Cerebrovascular Diseases Screening from Color Fundus Photography via Cross-View Fusion and Graph-Based Discrimination

Congyu Tian*[1,2], Shihao Zou*[1], Xiangyun Liao[1], Cheng Chen[3], Chubin Ou(✉)[4], Jianping Lv[5], Shanshan Wang[1], and Weixin Si(✉)[6]

[1] Shenzhen Institutes of Advanced Technology, Shenzhen, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] The University of Hong Kong, HKSAR, China
[4] Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China
[5] Guangzhou First People's Hospital, Guangzhou, China
[6] Shenzhen University of Advanced Technology, Shenzhen, China
cou@connect.ust.hk, siweixin@suat-sz.edu.cn

Abstract. Cerebrovascular diseases can occur suddenly and unpredictably, making it crucial to identify high-risk individuals through screening to prevent or mitigate its impact. However, digital subtraction angiography (DSA), the current gold-standard, is difficult to apply to large-scale screening or primary healthcare settings due to its high cost, complex operation, and invasive nature. In contrast, Color Fundus Photography (CFP) can reflect related cerebrovascular diseases through retinal microvascular changes while maintaining low-cost and risk-free advantages. Nevertheless, current CFP image-based methods for predicting cerebrovascular disease mostly focus on pixel-level image features only, ignoring the correlation between arteriovenous morphology, optic disc structure and disease risk. To address this gap, we propose CVGB-Net, a method that integrates a cross-view encoder to fuse high-level semantic features, primarily capturing vascular abnormalities in the retinal vasculature caused by cerebrovascular diseases, with low-level pixel features extracted by the foundation model, RetFound, designed for ocular tasks. The fused cross-view features for each sample are then processed through a graph-based discriminator, which utilizes a graph adapter to link disease-related features across the entire dataset. This approach further enhances the model's ability to differentiate between diseased and healthy cases. To validate our approach, we present a tailored CFP-Cerebrovascular diseases Screening (CCS) dataset with 2,338 expert-diagnosed cases. Experimental results demonstrate the effectiveness of our approach, highlighting its potential for cost-effective large-scale cerebrovascular diseases screening. https://github.com/glodxy/CVGB_net

Keywords: Color Fundus Photography · Cerebrovascular diseases Screening· Graph Adapter.

*Equally contribute to this paper.

1 Introduction

Cerebrovascular diseases remain a leading cause of global mortality and chronic disability, accounting for nearly 11% of worldwide deaths [4]. While conventional neuroimaging modalities such as Digital Subtraction Angiography (DSA) and Magnetic Resonance Angiography (MRA) are clinically effective for screening cerebrovascular diseases, their limitations, such as suboptimal detection efficiency, high costs, and technical complexity, severely restrict their use in large-scale screening programs and primary healthcare settings. Recent studies have highlighted that retinal microvascular changes can serve as indicators of cerebrovascular and cardiovascular diseases [16, 5], prompting the development of methods to detect related diseases through retinal imaging.

Color Fundus Photography (CFP), as one of the retinal imaging modalities, has gained significant attention for its low cost and risk-free nature, making it suitable for screening programs. Several methods have explored combining CFP images with other data modalities to predict cardiovascular disease [14, 7, 17]. Furthermore, Lin et al. [9] introduced a cross-laterality feature alignment pre-training scheme to integrate information from CFP images of both eyes, thereby improving cardiovascular disease prediction. However, these methods usually rely on additional input data, which limits their applicability for large-scale screening.

In the field of cerebrovascular disease prediction, most methods focus on extracting more detailed information from CFP images. For example, Luengnaruemitchai et al. [12] employed region selection and polar transformation, similar to Polar-Net [10], to enhance feature extraction from CFP images. Other methods integrate specialized modules to extract more intricate features from CFP images [11, 1]. Additionally, Zhou et al. [19] pre-trained a foundation model on a large-scale dataset to develop a more effective feature extractor. More recently, Xia et al. introduced CoAtt-Net [18] to extract disease-related features at multiple levels from CFP images for disease prediction. However, all these methods primarily focus on extracting low-level, pixel-based features, often overlooking the intrinsic morphological relationships between blood vessels, optic discs, and cerebrovascular disease. For instance, blurring of the optic disc (OD) margins is commonly associated with intracranial hypertension [15], while arteriovenous (AV) nicking and an increased arteriolar light reflex can suggest atherosclerosis [6], both of which are indicative of a higher risk of cerebrovascular disease.

Inspired by clinical observations, we propose CVGB-Net, comprising a Cross-View Encoder (CVE) and a Graph Adapter (GA). The CVE extracts low-level pixel features from CFP images and high-level semantic features from vessel and optic disc masks, which are fused into a cross-view representation. The GA then refines this representation to improve discrimination between healthy and diseased cases, especially under data imbalance. To support validation, we introduce the CFP-Cerebrovascular diseases Screening (CCS) dataset, containing 2,338 expert-diagnosed cases (2,205 healthy, 133 diseased). Our main contributions are summarized as follows: (1) the CCS dataset for cerebrovascular disease screening with low-cost CFP images; (2) the CVE for integrating pixel- and semantic-level features; and (3) the GA for enhanced category discrimination.

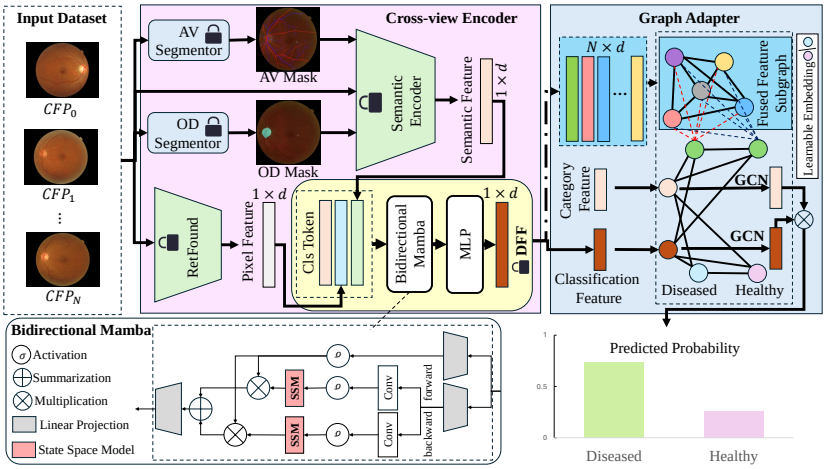

Fig. 1: Overview of our framework, which includes a Cross-View Encoder (CVE) and a Graph Adapter (GA). The CVE has two branches to extract low-level pixel and high-level semantic features, fused by a Dual Feature Fusion (DFF) module to produce cross-view classification features. The GA constructs subgraphs from all training samples' features to extract representative category feature nodes, which refine the input features via a graph neural network for improved discrimination. The lock icon denotes model freezing during training.

2 Method

The overall framework is illustrated in Fig. 1. First, the input CFP images are processed through the dual-branch architecture of the CVE, which captures both low-level pixel features and high-level semantic features to obtain cross-view classification feature. Next, the classification features are passed to the GA, where they undergo a similarity computation with reference category features to predict the most relevant clinical category.

2.1 Cross-view Encoder

Blurred OD margins in CFP images are often indicative of intracranial hypertension, while vascular features such as AV nicking and an increased arteriolar light reflex are strong indicators of atherosclerosis. These factors are closely linked to a higher risk of cerebrovascular diseases. To effectively integrate these key factors into our method, we propose combining the anatomic structure semantic view with the existing full-image pixel representation view to generate a cross-view feature for more accurate disease-related information capture. To achieve this, we introduce a novel Cross-view Encoder module, which incorporates a new branch designed to extract semantic-level features f_s from the segmented AV and OD masks using a pre-trained semantic encoder. The other branch, using a

fine-tuned RetFound [19], directly extracts pixel-level features f_p from the input CFP image.

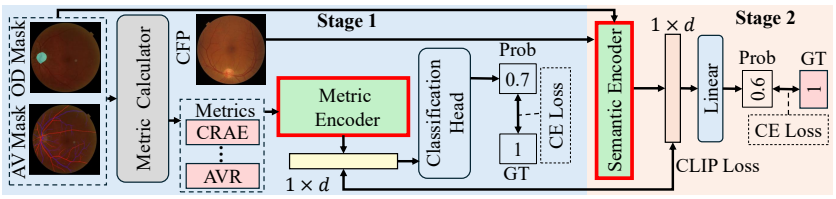

Fig. 2: The multi-stage training process of the Semantic Encoder involves two stages. In stage 1, we first train a Metric Encoder using cross-entropy (CE) loss. In stage 2, we align the features of the Semantic Encoder with those of the Metric Encoder by applying both CLIP loss and CE loss.

To ensure the features extracted by the semantic encoder contain essential vascular morphological information, we employ a Variational Autoencoder (VAE) as a metric encoder to assist in training the Semantic Encoder. Given that metrics like Arteriolar-to-Venular Ratio (AVR), Central Retinal Arteriolar Equivalent (CRAE), and Central Retinal Venular Equivalent (CRVE) are well-established retinal vascular indicators closely linked to cerebrovascular risk, we use the trained metric encoder to map these metrics to the feature vector f_m. The features f_s extracted by the Semantic Encoder are then aligned with f_m using CLIP loss. Additionally, we apply Cross-Entropy (CE) Loss to ensure that the features f_m and f_s are distinct for healthy and diseased cases. The multi-stage training process of the semantic encoder is illustrated in Fig. 2.

After training the RetFound and Semantic Encoder, we use the DFF module to fuse the features extracted from both branches into a cross-view representation. Specifically, inspired by the Vision Transformer (ViT) [3], we introduce a class token, t_{cls}, which is concatenated with the features from both branches and processed by efficient Mamba [2] to obtain the cross-view feature f_c. This cross-view feature f_c is then passed through an MLP to generate the cross-view classification features f_{cls} for each input CFP image. More specifically, during the training process of CVE, we first separately train the AV Segmentor and the OD Segmentor. After that, we freeze their parameters and train the Semantic Encoder as illustrated in Fig. 2. Meanwhile, we finetune RetFound on our dataset. Finally, we train the DFF module and fine-tune the entire CVE using a layer-wise learning rate decay strategy.

2.2 Graph Adapter

We introduce a new Graph Adapter module to enhance the model's ability to differentiate between healthy and diseased cases in cerebrovascular disease screening. Following the approach of Li et al. [8], we construct two sub-graphs,

$\{G_c, G_f\}$. The sub-graph G_c represents a class-specific sub-graph, composed of learnable nodes (embeddings) for each class, while G_f represents a feature sub-graph, constructed from the features extracted from all the training samples within each class.

For $G_c = \{\mathcal{N}_c, \mathcal{E}_c\}$, we treat each learnable embedding as a node to construct the node set $\mathcal{N}_c$. In our study, we have two embeddings of size 1024, representing the "healthy" and "diseased" classes, receptively. For the feature sub-graph $G_f = \{\mathcal{N}_f, \mathcal{E}_f\}$, we categorize all training samples into two groups (healthy and diseased) based on their labels and compute the average classification features for each group. These average features serve as the nodes in the node set $\mathcal{N}_f$. For both sub-graphs, we calculate the cosine similarity between all nodes to determine the edge weights, constructing the edge sets $\mathcal{E}_c, \mathcal{E}_f$.

After constructing the sub-graphs, we introduce additional nodes into the graph: the empty category features, f_{ref} (initialized with ones in our study), and the classification features, f_{cls}, extracted from the current batch. The edges of the graph are then updated accordingly. These nodes are processed through a Graph Convolutional Network (GCN) to aggregate information from both sub-graphs, G_c and G_f, yielding the final reference category feature, f'_{ref}, and the adjusted classification feature, f'_{cls}. A residual connection is applied afterward. The workflow is expressed as follows:

$$\begin{aligned} f'_{ref} &= \beta \cdot \mathrm{GCN}(f_{ref}, G_c, G_f) + (1-\beta) \cdot f_{ref}, \\ f'_{cls} &= \beta \cdot \mathrm{GCN}(f_{cls}, G_c, G_f) + (1-\beta) \cdot f_{cls}, \end{aligned} \tag{1}$$

where β is 0.2 in our implementation to control the weights of residual connections. We then calculate the similarity between f'_{cls} and f'_{ref} for each category and select the category with the highest similarity as the predicted label.

3 Experiment

3.1 Dataset

For this study, we constructed the new CCS dataset using color fundus photographs from 2205 healthy cases and 133 diseased cases collected in 2024 from the same medical institution, where each case was diagnosed as cerebrovascular disease when confirmed by angiography to fulfill any of the following conditions: major cerebral artery narrowed by more than 50%, large vessel occlusion, presence of aneurysm, presence of arteriovenous malformation and small-vessel disease. Cases free of the above conditions were defined as healthy. The average age is 50.4 (healthy) and 68 (diseased); the male/female ratio is 1017/1198 (healthy) and 57/79 (diseased). Common comorbidities include hypertension and diabetes. This study was approved by the institution ethics committee and informed consent was waived as no identifiable private information is involved. All color fundus images (original resolution: 2576×1934, see Fig. 3) are cropped and resized to 512×512 for practical use. The dataset is split into training, validation, and testing sets at a 3:1:1 ratio, comprising 1,323/441/441 healthy and 79/26/26 diseased cases, respectively.

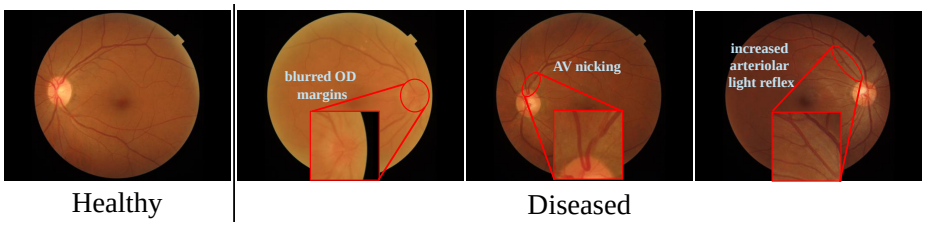

Fig. 3: Representative fundus color photographs from healthy cases and diseased cases. The three examples (from left to right) in the diseased cohort illustrate blurred OD margins, AV nicking, and an increased arteriolar light reflex.

Table 1: Classification results of various methods on our CCS dataset.

Method	Sensitivity	ROC-AUC	PR-AUC	F1-score
Vision Transformer [3]	0.1667	0.6833	0.5803	0.2308
Vision Mamba [20]	0.1389	0.6742	0.5867	0.2020
Polar-Net [10]	0.4444	0.7967	0.6014	0.3265
CoAtt-Net [18]	0.1806	0.8237	0.6681	0.2708
RetFound [19]	0.4444	0.8296	0.6645	0.3048
Ours w/o CVE	0.5634	0.8422	0.6886	0.3493
Ours w/o GA	0.5278	0.8510	0.6929	0.3393
Ours	**0.6389**	**0.8732**	**0.7088**	**0.3594**

3.2 Implementation Details

We implement our proposed method using PyTorch under Ubuntu 20.04 with four Nvidia RTX A6000 GPUs. For the Cross-View Encoder, we use the AdamW optimizer with a learning rate of 5e-4 and train it for 100 epochs with a batch size of 32. A layer-wise learning rate decay strategy is employed with a decay rate of 0.65. For the graph adapter, we use a learning rate of 5e-5 and train it for 50 epochs. The implementation and parameter settings for the comparison methods are consistent with the descriptions provided in their respective papers.

3.3 Evaluation and Interpretability Assessment

We evaluate the model on the test set using Sensitivity, ROC-AUC, PR-AUC and F1-score as performance metrics. For comparison, we benchmark our approach against two classical vision-domain methods (Vision Transformer [3], Vision Mamba [20]) and three methods specifically designed for retinal image-based prediction tasks (Polar-Net [10], CoAtt-Net [18], RetFound [19]). The comparative results are presented in Tab. 1. Our approach significantly outperforms the other methods in Sensitivity and achieves substantial improvements over RetFound [19] across all metrics, particularly in Sensitivity, ROC-AUC, PR-AUC, and F1-score, with gains of 19.45%, 4.36%, 4.43%, and 5.46%, respectively. Additionally, the results from the ablation experiments demonstrate that the introduction of the CVE module significantly enhances the model's performance

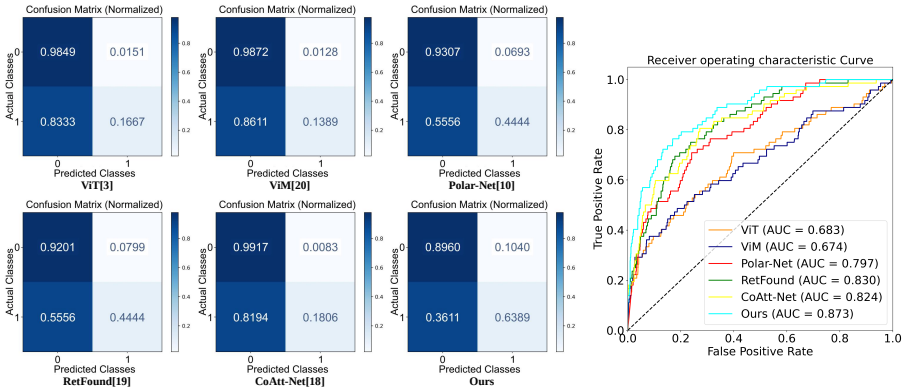

Fig. 4: Left: Normalized confusion matrices of difference methods. Category 0 represents healthy, category 1 represents disease. Right: Diagram of ROC curves for all methods.

across all metrics. The GA module helps balance the model's focus on features from different cases, slightly improving overall performance while notably increasing the F1 score.

To further validate the classification efficacy of our method, we visualize the confusion matrices for all comparative methods, as shown on the left side of Fig. 4. Our method achieves reliable identification of diseased cases, correctly classifying 63.89% of them, while other methods struggle to reliably detect diseased cases due to dataset imbalance. The effectiveness is further illustrated with the ROC curves for all compared methods on the right side of Fig. 4.

Additionally, we visualize the feature maps extracted by the pixel and semantic branches, as shown in Fig. 5. Existing studies suggest that microvascular dysfunction may increases the risk of cerebrovascular diseases [13]. From Fig. 5, it is evident that the pixel branch primarily targets intricate blood vessel networks in peripheral retinal regions, while the semantic branch concentrates on the optic disc and surrounding vessels, which have anatomical features with established clinical links to cerebrovascular disease.

Given the established link between a reduced AVR and an increased risk of cerebrovascular diseases, we compare the distribution of values derived from the CVE classification features with the distribution of values from the calculated AVR metrics. This comparison helps assess whether our method effectively utilizes vascular information. As shown in Fig. 5, the two distributions are similar, further supporting the idea that our method successfully retains relevant vascular information for extracting classification features.

4 Conclusion

In this paper, we propose a framework for cost-effective cerebrovascular disease screening based on fundus photography, which introduces a cross-view encoder

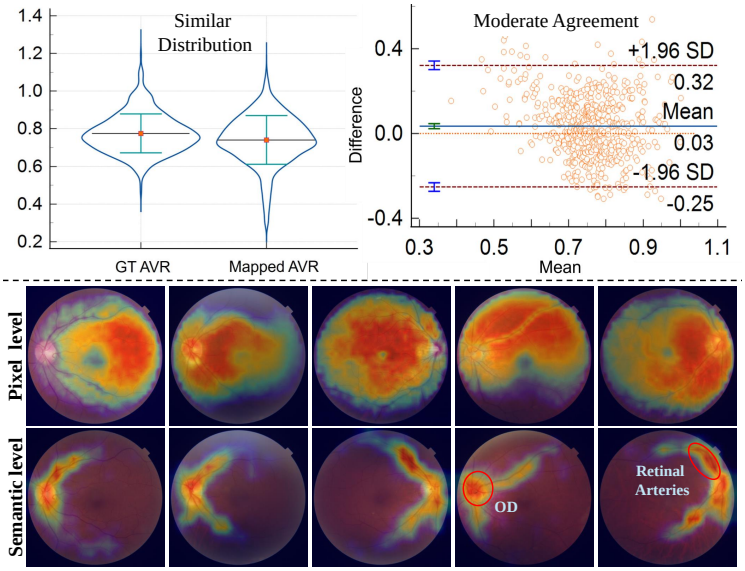

Fig. 5: Schematic of interpretability. The upper half shows a comparison between the categorized feature mapping values and the computed AVR values, while the lower half presents a visualization comparing the pixel feature branch to the semantic feature branch, where features from the OD and retinal arterial regions are identified as key factors associated with cerebrovascular diseases.

module that accurately integrates disease-related information. Additionally, we incorporate a graph adapter module to establish associations between classification features and category-specific embeddings, thereby enhancing the model's ability to discriminate features across different categories. We conduct experiments on our CCS dataset and perform an interpretability analysis of the results. Our method outperforms state-of-the-art approaches across multiple metrics, highlighting the potential of fundus photography for cerebrovascular disease screening.

Acknowledgements. This work was partially supported by a grant from the grants from National Natural Science Foundation of China (62372441, 82302300), in part by Guangdong Basic and Applied Basic Research Foundation (2023A1515 030268, 2023A0505030004), in part by Shenzhen Science and Technology Program (Grant No. RCYX20231211090127030) and in part by the Open Research Fund of State Key Laboratory of Digital Medical Engineering.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bang, S.Y.X., Le, K.N.T., Le, D.T., Choo, H.: Feature pool exploitation for disease detection in fundus images. In: 2023 17th International Conference on Ubiquitous

Information Management and Communication. pp. 1–4. IEEE (2023)
2. Dao, T., Gu, A.: Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In: International Conference on Machine Learning (2024)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
4. Feigin, V.L., Stark, B.A., Johnson, C.O., Roth, G.A., Bisignano, C., Abady, G.G., Abbasifard, M., Abbasi-Kangevari, M., Abd-Allah, F., Abedi, V., et al.: Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. The Lancet Neurology **20**(10), 795–820 (2021)
5. Ji, C., Li, J., Du, C., Lv, B., Wu, N., Li, H., Li, R., Hui, Y., Xie, G., Wu, S., et al.: Predicting cerebral small vessel disease through retinal scans and demographic data with bayesian feature selection. In: Medical Imaging 2024: Computer-Aided Diagnosis. vol. 12927, pp. 830–837. SPIE (2024)
6. Khazai, B., Adabifirouzjaei, F., Guo, M., Ipp, E., Klein, R., Klein, B., Cotch, M.F., Wong, T.Y., Swerdloff, R., Wang, C., et al.: Relation between retinopathy and progression of coronary artery calcium in individuals with versus without diabetes mellitus (from the multi–ethnic study of atherosclerosis). The American journal of cardiology **149**, 1–8 (2021)
7. Lee, Y.C., Cha, J., Shim, I., Park, W.Y., Kang, S.W., Lim, D.H., Won, H.H.: Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. npj Digital Medicine **6**(1), 14 (2023)
8. Li, X., Lian, D., Lu, Z., Bai, J., Chen, Z., Wang, X.: Graphadapter: Tuning vision-language models with dual knowledge graph. Advances in Neural Information Processing Systems **36** (2024)
9. Lin, Z., Shi, D., Zhang, D., Shang, X., He, M., Ge, Z.: Camera adaptation for fundus-image-based cvd risk estimation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 593–603. Springer (2022)
10. Liu, S., Hao, J., Xu, Y., Fu, H., Guo, X., Liu, J., Zheng, Y., Liu, Y., Zhang, J., Zhao, Y.: Polar-net: A clinical-friendly model for alzheimer's disease detection in octa images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 607–617. Springer (2023)
11. Liu, S., Wang, W., Deng, L., Xu, H.: Cnn-trans model: A parallel dual-branch network for fundus image classification. Biomedical Signal Processing and Control **96**, 106621 (2024)
12. Luengnaruemitchai, G., Sangchocanonta, S., Munthuli, A., Phienphanich, P., Puangarom, S., Jariyakosol, S., Hirunwiwatkul, P., Tantibundhit, C.: Automated alzheimer's, mild cognitive impairment, and normal aging screening using polar transformation of optic disc and central zone of fundus images. In: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1–4. IEEE (2024)
13. Ong, Y.T., De Silva, D.A., Cheung, C.Y., Chang, H.M., Chen, C.P., Wong, M.C., Wong, T.Y., Ikram, M.K.: Microvascular structure and network in the retina of patients with ischemic stroke. Stroke **44**(8), 2121–2127 (2013)
14. Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R.: Prediction of cardiovascular risk factors from retinal

fundus photographs via deep learning. Nature biomedical engineering **2**(3), 158–164 (2018)
15. Reier, L., Fowler, J.B., Arshad, M., Hadi, H., Whitney, E., Farmah, A.V., Siddiqi, J.: Optic disc edema and elevated intracranial pressure (icp): a comprehensive review of papilledema. Cureus **14**(5) (2022)
16. Scoles, D., McGeehan, B., VanderBeek, B.L.: The association of stroke with central and branch retinal arterial occlusion. Eye **36**(4), 835–843 (2022)
17. Wang, Y., Zhen, L., Tan, T.E., Fu, H., Feng, Y., Wang, Z., Xu, X., Goh, R.S.M., Ng, Y., Calhoun, C., et al.: Geometric correspondence-based multimodal learning for ophthalmic image analysis. IEEE Transactions on Medical Imaging **43**(5), 1945–1957 (2024)
18. Xia, X., Li, Y., Xiao, G., Zhan, K., Yan, J., Cai, C., Fang, Y., Huang, G.: Benchmarking deep models on retinal fundus disease diagnosis and a large-scale dataset. Signal Processing: Image Communication **127**, 117151 (2024)
19. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (2023)
20. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: Proceedings of the 41st International Conference on Machine Learning. vol. 235, pp. 62429–62442. PMLR (2024)