MICCAI

VAMPIRE: Uncovering Vessel Directional and Morphological Information from OCTA Images for Cardiovascular Disease Risk Factor Prediction

Lehan Wang[1], Hualiang Wang[1], Chubin Ou[2(✉)], Lushi Chen[3], Yunyi Liang[4], and Xiaomeng Li[1(✉)]

[1] The Hong Kong University of Science and Technology, Hong Kong, China
eexmli@ust.hk
[2] Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China
cou@connect.ust.hk
[3] Health Management Center, Foshan First People's Hospital
[4] Health Management Center, The Sixth Affiliated Hospital, School of Medicine, South China University of Technology

Abstract. Cardiovascular disease (CVD) remains the leading cause of death worldwide, requiring urgent development of effective risk assessment methods for timely intervention. While current research has introduced non-invasive and efficient approaches to predict CVD risk from retinal imaging with deep learning models, the commonly used fundus photographs and Optical Coherence Tomography (OCT) fail to capture detailed vascular features critical for CVD assessment compared with OCT angiography (OCTA) images. Moreover, existing methods typically classify CVD risk only as high or low, without providing a deeper analysis on CVD-related blood factor conditions, thus limiting prediction accuracy and clinical utility. As a result, we propose a novel multi-purpose paradigm of CVD risk assessment that jointly performs CVD risk and CVD-related condition prediction, aligning with clinical experiences. Based on this core idea, we introduce OCTA-CVD, the first OCTA dataset for CVD risk assessment, and a **V**essel-**A**ware **M**amba-based **P**rediction model with **I**nfo**r**mative **E**nhancement (**VAMPIRE**) based on OCTA enface images. Our proposed model aims to extract crucial vascular characteristics through two key components: (1) a Mamba-Based Directional (MBD) Module that captures fine-grained vascular trajectory features and (2) an Information-Enhanced Morphological (IEM) Module that incorporates comprehensive vessel morphology knowledge. Experimental results demonstrate that our method can surpass standard classification backbones, OCTA-based detection methods, and ophthalmologic foundation models. Our codes and the collected OCTA-CVD dataset are available at https://github.com/xmed-lab/VAMPIRE.

Keywords: OCTA · Mamba · Cardiovascular Disease Risk Prediction.
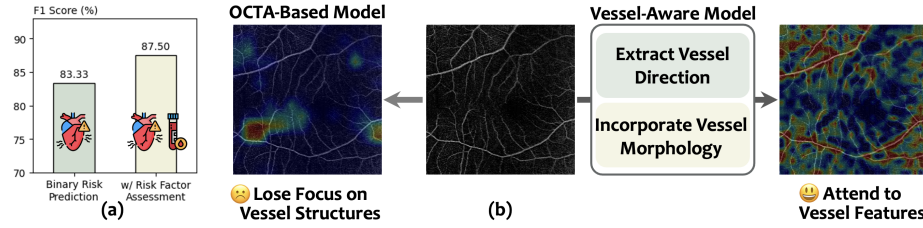
Fig. 1. Comparative Analysis of (a) Task Paradigm and (b) Model Capability. (a) Performance Comparison between binary risk classification and joint risk factor prediction. (b) Evaluation of vessel feature representation abilities between OCTA-Based Model and our proposed Vessel- Aware Model.

1   Introduction

Cardiovascular diseases (CVDs) remain a significant burden on public health and a leading cause of mortality worldwide. The World Health Organization (WHO) estimates that CVDs are responsible for nearly 18 million deaths each year, accounting for over 30% of all global deaths [1]. Thus, risk assessment of CVD is critical, as early intervention can prevent severe cardiovascular events.

Recent advances in retinal imaging and artificial intelligence (AI) [2,13,15,14] have the potential to revolutionize cardiovascular risk assessment into a non-invasive and automatic approach. The retinal vascular structure shares anatomical similarities with the coronary circulation, making it a promising biomarker for monitoring vascular health [6,23]. Several studies [3,4,11,19,21] have leveraged deep learning models for fundus photography or Optical Coherence Tomography (OCT) to predict CVD risk based on retinal vascular morphology and structural changes. Poplin *et al.* [21] first attempted to apply a deep learning model to fundus photos for predicting the risk of CVD events. Cheung *et al.* [3] automatically measured retinal-vessel calibre in retinal photographs and demonstrated its association with CVD risk factors. Other works [4,11] further complemented fundus photos with additional modalities, like cardiovascular magnetic resonance images and clinical risk factors, to further improve the risk prediction accuracy. Maldonado *et al.* [19] developed an AI model with OCT scans to leverage subtle abnormalities in retinal microstructure. These methods have achieved promising result in non-invasive CVD risk prediction, serving as a valuable alternative to conventional CVD risk assessment approaches and facilitating earlier detection.

Despite these advancements, two significant limitations hinder these models in providing more reliable and accurate risk prediction. First, the input modalities employed in the previous methods, fundus photography and OCT, capture insufficient vascular details for precise CVD risk assessment due to their limited resolution of microvascular structures, particularly compared with OCT angiography (OCTA) images that reflect capillary networks and vessel density more clearly. Second, existing approaches merely perform binary CVD risk detection without assessing fine-grained CVD risk factors, such as blood glucose

and cholesterol, which undermines both confidence and reliability. As proved by [28], these blood biomarkers are vital in determining CVD risk, and excluding these factors would notably reduce risk prediction accuracy. Additionally, simply providing binary risk classification (high/low risk) poses challenges for patients to understand and trust the results, preventing the implementation of timely and personalized intervention strategies.

To overcome these limitations, we propose a novel paradigm of utilizing OCTA images to predict four key CVD-related conditions, including high blood glucose, high blood cholesterol, high blood triglycerides, and high blood pressure, while jointly performing 10-year CVD risk prediction. Unlike existing methods that solely perform binary risk classification, our multi-task framework provides deeper insights into blood biomarker abnormalities, which align well with empirical clinical knowledge, thus substantially enhancing CVD risk prediction performance by over 4% in F1 score, see in Fig. 1 (a). Besides, the identification of blood factor conditions offers more clinically interpretable results could be delivered, thereby addressing a critical gap in current CAD risk assessment frameworks.

To fulfill this objective, we collect **the first OCTA enface dataset, OCTA-CVD, for joint CVD risk and factor estimation**. We first evaluate existing deep learning models developed to extract modality-specific features from OCTA images. [8,16,27,29]. However, our evaluation reveals that directly applying these methods inadequately captures continuous vascular structures, which is a crucial characteristic in OCTA imaging, as illustrated in Fig. 1 (b). Consequently, we draw inspiration from Mamba's capability in modeling long-range dependencies [7,9,26,31], which is suitable for capturing vascular trajectory. Nevertheless, Mamba-based models typically process images in patches, hindering the comprehensive understanding of continuous morphological features. Building upon these insights, we introduce a **V**essel-**A**ware **M**amba-based **P**rediction model with **I**nfo**r**mative **E**nhancement (**VAMPIRE**) for joint CVD risk and CVD-related factor estimation given OCTA enface images. The proposed VAMPIRE model consists of two components: ❶ **Mamba-Based Directional (MBD) Module** focusing on the intricate trajectories of inner vascular paths, and ❷ **Information-Enhanced Morphological (IEM) Module** to integrate comprehensive vessel shape information. These two components could collaboratively extract CVD-specific vessel features from OCTA enface images, thus effectively leveraging vascular structure information for accurate prediction of CVD risk and related factors. By combining both directional and morphological knowledge, our method could effectively attend to vessel features as in Fig. 1 (b), and outperform general classification baselines, OCTA-based detection models, and ophthalmological foundation models on multi-center validation datasets.

2 Methodology

2.1 Architecture Overview

An overview of our framework is presented in Fig. 2. To align with standard Mamba sequence processing, we first segment the input OCTA enface image
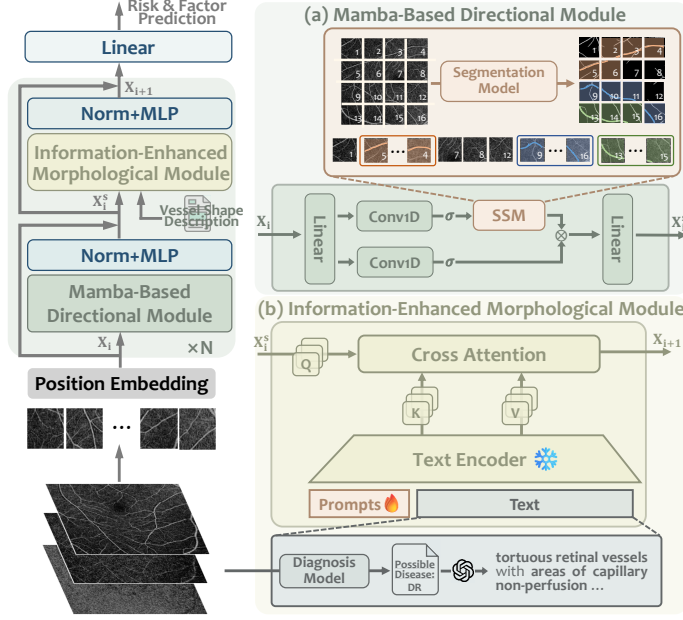
Fig. 2. Overview of the Proposed VAMPIRE Framework, which contains (a) Mamba-Based Directional (MBD) Module for vascular trajectory feature extraction and (b) Information-Enhanced Morphological Module for vessel shape knowledge integration.

$\mathbf{I} \in \mathbb{R}^{B \times H \times W}$ into flattened patches $\mathbf{P} \in \mathbb{R}^{B \times N \times h \times w}$, where B represents the batch size, N denotes the number of patches, and $(H, W)/(h, w)$ corresponds to the size of original image/patch. These 2D patches are then projected into 1D sequences and combined with position encodings, yielding $\mathbf{X} \in \mathbb{R}^{B \times N \times D}$, where D represents the dimension of patch embeddings. Following Vim [31], we also apply the class token $\mathbf{t}_{cls}$ to encode global knowledge. Next, the complete patch sequence $\mathbf{X_0} = [\mathbf{t}_{cls}; \mathbf{X}]$ is fed into the proposed sequential VAMPIRE blocks, which is composed of two parts: a Mamba-Based Directional (MBD) Module that effectively extracts vascular sequence features (§ 2.2), and an Information-Enhanced Morphological (IEM) Module that incorporates vessel morphology descriptions to enrich the model's representation (§ 2.3). Finally, the extracted features are input into a linear classification layer to obtain the final prediction for CVD risk and related conditions. The model is optimized with binary cross-entropy loss as the classification loss function. In the following sections, we will elaborate on the details of each constituent module.

2.2 Mamba-Based Directional Module

To effectively extract vessel trajectory features from OCTA enface images, we adopt the Mamba-Based Directional (MBD) Module for modeling vascular sequences, as demonstrated in Fig. 2 (a). The intricate structure of blood vessels

in OCTA images poses challenges for conventional Mamba scanning operations. The commonly applied linear- or cross-scanning approaches inevitably disrupt vessel continuity, thereby constraining the model's capability to capture intrinsic vascular structure. To address this limitation, we intuitively introduce vessel segmentation maps as directional guidance to formulate the scanning order within the State Space Model (SSM) module. Specifically, we employ SAM-OCTA [25], a pre-trained vessel segmentation network for OCTA images, to generate initial vessel maps, followed by noise filtering and morphological transformation to acquire well-defined vascular structures. Subsequently, we use depth-first search (DFS) to traverse the vessel tree in the segmentation map to obtain different continuous trajectories of large vessels, and generate patch sequences along each vascular path $\mathbf{V_i} = [v_{i,1}, ...v_{i,n_i}]$ based on the refined segmentation map. The complete scanning sequence is then constructed by concatenating vessel patches from different branches with intermediate background patches $\mathbf{B_{i,j}} = [b_{i,1}, ...b_{i,m_i}]$, which can be formulated as:

$$\mathbf{S} = \text{concat}_{i,j \in T}[\mathbf{V_i}; \mathbf{B_{i,j}}; \mathbf{V_j}] , \tag{1}$$

where T is the number of all the continuous vessel trajectories.

Through the vessel-following traversal paths, our proposed MBD module enables each OCTA patch to integrate contextual information from adjacent patches along the same vascular branch, thus encouraging the model to concentrate on CVD relevant vascular features.

2.3   Information-Enhanced Morphological Module

After obtaining the output of the MBD module, $\mathbf{X_i^s}$, we proceed to integrate vessel shape descriptive information through our subsequent Information-Enhanced Morphological (IEM) module. Considering that the patch-wise encoding mechanism in Mamba-based architecture would potentially constrain the model's ability to comprehend global vascular morphology, we propose to promote feature extraction through textual descriptions generated by the Multimodal Large Language Model (MLLM), as shown in Fig. 2 (b). However, we observe that directly feeding OCTA images into MLLMs yields generic and uninformative descriptions, such as "the retinal vessels are well-defined", providing limited morphological insights. Consequently, we first employ a classification model trained on the OCTA-500 dataset [12] to identify potential retinal diseases. Subsequently, we prompt GPT-4o [20] with the diagnostic results to generate descriptions $\mathbf{t}$ on possible vascular morphologies. Combined with learnable prompts $\mathbf{p}$, these descriptions are fed into a frozen text encoder Θ_T and then integrated with image representation $\mathbf{X_i^s}$ through the cross-attention mechanism. The process is denoted as:

$$\mathbf{T} = \Theta_T([\mathbf{p}; \mathbf{t}]) , \tag{2}$$

$$\mathbf{X_{i+1}} = \text{CrossAttention}(\mathbf{X_i^s}, \mathbf{T}, \mathbf{T}) , \tag{3}$$

Table 1. Data Statistics of the collected OCTA-CVD dataset.

Category	HR	HG	HC	HTG	HTN	Total
Patients	101	81	267	296	215	843
Images	195	160	525	580	420	1659

Through this approach, our model could effectively capture both fine-grained vessel trajectory features and global morphological information, enhancing its capability to identify CVD-associated microvascular patterns.

3   Experiments

3.1   Experimental Setup

Dataset. We collect an in-house OCTA dataset, OCTA-CVD[1], for joint CVD risk and factor estimation, which contains 1659 OCTA enface images of Superficial (ILM-IPL), Deep (INL-OPL), and Avascular (OPL-BM) layers from 843 patients in a local hospital's health check-up center from 2022 August to 2022 December. The dataset includes 74.3% female and 25.7% male, with a mean age of 45±12 years.. The OCTA images were captured in a 6×6 mm^2 area centering at fovea with a Velite C3000 OCTA device. Each patient is annotated with CVD High Risk (HR) and CVD-related conditions derived from blood test, including high blood glucose (HG), high blood cholesterol (HC), high blood triglycerides (HTG), and high blood pressure (hypertension, HTN). The data statistics is shown in Table 1. The dataset is divided using five-fold cross-validation, maintaining consistent class distributions in both training and test sets. To prevent data leakage, we ensured that data from individual patients appeared in only one subset.

Data Preprocessing. We first apply median filter for OCTA enface images for noise reduction, and then implement contrast-limited adaptive histogram equalization (CLAHE) to enhance vessel clarity. All the images are resized to 448×448 for input. We adopt data augmentation including random crop, flip, and rotation.

Implementation Details. During training, we also incorporate patient demographic information (age and gender) as additional input features to concatenate with the image features for final classification. The OCT layers are concatenated along the channel dimension to form a three-channel input. All the experiments are conducted on the NVIDIA RTX 3090 GPU. We employ AdamW optimizer with a base learning rate of $1e^{-4}$, scaled according to batch size, and a weight decay of $5e^{-2}$. A cosine annealing scheduler is applied to gradually reduce the learning rate to zero. For fair evaluation, we apply identical data preprocessing strategies and running epochs for all the experiments and report the average performance of all the categories at the final epoch across five-fold cross-validation. Our results are reported at eye level.

---

[1] Our OCTA-CVD dataset is released at https://github.com/xmed-lab/VAMPIRE

Table 2. Results on our collected OCTA-CVD dataset. The bold and underlined numbers indicate the best and second-best score, respectively. †: CNN backbones; ⋆: OCTA-based detection methods; ‡: Fine-tuned ophthalmologic foundation models.

Model	Precision	Recall	F1 Score	Accuracy	AUC	AUPR
ResNet[†][10]	0.4928 ± 0.0396	0.5750 ± 0.0387	0.5225 ± 0.0268	0.7576 ± 0.0244	0.7581 ± 0.0244	0.5295 ± 0.0224
EfficientNet[†][24]	0.4581 ± 0.0387	0.6057 ± 0.0542	0.5188 ± 0.0352	0.7485 ± 0.0179	0.7582 ± 0.0215	0.5207 ± 0.0340
ViT[†][5]	0.5058 ± 0.0593	0.5783 ± 0.0270	0.5374 ± 0.0214	0.7752 ± 0.0202	0.7778 ± 0.0222	0.5513 ± 0.0164
Swin V2[†][18]	0.5010 ± 0.0279	0.5592 ± 0.0589	0.5265 ± 0.0264	0.7782 ± 0.0282	0.7879 ± 0.0356	0.5582 ± 0.0357
Vim[†][31]	0.4636 ± 0.0137	0.5517 ± 0.0461	0.4851 ± 0.0161	0.7818 ± 0.0450	0.7971 ± 0.0263	0.5712 ± 0.0308
PolarNet[⋆][17]	$\underline{0.5758}\pm0.0376$	0.4142 ± 0.0341	0.4077 ± 0.0509	0.7788 ± 0.0043	0.7569 ± 0.0474	0.5291 ± 0.0124
Eye-AD[⋆][8]	0.5153 ± 0.0199	0.6349 ± 0.0132	$\underline{0.5754}\pm0.0102$	0.7594 ± 0.0126	0.7501 ± 0.0235	0.5307 ± 0.0143
RETFound[‡][30]	0.5614 ± 0.0077	0.5886 ± 0.0356	0.5407 ± 0.0116	$\underline{0.7885}\pm0.0116$	0.8024 ± 0.0190	0.5838 ± 0.0382
VisionFM[‡][22]	0.5610 ± 0.0355	$\underline{0.6351}\pm0.0439$	0.5522 ± 0.0255	0.7739 ± 0.0164	$\underline{0.8061}\pm0.0192$	$\underline{0.6009}\pm0.0341$
VAMPIRE	$\mathbf{0.6068}\pm\mathbf{0.0652}$	$\mathbf{0.6570}\pm\mathbf{0.0442}$	$\mathbf{0.6271}\pm\mathbf{0.0293}$	$\mathbf{0.8012}\pm\mathbf{0.0183}$	$\mathbf{0.8244}\pm\mathbf{0.0153}$	$\mathbf{0.6467}\pm\mathbf{0.0394}$

Table 3. Extended Comparison on Supplementary Dataset. The AUC metric for each category is presented.

Model	HR	HG	HC	NTG	HTN
ViT [5]	0.9430	0.6436	0.6068	0.6655	0.7371
RETFound [30]	0.9599	0.6584	0.6377	0.7094	0.7465
VisionFM [22]	0.9716	0.6632	0.6546	0.7220	0.7365
VAMPIRE	**0.9742**	**0.7649**	**0.7205**	**0.7462**	**0.7668**

Table 4. Ablation study of the proposed MBD and IEM modules in our method on the OCTA-CVD dataset.

Model	F1 Score	AUC	AUPR
Baseline	0.5522 ± 0.0255	0.8061 ± 0.0192	0.6009 ± 0.0341
w/ MBD	0.6162 ± 0.0322	0.8208 ± 0.0180	0.6301 ± 0.037
w/ IEM	0.6101 ± 0.0319	0.8195 ± 0.0193	0.6398 ± 0.0355
Both	$\mathbf{0.6271}\pm\mathbf{0.0293}$	$\mathbf{0.8244}\pm\mathbf{0.0153}$	$\mathbf{0.6467}\pm\mathbf{0.0394}$

3.2 Comparison with State-of-the-Art Models

To demonstrate the effectiveness of our proposed approach, we compare our model with Convolutional Neural Network (CNN) backbones [5,10,18,24,31], OCTA-based detection methods [8,17], and fine-tuned ophthalmologic foundation models [22,30]. As shown in Table 2, it is apparent that fine-tuned ophthalmologic foundation models can outperform both CNN backbones and OCTA-based models. This suggests that the inherent ophthalmologic knowledge facilitates the model to effectively adapt to the OCTA data distribution even without specific pre-training on OCTA images. Furthermore, the comparison results demonstrate that our proposed VAMPIRE model marginally surpasses all the other baselines. Notably, our approach achieves the improvement of 7.55% (64.67% *v.s.* 57.12%) and 11.6% (64.67% *v.s.* 53.07%) compared with the best AUPR of CNN backbones and OCTA-specific models. Additionally, our model improves the state-of-the-art fine-tuned ophthalmologic foundation model from 55.22% to 62.71% in F1 score and 60.09% to 64.67% in AUPR.

Extended Experiment. To further verify the effectiveness and generalizability of our method, we also collect a dataset from another hospital for evaluation, which contains 765 images from 379 patients. The AUC results of all the categories are shown in Table 3. Notably, our proposed VAMPIRE model consistently outperforms other baselines across all five categories, demonstrating its remarkable ability to interpret OCTA-specific features with a particular target on vessel characters.
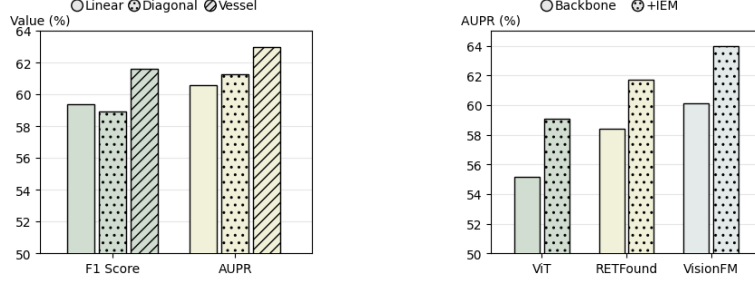
Fig. 3. Comparison between Different Scanning Strategies for the Mamba-Based Directional Module.

Fig. 4. The Results of Integrating the Information-Enhanced Morphological Module into Different Backbones.

3.3   Ablation Studies

Effectiveness of Proposed Modules. The improvement of our VAMPIRE model can be attributed to the proposed two components, the Mamba-Based Directional (MBD) module modeling the trajectories of inner blood vessels and the Information-Enhanced Morphological (IEM) module combining vessel morphology descriptions with the image features. From Table 4, we can conclude that applying MBD and IEM independently can still improve the overall result by 6.4%, 2.92% in F1 score and 2.92%, 3.89% in AUPR. This proves that both designs significantly enhance the model performance, and removing either module would lead to decreased results.

Analysis of Vessel-Following Scanning Strategies. To demonstrate the indispensability of our proposed Vessel-Following scanning strategy, we perform linear scanning and diagonal scanning for comparison. Fig. 3 presents the average F1 score and AUPR metric of these different implementations. It can be observed that other scanning methods show minimal contribution to performance enhancement. This underscores the unique advantage of our approach in extracting specific vascular features, which particularly aligns with the OCTA imaging characteristics.

Generalizability of Vessel Shape Information. To prove the value of descriptive vessel shape knowledge, we similarly integrate the IEM module into other backbones and assess the impact. Fig. 4 shows that our IEM module brings performance improvements regardless of the backbone architecture employed, demonstrating the effectiveness of incorporating external knowledge about vessel morphology. This suggests that leveraging vessel shape information as prior knowledge would further enhance the OCTA-specific feature extraction process.

4   Conclusion

In this paper, we formulate a multi-task paradigm that simultaneously predicts CVD risk and related blood conditions, substantially enhancing risk estimation

accuracy and reliability. We introduce OCTA-CVD dataset, the first collection of OCTA enface images for joint CVD risk and factor assessment, along with a Vessel-Aware Mamba-based Prediction model with Informative Enhancement (VAMPIRE). Our framework effectively concentrates on the vital vessel structure features within the OCTA images through two key components: a Mamba-Based Directional (MBD) Module that extracts vessel trajectory features, and an Information-Enhanced Morphological (IEM) Module that incorporates vessel shape knowledge. These components collaboratively extract CVD-specific vessel features from OCTA images, significantly improving CVD risk prediction and factor estimation accuracy.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cesare, M., Perel, P., Taylor, S., Kabudula, C., Bixby, H., et al.: The heart of the world (2024)
2. Chen, Z., Wang, H., Ou, C., Li, X.: Mutri: Multi-view tri-alignment for oct to octa 3d image translation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20885–20894 (2025)
3. Cheung, C.Y., Xu, D., Cheng, C.Y., Sabanayagam, C., Tham, Y.C., Yu, M., Rim, T.H., Chai, C.Y., Gopinath, B., Mitchell, P., et al.: A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. Nature biomedical engineering **5**(6), 498–508 (2021)
4. Diaz-Pinto, A., Ravikumar, N., Attar, R., Suinesiaputra, A., Zhao, Y., Levelt, E., Dall'Armellina, E., Lorenzi, M., Chen, Q., Keenan, T.D., et al.: Predicting myocardial infarction through retinal scans and minimal personal information. Nature Machine Intelligence **4**(1), 55–61 (2022)
5. 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. arXiv preprint arXiv:2010.11929 (2020)
6. Farrah, T.E., Webb, D.J., Dhaun, N.: Retinal fingerprints for precision profiling of cardiovascular risk. Nature Reviews Cardiology **16**(7), 379–381 (2019)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Hao, J., Kwapong, W.R., Shen, T., Fu, H., Xu, Y., Lu, Q., Liu, S., Zhang, J., Liu, Y., Zhao, Y., et al.: Early detection of dementia through retinal imaging and trustworthy ai. npj Digital Medicine **7**(1), 294 (2024)

9. Hatamizadeh, A., Kautz, J.: Mambavision: A hybrid mamba-transformer vision backbone. arXiv preprint arXiv:2407.08083 (2024)

10. 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. pp. 770–778 (2016)

11. 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)

12. Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q.: Octa-500: a retinal dataset for optical coherence tomography angiography study. Medical image analysis **93**, 103092 (2024)

13. Li, S., Zhang, D., Li, X., Ou, C., An, L., Xu, Y., Yang, W., Zhang, Y., Cheng, K.T.: Vessel-promoted oct to octa image translation by heuristic contextual constraints. Medical Image Analysis **98**, 103311 (2024)

14. Li, X., Hu, X., Qi, X., Yu, L., Zhao, W., Heng, P.A., Xing, L.: Rotation-oriented collaborative self-supervised learning for retinal disease diagnosis. IEEE Transactions on Medical Imaging **40**(9), 2284–2294 (2021)

15. Li, X., Jia, M., Islam, M.T., Yu, L., Xing, L.: Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. IEEE Transactions on Medical Imaging **39**(12), 4023–4033 (2020)

16. Liu, S., Hao, J., Liu, Y., Fu, H., Guo, X., Zhang, S., Zhao, Y.: Beyond the eye: A relational model for early dementia detection using retinal octa images. arXiv preprint arXiv:2408.05117 (2024)

17. 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)

18. Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., et al.: Swin transformer v2: Scaling up capacity and resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12009–12019 (2022)

19. Maldonado-Garcia, C., Bonazzola, R., Ferrante, E., Julian, T.H., Sergouniotis, P.I., Ravikumara, N., Frangi, A.F.: Predicting risk of cardiovascular disease using retinal oct imaging. arXiv preprint arXiv:2403.18873 (2024)

20. OpenAI: Gpt-4o system card. arXiv preprint arXiv:2410.21276 (2024)

21. 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)

22. Qiu, J., Wu, J., Wei, H., Shi, P., Zhang, M., Sun, Y., Li, L., Liu, H., Liu, H., Hou, S., et al.: Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. New England Journal of Medicine – Artificial Intelligence **1**(12), AIoa2300221 (2024)

23. Seidelmann, S.B., Claggett, B., Bravo, P.E., Gupta, A., Farhad, H., Klein, B.E., Klein, R., Di Carli, M., Solomon, S.D.: Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. Circulation **134**(18), 1328–1338 (2016)

24. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)

25. Wang, C., Chen, X., Ning, H., Li, S.: Sam-octa: A fine-tuning strategy for applying foundation model octa image segmentation tasks. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1771–1775. IEEE (2024)

26. Wang, H., Chen, Y., Chen, W., Xu, H., Zhao, H., Sheng, B., Fu, H., Yang, G., Zhu, L.: Serp-mamba: Advancing high-resolution retinal vessel segmentation with selective state-space model. arXiv preprint arXiv:2409.04356 (2024)

27. Wang, X., Li, H., Xiao, Z., Fu, H., Zhao, Y., Jin, R., Zhang, S., Kwapong, W.R., Zhang, Z., Miao, H., et al.: Screening of dementia on octa images via multi-projection consistency and complementarity. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 688–698. Springer (2022)

28. Wong, T.Y., Islam, F.A., Klein, R., Klein, B.E., Cotch, M.F., Castro, C., Sharrett, A.R., Shahar, E.: Retinal vascular caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of atherosclerosis (mesa). Investigative ophthalmology & visual science **47**(6), 2341–2350 (2006)

29. Zhao, Y., Hao, J., Kwapong, W., Xu, Y., Shen, T., Fu, H., Lu, Q., Liu, S., Liu, Y., Zhao, Y., et al.: Eye-ad: A graph-based model for early-onset alzheimer's disease and mild cognitive impairment detection based on retinal octa images (2023)

30. 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)

31. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: Forty-first International Conference on Machine Learning