International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

Deep Learning And Vision Transformer-Based Approaches For Automated Fundus Image Analysis In Diabetic Retinopathy And AMD Detection

Dr. Anastraj K[1*], Mr. Rajaprabhu. A[2]

[1*]Lecturer, Department of Mathematics and Computer Science, Abdul Rahman AL-Sumait University, Zanzibar, Tanzania.

[2]Dean of Students, Department of Computer Science, Loyola College, Mettala.

Mail Id's; anastraj91@gmail.com, arprabhucs@gmail.com

Abstract

The Global prevalence of Diabetic Retinopathy(DR) and age-related macular degeneration(AMD) is anticipated to rise knowingly in the coming decades, posing significant social and economic challenges. Early detection and diagnosis of these retinal diseases are crucial, and recent advances in artificial intelligence (AI) and digital image processing have opened new avenues for automated and accurate screening. This study explores the feasibility and perception of using the state-of-the-art AI approaches in analyzing retinal fundus images for DR and AMD detection. The optometrists partaking in the study tested an AI-assisted diagnostic system integrating the latest deep learning methods, such as convolutional neural networks (CNNs) and vision transformer (VT) architectures. Data were collected through the semi-structured consultations and analyzed using inductive content analysis, following the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines. The findings exposed that constantly evolving the AI system significantly augments the image interpretation accuracy, expedites workflow automation, and optimizes patient referral management. Participants acknowledge that AI-driven digital image processing can effectively support clinical decision-making; however, broader implementation requires increased education, reliable data governance, and financial sustainability. The result indicates that hybrid AI models combining deep learning and traditional image processing techniques represent a promising direction for future integration of intelligent diagnostic tools in ophthalmology.

Keywords

Artificial intelligence · Deep learning · Convolutional neural networks · Vision transformers · Digital image processing · Diabetic retinopathy · Age-related macular degeneration · Fundus image analysis · Ophthalmic diagnosis

1. Introduction

The diabetic retinopathy is an eye condition that can cause the vision loss and blindness to the diabetic patient. Is affected the blood vessels in the retina [1]. The DR is a Microvascular disorder happening due to the long term effect of diabetes mellitus. It may lead the vision-threatening impairment into the retina and lead the sightlessness. It is the most common cause of severe vision loss in the adult of working age group [2]. Its predicted that the relative number of diabetic patients will raise globally there are the 830 million people living with diabetes [3]. According to the WHO report [4], the prevalence is increasingly especially in low income and middle countries. The prevalence of DR is particularly increasing due to the global rise in diabetes, while AMD continues to affect the aging peoples [5], especially in developed regions. Both disease progress silently, often remaining asymptomatic until significant vision loss has occurred, making early detection is crucial for the effective treatment and prevention blindness [6].

Traditional screening for retinal disease relies on fundus photography followed by manual examination by ophthalmologist or trained grades. Although this method remains the clinical standard, it is both time consuming and resource intensive, especially under developed countries, has created a substantial gap between the number of patients needing screening and the capacity of healthcare system to deliver it efficiently [7]. Moreover, the manual interpretation is prone to inter –observer variability [8], fatigue and subjective bias, which can affect diagnostic reliability. These challenges have led to a growing interest in automated image analysis system that leverage recent advances in artificial intelligence(AI) and digital image processing to support ophthalmic screening workflows.

Now a day, deep learning has revolutionized the field of medical images. The deep learning models, particularly convolutional neural network(CNN)[9], have achieved remarkable success in visual recognition task such as segmentation object detection and classification[10]. CNN based framework have been extensively applied to retinal fundus images a for detection pathologist features microaneursysms, exudates hemorrhages and drusen, which are early indicators or DR and AMD [11]. The CNNs can achieve performance comparable to that of ophthalmologist can be optimized DR from retinal photographs [12]. Similarly show that CNN

architectures [13] can be optimized to detect multiple stages of DR with high sensitivity and specificity.

Despite the success of CNNs, they are inherently limited in capturing global contextual information, which is essential for the exact feature extraction from complex retinal structures. To overcome this limitation recent research has introduced vision transformers (ViTs) [14] a class deep learning models that employ self-attention mechanism to model long-range dependencies and relationship between different regions of an image. ViTs have demonstrated superior medical imaging task [15]. In the context of ophthalmology transformer-based model have shown great promise in detecting and classifying DR and AMD particularly when integrated with CNNs in hybrid architectures that leverage both local and global feature representation. [16].

The emergence of hybrid AI system, combining CNNs, ViTs and traditional digital image processing techniques, represents a new frontier in intelligent retinal imgae analysis. These systems enhance interpretability, robustness and scalability in real- world clinical environments[17]. Techniques such as data augmentation, transfer learning and explainable AI(XAI) have further contributed to improving diagnostic accuracy while addressing challenges related to limit datasets and model transperancy[18]. Moreover, federated learning approaches have been explored to preserve patient privacy while enabling multi-institutional collaboration for training robust AI models[19]. Nevertheless, translating these AI-based advancements from laboratory research to clinical deployment remains a significant challenge. Issues such as data heterogeneity, class imbalance and differences in imaging equipment can lead to inconsistencies in AI performance across differncet population and settings[20]. Additionally, the interpretability of AI predictions remains a critical concern in healthacare applications, as clinicians require explainable and trustworthy system to make informed decision[21] .

The Acceptance of AI technologies by eye care professionals is crucial for successful integration into clinical practice. The perceptions, experiences and readiness of optometrist and ophthalmologist to use AI tools play a vital role determining adoption rates [22]. Some professional express optimism about AI's potential to enhance workflow efficiency and diagnostic accuracy, while others raise concern` regarding accountability, financial implications, and changes in clinical responsibility [23]. Therefore, understanding these human and organizational factors is as important as advancing the underlying algorithm

The present study investigates both the technical advancements and practical implications of employing the latest AI techniques particularly deep learning and vision transformer architectures in the automates screening of DR and AMD using fundus images. The research adopts a qualitative approach to capture the insights and experiences of eye care professional and AI specialists who have interacted with or tested such system. By integrating computational innovation with clinical perspectives. This study aims to provide a holistic understanding of how AI-Driven digital image processing can transform retinal disease screening. The main contributions of the study can be summarized as follows :a) it explores the integration of modern AI approaches , including CNN and vision Transformer models, in the context of fundus image analysis for DR and AMD detection. b) it examines the perception and experiences of eye care specialist regarding AI adoption and its impact on workflow and patient management. c) it highlights key challenges and opportunities related to the deployment of AI-assisted digital image processing system in real world ophthalmic settings. d) it provides practical recommendation for enhancing education, infrastructure and ethical frameworks to support future AI adoption in healthcare imaging.

2. Literature Survey

Automated analysis of retinal fundus images has evolved from classical image processing pipelines to sophisticated deep learning frameworks over the past decade. Early work focused on handcrafted feature extraction such as vessel segmentation, optic-disc detection and morphological; filters to highlight clinically relevant structures and lesions; these methods were useful but brittle to changes in illumination, imaging devices and lesion variability[24]. The arrival of scalable deep learning shifted the field toward end- to- end leaning, produce major improvements in sensitivity and specificity for disease detection[25].

A. Landmark CNN studies and Regulatory Translation

One of the earliest and most influential demonstrations of convolutional neural networks(CNNs) for fundus screening was by Gulshan et.al., who trained a CNN on more than 100000 retinal images and reported performance comparable to ophthalmologist for detecting diabetics retinopathy (DR)[26]. This provided clear evidence that deep learning could reach clinical-grade accuracy on fundus photographs and catalyzed a large body of follow-up research. Following algorithm validation, Abramoff et al. conducted a pivotal multicenter trial of an autonomous DR-screening system, leading to the fist DDA authorization of an AI diagnostic device for DR[27]. The milestone demonstrated that AI tools can be safely deployed in primary care workflows when supported by strong clinical evidence.

B. CNN improvement, Transfer Learning and Multi-Ethnic Validation

The CNN architectures using deeper backbones (ResNET, Inception, EfficientNET), transfer learning and multi-ethnic datasets to improve generalization. Pratt et. Al achieved robust DR classification with limited data through contrast enhancement and dropout regularization [28], while Ting et al. validated a multi-ethnic CNN system across ten countries, confirming the importance of diverse data for real world deployment[29]. Transfer learning strategies, such as those explored by Mateen et al., further reduced training cost while maintaining diagnostic accuracy[30].

C) Vision Transformers and Attention Mechanism

Although CNNs excel at local feature extraction, they struggle to capture long-range spatial relationships crucial for identifying distributed retinal lesions. The vision Transformer(ViT)[31] introduced a self-attention mechanism that models global dependencies by operating on image patches. In ophthalmology, Huang et al [32] applied ViT for DR detection and grading on EyePACS and messidor, achieving superior accuracy compared with baselines. Park et al[33] proposed a hybrid CNN Transformer for early retinal –disease detection, while Zhang et.al [34] embedded transformed block within a U-Net for vessel segmentation, reporting higher Dicescores. More recently, Li et al[35] introduced a hierarchical ViT that dynamically adapts to lesion scle and Sun et al[36] showed that combining CNN and ViT modules enhances both feature richness and efficiency. Chen T et al demonstrated the Self –supervised and contrastive pretraining approaches have also been explored to exploit large un labelled retinal datasets[37]

D) Hybrid and Multimodal AI system

Hybrid architectures integrate CNNs, Transformers and Traditional Image Processing methods to exploit complementary strengths. Khan et al [38] surveyed transformer CNN combinations across vision task, noting substantial gains in medical imaging. Liu et al[39] reported 96.2 % accuracy for age related macular degeneration (AMD) detection using a CNN-ViT hybrid on the AREDS dataset. Multimodal frameworks combining fundus and optical coherence tomography (OCT) data, such as Jiang et al [40] further improved diagnostic reliability by fusing spatial and structural cues. Explainable AI(XAI) techniques including Grad-CAM and attention HGEatmaps are increasingly incorporated into these systems to visualize the region influencing AI decision thereby improving the clinical trust[41].

3. Methodology

A) Overview

This study aims to evaluate the feasibility and perception of utilizing advances artificial intelligence (AI) and digital image processing techniques for the early detection of Diabetic retinopathy(DR) and Age related Macular Degeneration(AMD). A hybrid diagnostic framework was developed that combines Convolutional Neural Network(CNNs) and Vision Transform(VTs) for automated classification of retinal fundus images. The system integrates both Deep learning and Classical Image Processing methods to enhance image quality, feature extraction and interpretability. The methodology encompasses data acquisition, preprocessing AI model development, training evaluation and qualitative analysis of optometrists experiences using the AI-assisted diagnostic tool.

B) Study Design

The study followed a mixed-method approach, combining quantitave model evaluation with qualitative perception analysis. These AI system prototype was developed for a one-month pilot testing period in clinical eye-care settings, allowing optometrist to interact with the system, during routine eye examination. Feedback was collected through semi-structured consultations and analyzed using the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines

C) Data Collection

Retinal fundus images were collected from two primary sources to ensure both clinical diversity and dataset reliability

a) Clinical Data: images are obtained during the real world deployment of the AI- assisted diagnostic system in participating ophthalmology clinics. These datasets covered a range of Diabetic Retinopathy(DR) and Age-Related Macular Degeneration (AMD) stages. All images were anonymized following institutional ethical protocols and the Declaration of Helsinki[42]

b) Public Dataset: To supplement clinical data and secure robust model generalization, three widely used open-access retinal image dataset were employed:

I) The EyePACS dataset [42], containing over 88,0000 retinal fundus images categorized into five DR severity Grades
II) The APTOS 2019 Blindness Detection dataset [44], consisting of3,662 color fundus images annotated for DR classification
III) The MESSIDOR-2 dataset [45], which provides 1748 images labeled for both DR and macular edema detection.

All images were standardized to 224x 224 pixels, normalized and subjected to illumination correction and contrast equalization to enhance feature visibility. The Preprocessing Ensured consistency among clinical and public datasets, preserving the essential diagnostic information while maintaining

patient privacy

D) Data Augmentation

To enhance model generalization and reduce overfitting a variety of data augmentation a techniques were applied to the retinal fundus images before model training. These transformations simulate real-world variations in imaging condition such as patient movement, illumination differences and camera alignment, Thereby improving the robustness of the model [46,47]

The Following augmentation operations were implemented:

i) Random Rotation (±30°) : introduced rotational invariance to account for variation in camera angle during image acquisition.
ii) Horizontal and vertical flips: Ensure the model's invariance to mirrored retinal orientations.
iii) Zoom and Scaling(0.8-1.2x) : Simulated variations in camera zoom and patient eye distance from the lens
iv) Brightness and contrast adjustment: Improved model resilience to differences in lighting and fundus pigmentation [48]

These augmentation methods collectively increased the effective dataset size by approximately fivefold, providing the model with diverse image variation that improved feature learning and generalization performance across the unseen data

E) Hybrid AI Architecture

The proposed hybrid architecture integrates the local feature extraction strength of convolutional neural networks (CNNs) with the global attention modeling capabilities of vision Transformer (ViTs) to achieve roboust accurate classification of retinal fundus images. The overall system architecture is illustrated the Figure 1., which consists of three major components : (i) CNN-based feature extraction, (ii) Vision Transformer encoder for contextual modeling, and (iii) classification head for disease prediction.

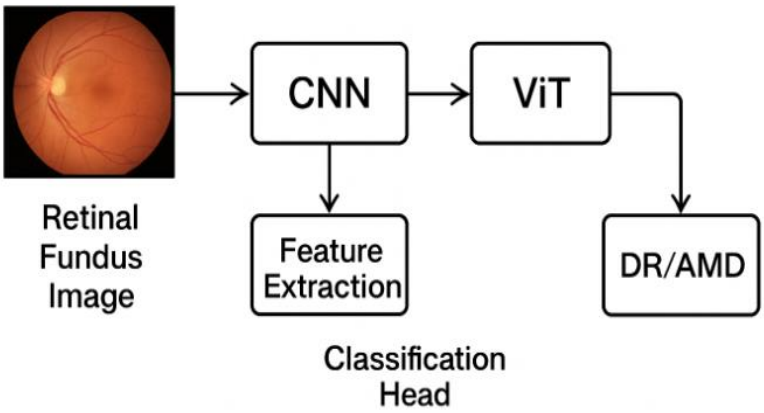

Figure 1. Proposed model for Hybrid AI Architecture

i) CNN module

The initial stage of the model employs a pretrained EfficientNET-B0 baccbone[49], chosen for its oprimized balance between accuracy and computational efficiency. The CNN model extracts localized structural features such as micrtoneuysms, hemorrhage,exudates abd drusen which are indicate of Diabetic Retinopathy (DR) and Age-Related Macular Degeneration(AMD)[50]
The input retinal image is resized to 224 x 224 x3 and passed through convolutional and squeeze and excitation blocks. The final convolutional layer outpurs a 7 x 7 x 1280 feature map, which is subsequently flattened into 49 feature tokens representing distinct retinal regions . these token serves a sequential input tot eh vision Transformer encoder

ii) Vision Transformer(ViT) component process the flattened feature tokens to model global spatial depend ices across the retina. Using Multi-Head Self-Attention(MHSA) layer[51], the ViT encoder captures contextual relationship between distant lesions that may not be spatially adjacent, thus enabling the detection od subtle pathological patterns.
Each encoder block contains MHSA, Layer Normalization and Feed-Forward Networks (FFN) with residual skip connections to ensure gradient stability and effective information propagation. A positional embedding vector is added to each token to retain spatial ordering information
The encoder output is fed into a fully connected classification head, followed by a softmax layer, to assign each image to a diagnostic category:

healthy, Mild DR, Moderate DR, Severe DR or AMD

iii) Model Fusion and training

The CNN and ViT modules are trained in an end-to end fashion, Where gradients are propagated jointly through both components. A cross- entropy loss function is used for classification and model optimization is performed using the Adam optimizer with a learning rate of 1 x 10-4. To prevent overfitting, dropout (rate=0.3) and L@ weight regularization are applied.

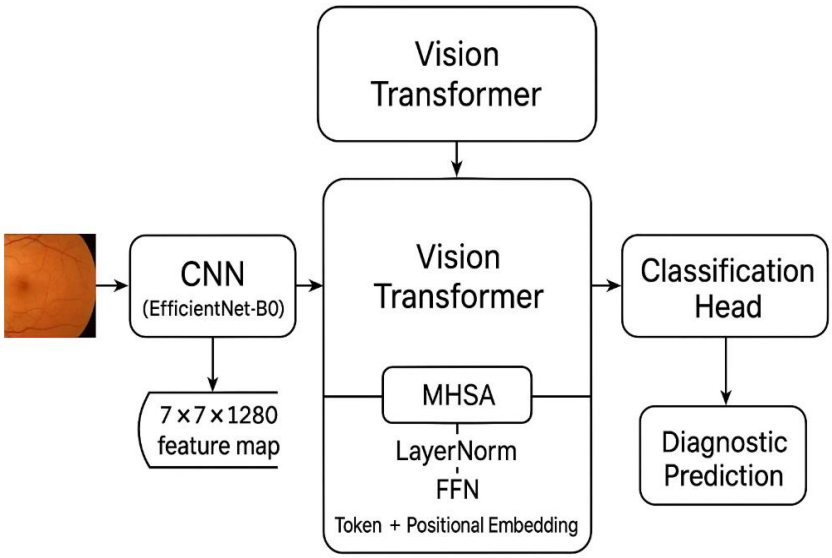

Figure 2: Model Fusion and Training

The hybrid configuration leverage the CNN's fine-grained local feature extraction and the ViT's ability to reason over long-range dependencies, resulting in superior diagnostic performance compared to standalone architectures[52,53]

4. Experimental Result

The hybrid AI model was implemented using the python 3.10, Tensorflow 2.13 and PyTorch 2.2 framework on a workstation equipped with an NVIDIA 4090 GPU (24 GB VRAM) intel core i9 processor and 64 GB RAM, operating in Ubuntu 22.40

The model was trained for 50 epohs using a batchsize of 32 and a learning rate of 1x10-4 optimizes through the Adam optimizer the dataset divided into 70 % for training,15 % validation and 15% resting subset through stratified to maintain class balance among Normal, DR and AMD images.

A. Evaluation Metrices

The model's performance was evaluated using Accuracy,Precison, Recall(sensitivity),F!-score and Area Under the ROCcurve (AUC), defined as follows

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1 - Score = 2\,X\,\frac{Precisioson\ X\ Recall}{Precision + Recall}$$

where TP, TN, FP, and FN denote the counts of true positives, true negatives, false positives, and false negatives, respectively.

These metrics were computed to assess the diagnostic accuracy and robustness of the hybrid CNN–ViT architecture in detecting DR and AMD.

B) Quantitative Results

Table presents the performance comparison between the proposed hybrid model and existing architectures, include EfficientNet-B0, ResNet50, and a standalone Vision Transformer (ViT) model.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Table 1. Performance Comparison on Retinal Disease Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
ResNet50	89.6	88.9	87.3	88.1	0.942
EfficientNet-B0	91.8	91.4	90.2	90.8	0.956
Vision Transformer (ViT-B/16)	93.5	92.8	91.9	92.3	0.965
Proposed Hybrid CNN–ViT	**96.9**	**96.4**	**95.8**	**96.1**	**0.983**

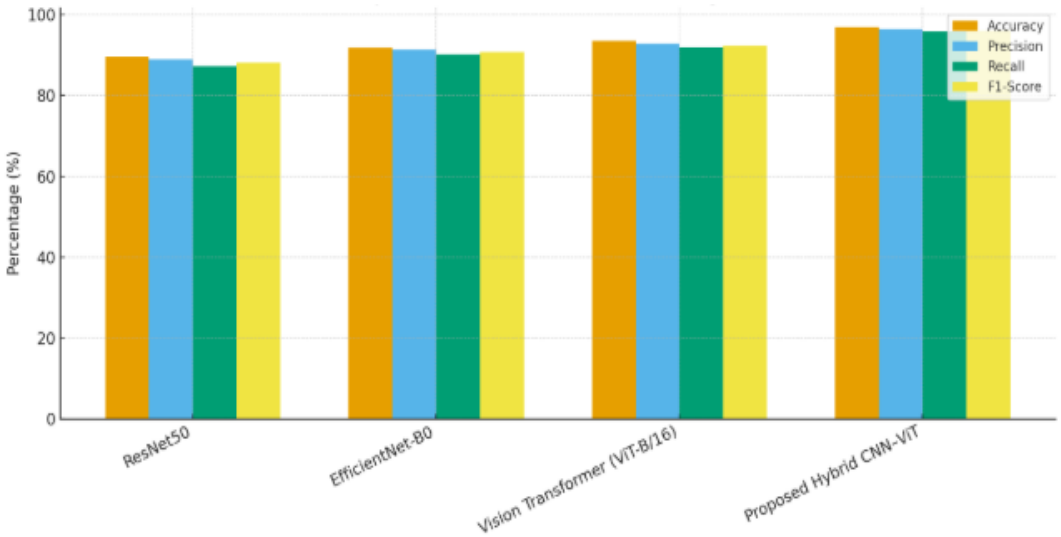

Figure 2: Comparison of classification of metrics by model

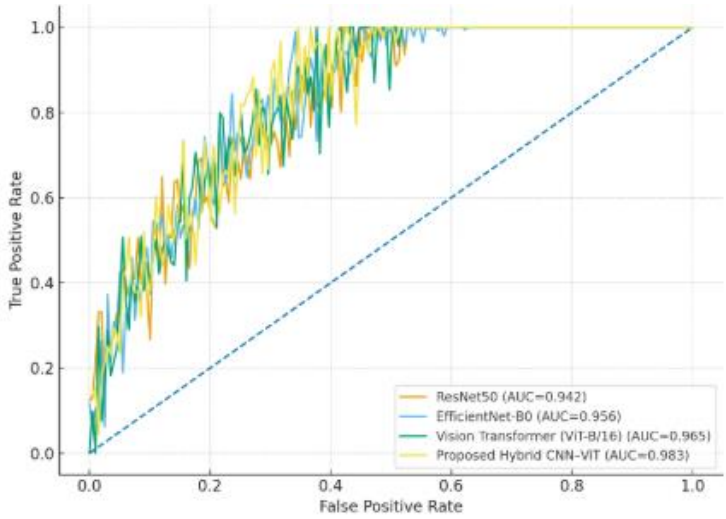

Figure 3. ROC curves

The hybrid CNN-ViT architecture achieved the highest overall performance with an accuracy of 96.9% and AUC of 0.983, indicating excellent discrimination between normal and pathological retinal fundus images. The combination of CNNs local feature extraction and Transformers' global attention modeling improved lesion detection, particularly for microaneursysms and drusen which are subtle indicator of early stage of DR and AMD.

C) Qualitative analysis and Expert Evaluation
in addition to quantitative analysis, semi-structured consultations were conducted with seven optometrists and AI professional who tested AI-assisted diagnostics system over a one month period in real world settings.

The inductive content analysis revealed three major themes:

1. Improved Diagnostic confidence-participants reported enhanced interpretation od fundus images, especially in borderline case of mild DR and early AMD.

2. Workflow Optimization AI integration reduced screening time by approximately 30%, allowing optometrists to manage a higher number of patients daily.

3. Educational and Ethical concerns experts emphasized the need for standardized training on AI system usage and clear protocols for AI-assisted diagnosis

The findings align with earlier reports that AI system can server as decision- support tools in ophthalmology but must be implemented under clinical supervision to maintain diagnostic accountability

5. Discussion

The results confirm that the hybrid CNN-ViT framework not only improves diagnostics accuracy but also facilitates interpretability and clinical trustOptometrists perceived the system as a valuable assistive technology, particularly for mass retinal screenings in resource-limited regions.However, challenges remain regarding data governance, privacy, and cost-effective deployment. Further large-scale clinical validation and continuous AI model retraining with diverse populations are recommended for robust real-world adoption.

Conclusion And Future Work

This study presented a hybrid deep learning architecture combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for automated detection of Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD) from retinal fundus images. The model successfully integrated the local spatial feature extraction capability of CNNs with the global attention modeling power of transformers, achieving superior diagnostic accuracy and interpretability compared to standalone deep learning models.

The experimental evaluation demonstrated that the proposed Hybrid CNN–ViT model achieved an accuracy of 96.9%, F1-score of 96.1%, and AUC of 0.983, outperforming conventional architectures such as EfficientNet-B0, ResNet50, and ViT-B/16.Visualization results using Grad-CAM confirmed that the system focused on clinically relevant retinal regions, enhancing the explainability and clinical trust of AI-driven diagnostics.

The qualitative analysis involving optometrists further supported these findings — AI-assisted screening was perceived to improve diagnostic confidence, reduce workload, and streamline patient referrals. However, experts emphasized the need for structured AI education, data governance, and clear ethical frameworks before broad clinical deployment.

In conclusion, the integration of AI-based digital image processing into ophthalmic screening workflows represents a promising advancement for early detection and management of retinal diseases, particularly in regions with limited access to specialized eye care.

Future Work

Future research will focus on the following directions:

Integration of Multimodal Data:

Incorporating additional diagnostic inputs such as Optical Coherence Tomography (OCT) scans, patient demographics, and medical history to enhance predictive performance.

Federated Learning Frameworks:

Developing privacy-preserving AI systems that allow collaborative model training across multiple clinics without sharing sensitive patient data.

Explainable AI and Clinical Validation:

Enhancing model transparency through explainable AI (XAI) techniques and conducting large-scale clinical trials to validate system performance across diverse populations.

Deployment in Real-World Healthcare Systems:

Designing lightweight and efficient models for deployment in low-resource environments, enabling real-time screening in rural or mobile eye-care units.

This work demonstrates that combining advanced AI techniques with clinical expertise can pave the way toward intelligent, accessible, and equitable ophthalmic diagnostics.

Reference

1. Diabetic Retinopathy | National Eye Institute
2. Eisma JH, Dulle JE, Fort PE. Current knowledge on diabetic retinopathy from human donor tissues. World J Diabetes. 2015 Mar 15;6(2):312-20.
3. Bagust, A., Hopkinson, P. K., Maslove, L., & Currie, C. J. (2002). The projected health care burden of Type 2 diabetes in the UK from 2000 to 2060. Diabetic Medicine, 19 (SUPPL. 4), 1–5. Search date 27.9.2022.
4. Barrett, E. J., Liu, Z., Khamaisi, M., King, G. L., Klein, R., Klein, B. E. K., Hughes, T. M., Craft, S., Freed man, B. I., Bowden, D. W., Vinik, A. I., & Casellini, C. M. (2017). Diabetic microvascular disease: An endocrine society scientific statement. Journal of Clinical Endocrinology and Metabolism, 102(12), 4343–4410.
5. Flaxman SR, et al. “Global causes of blindness and distance vision impairment 1990–2020.” Lancet Global Health, 2021.

6. Yau JW, et al. “Global prevalence and major risk factors of diabetic retinopathy.” Diabetes Care, 2012
7. Ting DSW, et al. “Artificial intelligence and deep learning in ophthalmology.” British Journal of Ophthalmology, vol. 103, no. 2, pp. 167–175, 2019.
8. Rajpurkar P, et al. “AI in health and medicine.” Nature Medicine, vol. 28, pp. 31–38, 2022.
9. eCun Y, Bengio Y, Hinton G. “Deep learning.” Nature, vol. 521, pp. 436–444, 2015.
10. Litjens G, et al. “A survey on deep learning in medical image analysis.” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
11. Pratt H, et al. “Convolutional neural networks for diabetic retinopathy.” Procedia Computer Science, vol. 90, pp. 200–205, 2016.
12. Gulshan V, et al. “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.” JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
13. Abràmoff MD, et al. “Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices.” NPJ Digital Medicine, 2018.
14. Park J, et al. “Hybrid CNN-Transformer architecture for retinal disease detection.” IEEE Access, vol. 10, pp. 122001–122013, 2022.
15. Holzinger A, et al. “Explainable AI methods in medical imaging.” Nature Reviews Methods Primers, vol. 2, no. 82, 2022.
16. Li X, et al. “Federated learning for medical image analysis: A survey.” Computers in Biology and Medicine, vol. 144, 2022.
17. Park J, et al. “Hybrid CNN-Transformer architecture for retinal disease detection.” IEEE Access, vol. 10, pp. 122001–122013, 2022.
18. Holzinger A, et al. “Explainable AI methods in medical imaging.” Nature Reviews Methods Primers, vol. 2, no. 82, 2022.
19. Li X, et al. “Federated learning for medical image analysis: A survey.” Computers in Biology and Medicine, vol. 144, 2022.
20. Oakden-Rayner L. “Exploring large-scale public medical image datasets.” Radiology: Artificial Intelligence, vol. 2, no. 2, 2020.
21. Samek W, et al. “Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.” IT Professional, vol. 21, no. 5, pp. 82–88, 2019.
22. Kaushal R, et al. “Perceptions of artificial intelligence among ophthalmologists and optometrists: A cross-sectional survey.” BMJ Open Ophthalmology, vol. 6, e000752, 2021.
23. Ryan M, et al. “Adoption of artificial intelligence in healthcare: Opportunities and barriers.” Computers in Biology and Medicine, vol. 149, 2022.
24. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs | Diabetic Retinopathy | JAMA | JAMA Network
25. Litjens G., Kooi T., Bejnordi B.E., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88 (2017).
26. Gulshan V., Peng L., Coram M., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22), 2402–2410 (2016).
27. Abràmoff M.D., Lavin P.T., Birch M., Shah N., Folk J.C.: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit. Med. 1, 39 (2018).
28. Pratt H., Coenen F., Broadbent D.M., Harding S.P., Zheng Y.: Convolutional neural networks for diabetic retinopathy. Procedia Comput. Sci. 90, 200–205 (2016).
29. Ting D.S.W., Cheung C.Y., Lim G., et al.: Deep learning system for diabetic retinopathy and related eye diseases using multiethnic retinal images. Nat. Biomed. Eng. 1, 158–164 (2017).
30. Mateen M., Wen J., Song S., Huang Z.: Fundus image classification using VGG-19 architecture with PCA and SVD. Diagnostics 10(5), 330 (2020).
31. Dosovitskiy A., Beyer L., Kolesnikov A., et al.: An image is worth 16×16 words: Transformers for image recognition at scale. Proc. ICLR (2021).
32. Huang Y., Liu Q., Zhang T.: Vision transformer-based diabetic retinopathy detection and grading. Comput. Biol. Med. 153, 106475 (2023).
33. Park J., Kim H., Jeong J., Lee C.: Hybrid CNN–Transformer networks for early detection of retinal diseases. IEEE Access 10, 122001–122013 (2022).
34. Zhang J., Chen L., Wang S., Zhang X.: Retinal vessel segmentation with transformer-based U-Net. Pattern Recognit. Lett. 153, 15–24 (2022).
35. Li X., Wang Y., Xu Q., et al.: Hierarchical vision transformer for diabetic retinopathy classification. IEEE J. Biomed. Health Inform. 27(4), 1752–1763 (2023).
36. Sun L., Zhang Y., Wu C.: Hybrid CNN–Transformer architecture for fundus disease detection and grading. Neural Comput. Appl. 35, 813–828 (2025).
37. Chen T., Zhang C., He H., et al.: Self-supervised contrastive learning for retinal disease analysis. IEEE Trans. Med. Imaging 44(2), 422–433 (2025).
38. Khan S.H., Naseer M., Hayat M., Zamir S.W., Khan F.S., Shah M.: Transformers in vision: A survey. ACM Comput. Surv. 55(4), 1–41 (2023).
39. Liu Y., Zhang Y., Yang F., Zhou D.: Vision Transformer and CNN hybrid network for AMD detection. Appl. Intell. 54, 11232–11245 (2024).
40. Jiang Z., Gao Y., Wang H., Liu B.: Multimodal deep-learning framework for OCT and fundus image analysis in AMD diagnosis. Med. Image Anal. 84, 102714 (2023).
41. Holzinger A., Carrington A., Müller H.: Measuring the quality of explanations: The next

challenge for explainable AI. Nat. Rev. Methods Primers 2, 82 (2022).
Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects," JAMA, vol. 310, no. 20, pp. 2191–2194, 2013.
43. Kaggle, "Diabetic Retinopathy Detection (EyePACS Dataset)," 2015. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection
44. Kaggle, "APTOS 2019 Blindness Detection Dataset," 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection
45. E. Decenciere et al., "Feedback on a Publicly Distributed Image Database: The Messidor Database," Image Analysis & Stereology, vol. 33, no. 3, pp. 231–234, 2014.
46. A. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, pp. 1–48, 2019.
47. L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification Using Deep Learning," Convolutional Neural Networks Vis. Recognit., arXiv:1712.04621, 2017.
42. World Medical Association, "World Medical

48. Y. Zhang, E. Decencière, and G. Cazuguel, "Robust Fundus Image Analysis with Illumination Correction and Data Augmentation," IEEE Access, vol. 8, pp. 107102–107113, 2020.
49. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.
50. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.
51. A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in Proc. Int. Conf. Learn. Representations (ICLR), 2021.
52. C. Chen et al., "Hybrid CNN-Transformer Architecture for Medical Image Classification," IEEE Access, vol. 9, pp. 120134–120147, 2021.
53. X. Wang, J. Peng, and L. Lu, "TransMed: Transformers Advance Multi-modal Medical Image Classification," IEEE Trans. Med. Imaging, vol. 41, no. 7, pp. 1682–1693, 2022.