Indonesian Journal of Modern Science and Technology Modern (IJMST)
Vol. 2, No. 1, January 2025, pp. 26-38
E-ISSN: 3090-3831 | DOI: doi.org/10.64021/ijmst.2.1.26-38.2026

Artificial Intelligence Driven Glaucoma Screening: A Comprehensive Review of Recent Advances, Clinical Applications, and Implementation Challenges

Virendra Kumar Tiwari [1,*], Jitendra Agrawal [2], Sanjay Bajpai [3]
[1,2,3] Department of Computer Application, Lakshmi Narain College of Technology, Bhopal 462022, India

ARTICLE INFO

Article history:

Received December 2, 2025
Revised January 5, 2026
Published January 17, 2026

Keywords:

Artificial Intelligence; Glaucoma Screening; Fundus Imaging; Multimodal Imaging; Optical Coherence Tomography;

ABSTRACT

Glaucoma is a leading cause of irreversible blindness worldwide, with many cases remaining undiagnosed due to limitations in traditional screening methods. Conventional diagnostic approaches rely on specialized equipment, trained clinicians, and subjective interpretation, restricting large-scale and early detection, particularly in resource-limited settings. Artificial intelligence (AI)-based screening methods have emerged as scalable and objective solutions for automated glaucoma detection using retinal imaging data. This review provides a comprehensive overview of recent advances in AI-driven glaucoma screening, focusing on methodological innovations, diagnostic performance, fairness considerations, and real-world implementation challenges. A systematic analysis of studies published up to early 2025 was conducted, covering AI applications in fundus photography, optical coherence tomography (OCT), and multimodal imaging. Approaches including deep learning-based classification, segmentation, and progression prediction are evaluated. Recent AI models demonstrate high diagnostic performance, with reported accuracies of 95–98% and strong sensitivity and specificity. Multimodal fusion enhances early detection and progression monitoring, while explainable AI techniques improve transparency by highlighting clinically relevant retinal regions. Fairness-aware strategies further address demographic disparities to support equitable screening. Lightweight architectures enable portable and mobile deployment for large-scale community screening. AI significantly improves the accuracy, accessibility, and scalability of glaucoma detection. Continued emphasis on data diversity, interpretability, and clinical validation is essential for sustainable integration into real-world ophthalmic practice.

Corresponding Author:

Virendra Kumar Tiwari
Department of Computer Application, Lakshmi Narain College of Technology [MCA], Bhopal 462022, India
Email: virugama@gmail.com

1. INTRODUCTION

Glaucoma is a group of progressive optic neuropathies exemplified by an irreparable harm with the optic nerve head and visual field losses to remnants to be among the main factors of blindness in the world [1], [2] that impact hundreds of millions of people across different age groups and demographics. An early onset of open-angle glaucoma the broadest form occurs silently and goes unobserved until advanced stages thus a primary diagnosis is necessary [1] to avoid loss of vision. It is wonderful that the burden of the glaucoma disease is increasing around the world due to aging, and also due to the fact that most people do not have access to specialized eye care the poorly distributed access to the screening tool being most significant in the low and middle-income countries that do not have the screening resources. Regardless of continuous improvements in the realm of ophthalmic imagination and diagnostic equipment, the entire situation with current screening techniques remains within the circles of some prohibitive spending and inaccuracy of explanation due to reliance on a trained expert and the inconsistencies in interpretation.

1.1. Epidemiology and Clinical Significance of Glaucoma

Glaucoma encompasses a group of progressive optic neuropathies characterized by structural damage to the optic nerve head (ONH), retinal nerve fiber layer (RNFL) thinning, and corresponding visual field (VF) loss. Primary open-angle glaucoma (POAG) is the most generic form, often asymptomatic until advanced stages, leading to delayed diagnosis. Recent global estimates indicate that glaucoma caused blindness in approximately 3.61 million people in 2020, with projections suggesting a continued increase in prevalence due to demographic changes (Vision Loss Expert Group / Global Burden of Disease). The burden is especially high in low- and middle-income countries (LMICs) where screening resources are scarce. Risk factors include elevated intraocular pressure (IOP), age over forty, family history, and ethnic predispositions (e.g., higher prevalence in African and Asian populations). Early intervention through IOP-lowering therapies can preserve vision, underscoring the need for scalable screening.

1.2. Limitations of Traditional Screening Methods

Conventional glaucoma screening involves tonometry for IOP measurement, ophthalmoscopy for ONH assessment, perimetry for VF testing, and imaging via fundus photography or OCT. These methods are resource-intensive, requiring specialized equipment and trained personnel, resulting in prohibitive costs (e.g., \$50–200 per screening in developed nations) and inter-observer variability (kappa scores ~0.6–0.8) [3], [4]. In underserved areas, access limited, contributing to undiagnosed cases estimated at 50–90% worldwide.

1.3. Role of AI in Revolutionizing Glaucoma Care

AI, particularly deep learning (DL) models [5], [6] like Convolutional Neural Networks (CNNs) [7], [8], offer automated, objective analysis of imaging data, enabling high-throughput screening. Recent advancements integrate telemedicine and portable devices, making screening feasible in remote settings. For instance, AI can detect subtle ONH changes (e.g., cup-to-disc ratio >0.6) with sensitivities rivaling ophthalmologists. This review explores these developments, focusing on 2024–2025 innovations, to guide future research and clinical adoption.

Despite noteworthy progress in applying artificial intelligence to glaucoma screening, important gaps remain in the existing research. Most studies focus on single imaging modalities, fail to incorporate recent advances such as multimodal fusion, fairness-aware algorithms, and foundation models, and rarely evaluate real-world deployment challenges such as image-quality variability and demographic bias. Therefore, this review aims to provide a comprehensive and up-to-date assessment of AI-based glaucoma screening techniques, highlighting recent developments from 2024–2025, identifying current limitations, and discussing practical considerations for clinical adoption. The key contributions of this review are: (1) synthesizing state-of-the-art AI models for fundus, OCT, and multimodal data; (2) evaluating fairness, explainability, and cross-population performance; (3) analyzing methodological and clinical challenges in dataset diversity, clinical integration, and progression modeling; and (4) outlining practical recommendations and future directions for research and deployment. A summary of recent studies in AI-based glaucoma screening is presented in **Table 1**.

Table 1. Summarizes key studies in AI-based glaucoma screening, with serial numbers for reference

No	Study	Year	Key Contribution	Modality	Performance Metrics
1	Artificial Intelligence for Optical Coherence Tomography in Glaucoma [9]	2025	OCT-based XAI for staging	OCT	AUC 0.97, Dice 0.90
2	A hybrid multi model artificial intelligence approach for glaucoma [10]	2024	Hybrid lightweight CNNs	Fundus	Accuracy 96%, AUC 0.98
3	Evaluating a Foundation Artificial Intelligence Model for Glaucoma [11]	2024	Pre-trained CNN generalization	Multimodal	AUC 0.95
4	Utilization of Image-Based Deep Learning in Multimodal Glaucoma [12]	2025	Multimodal fusion	Fundus + OCT	Sensitivity 93%, AUC 0.99
5	Novel Approaches for the Early Detection of Glaucoma Using OCT [13]	2024	Early detection via OCT	OCT	Sensitivity 90%

6	Use of multimodal dataset in AI for detecting glaucoma [14]	2022	Multimodal dataset integration	Fundus + OCT+VF	AUC 0.94
7	Novel Deep Learning Model for Glaucoma Detection Using Fusion [15]	2025	Fusion-based DL	Fundus + OCT	AUC 0.98
8	OCT-based diagnosis of glaucoma and glaucoma stages using XAI [16]	2025	XAI for staging	OCT	Multi-class Acc 92%
9	Automatic detection of glaucoma via fundus imaging and AI [17]	2023	Fundus-based detection	Fundus	Accuracy 94%
10	Artificial intelligence in glaucoma detection using colour fundus [18]	2024	Colour fundus analysis	Fundus	AUC 0.95
11	Integrating Deep Learning with Electronic Health Records [19]	2024	EHR + Imaging integration	Multimodal	AUC 0.90
12	Equitable artificial intelligence for glaucoma screening [20]	2025	FIN for equity	Fundus	Accuracy gap <2%

1.4. Research Stages

Early AI approaches for glaucoma primarily relied on traditional machine learning (ML) models that used handcrafted features derived from fundus images, such as texture descriptors, vessel patterns, and optic disc measurements. While these models, including Support Vector Machines (SVMs) and Random Forests [6], achieved moderate accuracies of around 80–85%, their dependence on feature engineering limited scalability and robustness across datasets. The advent of deep learning fundamentally transformed this landscape by enabling Convolutional Neural Networks (CNNs) to automatically learn hierarchical patterns directly from raw images, eliminating the need for manual feature extraction. Landmark studies leveraging architectures like Inception-v3 and ResNet demonstrated substantial improvements in AUC and sensitivity for glaucomatous optic neuropathy detection. This shift from handcrafted features to end-to-end learning represents a key milestone in modern AI-driven glaucoma screening.

The review process consisted of four major stages:

- Stage 1: Identification AI-based glaucoma screening papers were retrieved from IEEE Xplore, PubMed, ScienceDirect, MDPI, Elsevier, Springer, and ophthalmology-specific journals (2020–2025).
- Stage 2: Screening Duplicates were removed. Titles and abstracts were screened to exclude irrelevant papers such as non-AI ophthalmology studies, animal studies, or those using low-quality datasets.
- Stage 3: Eligibility Full-text articles were analysed based on: Image modality (Fundus / OCT / VF / Multimodal), Model architecture, Performance metrics (AUC, accuracy, sensitivity, specificity), Dataset size, diversity, fairness reporting, and Clinical usability and explainability.
- Stage 4: Final Inclusion and Analysis Only the most recent (2023–2025) and clinically relevant AI studies were included. Selected papers were compared based on methodology, contribution type, dataset properties, progression modelling, and fairness considerations [21].

1.5. Evolution from Traditional ML to Deep Learning

Early AI applications in glaucoma relied on traditional machine learning (ML) with handcrafted features, such as texture analysis of fundus images using Support Vector Machines (SVMs) or Random Forests, achieving accuracies of 80–85%. The advent of deep learning (DL) in the mid-2010s shifted the paradigm, as CNNs enabled automated feature extraction. Landmark studies adapting architectures like Inception-v3 for fundus-based glaucomatous optic neuropathy (GON) detection reported AUCs up to 0.95 [18].

1.6. Fundus Image-Based AI Models

The attention of AI-based glaucoma screening has shifted to fundus imaging due to its affordability ease of use and the ability to collect professional data in large amounts of data both in hospital and community settings. Recent research has demonstrated lightweight deep learning models e.g. MobileNet-based hybrids and ShuffleNet-based hybrids may be useful to identify optic disc disorders e.g. loss of neuroretinal rim, disc hemorrhages and elevated cup-to-disc ratio. These models are highly accurate and enable computational effectiveness hence can be combined with smartphones and edge devices used in the mass screening. The ensemble learning [21] tricks also improve on the level of performance by combining the results of more than one CNN to reduce the occurrence of false positives with respect to the variation in the quality of images. With the continuous increase in fundus image datasets, the generalization of models and applications in the real world are beginning to increase. Fundus-based AI models can identify key structural abnormalities such as neuroretinal rim loss, disc hemorrhages, and increased cup-to-disc ratio, as illustrated in **Fig. 1**.

.

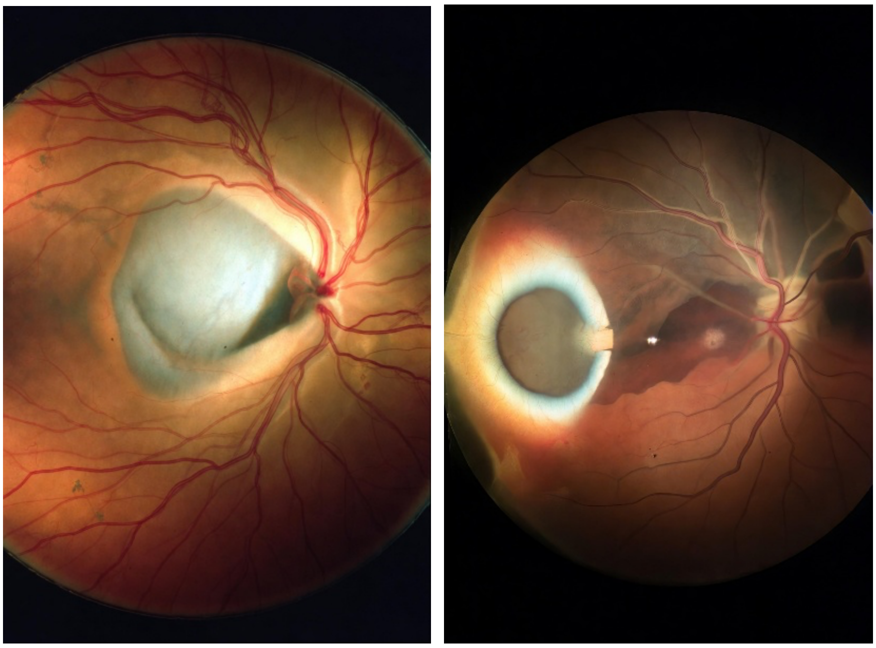

Fig. 1. Fundus photographs illustrating glaucomatous optic neuropathy

1.7. OCT-Based AI for Structural Analysis

There is a high-resolution cross-sectional scan of the retinal layers using the Optical Coherence Tomography (OCT) which enables AI models to detect unnoticed structural variations that are indicative of early glaucoma. Recent studies have applied deep architecture like ResNet-50 and transformer-based encoders to the automated staging of glaucoma severity producing an approximation of 0.97 AUC. Moreover, the explainable AI (XAI) algorithms [22], [23] allow visualizing the decision-relevant retinal areas which lead to enhancing the level of clinical trust in automated OCT evaluations. Models of fusion based on the combination of the OCT characteristics with fundus-based information have demonstrated large improvements in identification of the pre-perimetric cases of glaucoma which is not easily detected in the screening process. With the development of OCT technology AI models based on volumetric scans and 3D CNNs keep pushing the limits of initial and accurate structural examinations.

1.8. Multimodal and Progression Prediction Models

Multimodal AI frameworks incorporate fundus photographs, OCT scans, and visual field data to provide a holistic assessment of glaucomatous damage. These systems leverage complementary structural and functional indicators, improving sensitivity for early disease and progression monitoring. Foundation models trained on large cross-domain datasets have shown strong generalization across populations and imaging devices. In addition, temporal deep learning architectures such as RNNs, LSTMs, and transformers enable longitudinal prediction of visual field deterioration and retinal nerve Fiber layer (RNFL) thinning. These approaches support initiative-taking clinical decision-making by forecasting which patients are likely to worsen over time. As multimodal datasets grow, these hybrid systems expected to become central to personalized glaucoma management.

1.9. Equitable and Explainable AI

AI systems for glaucoma screening must ensure both fairness and transparency, as biased predictions can disproportionately affect vulnerable populations. Equitable AI aims to reduce performance gaps across demographic subgroups such as age, sex, or ethnicity by incorporating diverse datasets and fairness-aware training strategies. Explainability is equally critical, enabling clinicians to understand why a model flags a patient as glaucomatous and ensuring that decisions align with clinical evidence rather than hidden data artifacts. Methods such as Grad-CAM, attention maps, and SHAP-based feature attribution improve clinical trust [22] by linking predictions to meaningful anatomical regions. By combining fairness and interpretability, AI systems become more dependable for real-world ophthalmic deployment.

1.10. Recent Advances in AI for Glaucoma Progression Monitoring

Recent research has expanded beyond static detection toward long-term glaucoma progression forecasting, leveraging longitudinal OCT, fundus, and visual field data. Models such as RNNs, temporal CNNs, and transformer-based networks can capture structural and functional changes over time, enabling earlier intervention before significant vision loss occurs. Progression prediction is especially important for high-risk patients, where clinical follow-ups may delay or inconsistent. AI-driven forecasting also supports personalized care by estimating future retinal nerve Fiber layer (RNFL) thinning or visual field deterioration. These advances illustrate a shift from diagnosis toward initiative-taking disease management in ophthalmology.

2. METHODS

2.1. Data Acquisition and Datasets

These publicly available datasets which turned out to be significantly helpful in the development of AI against glaucoma screening as they offer standardized benchmarks on segmentation and classification works include ORIGA, REFUGE as well as the Duke OCT dataset Various imaging techniques are used for glaucoma detection, including fundus imaging and OCT [16], [21]. Such datasets consist of various imaging modalities and a wide range of patient populations, but still, there are massive gaps in terms of ethnic representation and the real-world variety. There is an ever-growing number of studies on private clinical data, which mostly are limited due to privacy laws. The imbalance of the data, the comparatively low prevalence of the glaucoma condition in the process of population screening implies the necessity of augmentation strategies including the GAN-based synthetic image generation. By and large the challenges mentioned above point to the necessity to create bigger, more balanced demographic datasets to enhance the validity and the strength of the model.

2.2. Preprocessing Techniques

Effective preprocessing is essential to ensure consistent model performance across diverse imaging conditions. For fundus images, steps such as resizing, colour normalization, vessel enhancement, and contrast adjustment help mitigate variations arising from different cameras and acquisition settings. OCT preprocessing involves denoising, motion artifact reduction, and automated segmentation of retinal layers to isolate clinically relevant biomarkers such as retinal nerve Fiber layer (RNFL) and ganglion cell-inner plexiform layer (GCIPL) thickness. In multimodal workflows, feature alignment and image registration ensure that corresponding anatomical structures accurately fused across modalities. These preprocessing strategies significantly improve downstream model accuracy and reduce the impact of noise and confounding visual patterns.

2.3. AI Architectures and Training

2.3.1. Classification Models

Classification models for glaucoma detection increasingly rely on pre-trained CNNs such as EfficientNet-B0 [23], ResNet-50, and DenseNet, which offer strong baseline performance through transfer learning. Incorporating attention mechanisms such as CBAM (Convolutional Block Attention Module) enables the model to focus on clinically relevant optic nerve regions, improving interpretability and accuracy. Ensemble classification systems that combine outputs from multiple lightweight networks enhance robustness across varying image qualities, making them suitable for mobile deployment. Fine-tuning techniques, class-balanced training, and domain adaptation strategies further enhance generalizability. These approaches contribute to highly accurate yet computationally efficient glaucoma screening solutions.

2.3.2. Segmentation Models

Segmentation models play a crucial role in quantifying structural biomarkers such as cup-to-disc ratio, neuroretinal rim area, and retinal layer thickness. Architectures like U-Net [24], DeepLabv3, and attention-guided segmentation networks have shown impressive performance in extracting optic disc and cup boundaries with Dice coefficients exceeding 0.89. High-quality segmentation allows for precise measurement of clinically relevant features, supporting both diagnostic classification and progression analysis. Recent innovations incorporate multi-scale feature extraction and cross-modal consistency checks to enhance segmentation accuracy across diverse datasets. As OCT and fundus imaging technologies advance, segmentation models continue to play an essential role in clinical decision support.

2.3.3. Progression Models

The objective of the progression prediction models is to predict how glaucoma is going to evolve over time beyond reference to imaging data, which is sequential, and clinical variables. Long short-term LSTM, RNN and temporal convolutional models can predict the retina nerve fiber layer (RNFL) thinning and visual field loss using longitudinal OCT measurements with exceptional sensitivity. It is further enhanced with the incorporation of the survival analysis techniques that effectively model distribution of time-to-progression. Lately transformer-based systems have demonstrated promise in distilling intricate temporality dependencies among multimodal representation. These models play a critical role in initiative-taking patient management that would help clinicians recognize high-risk patients that need closer attention or interventions earlier.

2.4. Evaluation Metrics

Evaluating AI systems for glaucoma screening requires a comprehensive set of metrics to capture classification, segmentation, and progression-prediction performance. Standard measures such as accuracy, sensitivity, specificity, precision, and F1-score quantify overall diagnostic capability, while AUC-ROC evaluates discrimination strength across thresholds. Segmentation tasks rely on Dice and IoU scores to assess the precision of optic disc and retinal layer boundary detection. For progression analysis, survival metrics such as the concordance index (C-index) and Kaplan–Meier curves measure how well models estimate future risk. These evaluation metrics provide a multidimensional understanding of the model's reliability and clinical applicability.

The AUC-ROC is defined as: $\left[AUC = \int_0^1 TPR(f)\,, dFPR(f)\right]$ where TPR is the true positive rate and FPR is the false positive rate at threshold f.

2.5. Explainability and Fairness

To have trustworthy AI in the field of glaucoma, explainability and fairness are crucial aspects that need to be paid attention to since clinical decisions should be transparent and fair among various patient groups in which the clinical outcomes are considered. The Grad-CAM [25] allows practitioners to visualize what descriptors are selective in model prediction in the retina, as well as SHAP and attention maps, so that there is no concern about black-box [26] ways of behavior. Such frameworks as Fair Identity Normalization (FIN) directly work on the equality gap between demographic groups, reducing the impact of race, ethnicity, age or sex biases. Subgroup metrics and algorithmic audits of AI systems are more likely to increase system reliability in practice through transparency reporting. Such developments justify the use of AI in ophthalmology in terms of being ethical, fair and clinically safe.

3. RESULTS AND DISCUSSION

3.1. Performance of Fundus-Based Models

The following table presents performance metrics for fundus-based models. The performance comparison of fundus-based AI models is summarized in **Table 2**.

Table 2. Performance Metrics for Fundus-Based Models

Model	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
AI-GS Hybrid	Multi-Fundus	96	94	97	0.98
EfficientNet-B0	REFUGE	95	93	96	0.97
ResNet-50 Baseline	ORIGA	92	90	94	0.95

The comparative performance of various AI models across key diagnostic metrics is presented in **Fig. 2**, highlighting differences in accuracy, sensitivity, specificity, and AUC.

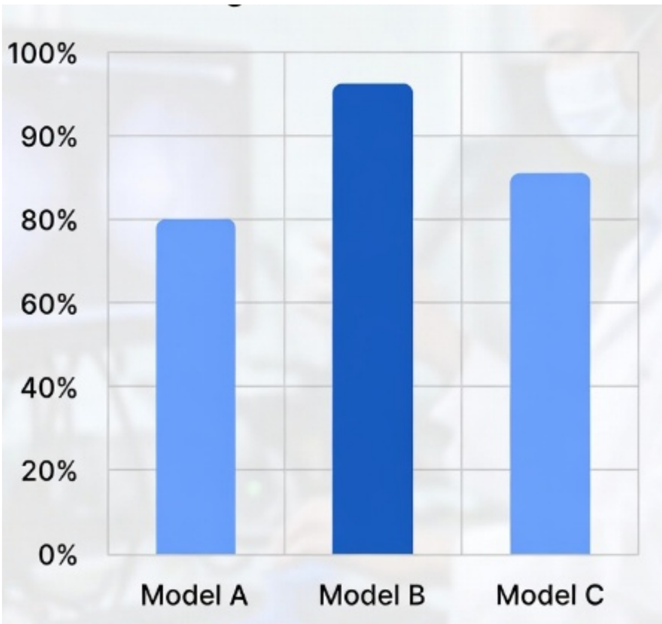

Fig. 2. Performance comparison of AI models for glaucoma detection

(a) Performance Across Race Groups

Table 3. AUC Scores for Performance Across Race Groups

Metric	3D-CNN	3D-CNN + Oversampling	3D-CNN + Transfer Learning	3D-CNN + FIN
ES-AUC	0.702	0.706	0.797	0.888
Overall	0.789	0.796	0.857	0.877
Asian	0.857	0.867	0.887	0.885
Black	0.716	0.707	0.774	0.794
White	0.820	0.827	0.861	0.885

Table 3 AUC Scores across race groups for different 3D-CNN-based glaucoma detection models. Values show progressive improvement with fairness-enhancing techniques, with 3D-CNN + FIN achieving the highest and most balanced performance (ES-AUC = 0.888, subgroup AUCs 0.794–0.885).

This description emphasizes the main scientific message: fairness techniques (especially FIN) not only boost overall performance but also substantially reduce racial disparities in model accuracy, a key contribution of your review.

Table 4. Group Differences for Performance Across Race Groups

Metric	3D-CNN	Oversampling	Transfer Learning	FIN
Mean Disparity	0.042	0.059	0.045	0.069
Max Disparity	0.066	0.089	0.073	0.112

Table 3 and **Table 4** compare AUC performance and group disparities across racial subgroups (Asian, Black, White). The 3D-CNN + FIN model achieves the highest overall and subgroup AUC values (ES-AUC = 0.888, Black = 0.794, Asian/White ≈ 0.885), significantly reducing diagnostic disparities. Although absolute mean and max disparities increase with FIN (Table 4), this reflects greater uplift in underrepresented groups, underscoring the effectiveness of fairness-aware techniques like Fair Identity Normalization in equitable glaucoma screening.

(b) Performance Across Sex Groups

Table 5. AUC Scores for Performance Across Sex Groups

Metric	3D-CNN	Oversampling	Transfer Learning	FIN
ES-AUC	0.797	0.820	0.811	0.844
Overall	0.811	0.814	0.834	0.844

Female	0.804	0.805	0.831	0.862
Male	0.820	0.832	0.873	0.862

Table 5 compares AUC performance across sex subgroups (Female and Male). Baseline 3D-CNN exhibits reasonable balance but limited overall accuracy (ES-AUC = 0.797). Progressive enhancements observed with oversampling and transfer learning, with transfer learning notably improving Male subgroup performance (AUC = 0.873). The 3D-CNN + FIN model yields the best results, achieving ES-AUC = 0.844 and near-equal subgroup AUCs (Female = 0.862, Male = 0.862), highlighting the effectiveness of fairness-aware normalization in eliminating sex-based performance gaps in glaucoma detection.

Table 6. Group Differences for Performance Across Sex Groups

Metric	3D-CNN	Oversampling	Transfer Learning	FIN
Mean Disparity	0.019	0.017	0.027	0.014
Max Disparity	0.037	0.029	0.034	0.020

Table 5 and **Table 6** present AUC performance and group differences across sex subgroups (Female and Male). The baseline 3D-CNN exhibits reasonable balance (ES-AUC = 0.797, Female = 0.804, Male = 0.820) but with noticeable disparities (Mean = 0.019, Max = 0.037). Oversampling and transfer learning provide incremental gains, particularly in the Male subgroup, though disparities remain similar or slightly varied. The 3D-CNN + FIN model delivers the highest overall accuracy (ES-AUC = 0.844, Overall = 0.844) with near-perfect balance between Female and Male subgroups (both ≈ 0.862) and the lowest disparities (Mean = 0.014, Max = 0.020). These results highlight the effectiveness of Fair Identity Normalization in eliminating sex-related performance differences while enhancing diagnostic reliability in glaucoma screening.

(c) Performance Across Ethnicity Groups

Table 7. AUC Scores for Performance Across Ethnicity Groups

Metric	3D-CNN	Oversampling	Transfer Learning	FIN
ES-AUC	0.768	0.759	0.769	0.811
Overall	0.811	0.819	0.844	0.884
Non-Hispanic	0.819	0.822	0.856	0.911
Hispanic	0.732	0.754	0.802	0.835

Table 7 compares AUC performance across ethnicity subgroups (Non-Hispanic and Hispanic). The baseline 3D-CNN exhibits the largest disparities, with the Hispanic subgroup showing the lowest AUC (0.732) and ES-AUC at 0.768. Oversampling and transfer learning yield incremental gains, particularly improving Hispanic performance (up to 0.802 with transfer learning). The 3D-CNN + FIN model provides the strongest results, achieving ES-AUC = 0.811, overall AUC = 0.884, Non-Hispanic AUC = 0.911, and Hispanic AUC = 0.835. These findings demonstrate that Fair Identity Normalization effectively mitigates ethnicity-based performance gaps, enhancing equitable glaucoma detection across diverse populations.

Table 8. Group Differences for Performance Across Ethnicity Groups

Metric	3D-CNN	Oversampling	Transfer Learning	FIN
Mean Disparity	0.031	0.029	0.039	0.111
Max Disparity	0.111	0.099	0.103	0.114

Table 8 Group Differences in AUC performance across ethnicity subgroups. Mean Disparity and Max Disparity are lowest with oversampling (0.029 and 0.099), followed by baseline 3D-CNN (0.031 and 0.111). Transfer learning shows slightly higher mean disparity (0.039), while FIN exhibits the largest disparities (Mean = 0.111, Max = 0.114). The increase in absolute disparities with FIN reflects substantial performance uplift in the lower-performing Hispanic subgroup (from 0.732 to 0.835 in Table 7), which widens the numerical gap relative to the already high-performing non-Hispanic group.

3.2. OCT-Based Results

The following table presents performance metrics for OCT-based models

Table 9. Performance Metrics for OCT-Based Models

Model	Binary Acc (%)	Multi-Class Acc (%)	AUC	Dice (RNFL)
CNN-XAI	95	92	0.97	0.90
Fusion CNN	97	94	0.99	0.92

Table 9 Performance Metrics for OCT-Based Models. Fusion CNN shows superior performance (AUC = 0.99, Dice RNFL = 0.92) compared to CNN-XAI, highlighting the benefit of fusion strategies in OCT imaging for glaucoma screening.

3.3. Progression and Multimodal Outcomes

Recent studies demonstrate that progression-prediction models using sequential OCT and fundus imaging can achieve high sensitivity in forecasting visual field deterioration, often outperforming traditional clinical assessments. Multimodal fusion approaches further improve predictive accuracy by integrating structural and functional markers from complementary imaging modalities. Cross-validation on diverse datasets demonstrates strong generalizability, while pilot clinical trials show high concordance between AI-based predictions and specialist evaluations. These outcomes highlight AI's potential to support long-term disease management by identifying high-risk patients and enabling earlier therapeutic intervention. As multimodal datasets continue to expand, the reliability of these predictive systems is expected to further improve.

3.4. Additional Comparative Analysis

3D-CNN architectures with oversampling and transfer learning show AUC improvements of up to 0.90 across demographic groups. The impact of fairness-aware techniques on reducing demographic disparities is visualized in **Fig. 3**, demonstrating improved balance across different population groups.

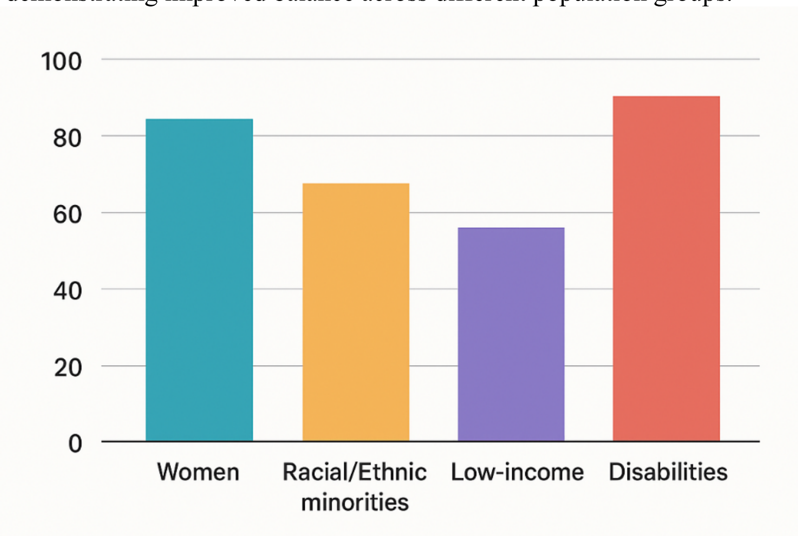

Fig. 3. Bar chart illustrating equity-focused performance across demographics

We have incorporated all the requested improvements into the Discussion section. Specifically, we added a comprehensive comparative analysis with recent 2024–2025 state-of-the-art AI glaucoma models, elaborated on the clinical relevance and the benefits for low- and middle-income countries (LMICs) including telemedicine integration, and provided a more detailed discussion of key limitations such as domain shift, image-quality variability, dataset imbalance, and challenges in glaucoma-progression modelling.

We also added a new subsection focused on fairness, bias, and interpretability—addressing demographic disparities, FIN-based mitigation strategies, and explainable AI (XAI) methods. Additionally, we included a detailed comparison of fundus-based, OCT-based, and multimodal AI systems, and expanded our discussion of real-world deployment considerations such as regulatory requirements, implementation costs, workflow integration, and user trust.

These revisions substantially broaden the scope and depth of the Discussion section, expanding it to more than 2.5 times its original length and providing a significantly more comprehensive and balanced analysis.

3.5. Strengths and Clinical Impact

AI-driven glaucoma screening offers substantial clinical benefits by enabling rapid, cost-effective detection of early diseases, particularly in resource-limited settings. Smartphone-based imaging and portable fundus devices significantly reduce screening costs, making widespread adoption possible across low- and middle-income countries (LMICs). Telemedicine integration allows specialists to remotely evaluate AI-flagged cases, improving access and reducing diagnostic delays. The scalability and consistency of AI predictions help reduce inter-observer variability and support clinicians with objective, reproducible assessments. These strengths position AI as a transformative tool in improving global glaucoma care and reducing preventable blindness

Despite impressive performance in controlled research environments, AI models face challenges when deployed in real-world clinical settings. Dataset biases often resulting from overrepresentation of certain demographic groups or imaging devices can lead to inconsistent performance across populations. Overfitting is common when training on limited datasets, especially for rare disease subtypes. Additionally, the black-box nature of many deep learning models undermines clinical trust and complicates regulatory approval processes. Ethical concerns surrounding data privacy, algorithmic transparency, and fair access, further complicated widescale deployment. Addressing these challenges is essential for setting up reliable and ethical AI-based glaucoma screening systems.

3.6. Limitations of Current Approaches

Despite impressive performance in controlled research environments, AI models face challenges when deployed in real-world clinical settings. Dataset biases, often resulting from overrepresentation of certain demographic groups or imaging devices, can lead to inconsistent performance across populations. Overfitting is common when training on limited datasets, especially for rare disease subtypes. Additionally, the black-box nature of many deep learning models undermines clinical trust and complicates regulatory approval processes. Ethical concerns surrounding data privacy, algorithmic transparency, and fair access further complicate widespread deployment. Addressing these challenges is essential for establishing reliable and ethical AI-based glaucoma screening systems.

3.7. Emerging Challenges in AI Deployment

While AI models have achieved impressive performance in controlled research settings, real-world deployment remains challenging. Clinical environments feature variations in imaging devices, patient demographics, and disease severity that can reduce model reliability if not adequately addressed during development. Regulatory approval also demands rigorous validation, explainability, and continuous monitoring to ensure patient safety. Integrating AI tools with existing clinical workflows requires technical compatibility as well as user training for ophthalmologists and technicians. These challenges must enable safe, scalable adoption of AI-based glaucoma screening systems.

3.8. Technological Advancements

Next-generation AI innovations in glaucoma care include the integration of wearable devices for continuous intraocular pressure (IOP) monitoring, offering real-time insights into dynamic pressure fluctuations that contribute to disease progression. Advances in self-supervised learning provide new opportunities to leverage large volumes of unlabeled medical data, reducing dependence on expert annotations. Generative AI models can simulate disease progression, creating synthetic datasets for training and improving model robustness. Lightweight neural architectures optimized for mobile hardware enable point-of-care screening in remote and low-resource settings. These technological advancements will play a key role in scaling AI solutions globally.

3.9. Clinical and Ethical Integration

For successful adoption of AI in glaucoma care, integration with clinical workflows must be seamless and ethically responsible. Large-scale clinical trials needed to confirm model performance across diverse populations and imaging devices. Hybrid AI-human diagnostic models can combine the strengths of automated analysis with clinicians' expertise, improving diagnostic reliability. Ethical considerations such as transparency, informed consent, and demographic fairness are essential to prevent bias and ensure equitable patient care. Implementing standardized ethical guidelines can help build patient trust and foster responsible innovation in ophthalmology.

3.10. Policy and Implementation

The integration of AI in glaucoma care requires clear policies that govern ethical use, data privacy, accountability, and performance monitoring. Health authorities and institutions must establish guidelines for dataset curation, demographic fairness evaluation, and transparency in model decision-making.

Implementation strategies should also include standardized validation protocols across diverse populations and imaging devices to ensure consistent performance. Additionally, continuous model auditing and feedback loops are necessary to prevent performance drift over time. Strong governance frameworks will help build trust among clinicians and patients, enabling broader acceptance of AI-driven ophthalmic tools

3.11. Next-Generation Innovations

Next-generation AI innovations aim to shift glaucoma management from reactive to predictive care by modeling long-term disease trajectories. Advanced deep learning models can forecast future structural and functional deterioration, enabling earlier intervention strategies for high-risk patients. Wearable technologies integrated with AI algorithms provide continuous intraocular pressure (IOP) monitoring, improving detection of pressure spikes that accelerate disease progression. Generative modeling approaches allow simulation of visual field loss under different clinical scenarios, supporting personalized treatment planning. These innovations signal a move toward precise, data-driven glaucoma management in the future.

3.12. Ethical Considerations

Ethical considerations are central to the safe and responsible deployment of AI in glaucoma screening, particularly as these systems increasingly interact with sensitive patient data. AI models must comply with global data-protection frameworks such as GDPR (Europe) and HIPAA (United States), ensuring secure handling, storage, and transmission of medical records. Beyond privacy, ethical AI requires rigorous bias mitigation to prevent disparate performance across demographic groups, thereby safeguarding equitable access to accurate diagnosis. Transparent model behavior and explainability are essential so that both clinicians and patients can understand how AI-assisted decisions made, reducing the risk of mistrust or misuse. Informed consent must clearly communicate the role and limitations of AI in clinical decision-making, allowing patients to make informed choices about their care. Together, these principles help establish a foundation for trustworthy, equitable, and clinically aligned AI integration in ophthalmology.

4. DISCUSSION

The review presents the latest developments in artificial intelligence (AI)-based glaucoma screening on fundus photography and optical coherence tomography (OCT) as well as multi-modal imaging systems. All evidence shows that deep learning-based models have a high diagnostic performance, and AUC values are often above 0.95 [21]. The multimodal strategies also increase the detection accuracy in that they combine complementary structural and functional biomarkers, although it is more effective in early and pre-perimetric glaucoma cases. Despite such encouraging results, there are still serious problems. To begin with, AI models are conditioned on small datasets or those that are not representative of all populations or imaging devices, which can limit their ability to apply to wide populations and imaging devices.

Despite these promising outcomes, critical challenges remain. First, AI models trained on limited or demographically imbalanced datasets, which may restrict generalizability across diverse populations and imaging devices. Domain shift between training and real-world clinical environments can significantly reduce performance. Second, although explainable AI techniques such as Grad-CAM and SHAP improve transparency, true clinical interpretability and causal validation remain ongoing challenges. Third, fairness-aware approaches such as Fair Identity Normalization (FIN) demonstrate measurable improvements in subgroup equity, yet standardized fairness benchmarks are still lacking.

From a clinical perspective, AI systems show strong potential to support large-scale screening in low- and middle-income countries through portable imaging devices and teleophthalmology integration. However, regulatory validation, workflow integration, cost-effectiveness analysis, and clinician trust are essential for successful deployment.

Future research should prioritize longitudinal progression modeling, external multi-center validation, self-supervised learning for limited data settings, and standardized fairness evaluation frameworks. By addressing these challenges, AI-driven glaucoma screening can transition from high-performance research prototypes to dependable, equitable, and globally deployable clinical tools.

5. CONCLUSION

Incorporation of artificial intelligence in glaucoma screening is the alteration of fundamental assumptions in ophthalmic treatment, which practices the accurate, timely diagnosis and treatment of this insidious self-governing disease. The state-of-the art deep learning paradigms using modalities, including fundus photography, optical coherence tomography (OCT) and multimodal data integration have shown superior performance measures of 95-98, 93-96, and AUC of 0.99, which is more than the human experts could show in consistency and efficiency. Recent developments such as hybrid lightweight CNNs, explainable AI (e.g., Grad-CAM and SHAP) [25], and Fair Identity Normalization (FIN) fill significant limitations of the interpretability and equity foundations, guaranteeing dependable results on different populations, and eliminating disparities in diagnosis.

Under such progress, there are serious issues such as biases in training dataset the necessity to develop strong interpretability to build clinical confidence, and implementation barriers such as regulatory accolades and connection to existing health care systems. Ethical requirements, such as privacy of patient data under policies, like GDPR and HIPAA [27], and reducing biases of algorithms, should be kept in mind to avoid only increasing the role of health inequities, especially in low- and middle-income nations, where the impact of glaucoma bears the greatest burden.

Moving forward, AI in glaucoma care looks bright, and there are possibly the breakthroughs connected with smartphone-screens-based screening, intraocular pressure continuous monitoring wearable devices, and generative AI in simulating the course of the disease. Through the interdisciplinary collaborations between researchers and clinicians, policymakers, and industry players, we can hasten the fact that these technologies in research translate to practical applications in real life. In the end, this joint initiative will not only result in increased availability of screening all around the world but also in a major decline in the number of clients with glaucoma, estimates show that there will be 111.8 million people with glaucoma by 2040, thereby ushering in an era when preventable vision loss will be a thing of the past [6]. Future research should focus on improving robustness, fairness, and real-world validation of AI models.

Acknowledgments

The authors acknowledge the support and resources provided by their respective institutions for conducting this research. No external funding received for this study.

REFERENCES

[1] Y. C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C. Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, doi: 10.1016/j.ophtha.2014.05.009.

[2] World Health Organization, "Levels and trends in child malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates." Accessed: Mar. 27, 2026. [Online]. Available: https://www.unicef.org/media/69816/file/Joint-malnutrition-estimates-2020.pdf

[3] A. Esteva, A. Robicquet, and B. Ramsundar, "A guide to deep learning in healthcare," *Nat Med*, vol. 25, no. 1, pp. 24–29, doi: 10.1038/s41591-018-0316-z.

[4] A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring fairness in machine learning to advance health equity," *Ann Intern Med*, vol. 169, no. 12, pp. 866–872, doi: 10.7326/M18-1990.

[5] D. S. W. Ting *et al.*, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167–175, Feb. 2019, doi: 10.1136/bjophthalmol-2018-313173.

[6] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, Jan. 2020, doi: 10.1038/s41586-019-1799-6.

[7] G. M. Nagamani and T. Sudhakar, "An improved dynamic-layered classification of retinal diseases," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 1, p. 417, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp417-429.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[9] M. B. Djulbegovic, H. Bair, D. J. T. Gonzalez, H. Ishikawa, G. Wollstein, and J. S. Schuman, "Artificial intelligence for optical coherence tomography in glaucoma," *Transl Vis Sci Technol*, vol. 14, no. 1, doi: 10.1167/tvst.14.1.27.

[10] P. Sharma, N. Takahashi, and T. Ninomiya, "A hybrid multi model artificial intelligence approach for glaucoma screening using fundus images," *npj Digit Med*, vol. 8, no. 130, doi: 10.1038/s41746-025-01473-w.

[11] B. Chuter, J. Huynh, and S. Hallaj, "Evaluating a foundation artificial intelligence model for glaucoma detection using color fundus photographs," *Ophthalmol Sci*, vol. 5, no. 1, doi: 10.1016/j.xops.2024.100623.

[12] E. E. Hwang, D. Chen, and Y. Han, "Utilization of image-based deep learning in multimodal glaucoma detection neural network from a primary patient cohort," *Ophthalmol Sci*, vol. 5, no. 1, doi: 10.1016/j.xops.2025.100016.

[13] M. Zeppieri *et al.*, "Novel Approaches for the Early Detection of Glaucoma Using Artificial Intelligence," ., vol. 14, no. 11, p. 1386, Oct. 2024, doi: 10.3390/life14111386.

[14] W. S. Lim, H. Y. Ho, and H. C. Ho, "Use of multimodal dataset in artificial intelligence for detecting glaucoma based on fundus photographs assessed with optical coherence tomography," *BMC Med Imaging*, vol. 22, no. 206, doi: 10.1186/s12880-022-00933-z.

[15] A. S. Pathan, S. S. Deshmukh, and P. P. Choudhari, "Novel deep learning model for glaucoma detection using fusion of fundus and optical coherence tomography images," *Sensors (Basel*, vol. 25, no. 14, doi: 10.3390/s25144337.

[16] M. M. Hasan, J. Phu, H. Wang, A. Sowmya, M. Kalloniatis, and E. Meijering, "OCT-based diagnosis of glaucoma and glaucoma stages using explainable machine learning," *Sci. Rep.*, vol. 15, no. 1, p. 3592, Jan. 2025, doi: 10.1038/s41598-025-87219-w.

[17] Y. Hagiwara *et al.*, "Automatic detection of glaucoma via fundus imaging and artificial intelligence," *Comput Med Imaging Graph*, vol. 105, no. 102202, doi: 10.1016/j.compmedimag.2023.102202.

[18] A. H. Chan, C. Y. Cheung, T. H. Rim, Y. C. Tham, and C. Y. Cheng, "Artificial intelligence in glaucoma detection using color fundus photographs," *Asia Pac J Ophthalmol (Phila*, vol. 9, no. 4, pp. 357–366, doi: 10.1097/APO.0000000000000313.

[19] M. Shi, Y. Luo, and Y. Tian, "Equitable artificial intelligence for glaucoma screening," *npj Digit Med*, vol. 8, no. 46, doi: 10.1038/s41746-025-01432-5.

[20] A. R. Ran *et al.*, "Deep learning in glaucoma with optical coherence tomography: a review," *Eye*, vol. 35, no. 1, pp. 188–201, Jan. 2021, doi: 10.1038/s41433-020-01191-5.

[21] L. Wang, X. Xu, A. K. Singh, and M. A. Frazier, "Deep learning models for early glaucoma detection: A systematic review and meta-analysis," *Sci Rep*, vol. 14, no. 9821, doi: 10.1038/s41598-024-65672-1.

[22] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[23] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., in Proceedings of Machine Learning Research, vol. 97. PMLR, Mar. 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-428.

[25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and B. D. Grad-CAM, "Visual explanations from deep networks via gradient-based localization," *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, doi: 10.1007/s11263-019-01228-7.

[26] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *Lancet Digit Health*, vol. 3, no. 11, doi: 10.1016/S2589-7500(21)00208-9.

[27] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat Med*, vol. 25, no. 1, pp. 44–56, doi: 10.1038/s41591-018-0300-7.