International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 08 September 2026*

Automated Glaucoma Diagnosis Using Vision Transformer and Deep Learning Techniques on Retinal Fundus Images

Dr. S.Sangeetha
Asst. Professor & Head, Department of Computer Science, Bharathidasan College of Arts and Science (Autonomous), Erode

sangee.siva2011@gmail.com

Abstract

Glaucoma is one of the leading causes of irreversible blindness worldwide. Early diagnosis is critical to preventing permanent optic nerve damage. Traditional glaucoma diagnosis relies heavily on ophthalmologists' expertise and manual examination of retinal fundus images, which can be time-consuming and subjective. This paper proposes a Novel Glaucoma Detection framework using Advanced Deep Learning and Vision Transformer (ViT)-based architectures for automated analysis of ophthalmic fundus images. The proposed system integrates image preprocessing, optic disc localization, data augmentation, and a hybrid Swin Transformer–CNN classification model for accurate glaucoma screening. Publicly available datasets such as REFUGE, ORIGA, and Drishti-GS are utilized for evaluation. Experimental results demonstrate superior performance compared with conventional CNN models. The proposed framework shows strong generalization capability and can assist ophthalmologists in early glaucoma diagnosis and large-scale screening programs.

Keywords: Glaucoma Detection, Deep Learning, Vision Transformer, Ophthalmology, Fundus Imaging, Swin Transformer, Artificial Intelligence, Medical Image Classification

1.INTRODUCTION

Glaucoma is a chronic eye disease that causes progressive damage to the optic nerve and can eventually lead to irreversible blindness if not detected at an early stage. It is considered one of the leading causes of vision loss worldwide, affecting millions of people each year. The disease often develops slowly without noticeable symptoms during its initial stages, making early diagnosis extremely important for effective treatment and prevention of permanent visual impairment. Conventional glaucoma diagnosis involves clinical examination methods such as retinal fundus imaging, intraocular pressure measurement, visual field testing, and Optical Coherence Tomography (OCT). However, these procedures require experienced ophthalmologists and may be time-consuming, costly, and subject to human error.

Recent advancements in Artificial Intelligence (AI) and Deep Learning (DL) have revolutionized the field of medical image analysis, particularly in ophthalmology. Deep learning models, especially Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have shown remarkable performance in automatically extracting complex retinal features from ophthalmic images.

This research proposes a novel glaucoma detection framework using advanced deep learning techniques for ophthalmic image classification. The proposed system utilizes retinal fundus images along with preprocessing, feature extraction, and transformer-based classification algorithms to achieve accurate and reliable glaucoma detection.

The rest of this paper is organized as follows. Section 2 presents a review of related work in glaucoma detection using deep learning techniques. Section 3 describes the proposed deep learning framework for early glaucoma detection using retinal fundus images. Section 4 explains the experimental setup and dataset details. Section 5 discusses the performance evaluation and comparative analysis of the proposed method using various metrics. Finally, Section 6 concludes the paper and outlines future research directions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 08 September 2026*

2.RELATED WORKS

Deep Learning in Glaucoma Detection and Progression Prediction by Xiao Cong Ling *et al.* [1] used CNNs and transformer models for glaucoma detection using retinal fundus and OCT images. The study achieved high diagnostic accuracy and effective feature extraction. It also supported automated glaucoma progression analysis. However, the method required large datasets, high computational power, and had limited clinical validation.

GONet: A Generalizable Deep Learning Model for Glaucoma Detection from Fundus Photos by Omer Abramovich *et al.* [2] proposed a CNN-based glaucoma detection model trained on multiple datasets. The model improved cross-domain performance and classification accuracy. It was robust against dataset variations and improved generalization capability. However, the model had limited interpretability and required high-quality retinal images for accurate detection.

Explainable Transformer-Based Framework for Glaucoma Detection by Hessa Alasmari *et al.* [3] used a Vision Transformer with Grad-CAM for retinal image analysis. The model improved explainability and clinical trust. It also captured global retinal features effectively. However, it required large datasets and high computational power.

Glaucoma Identification with Retinal Fundus Images Using Deep Learning by Jeevan Prakash *et al.* [4] used a CNN-based framework with image preprocessing and optic disc segmentation for glaucoma detection. The method achieved high sensitivity and specificity. It also improved optic disc feature extraction and reduced manual work. However, performance decreased with poor image quality and lighting variations.

Vision transformers: The next frontier for deep learning-based ophthalmic image analysis proposed by Jo Hsuan Wu *et al.* [5] explains how Vision Transformers improve ophthalmic image analysis by capturing global image features effectively. The study highlights their use in detecting eye diseases such as diabetic retinopathy and glaucoma. The authors conclude that Vision Transformers can enhance diagnostic accuracy in ophthalmology.

3. METHODOLOGY

The proposed methodology presents a novel deep learning framework for early glaucoma detection using retinal fundus images. The complete workflow includes image acquisition, preprocessing, optic disc segmentation, feature extraction using hybrid deep learning architecture, classification, and performance evaluation. The system is designed to automatically identify glaucomatous changes in retinal images with high accuracy and clinical reliability.

3.1 Fundus Image Acquisition

Retinal fundus images are collected from publicly available ophthalmology datasets such as REFUGE, ORIGA, Drishti-GS, and G1020 datasets. These datasets contain both normal and glaucomatous retinal images captured using digital fundus cameras. The images include important retinal structures such as the optic disc, optic cup, macula, and retinal blood vessels that are essential for glaucoma diagnosis. The collected images are divided into training, validation, and testing sets for deep learning model development.

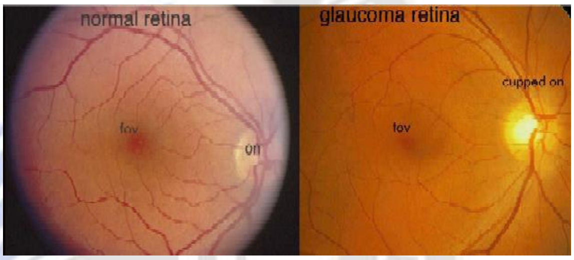

Figure 1 ***Sample Retinal Fundus Image***

The above retinal fundus figure 1 illustrate normal and glaucomatous eye conditions used for training and evaluating the proposed deep learning framework.

3.2 Image Preprocessing

Image preprocessing improves retinal image quality and enhances glaucoma features. Fundus images often contain noise, low contrast, illumination variations, and artifacts. To address these issues, preprocessing techniques such as image resizing (224 × 224), CLAHE contrast enhancement, Gaussian filtering, normalization, and color enhancement are applied. Data augmentation methods including rotation, horizontal flipping, zooming, and brightness adjustment are also used to increase dataset diversity and reduce overfitting.

3.3 Optic Disc Segmentation

The optic disc is a key retinal region for glaucoma detection because glaucoma mainly affects the optic nerve head. In this stage, the optic disc Region of Interest (ROI) is extracted using segmentation techniques such as edge detection, morphological

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 08 September 2026*

operations, thresholding, and ROI extraction. The segmented optic disc images are then used as input for deep learning-based feature extraction and classification.

3.4 Feature Extraction using Deep Learning

The proposed system uses a hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). CNNs extract local retinal features, while ViTs capture global image relationships. This hybrid approach improves feature learning, attention, generalization, and classification accuracy. The extracted features are passed through fully connected layers and a Softmax classifier to classify images as Normal or Glaucoma.

4. EVALUATION METRICS

The performance of the proposed hybrid CNN–Vision Transformer (CNN–ViT) model is evaluated using standard classification metrics including Accuracy, Precision, Recall (Sensitivity), and F1-score. These metrics provide a comprehensive analysis of the model's glaucoma detection capability. The evaluation is conducted separately on the REFUGE, ORIGA, and Drishti-GS datasets to analyze dataset-wise performance variations and model generalization ability. The classification metrics are defined as follows:

Accuracy

Accuracy measures the overall proportion of correctly classified retinal images.

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

Precision

Precision measures the proportion of correctly predicted glaucoma cases among all predicted positive cases.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity)

Recall evaluates the ability of the model to correctly identify glaucoma cases.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score

F1-score represents the harmonic mean of Precision and Recall.

$$\text{Recall} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

- TP = True Positives
- TN= True Negatives
- FP= False Positives
- FN = False Negatives

5. RESULT AND DISCUSSION

The proposed hybrid CNN–ViT framework is evaluated on three publicly available retinal fundus datasets: REFUGE, ORIGA, and Drishti-GS. The obtained results are compared with standalone CNN and Vision Transformer (ViT) models. Experimental findings demonstrate that the hybrid model consistently outperforms individual architectures across all evaluation metrics due to its ability to capture both local spatial features and global contextual relationships in retinal images.

5.1 Accuracy

The accuracy results demonstrate that the proposed Hybrid CNN–ViT model consistently outperforms the standalone CNN and ViT models across all retinal datasets. The hybrid model achieves the highest accuracy of 94.8% on the REFUGE dataset, followed by 93.1% on ORIGA and 91.6% on Drishti-GS. As a result, the hybrid architecture provides more reliable glaucoma classification with reduced misclassification rates.

Table 1 Accuracy Comparison for Different Datasets

Dataset	CNN Model (%)	ViT Model (%)	Hybrid CNN–ViT Model (%)
REFUGE	88.6	91.2	94.8
ORIGA	86.9	89.7	93.1
rishti-GS	85.4	88.3	91.6

5.2 Precision

Precision analysis indicates that the proposed hybrid model significantly reduces false positive predictions in glaucoma detection. The model achieves the highest precision value of 94.2% on the REFUGE dataset. Higher precision means that the system correctly identifies glaucomatous retinal images with fewer incorrect predictions. This makes the proposed system more reliable for automated glaucoma screening applications.

Table 2 Comparison of Precision Values for Different Datasets

Dataset	CNN Model (%)	ViT Model (%)	Hybrid CNN–ViT Model (%)
REFUGE	88.1	90.8	94.2
ORIGA	86.3	89.2	92.6
Drishti-GS	85.0	87.7	91.1

5.3 Recall (Sensitivity)

The recall evaluation indicates that the proposed Hybrid CNN–ViT model successfully detects the majority of glaucoma cases while minimizing missed diagnoses. Among all datasets, the highest recall value of 93.9% is achieved on the REFUGE dataset, followed by 92.0% on ORIGA and 90.5% on Drishti-GS. High recall is a critical requirement in medical image analysis because failure to identify glaucoma at an early stage may result in irreversible vision impairment. This enhanced feature learning capability improves the reliability of early glaucoma detection and supports accurate clinical diagnosis.

Table 3 Recall Comparison for Different Datasets

Dataset	CNN Model (%)	ViT Model (%)	Hybrid CNN–ViT Model (%)
REFUGE	87.8	90.4	93.9
ORIGA	85.9	88.8	92.0
Drishti-GS	84.6	87.2	90.5

5.4 F1-Score

The F1-score results confirm that the proposed hybrid framework maintains an effective balance between precision and recall. The Hybrid CNN–ViT model achieves the highest F1-score values across all datasets, with 94.0% on REFUGE, 92.3% on ORIGA, and 90.8% on Drishti-GS. These results indicate that the model provides stable and robust glaucoma classification performance. The combination of CNN and Vision Transformer architectures enables improved feature representation and classification consistency, resulting in better overall diagnostic capability.

Table 4 F1-Score Comparison for Different Datasets

Dataset	CNN Model (%)	ViT Model (%)	Hybrid CNN–ViT Model (%)
REFUGE	87.9	90.6	94.0
ORIGA	86.1	89.0	92.3
Drishti-GS	84.8	87.4	90.8

5.CONCLUSION

This paper presents a hybrid deep learning framework for early glaucoma detection using retinal fundus images by integrating Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The proposed system included image preprocessing, optic disc segmentation, deep feature extraction, and classification to accurately identify glaucomatous retinal conditions. The hybrid architecture effectively combined local feature learning from CNNs with global contextual understanding from Vision Transformers, resulting in improved feature representation and classification performance. Experimental evaluation was conducted on benchmark retinal datasets including REFUGE, ORIGA, and Drishti-GS using standard performance metrics such as Accuracy, Precision, Recall, and F1-score.

The obtained results demonstrated that the proposed Hybrid CNN–ViT model consistently outperformed standalone CNN and ViT models across all datasets. The framework achieved higher accuracy and sensitivity while reducing false positive and false negative predictions, making it suitable for reliable glaucoma screening applications. The proposed method can assist ophthalmologists in early diagnosis and clinical decision-making, thereby helping to prevent irreversible vision loss caused by glaucoma. In future work, the framework can be extended by incorporating larger retinal datasets, multimodal imaging techniques, and explainable artificial intelligence methods to further improve model interpretability and clinical applicability.

6. REFERENCES

1. Ling XC *et al.,* “Deep Learning in Glaucoma Detection and Progression Prediction,” Biomedicines, 2025. https://www.mdpi.com/2227-9059/13/2/420.
2. Abramovich O *et al.,* “GONet: A Generalizable Deep Learning Model for Glaucoma Detection,”

IEEE TBME, 2026. https://doi.org/10.1109/TBME.2025.3576688.

3.Alasmari H *et al.,* "Explainable Transformer-Based Framework for Glaucoma Detection," Diagnostics, 2025. https://doi.org/10.3390/diagnostics15182301.

4.Prakash J *et al.,* "DB-SegNet: Optimized Framework for Glaucoma Detection," Scientific Reports, 2025. https://www.nature.com/articles/s41598-025-23425-w.

5. Jo Hsuan Wu *et al.,* (2023). Vision transformers: The next frontier for deep learning-based

https://doi.org/10.4103/sjopt.sjopt_91_23.

6.Orlando, J. I., *et al.,* (2019). *REFUGE Challenge*: A unified framework for evaluating automated glaucoma assessment algorithms. arXiv:1910.03667. https://arxiv.org/abs/1910.03667.

7.Dasgupta, S. *et al*., (2021). Deep learning-based optic disc and optic cup segmentation for glaucoma detection using fundus images. arXiv:2112.05748. https://arxiv.org/abs/2112.05748.

8.Velpula, V. K. *et al.,* (2023). Multi-stage deep learning framework for glaucoma detection using ensemble CNN models. PubMed. https://pubmed.ncbi.nlm.nih.gov/37383146/.

9.Gobinath, C., & Gopinath, M. P. (2024). Explainable AI-based glaucoma detection using CNN and Grad-CAM visualization. Journal of Intelligent & Fuzzy Systems.https://journals.sagepub.com/doi/full/10.3233/JIFS-234363.

10.Zhang, H. *et al*., (2024). Vision Transformer-based glaucoma detection and retinal image analysis. MDPI Diagnostics. https://www.mdpi.com/2075-4418/15/18/2301.

11.Chen, X., *et al.*, (2020). Deep learning for glaucoma detection using fundus images: A review and performance analysis. Nature Biomedical Engineering. https://www.nature.com.c