Swin-MMC: Swin-Based Model for Myopic Maculopathy Classification in Fundus Images

Li Lu[1], Xuhao Pan[2], Panji Jin[1], and Ye Ding[1(✉)]

[1] School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China
{221115220,dingye}@dgut.edu.cn, 1340880735@qq.com
[2] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China

Abstract. Myopic maculopathy is a highly myopic retinal disorder that often occurs in highly myopic patients, serving as a major cause of visual impairment and blindness in numerous countries. Currently, fundus images serve as a prevalent diagnostic tool for myopic maculopathy. However, its efficacy relies on the expertise of clinicians, making the process labor-intensive. Thus, we propose a model specifically designed for the image classification of myopic maculopathy, named Swin-MMC, based on the Swin Transformer model architecture, which achieves outstanding performance on the test dataset. To achieve a finer-grained classification of myopic maculopathy in fundus images, we have innovatively and for the first time proposed the use of enhanced ArcFace loss in medical image classification. Then, based on the Swin-MMC model, we introduce a weak label strategy that effectively mitigates overfitting. Our approach achieves significantly improved results on the test dataset and can be easily used for various datasets and classification tasks. We conduct a series of experiments in the MMAC2023 challenge. In the testing phase, our average performance metric reaches 86.60%. In the further testing phase, our model's performance improves to 88.23%, ultimately securing the championship in the MMAC2023 challenge. The codes allowing replication of this study have been made publicly available at https://github.com/LuliDreamAI/MICCAI_TASK1.

Keywords: Medical image classification · Myopic maculopathy · Fundus images

1 Introduction

Myopic maculopathy stands as one of the most severe threats to global visual health, also known as myopic maculopathy degeneration [9]. In numerous regions, myopic maculopathy degeneration serves as a prominent cause of visual impairment and blindness. Moreover, the application scope of fundus images is remarkably extensive, they not only assist in the treatment of age-related macular

degeneration (AMD) [17] but also are employed in the detection and screening of glaucoma [23] while playing a crucial role in the identification and classification of diabetic retinopathy (DR) [2]. Medical practitioners diagnose the presence of myopic maculopathy through color fundus images or aim to prevent its further deterioration [16,21]. In the field of medicine, determining whether an individual is afflicted with myopic maculopathy often requires specialized ophthalmologists to meticulously examine images. This process not only consumes a substantial amount of manpower and time but is also relatively slow. However, with the advancement of artificial intelligence (AI), deep learning plays a crucial role in automating clinical data processing, it is now possible to discern the presence of macular lesions in fundus images without the need for extensive human resources and time investment [14].

In the early stages, the detection of ocular diseases primarily relied on the binary classification of fundus images, which merely distinguished between abnormal and normal images. With the advancement of deep neural network technologies, Khan et al. [11] combined popular models ResNet50 and InceptionResNetV2 to create an ensemble model, achieving an accuracy of 86.08% in binary classification tasks. Chen et al. [3] enhanced accuracy to 90.56% in binary datasets by introducing hybrid units in the dense layer. However, due to the complexity of ocular structures, the classification of fundus images in recent years has moved beyond simply separating pathological from normal images. It now focuses more on finer and more precise categorizations of fundus pathologies and classifications of various types of ocular diseases. Liu et al. [8] focused their research on developing a model based on the Vision Transformer (ViT) [6], aimed at classifying a dataset containing a diverse range of abnormal images. These abnormalities may originate from any of six different ocular diseases, including age-related macular degeneration, diabetes, and glaucoma, among others. The study employed Vision Transformers of varying layer counts, with the objective of achieving accurate classification across seven distinct labels (healthy and six different diseases). Utilizing a 14-layer Vision Transformer model, the model demonstrated optimal performance, evidenced by an F1-score of 83.49%, sensitivity of 84.00%, precision of 83.00%, and a Kappa score of 0.802. However, it is noteworthy that the study did not venture into a more detailed subdivision within each disease category. Compared to professional ophthalmologists, AI models based on deep neural networks have shown exceptional efficiency and outstanding performance in large-scale medical analysis [7]. Therefore, exploring how to efficiently apply neural network models for more precise classification of ocular diseases represents a highly pertinent direction. Additionally, several deep learning algorithms have already been effectively utilized for screening and classification tasks related to diabetic retinopathy [4,12,18] and glaucoma [15,20].

To address the classification of the myopic maculopathy problem, Sun et al. [19] proposed a feature fusion framework that comprises a prior knowledge extraction module and a feature fusion module, and the model achieves an AUC value of 0.998 on the test dataset. Wang et al. [21] developed a deep learning model for detection and classification and achieved high sensitivities, specificities,

and reliable Cohen's kappa. In recent years, with the popularity of Vision Transformer [6] and Swin Transformer [13], many researchers and scholars began to use them widely. Hossain et al. [10] proposed the Swin-FSR model, which employs the Swin Transformer with spatial and depth-wise attention mechanisms for fundus image super-resolution. This ensures that important fine details are preserved during the compression and decompression processes of super-resolved images. The ViT is commonly employed in medical image classification and segmentation. However, the performance of ViT significantly deteriorates when subjected to adversarial attacks. Almalik et al. [1] introduced a novel self-ensembling approach to enhance the robustness of the ViT model against adversarial attacks. The structure of Swin Transformer is shown in Fig. 1, where SwinTB represents Swin Transformer blocks. Transformer-based models have surpassed traditional Convolutional Neural Network (CNN) architectures in many image classification tasks. Considering computational resources and inference speed, we choose the Swin-base as our baseline model. Building on the Swin Transformer model, we introduce a novel classification model named Swin-MMC to specifically address myopic maculopathy classification. The performance details of Swin-MMC will be discussed in the next section.
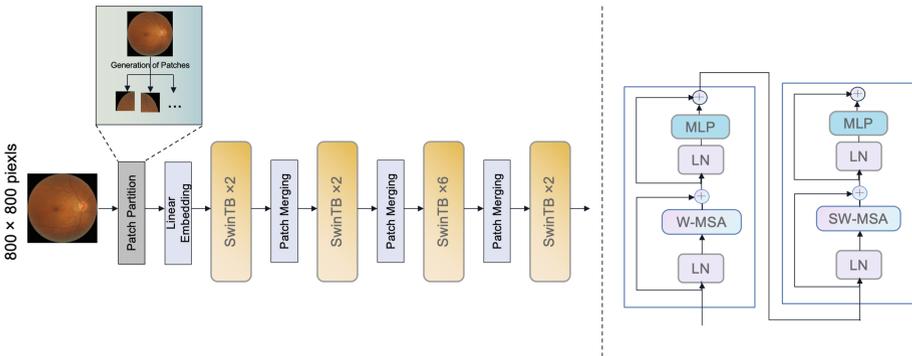
Fig. 1. The architecture of a Swin Transformer (Swin-base).

In this paper, our primary contributions are summarized as follows:

- We propose a Swin-based framework, named Swin-MMC, that is adaptive to medical image classification and demonstrates superior performance on the test dataset.
- We use the enhanced ArcFace loss with 3 sub-centers (En-ArcFace loss) as the model's loss function for the first time in classification tasks in the field of medical image processing.
- We design a weak label strategy based on our Swin-MMC model that can generate high-quality weak labels and make inferences efficiently simultaneously.

The rest of this paper is organized as follows. In Sect. 2, we will provide a detailed introduction to our classification method. Next, in Sect. 3, we will provide a detailed introduction to the dataset, as well as specific details about

the implementation of the experiments. Then, in Sect. 4, we will discuss the results and the ablation study, as well as our future work. Finally, in Sect. 5, we will conclude the paper.

2   Method

In this section, we will present our proposed method for classifying myopic maculopathy in detail. The architecture of our model is shown in Fig. 2. The model framework consists of four components: data augment, Swin-base, enhanced Arc-Face loss with 3 sub-centers, and weak label. Initially, in the supervised training phase, color fundus images are subjected to data augmentation, and the augmented images, along with their corresponding ground truth labels, are input to the Swin-base module for training, resulting in the initial Swin-MMC model. Then, in the semi-supervised training phase, the same model is applied once again to all fundus images to train the Swin-MMC model. Unlike other methods, we combined the training and validation sets to form a new image dataset by merging the true labels of the training data with those of the validation data to create the ground truth for the new dataset. This new dataset is used for three rounds of training, ultimately yielding the final Swin-MMC model. Throughout this process, the loss function employed is the En-ArcFace loss with 3 sub-centers.
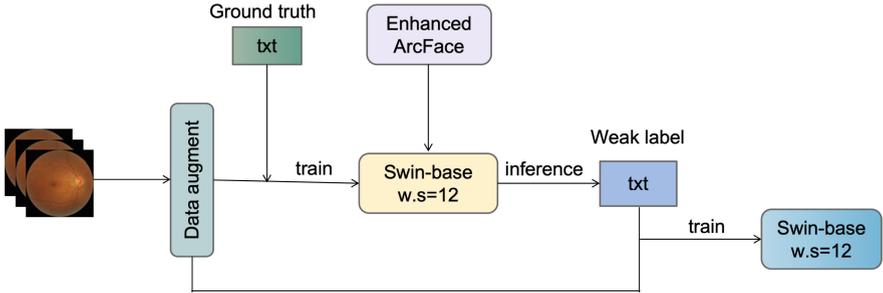
Fig. 2. The architecture of a Swin-MMC.

2.1   Enhanced ArcFace Loss with 3 Sub-centers

In the field of medical image processing, we applied the sub-center ArcFace loss function for the first time to calculate the loss. The enhanced ArcFace loss with 3 sub-centers is an improved version of the ArcFace loss, and it currently presents the best performance during the further test phase of MMAC Task 1. The En-ArcFace loss is widely used in face recognition tasks, capable of handling numerous facial categories [5]. This loss function can obtain more discriminative features compared to using softmax + cross-entropy because it calculates

geodesic distance on a high-dimensional hypersphere, rather than Euclidean distance. Therefore, utilizing ArcFace loss as the loss function in the model enhances the ability to accurately discern variations among different lesion categories in fine-grained color fundus image classification tasks. This implies that the model is better equipped to accurately capture the differences between various lesion categories. Similar to many face recognition tasks, the myopic maculopathy classification task typically encompasses multiple categories. The En-ArcFace loss is specifically designed to handle a large number of categories and, after fine-tuning, proves to be well-suited for the task of myopic maculopathy classification. Our claims have also been validated through ablation experiments, as detailed in Table 5.

The most widely used softmax loss function for classification tasks is presented as follows:

$$L = -\log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{N} e^{W_j^T x_i + b_j}}, \tag{1}$$

where $x_i \in R^d$ stands for the deep feature of the i-th sample, belonging to the y_i-th class, d stands for dimension. $W_j \in R^d$ the j-th column of the weight $W \in R^{d \times N}$. The number of classes in the classification task is N. $b_j \in R^N$ is the bias term. Then, by replacing $W_j^T x_i = \|W_j\| \|x_i\| \cos\theta_j$, We can now transform the logistic function into an ArcFace loss function, and s is the radius of the hypersphere.

The enhanced ArcFace loss [5] is presented as follows:

$$L' = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{N} e^{s\cos\theta_j}}. \tag{2}$$

where θ_j is the angle between the weight W_j and the feature x_i. m is an additive angular margin penalty between x_i and W_{y_i}, $\theta_j = arccos\left(\max_k\left(W_{j_k}^T \mathbf{x}_i\right)\right), k \in \{1, \cdots, K\}$. In our paper, k is 3.

2.2   Weak Label

Weak Label Scheme Our weak label strategy includes the following steps:

- In the first stage, a special strategy different from the usual data augmentation methods was employed. Initially, the original training dataset and the validation dataset were merged into a new, larger dataset. The real labels of the original training data were combined with the real labels of the validation data to create the labels for the new dataset. Preprocessing of the dataset was carried out during this stage.
- In the second stage, the preprocessed fundus image data, along with the real labels, were input into the Swin-base model for training. This stage resulted in the initial Swin-MMC model and weak labels for the training dataset.

- In the third stage, the dataset obtained in the first stage was used as input images, and the weak labels generated in the second stage were used as input labels. They were input into the Swin-base model for further training.
- In the fourth stage, the operations from the second stage and third stages were repeated three times, ultimately leading to the final Swin-MMC model.

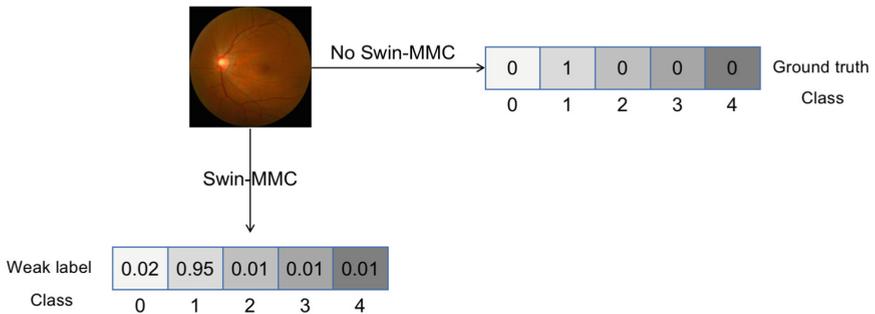**Advantages of Weak Label**

- We create a larger dataset by merging the original training and validation data. This step aims to provide the model with a more diverse set of samples, enhancing its ability to learn image features and patterns effectively.
- Second, as depicted in Fig. 3, we employ the Swin-MMC model to transform the initial consistent labels into probability values representing each image's likelihood of belonging to different categories. These probability values are then used as training labels. The approach helps improve the model's generalization, enabling it to better handle noise and uncertainty in real-world applications.
- Furthermore, our weak label strategy involves a multi-stage training process, introducing more variations and diversity. This allows the model to better capture complex relationships and features within the data, thereby enhancing its robustness.
- Finally, the weak label strategy can be easily applied to other datasets and models.

In conclusion, the weak label strategy offers benefits such as dataset expansion, enhanced generalization, improved robustness, and adaptability to various datasets and models.

3 Experiments

3.1 Dataset and Evaluation Measures

In task 1 of the MMAC competition, the objective is to classify myopic maculopathy. The dataset for this task consists of a comprehensive collection of color

Fig. 3. The advantage of weak label.

fundus images. The training set comprises a total of 1143 images, sourced from two distinct data centers. Among these, 990 images originate from Data Center 1, while 153 images are obtained from Data Center 2. The validation set includes 248 color fundus images, with 215 images from Data Center 1 and an additional 33 images from Data Center 2. The test set contains 915 color fundus images, with 783 images originating from Data Center 1 and 132 images from Data Center 2. The detailed description of the data set is shown in Table 1.

Table 1. Data set descriptions.

Data set	Total images	Data center 1	Data center 2
Training set	1143	990 (87%)	153 (13%)
Validation set	248	215 (87%)	33 (13%)
Testing set	915	783 (86%)	132 (14%)

Furthermore, the dataset is enriched with essential patient metadata associated with each image, providing insights into age, sex, height (measured in centimeters, cm), and weight (measured in kilograms, kg). It is important to acknowledge the possibility of missing metadata values in some instances. This comprehensive dataset serves as the foundation for evaluating and advancing the performance of algorithms in the task of myopic maculopathy classification.

MMAC Task 1 aims to accurately determine whether a color fundus image falls into one of five categories: no macular lesions, tessellated fundus, diffuse chorioretinal atrophy, patchy chorioretinal atrophy, and macular atrophy. These categories are labeled with natural numbers: 0, 1, 2, 3, and 4, with higher numbers indicating more severe conditions. Consequently, the metrics calculation code utilizes quadratic weighted kappa (QWK), a widely employed evaluation metric for various medical imaging problems. QWK is a statistical measure designed to assess the agreement between two annotators. To ensure fairness, the evaluation metrics also include the F1 score and macro specificity, and individual metric scores are separately calculated on all test cases, and the final ranking score is obtained by averaging the scores of all metrics.

3.2   Image Preprocessing and Augmentation

In the preprocessing and augmentation phase, fundus images underwent normalization using the mean and variance extracted from the Imagenet dataset. This strategy is rooted in the presumption that leveraging the pretrained weights from Imagenet, in conjunction with its statistical characteristics, would maintain inherent attributes, including spatial locality and translational equivariance, acquired during model training. A thorough analysis of medical literature revealed that myopic maculopathy predominantly appears in the central zone of fundus images, with only occasional presence towards the periphery. To accentuate this central significance, images were first resized to 416×416 pixels and

subsequently subjected to a central crop, yielding a resolution of 384×384 pixels to align with the model's input requirements. This procedure not only amplifies data heterogeneity but also minimizes peripheral noise. Given the standardized procedure of fundus image acquisition in clinical contexts, certain augmentations like vertical flipping or Gaussian blurring were considered redundant. Therefore, the only adopted augmentation was horizontal flipping, executed with a probability of 0.5, to preserve the diagnostic essence of the images.

3.3   Implementation Details

The configuration of our experimental setups, including the development environments and requirements, can be found in Table 2. Detailed procedures employed for training the Swin-MMC model are delineated in Table 3.

Table 2. Development environments and requirements.

Ubuntu version	Linux release
CUDA version	11.3
CPU	15 vCPU AMD EPYC 7543 32-Core
GPU (number and type)	1× NVIDIA A40 48GB Tensor Core GPU
Programming language	Python 3.8.0
Deep learning framework	Pytorch (Torch 1.11.0)
Specific dependencies	mmpretrain 1.0.0rc8
Code	https://github.com/LuliDreamAI/MICCAI_TASK1

In order to optimize model convergence and enrich the feature learning process during training, we incorporated a two-phase learning rate scheduling approach:

LinearLR Increasing Strategy. The initial learning rate is 0, and the strategy linearly increases the learning rate to 0.000125 for the first 5 epochs. After this period, the learning rate stabilizes, ensuring no further modifications. This method is predicated on the idea of hastening convergence in the nascent epochs by leveraging a relatively augmented learning rate.

CosineAnnealingLR Strategy. Commencing post the 5th epoch, this strategy employs a cosine annealing approach to the learning rate, setting a floor value at 1e-5. This progressive decrement of the learning rate, characterized by its cosine nature, facilitates intricate model tuning during the concluding training phases, steering the model to a more refined convergence point.

Table 3. Training protocols.

Basic network	Swin Transformer (base)
Network initialization	Pretrained weight in Imagenet1K and 21K
Batch size	32
Window size	12×12
Optimizer	Warmup with betas(0.9, 0.99)
Loss	Enhanced ArcFace loss with subcenter=3
Weight decay	0.05
Initial learning rate (lr)	0.000125
Training time per iteration	0.68 s

4   Results and Discussion

4.1   Results on Testing Set

In the testing set, the challenge uses several metrics, namely quadratic weighted kappa, macro F1, macro specificity, and average, for evaluation. It's important to note that the "average" metric is calculated as the mean value obtained from quadratic weighted kappa, macro F1, macro specificity, and specificity. In the testing phase, our method achieved second place with an average of 86.60%. In the further testing phase, it performed even better, improving by 1.63% and ultimately securing first place with an average surpassing the original first place by 0.71%. Detailed evaluation metric scores for our approach in both the testing phase and the further testing phase can be found in Table 4.

Table 4. The quantitative results of the test phase and further test phase.

Phase	Model	Volumetric(%) ↑			Average(%) ↑
		QWK	Macro F1	Macro Specificity	
Test	Swin-MMC	88.93	76.81	94.06	86.60
Further Test	Swin-MMC+Weak label	89.81	80.42	94.48	88.23

4.2   Visualization Heatmap Analysis

We input different original fundus images, as shown in Fig. 4(A), into the Swin-MMC model. The model generates a color fundus heatmap through the visualization layer, highlighting areas it deems most critical and displaying them using a color scale as depicted in Fig. 4(B). On the right side of Fig. 4, the cylinder corresponds to different colors as the numbers increase, indicating the regions that the model pays increasing attention to. We can observe that Swin-MMC is

capable of learning and distinguishing the lesion areas of different categories of myopic maculopathy degeneration and correctly classifying the images into their respective.
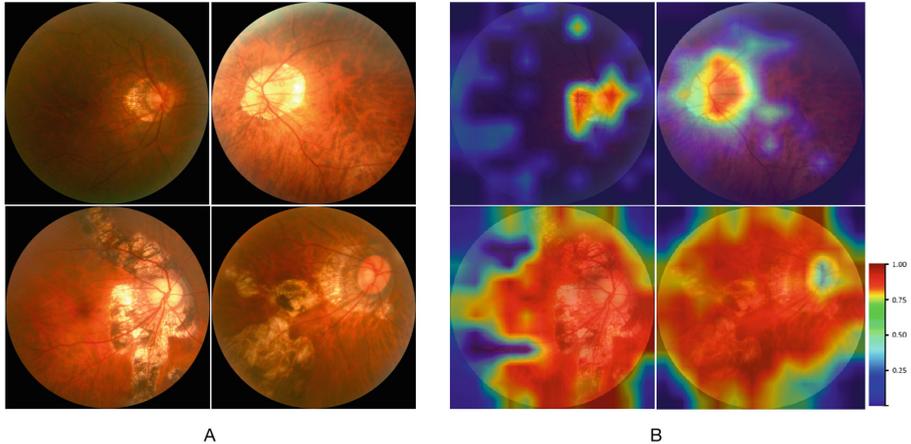
Fig. 4. Visualization of Swin-MMC for classifying the category of myopic maculopathy. **(A)** The different categories (Category 1 - Category 4) of myopic maculopathy in the original images. **(B)** Heatmaps generated on the deep features of the original images. Typical myopic maculopathy lesions were observed in hot regions.

4.3   Ablation Study in Further Test Phase

To validate the effectiveness of the En-ArcFace loss function and the weak label strategy in classifying Myopic Maculopathy, we conducted ablation experiments during the further test phase on both the Swin-MMC and Swin-MMC+Weak label models, as detailed in Table 5. For the Swin-MMC model, in the absence of the En-ArcFace loss function, the performance was at 87.04%. Introducing the En-ArcFace loss function improved the performance to 87.14%. Importantly, when incorporating the En-ArcFace loss function and combining it with the weak label strategy, the model's performance reached 88.23%. This represents a significant improvement of 1.63% compared to the previous 86.60% observed during the test phase. These results indicate that the En-ArcFace loss function and the weak label training strategy enable the model to capture the features of fundus images more accurately, thereby enhancing the overall model performance.

Table 5. Ablation experiments of En-ArcFace loss and weak label strategies.

Phase	Model	Loss		Volumetric(%) ↑			Average(%) ↑
		ArcFace	En-ArcFace	QWK	Macro F1	Macro Specificity	
Further Test	Swin-MMC	✓		89.87	77.04	94.22	87.04
			✓	89.88	77.24	94.32	87.14
	Swin-MMC+ Weak label	✓		**90.24**	79.67	94.41	88.11
			✓	89.81	**80.42**	**94.48**	**88.23**

4.4 Limitation and Future Work

Based on the Swin-base model, we have proposed a simple and effective weak label strategy, combined with the En-ArcFace loss function. Up to this point, we have achieved the best performance on the test data.

Moreover, we are confident that there is further potential for enhancement in our model. Firstly, self-supervised pre-training has achieved significant success in the field of image classification, enabling the learning of domain-level features from images. However, self-supervised pre-training requires a large volume of images as a foundation, and the limited availability of medical images constrains the ability to perform self-supervised training. Therefore, we aim to leverage weights pre-trained on ImageNet with self-supervised methods such as SimMIM [22] for continued training on fundus images, followed by fine-tuning with the pre-trained weights, which could yield promising results. Secondly, we can explore novel data augmentation techniques tailored to our retinal image dataset. Appropriate data augmentation strategies can enhance our performance when employing techniques like Test-Time Augmentation (TTA). Lastly, if there is no strict requirement for inference time, we may consider employing an ensemble approach with multiple models to enhance model performance.

5 Conclusion

In the classification of myopic maculopathy degeneration, our approach has successfully achieved high-precision recognition capability. There are two primary challenges facing this task: firstly, fundus images and general real-world images belong to different domains; secondly, the color representation of different diseases in the images is similar, making accurate classification particularly challenging.

To address these challenges, in comparison to other techniques, we innovatively proposed the use of En-ArcFace loss and weak label to improve the performance of our image classification algorithm. Specifically, the En-ArcFace loss provides a tighter feature embedding for classification tasks, aiding in distinguishing categories that look incredibly alike. Meanwhile, because the true label is too sharp, the weak label strategy offers an opportunity for the model to learn a more precise feature, further enhancing the model's generalization capability.

Through these methods, we have not only achieved outstanding results on the initial training dataset but also demonstrated superior performance on further

test dataset, far surpassing traditional methods. This offers ophthalmologists a powerful tool to more accurately diagnose and differentiate types of myopic maculopathy degeneration, leading to more precise treatment recommendations for patients.

References

1. Almalik, F., Yaqub, M., Nandakumar, K.: Self-ensembling vision transformer (SEViT) for robust medical image classification. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2022. MICCAI 2022. LNCS, vol. 13433, pp. 376–386. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_36

2. Bilal, A., Zhu, L., Deng, A., Lu, H., Wu, N.: Ai-based automatic detection and classification of diabetic retinopathy using u-net and deep learning. Symmetry **14**(7), 1427 (2022)

3. Chen, R., et al.: Automatic recognition of ocular surface diseases on smartphone images using densely connected convolutional networks. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 2786–2789. IEEE (2021)

4. Dai, L., et al.: A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nat. Commun. **12**(1), 3242 (2021)

5. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)

6. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

7. Fang, Z., Xu, Z., He, X., Han, W.: Artificial intelligence-based pathologic myopia identification system in the ophthalmology residency training program. Front. Cell Dev. Biol. **10**, 1053079 (2022)

8. Gummadi, S.D., Ghosh, A.: Classification of ocular diseases: a vision transformer-based approach. In: Roy, S., Sinwar, D., Dey, N., Perumal, T., Tavares, J.M.R.S. (eds.) Innovations in Computational Intelligence and Computer Vision. ICICV 2022, LNNS, vol. 680, pp. 325–337. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-2602-2_25

9. Holden, B.A., et al.: Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology **123**(5), 1036–1042 (2016)

10. Hossain, K.F., Kamran, S.A., Ong, J., Lee, A.G., Tavakkoli, A.: Revolutionizing space health (Swin-FSR): advancing super-resolution of fundus images for SANS visual assessment technology. In: Greenspan, H., et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2023, MICCAI 2023, LNCS, vol. 14226, pp 693–703. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43990-2_65

11. Khan, I.A., Sajeeb, A., Fattah, S.A.: An automatic ocular disease detection scheme from enhanced fundus images based on ensembling deep CNN networks. In: 2020 11th International Conference on Electrical and Computer Engineering (ICECE), pp. 491–494. IEEE (2020)

12. Liu, R., et al.: Deepdrid: diabetic retinopathy-grading and image quality estimation challenge. Patterns **3**(6), 100512 (2022)

13. Liu, Z., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

14. Lu, L., et al.: 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 (2021)

15. Nawaz, M., et al.: An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization. Sensors **22**(2), 434 (2022)

16. Ohno-Matsui, K., et al.: International photographic classification and grading system for myopic maculopathy. Am. J. Ophthalmol. **159**(5), 877–883 (2015)

17. Potapenko, I., et al.: Automated artificial intelligence-based system for clinical follow-up of patients with age-related macular degeneration. Acta Ophthalmol. **100**(8), 927–936 (2022)

18. Sebastian, A., Elharrouss, O., Al-Maadeed, S., Almaadeed, N.: A survey on deep-learning-based diabetic retinopathy classification. Diagnostics **13**(3), 345 (2023)

19. Sun, Y., et al.: A deep network using coarse clinical prior for myopic maculopathy grading. Comput. Biol. Med. **154**, 106556 (2023)

20. Velpula, V.K., Sharma, L.D.: Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion. Front. Physiol. **14**, 1–117588 (2023)

21. Wang, R., et al.: Efficacy of a deep learning system for screening myopic maculopathy based on color fundus photographs. Ophthalmol Therapy **12**(1), 469–484 (2023)

22. Xie, Z., et al.: Simmim: a simple framework for masked image modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9653–9663 (2022)

23. Zang, P., Hormel, T.T., Hwang, T.S., Bailey, S.T., Huang, D., Jia, Y.: Deep-learning-aided diagnosis of diabetic retinopathy, age-related macular degeneration, and glaucoma based on structural and angiographic oct. Ophthalmol. Sci. **3**(1), 100245 (2023)