JTE HCM-UTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Engineering
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Retinal Diseases Classification From OCT Images Using Pretrained Dual-Encoder Architecture

Ngo Quang Huy[2], Hoang Thai Xuan Khoa[1], Le Van Vinh[1*]
[1]*Ho Chi Minh City University of Technology and Engineering, Vietnam*
[2]*International University, VNU-HCM, Vietnam*

[*]Corresponding author. *Email: vinhlv@hcmute.edu.vn*

ARTICLE INFO

Received: 29/12/2025
Revised: 03/02/2026
Accepted: 23/02/2026
Published: 28/02/2026

KEYWORDS

Multimodal;
Classification;
Retinal diseases;
OCT images;
Deep learning.

ABSTRACT

Retinal diseases, such as age-related macular degeneration (AMD), diabetic retinopathy (DR), and glaucoma, are leading causes of irreversible vision loss, making early and accurate diagnosis essential for effective treatment. Optical coherence tomography (OCT) provides high-resolution cross-sectional retinal images that support disease assessment; however, many challenges still remain due to noise and artifacts in images or complex retinal structures. In this study, we propose a dual-encoder framework for retinal disease classification from OCT B-scan images by jointly leveraging two pretrained foundation models: RETFound and MIRAGE. Following standardized preprocessing and resampling, high-quality features extracted from the encoders are combined for the final classification tasks. To mitigate overfitting on limited medical data, the RETFound encoder is frozen during training to preserve general visual features, whereas the MIRAGE encoder is fine-tuned to adapt to specific classification objectives. Extensive experiments conducted on seven public OCT benchmark datasets demonstrate that the proposed method outperforms single-encoder baselines on the majority of benchmarks. The framework achieved an average balanced accuracy (BAcc) of 89.8%, an F1-score of 90.7%, and a Matthews Correlation Coefficient (MCC) of 83.9%. These results confirm the effectiveness of combining complementary pretrained encoders for robust and generalizable retinal disease classification in clinical settings.

Doi: https://doi.org/10.54644/jte.2026.2062

1. Introduction

Retinal diseases such as age-related macular degeneration (AMD), diabetic retinopathy (DR), and glaucoma are major causes of vision impairment and blindness [1]. Thus, early diagnosis and monitoring are beneficial in preventing permanent vision loss. The classification of retinal diseases from images significantly supports the disease detection and helps clinicians identify related abnormalities more efficiently. Various imaging techniques are used for the diagnosis and study of retinal diseases, including color fundus photography (CFP), scanning laser ophthalmoscopy (SLO), and three-dimensional optical coherence tomography (OCT) [2]. While CFP and SLO are mainly used to identify characteristic surface lesions such as microaneurysms, hemorrhages, and abnormal blood vessels, OCT provides high-resolution images (B-scans) supporting the assessment of retinal layer structures.

Early approaches for the classification of retinal diseases in images are mainly based on classical machine learning methods. A classification method using Support Vector Machines and Histogram of Oriented Gradients (HOG) descriptors was proposed in [3]. Similarly, the work in [4] is based on a Bayesian network classifier to analyze OCT to detect AMD disease. However, a major limitation of these methods is the dependence on manually designed features tailored to specific diseases, and the sensitivity to noise.

Deep learning techniques enable the ability to automatically learn feature representations from data, providing high quality performance in medical image analysis. Convolutional neural network (CNN)

architectures such as VGG-16 [5], ResNet [6], Inception-V3 [7], Inception-ResNet-v2 [8] and OCTNET [9] are able to perform disease classification tasks for images effectively. Furthermore, deep learning has demonstrated significant success in other complex medical imaging tasks, such as COVID-19 chest CT segmentation using attention mechanisms [10] and robust unified frameworks for liver tumor analysis [11]. However, those techniques mainly focus on local patterns of images and lack global attention. Some approaches, e.g., [12], [13], apply Vision Transformer (ViT) [14] for the retinal classification. The model uses a self-attention mechanism that allows capturing global context across the entire retinal image. A group of studies (e.g., RETFound [15], REMEDIS [16]) has focused on the development of Foundation Models for a wide range of image analysis tasks. These models are trained based on different techniques such as masked autoencoder [17], contrastive learning [18], or self-distillation [19], [20] which are designed to learn features from unlabeled data to create generalizable models. RETFound is based on a masked autoencoder where the encoder uses a large Vision Transformer. Otherwise, REMEDIS applies a contrastive learning algorithm in its classification process.

Recent studies use multimodal learning models that combine multiple types of input data, such as OCT, CFP, SLO images, disease descriptions, and expert knowledge [21], [22], [23], [24], [25]. However, existing multimodal models are unstable when one input modality is missing. MIRAGE model [26] addresses the issue through a Paired Multimodal MAE (Paired MultiMAE) mechanism [27], which is designed to learn detailed representations from OCT, SLO, and segmentation maps (pseudo-labels), and still works effectively when some input modalities are absent.

Despite the efforts, the noise and artifacts, or the complex structure of the retina still pose many research challenges. Motivated by recent studies, we propose a classification method for retinal diseases from OCT images which is based on a pretrained dual-encoder architecture. The proposed method jointly combines two pretrained encoders of MIRAGE and RETFound to utilize the strength of two models. Output features from the encoders are concatenated and passed through a fully connected layer for predicting the disease labels. Experimental results on available benchmarking datasets show the strength of the proposed method compared to baseline approaches.

The rest of this paper is organized as follows. Section 2 describes the proposed method, including data preprocessing and the dual-encoder architecture combining RETFound and MIRAGE. Section 3 presents the experimental results. The final section is for conclusion.

2. Method

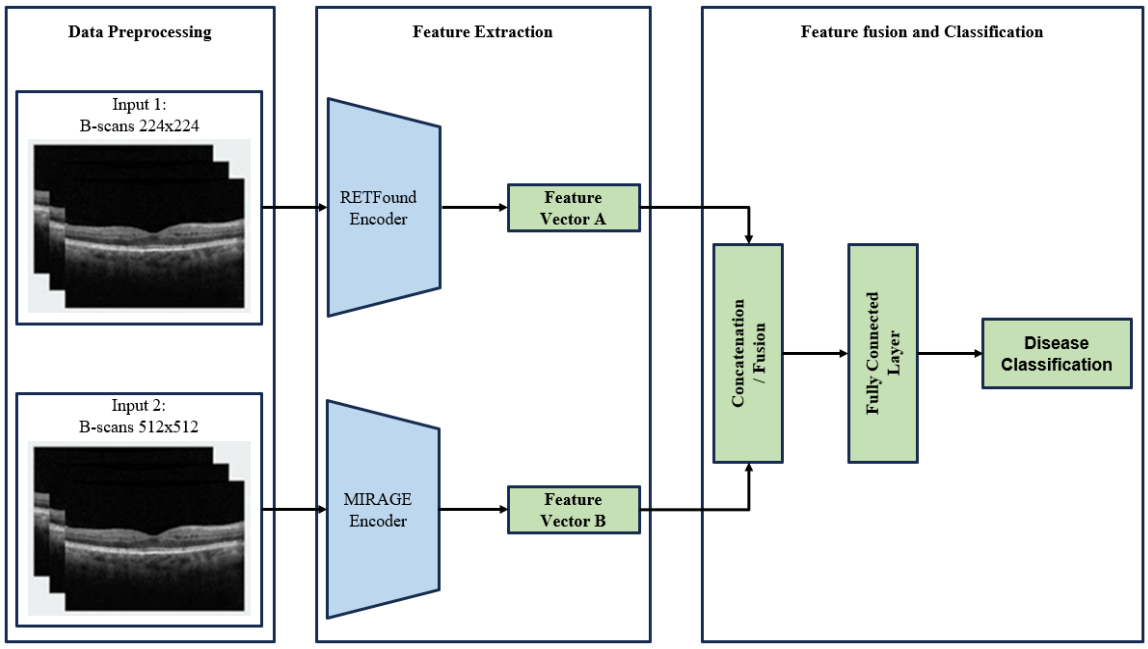

Figure 1. *Overall architecture of the proposed method.*

The proposed model for retinal disease classification from OCT B-scan images, presented in Figure 1, utilizes a dual-encoder architecture of two pretrained encoders, RETFound and MIRAGE. Feature vectors extracted from the encoders are then aggregated to form a meaningful representation before being fed into the final classification layer. The proposed method consists of three phases: (1) Data Preprocessing, (2) Feature Extraction, and (3) Feature fusion and Classification.

2.1. Data preprocessing

In order to provide suitable inputs for the dual-encoder architecture, we apply a preprocessing stage consisting of two steps: (1) Standardization, and (2) Resampling.

Step 1: Standardization

In order to eliminate non-informative noisy regions, such as excessive black borders or edge artifacts, and focus on the central structures of the retina, we crop the images to extract the largest possible square region from each original image. In particular, with an input image I_{raw} with dimensions W x H, the crop size is defined as $S = \min(W, H)$ - b, where b represents a margin parameter. This technique allows us to extract a region of interest I_{crop} of size $S \times S$ at the center of the image. It ensures that essential information remains while normalizing the ratio to 1:1.

Step 2: Resampling

In this step, we resample the cropped image I_{crop} using Lanczos method [28] which uses a windowed Sinc function and aims to keep edges and high-frequency components, helping preserve the sharpness of the image. The method is used separately to resize the images into two target sizes: 224 × 224 pixels is for the RETFound encoder and 512 × 512 pixels for MIRAGE.

2.2. Feature extraction, Fusion and Classification

This phase uses an architecture of two streams of pretrained encoders.

The first stream uses RETFound encoder. The encoder is based on a large Vision Transformer (ViT-large) consisting of 24 Transformer blocks. It has been trained on 1.6 million retinal images using the Masked Autoencoder (MAE), which enables the model to learn major structural features of the retina.

In the second stream, the encoder from the MIRAGE model (base version) is used. Different from RETFound encoder, MIRAGE model applies a paired multimodal pre-training on OCT B-scans, SLO images, and retinal layer segmentation images (pseudo-labels). The architecture is also designed to handle the case that only one modality is available. When only B-scans images are provided, MIRAGE encoder utilizes the latent space previously learned from cross-modal training to construct semantic representations. As a result, the output feature vector V_{MIR} could be more robust to noise.

Two output vectors from the two encoders V_{RET} and V_{MIR} are normalized and concatenated to construct joint feature vector V_{Joint}:

$$V_{Joint} = Concat(V_{RET}, V_{MIR}) \quad (1)$$

V_{Joint} is fed into a fully connected layer to map the features into the disease labels, and thus producing the final prediction.

2.3. Training strategy and Evaluation metrics

Between two encoders of the dual encoder architecture, RETFound is a larger model with ~303 million parameters, while MIRAGE has a smaller structure of ~86 million parameters. Thus, in order to help the proposed method preserve general visual features from RETFound encoder, as well as prevent overfitting when continuing training on small data, its parameters are frozen during the fine-tuning process. In contrast, MIRAGE encoder stream parameters are updated so that the model can adapt to the classification objective effectively.

This work uses different metrics to assess the performance of classification methods including: Balanced Accuracy (BAcc), F1-score, Matthews Correlation Coefficient (MCC) [29], [30], [31].

Let C be the number of classes, and $Recall_i$ be calculated as follows:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

BAcc is the mean of the recall achieved for each class and calculated as follows [32]:

$$BAcc = \frac{1}{C}\sum_{i=1}^{C} Recall_i \quad (3)$$

Besides, F1-score is the mean of precision and recall that balances false positives and false negatives. The MCC shows overall classification quality of the models using a confusion matrix.

3. Experiments and Results

3.1. Datasets and Experimental Setups

The proposed method and baseline methods are evaluated on seven benchmarking datasets which are used in previous study [26], including Duke iAMD, GAMMA, Harvard Glaucoma, Noor Eye Hospital, OCTDL, OCTID, and OLIVES. Table 1 provides detailed descriptions of these datasets. Each dataset, containing OCT images, is split into training, validation, test subsets following the same split configuration as in MIRAGE.

Table 1. *Benchmarking datasets.*

Datasets	Train/Val/Test	No. images	Dimensions (pixels)	No. of classes	Retinal diseases
Duke iAMD	229/77/77	383	~1000x512	2	AMD, Control
GAMMA	60/20/20	100	~512x992	3	Glaucoma: early, mid_advanced, non
Harvard Glaucoma	479/121/400	1000	~300x200	2	Glaucoma: glaucoma, non
Noor Eye Hospital	88/30/30	148	~512x496	3	AMD, DME, Normal
OCTDL	1388/343/332	2063	Variable	7	AMD, DME, ERM, NO, RAO, RVO, VID
OCTID	316/82/174	572	~700x500	5	Anormal, ARMD, CSR, Diabetic retinopathy (DR), Macular Hole
OLIVES	1107/274/209	1590	~504x400	2	DME, DR

The model is trained using the cross-entropy loss function incorporating a label smoothing technique with a coefficient of ϵ=0.1 to prevent overfitting. AdamW optimizer algorithm is used to optimize the model with a learning rate of 1e-5 for the case that the encoders are unfrozen, and lr=1e-3 for the case of the frozen encoders (applied to the trainable classification head). Besides, weight decay is 0.01 and the batch size is set to 32. The training process runs for a maximum of 100 epochs with an early stopping mechanism and using a patience of 15 epochs if the balanced accuracy values on the validation set do not improve.

The proposed method is compared with MIRAGE and RETFound. The parameters of the two baseline models are unfrozen and updated during the fine-tuning process with the benchmarking datasets. All experiments were conducted on a server system equipped with an Intel Core i9-13900K CPU and 128 GB of RAM, ensuring efficient data loading and processing capabilities. Besides, the model training and inference processes were accelerated using an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM.

3.2. Results

Table 2 presents the average results of the proposed method and two baseline methods on seven datasets. It can be seen that the proposed method achieves the best overall performance compared with the remaining methods, while RETFound gets the lowest for all metrics. The proposed model gets higher average BAcc from 1.8% to 7.7%, average F1-score 1.6% to 7.1%, and average MCC 2.7% to 10.5% compared with the other models.

Table 2. *Average results of the proposed method and baseline methods.*

Methods	Average BAcc	Average F1-score	Average MCC
RETFound	82.1%	83.6%	73.4%
MIRAGE	88.0%	89.1%	81.2%
Proposed method	**89.8%**	**90.7%**	**83.9%**

Table 3, Table 4, and Table 5 present BAcc, F1-score, and MCC values, respectively, of the three methods on seven datasets. It can be seen from the tables that the proposed model has the highest BAcc, F1-score, and MCC values across five, six, and four of the seven datasets, respectively.

Separately compared with the MIRAGE model, the proposed method gets better BAcc values on 4 out of 7 datasets (GAMMA, Harvard Glaucoma, OCTDL, and OLIVES), and has a significant improvement on the GAMMA dataset by approximately 11% (from 68.1% to 79.2%). On the Duke iAMD and Noor Eye Hospital datasets, the two models achieved comparable BAcc values. OCTID is the only where the proposed model gets lower BAcc, F1-score, and MCC values compared with MIRAGE. Dataset OCTID is a small dataset (316 training images) but more complex than other datasets with 5 categories, while datasets like Duke iAMD, GAMMA, Harvard Glaucoma, Noor Eye Hospital, and OLIVES have only 2 to 3 classes. It may lead to the fact that the single model of MIRAGE is better than the proposed method which is affected by pre-trained RETFound encoder on the dataset.

Table 3. *BAcc values of RETFound, MIRAGE, and the proposed method.*

Model	Duke iAMD	GAMMA	Harvard Glaucoma	Noor Eye Hospital	OCTDL	OCTID	OLIVES
RETFound	**96.9%**	45.8%	75.3%	86.7%	84.9%	91.4%	93.5%
MIRAGE	**96.9%**	68.1%	74.6%	**96.7%**	88.7%	**95.4%**	95.7%
Proposed method	**96.9%**	**79.2%**	**75.7%**	**96.7%**	**90.2%**	92.6%	**97.3%**

Table 4. *F1-score values of RETFound, MIRAGE, and the proposed method.*

Model	Duke iAMD	GAMMA	Harvard Glaucoma	Noor Eye Hospital	OCTDL	OCTID	OLIVES
RETFound	**97.4%**	47.9%	75.8%	86.4%	89.9%	93.7%	93.8%
MIRAGE	**97.4%**	69.8%	74.8%	**96.7%**	92.5%	**96.6%**	95.7%
Proposed method	**97.4%**	**80.6%**	**76.0%**	**96.7%**	**93.2%**	94.2%	**97.1%**

Table 5. *MCC values of RETFound, MIRAGE, and the proposed method.*

Model	Duke iAMD	GAMMA	Harvard Glaucoma	Noor Eye Hospital	OCTDL	OCTID	OLIVES
RETFound	**93.8%**	24.6%	51.2%	80.4%	84.9%	91.8%	87.3%
MIRAGE	**93.8%**	54.8%	49.1%	**95.2%**	89.1%	**95.5%**	91.2%
Proposed method	**93.8%**	**70.7%**	**51.5%**	**95.2%**	**89.5%**	92.5%	**94.2%**

Considering the effectiveness of training strategy into the proposed method, we compare the model with its variants Ours_MF_RU, and Ours_MU_RU with different strategies of training, presented in Table 6. The two variants are for the case that RETFound encoder parameters are unfrozen. However, Ours_MF_RU uses MIRAGE encoder parameters that are frozen, while the ones of Ours_MU_RU are unfrozen during the fine-tuning process.

Table 6. *Training strategies of proposed method and its variants*.

Methods	MIRAGE Encoder	RETFound Encoder
Ours_MF_RU	Frozen	Unfrozen
Ours_MU_RU	Unfrozen	Unfrozen
Proposed method	Unfrozen	Frozen

The bar chart in Figure 2 shows that the two variants get lower average results for all the metrics compared with the proposed method. This suggests that allowing updating parameters of the large model, RETFound, with small data leads to a reduction in the effectiveness gained from pretraining on a large-scale dataset. In contrast, the case of updating MIRAGE encoder parameters (Ours_MU_RU) leads to improved classification performance compared with the case that the encoder parameters are frozen.

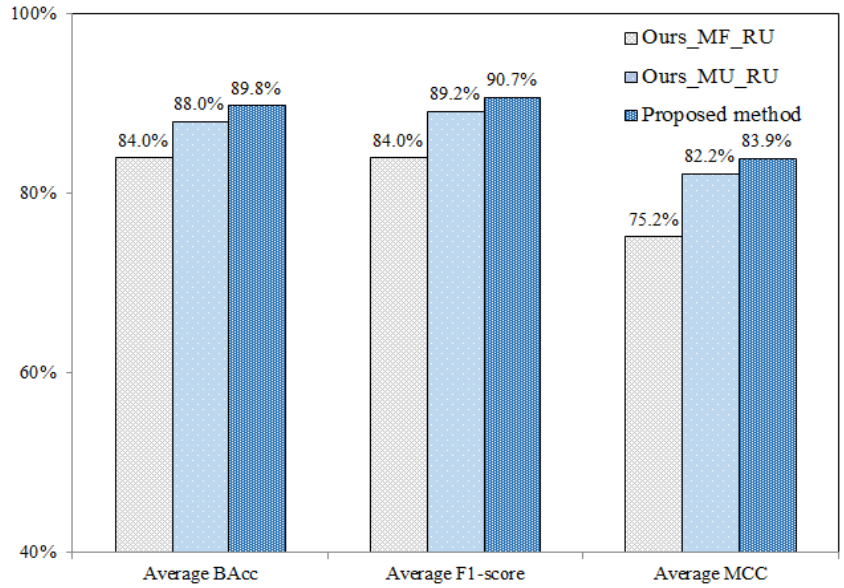

Figure 2. *Performance comparison of the proposed method and its variants across different evaluation metrics*.

Considering computational performance, the proposed method requires a higher number of parameters (393.68 M) compared with MIRAGE (86.04 M) and RETFound (303.30 M). Besides, our method maintains a throughput of 139.23 FPS compared with 578.46 FPS and 301.93 FPS of MIRAGE and RETFound, respectively.

4. Conclusion

This study proposes a classification method for retinal diseases from OCT images which is based on the fusion of two pretrained encoders of RETFound and MIRAGE. The proposed model is able to extract high-quality features by combining the strengths of the single foundation models supporting the final labeling task. Experimental results on seven public datasets show that the proposed method achieves overall performance improvements compared to using a single model for most cases. However, a drawback of the proposed method is requiring high computation cost due to using the dual-encoder. Some techniques could be applied in the future to make the model more lightweight. Furthermore,

exploring other advanced feature fusion mechanisms to better select useful information instead of simple concatenation is also our future research direction.

Acknowledgments

The authors would like to thank the Faculty of Information Technology, HCMC University of Technology and Engineering for providing facilities for this study. The applications presented in this paper were tested in the High Performance Computing system of the faculty.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

[1] M. J. Burton *et al*., "The Lancet Global Health Commission on global eye health: Vision beyond 2020," *The Lancet Global Health*, vol. 9, pp. e489–e551, 2021.
[2] J. S. Schuman, J. G. Fujimoto, J. J. Duker, and H. Ishikawa, *Optical Coherence Tomography of Ocular Diseases*. Boca Raton, FL, USA: CRC Press, 2024.
[3] V. Srinivasan *et al*., "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical Optics Express*, vol. 5, no. 10, pp. 3568–3577, 2014.
[4] A. Albarrak, F. Coenen, and Y. Zheng, "Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction," in *Proc. Int. Conf. Medical Image Understanding and Analysis (MIUA)*, 2013, pp. 59–64.
[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.
[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
[8] A. Kamble *et al*., "Automated diabetic macular edema (DME) analysis using fine-tuning with Inception-ResNet-v2 on OCT images," in *Proc. IEEE-EMBS Int. Conf. Biomedical Engineering and Sciences (IECBES)*, 2018, pp. 442–446.
[9] O. Perdomo *et al*., "Oct-net: A convolutional network for automatic classification of normal and diabetic macular edema using SD-OCT volumes," in *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI)*, 2018, pp. 1423–1426.
[10] N. Tran, "Segmentation on chest CT imaging in COVID-19 based on the improved attention U-Net model," in *Proc. Int. Conf. Intelligent Software Methodologies, Tools, and Techniques (SOMET)*, 2022, pp. 596–606.
[11] V. Pham *et al*., "Robust engineering-based unified biomedical imaging framework for liver tumor segmentation," *Current Medical Imaging*, vol. 19, no. 1, pp. 37–45, 2023.
[12] W. Nazih, A. O. Aseeri, O. Y. Atallah, and S. El-Sappagh, "Vision transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images," *IEEE Access*, vol. 11, pp. 117546–117561, 2023.
[13] D. Philippi, K. Rothaus, and M. Castelli, "A vision transformer architecture for the automated segmentation of retinal lesions in spectral-domain optical coherence tomography images," *Scientific Reports*, vol. 13, no. 1, art. no. 517, 2023.
[14] A. Dosovitskiy *et al*., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.
[15] Y. Zhou *et al*., "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156–163, Sep. 2023.
[16] S. Azizi *et al*., "Robust and efficient medical imaging with self-supervision," *arXiv preprint* arXiv:2205.09723, 2022.
[17] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16000–16009.
[18] Y. Liu *et al*., "Contrastive predictive coding with transformer for video representation learning," *Neurocomputing*, vol. 482, pp. 154–162, 2022.
[19] L. Zhang, C. Bao, and K. Ma, "Self-distillation: Towards efficient and compact neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4388–4403, Aug. 2022.
[20] M. Oquab *et al*., "DINOv2: Learning robust visual features without supervision," *Transactions on Machine Learning Research*, 2024.
[21] X. Li *et al*., "Multi-modal multi-instance learning for retinal disease recognition," in *Proc. 29th ACM Int. Conf. Multimedia (MM'21)*, 2021, pp. 2474–2482.
[22] W. Wang *et al*., "Learning two-stream CNN for multi-modal age-related macular degeneration categorization," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5565–5576, Nov. 2022.
[23] Y. Li *et al*., "Multimodal information fusion for glaucoma and diabetic retinopathy classification," in *Ophthalmic Medical Image Analysis (OMIA)*. Cham, Switzerland: Springer, 2022, pp. 53–62.
[24] D. Shi *et al*., "Eyefound: A multimodal generalist foundation model for ophthalmic imaging," *arXiv preprint* arXiv:2405.11338, 2024.
[25] D. Shi *et al*., "A multimodal visual–language foundation model for computational ophthalmology," *npj Digital Medicine*, vol. 8, no. 1, art. no. 381, 2025.
[26] J. Morano *et al*., "Multimodal foundation model and benchmark for comprehensive retinal OCT image analysis," *npj Digital Medicine*, vol. 8, no. 1, art. no. 576, Sep. 2025.
[27] R. Bachmann, D. Mizrahi, A. Atanov, and A. Zamir, "MultiMAE: Multi-modal multi-task masked autoencoders," in *Proc. Eur. Conf. Computer Vision (ECCV)*. Cham, Switzerland: Springer, 2022, pp. 348–367.
[28] C. Lanczos and J. Boyd, *Discourse on Fourier Series*. Philadelphia, PA, USA: SIAM, 2016.
[29] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," in *Proc. Workshop on Evaluation and Comparison of NLP Systems*, 2020, pp. 79–91.

[30] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, art. no. 6, 2020.
[31] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[32] L. Mosley, *A Balanced Approach to the Multi-Class Imbalance Problem*, Ph.D. dissertation, Iowa State University, Ames, IA, USA, 2013.

Ngo Quang Huy received the bachelor's degree in Information Technology from University of Science (Vietnam National University, Ho Chi Minh City) in 2007. Currently, he is working at the International University - Vietnam National University HCM City.
Email: nqhuy@hcmiu.edu.vn. ORCID: https://orcid.org/0009-0004-9538-4111

Hoang Thai Xuan Khoa received the bachelor's degree in Information Technology, and MSc degree in Computer Science from HCMC University of Technology and Education (currently Ho Chi Minh City University of Technology and Engineering) in 2012 and 2025, respectively. Currently, he is working at the Ho Chi Minh City University of Technology and Engineering, Vietnam. His research interests include Bioinformatics, Deep learning.
Email: khoahtx@hcmute.edu.vn. ORCID: https://orcid.org/0009-0006-8912-297X

Le Van Vinh received the bachelor's degree in Information Technology, and MSc degree in Computer Science from University of Science (Vietnam National University, Ho Chi Minh City) in 2005 and 2009, respectively. He received the PhD degree in Computer Science from HCMC University of Technology (Vietnam National University, Ho Chi Minh City) in 2017. Currently, he is working at the faculty of Information Technology, Ho Chi Minh City University of Technology and Engineering, Vietnam. His research interests include Bioinformatics, Deep learning, Computer vision, and High-performance computing.
Email: vinhlv@hcmute.edu.vn. ORCID: https://orcid.org/0000-0001-5218-3089