. All models are trained and validated using five-fold cross-validation and assessed on an independent test dataset. Evaluation metrics include accuracy, sensitivity, specificity, F1-Score, weighted Cohen's Kappa, and AUC, all reported with 95% CI. The models are then combined using a weighted voting ensemble approach. Table 2 presents the performance of each model as well as the results of the weighted voting ensemble.

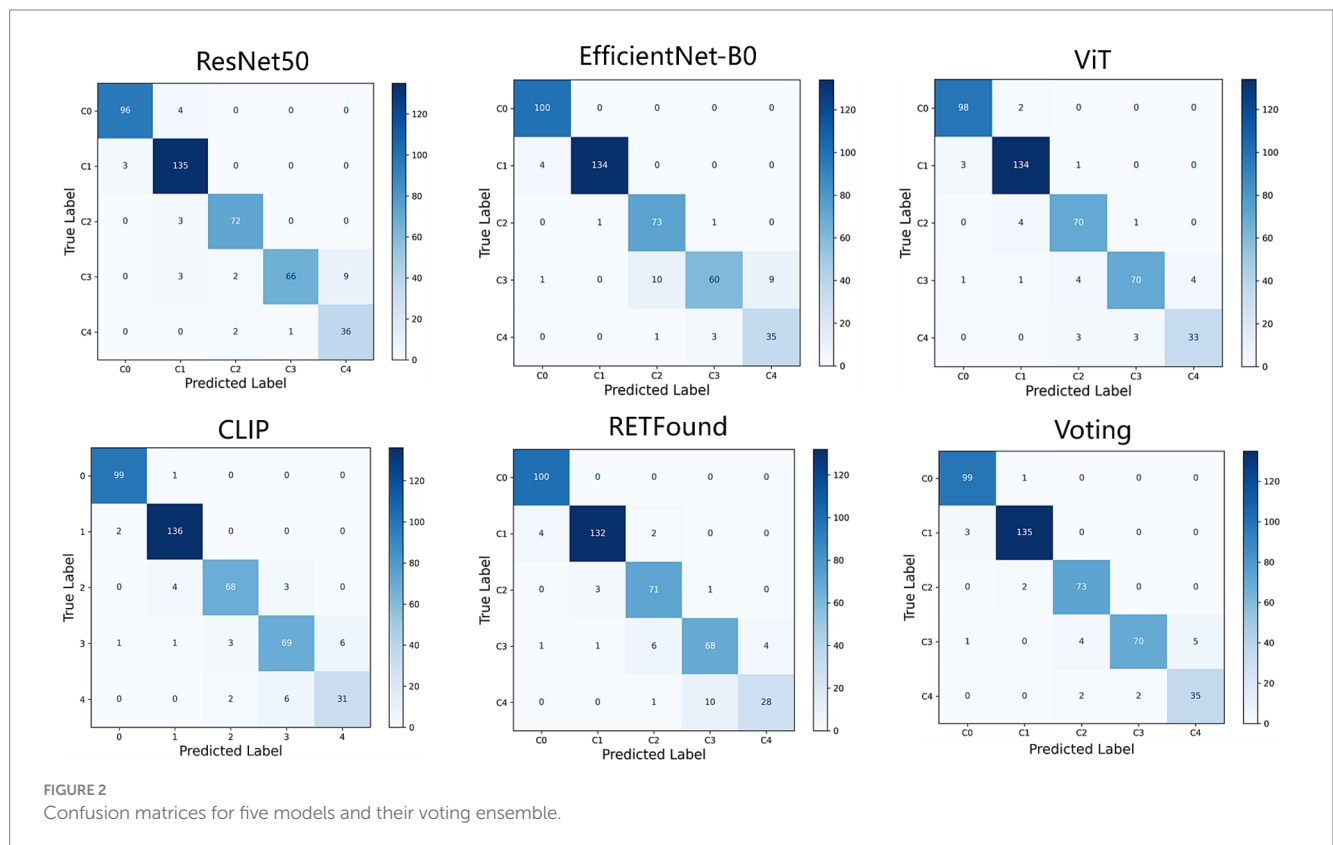
The voting ensemble algorithm shows the best performance across all evaluation metrics. Specifically, the voting algorithm achieves an accuracy of 95.4% (95% CI: 93.0–97.0%), sensitivity of 95.4% (95% CI: 86.8–97.5%), specificity of 98.9% (95% CI: 97.1–99.5%), F1-Score of 95.3% (95% CI: 93.2–97.2%), Kappa value of 0.976 (95% CI: 0.957–0.989), and AUC of 0.995 (95% CI: 0.992–0.998).

Figures 2, 3 display the performance of the five deep learning models and their voting ensemble results on the test set. From the confusion matrix (Figure 2), EfficientNet-B0 and RETFound models achieve the highest classification accuracy in the C0 class; the CLIP model performs best in the C1 class; EfficientNet-B0 model and voting strategy excel in the C2 class; ViT model and voting strategy are the best in the C3 class; and ResNet50 performs best in the C4 class. Although the voting strategy does not show the highest accuracy in every category, its accuracy is consistently high across categories, demonstrating strong robustness. Furthermore, from the ROC curves for each category (Figure 3), the voting ensemble method also performs better, with the highest AUC values for all categories, further confirming its effectiveness in complex lesion classification tasks. Additionally, it can be observed that most of the misclassified images

TABLE 2 Performance of individual models and their voting ensemble for classifying myopic maculopathy.

Model	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1-Score (95% CI)	Kappa (95% CI)	AUC (95% CI)
ResNet50	91.8%	91.8%	97.9%	91.7%	0.958	0.991
	(90.5, 92.8%)	(87.4, 93.0%)	(97.2, 98.5%)	(90.5, 92.8%)	(0.948, 0.966)	(0.988, 0.993)
EfficientNet-B0	91.2%	91.2%	97.8%	91.0%	0.964	0.990
	(89.9, 92.3%)	(86.4, 92.0%)	(97.0, 98.4%)	(89.8, 92.3%)	(0.956, 0.970)	(0.987, 0.992)
ViT	92.4%	92.4%	98.1%	92.3%	0.960	0.986
	(91.2, 93.5%)	(86.9, 92.7%)	(97.3, 98.6%)	(91.2, 93.5%)	(0.950, 0.968)	(0.982, 0.989)
CLIP	91.7%	91.7%	97.9%	91.6%	0.961	0.989
	(90.5, 92.8%)	(85.4, 91.4%)	(97.2, 98.5%)	(90.3, 92.8%)	(0.952, 0.968)	(0.987, 0.992)
RETFound	91.8%	91.8%	97.9%	91.6%	0.962	0.990
	(90.5, 92.8%)	(85.6, 91.4%)	(97.2, 98.5%)	(90.4, 92.7%)	(0.954, 0.970)	(0.987, 0.992)
Voting	95.4%	95.4%	98.9%	95.3%	0.976	0.995
	(93.0, 97.0%)	(86.8, 97.5%)	(97.1, 99.5%)	(93.2, 97.2%)	(0.957, 0.989)	(0.992, 0.998)

Bold indicates the best result.



were assigned to adjacent categories, which were largely due to the visual similarity between categories, making it difficult for the model to distinguish between them.

The inference time of the ensemble model is the sum of the inference times of the individual models, plus the time required for the ensemble process following each model's prediction. Table 3 provides detailed timing for each fold of every model when processing individual images. The ensemble process itself takes approximately 5.7 ms, demonstrating that the computational cost and time required for weighted voting in the ensemble are minimal.

t-SNE (34) is a commonly used technique for dimensionality reduction and visualization of high-dimensional data, helping us to visually observe the distribution of different categories in the feature space. Figure 4 shows the t-SNE plots for each model, which reveal that the scatter points for each category are relatively concentrated, with some overlap between adjacent categories. For example, the red points (C4) contain many yellow points (C3), indicating that the C3 class is prone to being misclassified as C4. However, the t-SNE plot for the CLIP model shows that points of the same category also appear in multiple clusters. Particularly, the boundary between red (C4) and

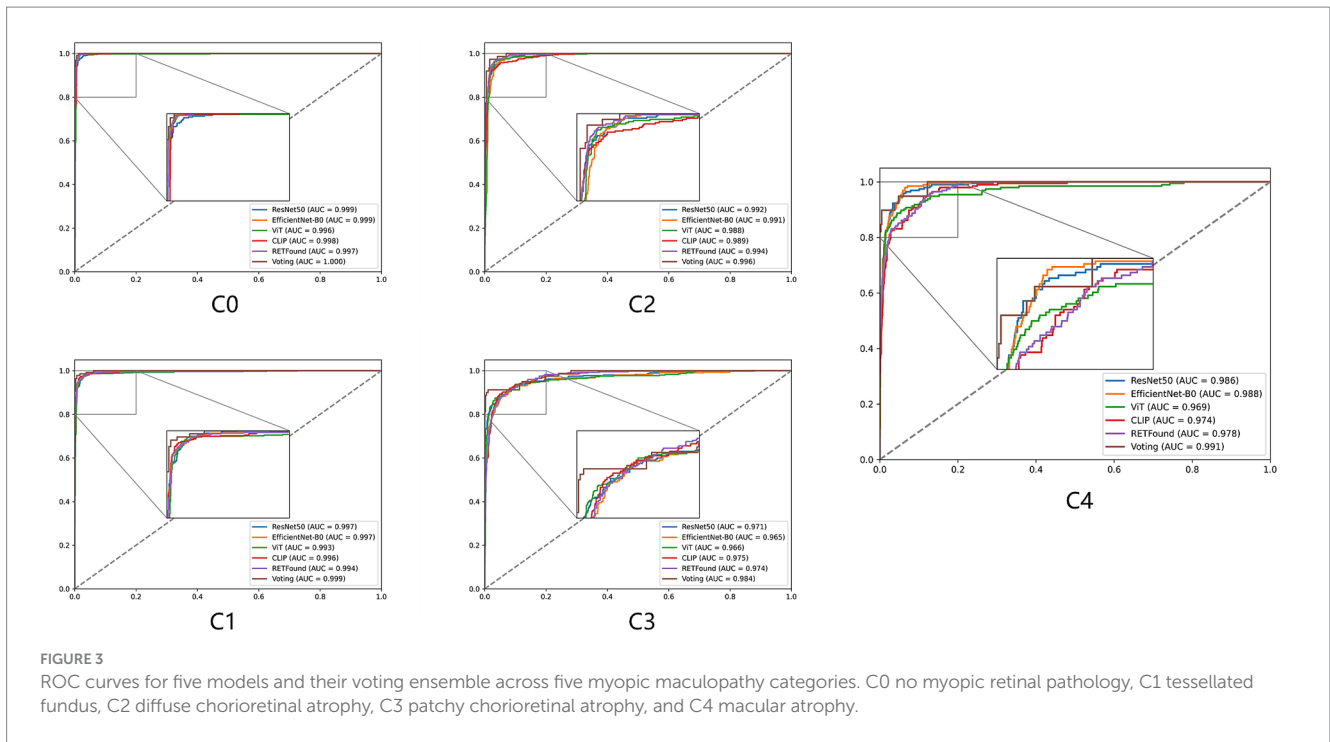


TABLE 3 Inference times for individual models (per image).

Model	Inference time per image (ms)
ResNet50	46.2
EfficientNet-B0	33.1
ViT	40.6
CLIP	49.6
RETFound	97.8

yellow (C3) points is not very clear, suggesting that distinguishing between these two categories is challenging. This overlap suggests that the visual similarity between these categories affects the model’s performance.

Additionally, the dataset predominantly contains images with single lesions, which may limit the model’s ability to generalize to cases with coexisting multiple lesions. Expanding the dataset to include such cases would improve model robustness and applicability in real-world clinical scenarios.

3.2 Classification errors

The test set contains 432 images, with 20 images (4.63% of the total) showing inconsistencies between the Voting ensemble results and the reference standards. Specifically, the voting ensemble method misclassified 1 image from class C0, 3 images from class C1, 2 images from class C2, 10 images from class C3, and 4 images from class C4. These errors were mainly due to the visual similarity of features between certain classes, making it challenging for the model to distinguish them. Figure 5 provides

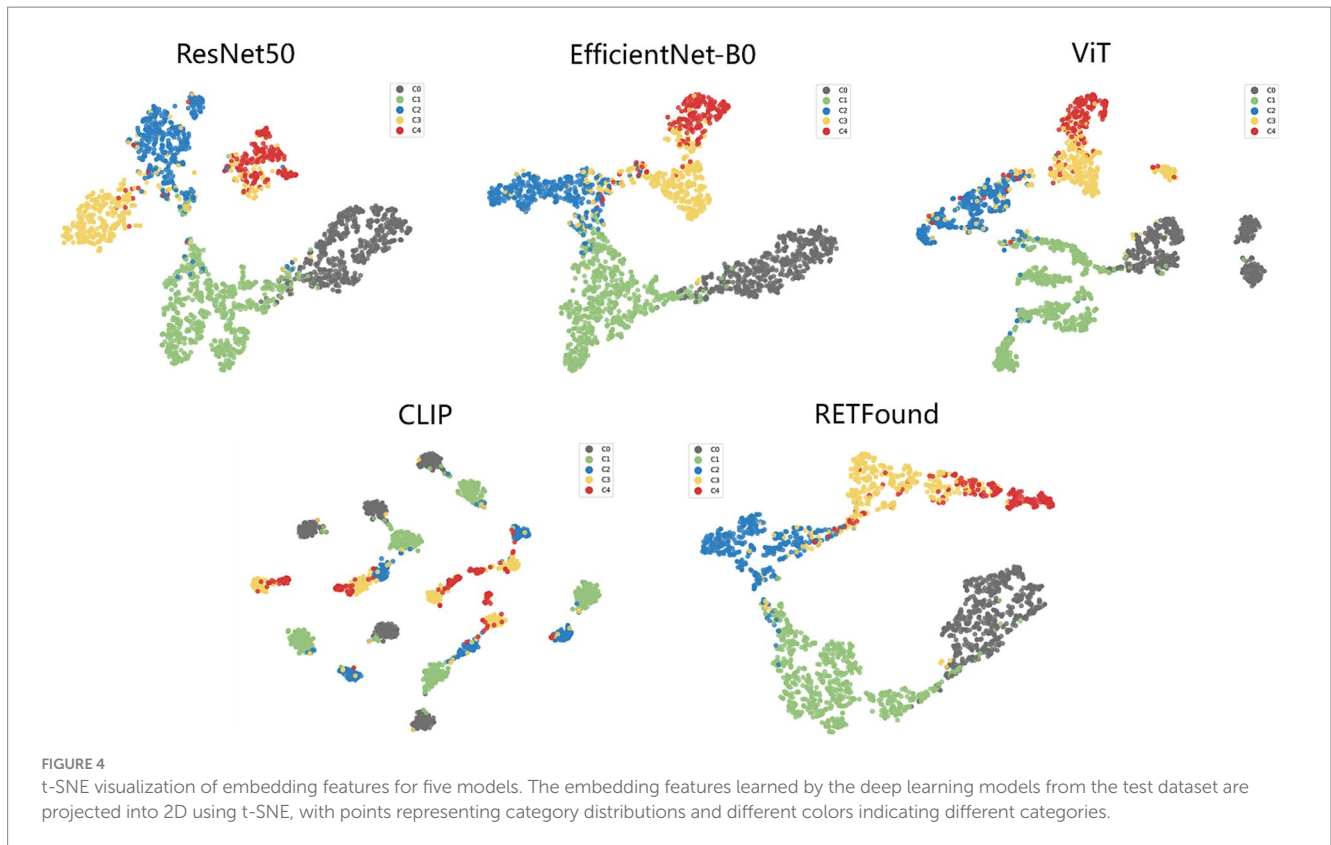
examples of typical images incorrectly classified by the voting ensemble.

3.3 Visual interpretation of models

Grad-CAM works by calculating feature maps from convolutional layers and highlighting the areas the model focuses on through weighted summation, which means it only applies to convolutional neural network (CNN) models. In this study, ResNet50 and EfficientNet-B0 are the CNN models used. We input original images of different types of MM into these models and use the Grad-CAM++ algorithm to generate heatmaps. The heatmaps highlight the areas most important for classification with bright colors like red and yellow. The study shows that the heatmaps effectively highlight lesion areas in the fundus images, such as retinal vessels, choroidal atrophy, and the macula. Figure 6 shows representative examples of heatmaps for MM levels C0-C4.

4 Discussion

Based on fundus images, this study developed artificial intelligence models to identify no myopic retinopathy, tessellated fundus, diffuse chorioretinal atrophy, patchy chorioretinal atrophy, and macular atrophy. The outputs of these models are fused using the weighted voting method from ensemble learning. Subsequently, all models and their ensemble results are evaluated. Our findings reveal that after ensembling, the models outperform all individual deep learning models (ResNet50, EfficientNet-B0, ViT, CLIP, and RETFound), demonstrating robustness and generalization ability.



The advantage of this work lies in the adoption of multiple advanced classification models. For instance, ResNet50, a widely used and well-performing CNN model, is extensively applied in various visual tasks and demonstrates high accuracy and stability. EfficientNet is a lightweight model that enhances performance through compound scaling methods without increasing model complexity, making it particularly suitable for resource-constrained environments. ViT is a model based on the self-attention mechanism that challenges the dominance of traditional CNNs in visual tasks, effectively capturing global features in images and showing exceptional performance, especially when handling large-scale datasets. CLIP is a multimodal model capable of processing both image and text data. Trained on a large-scale image-text paired dataset, it exhibits strong cross-modal transfer learning capabilities and can be applied to various downstream visual tasks. RETFound is a recently proposed model trained using self-supervised learning on 1.6 million unlabeled retinal images, then adapted to disease monitoring tasks with specific labels, making it particularly suitable for medical image analysis. Finally, this work innovatively employs a voting method in ensemble algorithms, integrating these different types of models, thereby enhancing the robustness and generalization of classification by ensuring diversity and complementarity among models.

During the experiments, we observed that the model performed best on the training set, with performance on the validation set being similar to that on the test set. This consistency indicates that the model has a strong understanding of the features needed to classify myopic maculopathy. It means that the model can generalize to new data, which is important for clinical applications. Additionally, five-fold

cross-validation helps ensure stable performance across different data splits.

In analyzing the misclassified images, we noticed that many shared visual similarities with the incorrect categories. This overlap makes it hard for the model to distinguish between certain lesions. To tackle this, future research will focus on strategies like contrastive learning to improve classification of these challenging samples. Additionally, using hard sample mining could help the model better differentiate between similar categories, leading to improved accuracy in clinical applications.

Grad-CAM was applied to visualize the decision-making process. This technique helps identify the areas of the input image that the model focuses on when making predictions. This provides interpretability by revealing how models make decisions, helping clinicians understand and trust AI predictions, which is crucial for clinical adoption.

This study has some limitations. First, in the dataset used for the research, fundus images labeled as C0 (no myopic retinopathy) were defined as free of any lesions. However, in a real clinical setting, some eyes may not have high myopia-related maculopathy but may still have other types of retinal diseases. Similarly, for images of other disease levels, patients in the sample only had a specific single disease, meaning the model did not encounter cases with multiple coexisting retinal diseases during training. This may lead to inaccurate classifications in practical applications. Secondly, our dataset is sourced from a single device, performance may decline when validating images from different fundus cameras due to variations in imaging protocols and quality, which can affect the model's ability to generalize. Additionally, although quality control

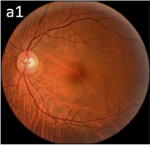
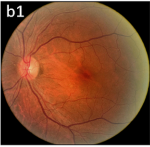

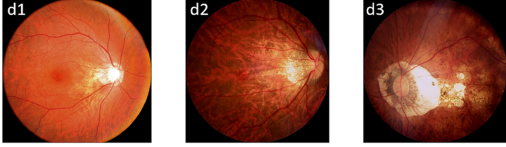
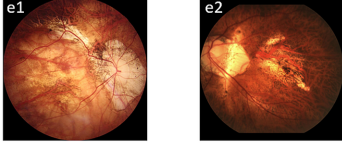
<p>a C0 incorrectly classified as other type of MM</p> 	<p>a1, C0 misclassified as C1</p>
<p>b C1 incorrectly classified as other type of MM</p> 	<p>b1, C1 misclassified as C0</p>
<p>c C2 incorrectly classified as other type of MM</p> 	<p>c1, C2 misclassified as C1</p>
<p>d C3 incorrectly classified as other types of MM</p> 	<p>d1, C3 misclassified as C0 d2, C3 misclassified as C2 d3, C3 misclassified as C4</p>
<p>e C4 incorrectly classified as other types of MM</p> 	<p>e1, C4 misclassified as C2 e2, C4 misclassified as C3</p>

FIGURE 5 Representative examples of misclassified images by voting ensemble. (a–e) Represent the misclassified images of C0–C4, respectively.

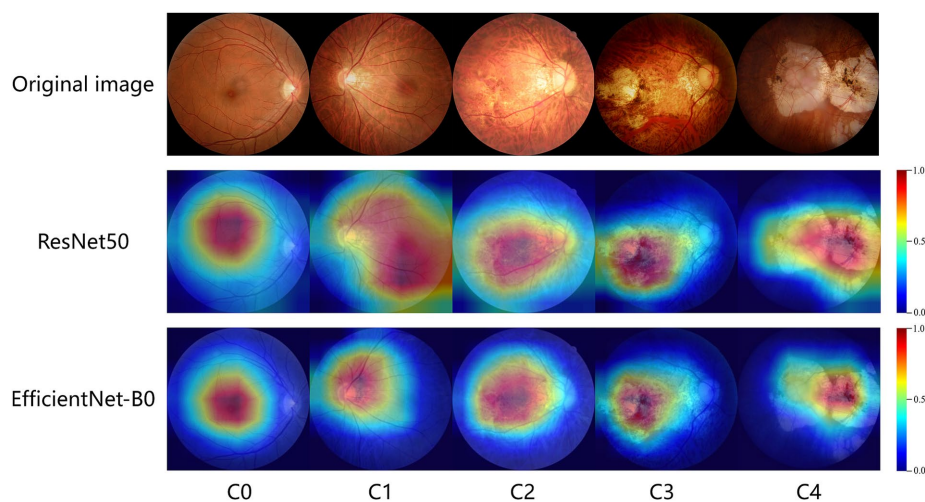


FIGURE 6 GradCam visualizations of ResNet50 and EfficientNet-B0 diagnoses for myopic maculopathy levels of C0–C4. The redder the area in the heatmap, the more it contributes to the model’s decision-making.

measures were implemented in the study to exclude low-quality images, such as image quality issues remain common in real-world scenarios. Finally, the application of automatic image quality assessment techniques (31, 32) is crucial, as they can help identify substandard images and alert operators. Finally, while the ensemble method improved diagnostic performance, integrating multiple models also reduced efficiency (33).

Future improvements could include the following: (1) expanding the dataset to enhance its diversity and improve the model's generalization ability by acquiring fundus photographs from different imaging devices and including cases with various coexisting retinal diseases. Additionally, incorporating external validation is crucial to ensure that the model performs reliably across diverse clinical settings; (2) incorporating automatic image quality assessment techniques to automatically exclude substandard images; (3) optimizing the ensemble model's algorithm by removing models that do not significantly contribute to the ensemble results or replacing them with more efficient models.

5 Conclusion

In conclusion, our study successfully developed an artificial intelligence model capable of automatically identifying no myopic retinopathy, tessellated fundus, diffuse chorioretinal atrophy, patchy chorioretinal atrophy, and macular atrophy from fundus images. By utilizing various advanced deep learning models, including ResNet50, EfficientNet-B0, ViT, CLIP, and RETFound, and innovatively employing a weighted voting ensemble algorithm, we significantly enhanced the model's classification accuracy and robustness. The results demonstrate that the ensemble model outperforms individual models across multiple metrics, particularly exhibiting strong robustness and generalization ability in the analysis of complex fundus images. This system has the potential to assist ophthalmologists in accurately and promptly identifying the causes of myopic macular lesions, thereby improving patient visual outcomes by enabling targeted treatment at an early stage.

Data availability statement

The datasets presented in this article are not readily available because the data is not public for ethical reasons. Research related requests to access the datasets should be directed to the corresponding author.

References

- Dolgin E. The myopia boom. *Nature*. (2015) 519:276–8. doi: 10.1038/519276a
- Silva R. Myopic maculopathy: a review. *Ophthalmologica*. (2012) 228:197–213. doi: 10.1159/000339893
- Rudnicka AR, Kapetanakis VV, Wathern AK, Logan NS, Gilmartin B, Whincup PH, et al. Global variations and time trends in the prevalence of childhood myopia, a systematic review and quantitative meta-analysis: implications for aetiology and early prevention. *Br J Ophthalmol*. (2016) 100:882–90. doi: 10.1136/bjophthalmol-2015-307724
- Ohno-Matsui K, Kawasaki R, Jonas JB, Cheung CMG, Saw S-M, Verhoeven VJ, et al. International photographic classification and grading system for myopic maculopathy. *Am J Ophthalmol*. (2015) 159:877–883.e7. doi: 10.1016/j.ajo.2015.01.022
- Hayashi K, Ohno-Matsui K, Shimada N, Moriyama M, Kojima A, Hayashi W, et al. Long-term pattern of progression of myopic maculopathy: a natural history study. *Ophthalmology*. (2010) 117:1595–1611.e4. doi: 10.1016/j.ophtha.2009.11.003
- Baird PN, Saw S-M, Lanca C, Guggenheim JA, Smith EL III, Zhou X, et al. Myopia. *Nat Rev Dis Prim*. (2020) 6:99. doi: 10.1038/s41572-020-00231-4
- Resnikoff S, Lansingh VC, Washburn L, Felch W, Gauthier T-M, Taylor HR, et al. Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? *Br J Ophthalmol*. (2020) 104:588–92. doi: 10.1136/bjophthalmol-2019-314336
- Chen X, Wang X, Zhang K, Fung K-M, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal*. (2022) 79:102444. doi: 10.1016/j.media.2022.102444

Ethics statement

The studies involving humans were approved by institutional review board of Shenzhen Eye 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

ZZ: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Investigation. QG: Methodology, Writing – original draft, Writing – review & editing, Investigation, Software, Validation, Visualization. DF: Writing – review & editing. AM: Writing – review & editing. LC: Writing – review & editing. WL: Writing – review & editing. YW: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the Shenzhen Science and Technology Program (KCXFZ20211020163813019). This research was supported by the Pazhou Laboratory's basic cloud computing platform.

Conflict of interest

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

Publisher's note

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

9. Dhar T, Dey N, Borra S, Sherratt RS. Challenges of deep learning in medical image analysis—improving explainability and trust. *IEEE Trans Technol Soc.* (2023) 4:68–75. doi: 10.1109/TTS.2023.3234203
10. Li Z, Xie H, Wang Z, Li D, Chen K, Zong X, et al. Deep learning for multi-type infectious keratitis diagnosis: a nationwide, cross-sectional, multicenter study. *NPJ Digit Med.* (2024) 7:181. doi: 10.1038/s41746-024-01174-w
11. Li Y, Foo L-L, Wong CW, Li J, Hoang QV, Schmetterer L, et al. Pathologic myopia: advances in imaging and the potential role of artificial intelligence. *Br J Ophthalmol.* (2023) 107:600–6. doi: 10.1136/bjophthalmol-2021-320926
12. Prashar J, Tay N. Performance of artificial intelligence for the detection of pathological myopia from colour fundus images: a systematic review and meta-analysis. *Eye.* (2024) 38:303–14. doi: 10.1038/s41433-023-02680-z
13. Choi JY, Kim H, Kim JK, Lee IS, Ryu IH, Kim JS, et al. Deep learning prediction of steep and flat corneal curvature using fundus photography in post-COVID telemedicine era. *Med Biol Eng Comput.* (2024) 62:449–63. doi: 10.1007/s11517-023-02952-6
14. Cen L-P, Ji J, Lin J-W, Ju S-T, Lin H-J, Li T-P, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun.* (2021) 12:4828. doi: 10.1038/s41467-021-25138-w
15. Li M, Liu S, Wang Z, Li X, Yan Z, Zhu R, et al. MyopiaDETR: end-to-end pathological myopia detection based on transformer using 2D fundus images. *Front Neurosci.* (2023) 17:1130609. doi: 10.3389/fnins.2023.1130609
16. Huang S-C, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digit. Med.* (2023) 6:74. doi: 10.1038/s41746-023-00811-0
17. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. *Nature.* (2023) 622:156–63. doi: 10.1038/s41586-023-06555-x
18. Namamula LR, Chaytor D. Effective ensemble learning approach for large-scale medical data analytics. *Int J Syst Assur Eng Manag.* (2022) 15:13–20. doi: 10.1007/s13198-021-01552-7
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE (Institute of Electrical and Electronics Engineers) (2016). 770–8.
20. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks In: International Conference on Machine Learning. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, New York: ACM (Association for Computing Machinery) (2019). 6105–14.
21. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst.* (2017) 30:6000–10. doi: 10.48550/arXiv.1706.03762
22. Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., et al. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning (ICML)*. New York: ACM (Association for Computing Machinery)
23. Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. *Front Comput Sci.* (2020) 14:241–58. doi: 10.1007/s11704-019-8208-z
24. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. *ICLR 2021 (International Conference on Learning Representations)*. Washington DC: Curran Associates, Inc.
25. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE (2009). 248–55.
26. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst.* (2019) 32: 8026–8037. doi: 10.48550/arXiv.1912.01703
27. Loshchilov I., Hutter F. (2019). Decoupled weight decay regularization. *conference is ICLR 2017*. Washington DC: Curran Associates, Inc.
28. Austern M., Syrgkanis V. (2020). Asymptotics of the empirical bootstrap method beyond asymptotic normality. arXiv. [Preprint].
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization In: Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: (2017). 618–26.
30. Chattopadhyay A., Sarkar A., Howlader P., Balasubramanian V.N. (2018). GRad-CAM++: improved visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (Piscataway, NJ: WACV). pp. 839–847.
31. Fu H, Wang B, Shen J, Cui S, Xu Y, Liu J, et al. Evaluation of retinal image quality assessment networks in different color-spaces In: D Shen, T Liu, TM Peters, LH Staib, C Essert and S Zhouet al, editors. Medical image computing and computer assisted intervention – MICCAI 2019, Lecture Notes in Computer Science. Cham: Springer International Publishing (2019). 48–56.
32. Shen Y, Sheng B, Fang R, Li H, Dai L, Stolte S, et al. Domain-invariant interpretable fundus image quality assessment. *Med Image Anal.* (2020) 61:101654. doi: 10.1016/j.media.2020.101654
33. Omar R, Bogner J, Muccini H, Lago P, Martínez-Fernández S, Franch X. The more the merrier? Navigating accuracy vs. energy efficiency design trade-offs in ensemble learning systems. *arXiv preprint arXiv:2407.02914.* (2024). doi: 10.48550/arXiv.2407.02914
34. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *JMLR.* (2008) 9.