Artificial Intelligence Review (2026) 59:155
https://doi.org/10.1007/s10462-026-11551-4

Explainable AI for glaucoma detection and classification: a comprehensive review

Ahmed M. Abd El-Gawad[1] · Sarah Hassan[2] · Mohamed Elsharkawy[2] · Shahad Al Hamadani[2] · Aliyah Shivel[3] · Tracy Couch[4] · Ibrahim Saleh[5] · Eman A. Atallah[6] · Mohammed Ghazal[7] · Guruprasad Giridharan[2] · Hanan M. Amer[8] · Abeer Twakol Khalil[8] · Ayman El-Baz[2]

Received: 9 October 2025 / Accepted: 26 March 2026 / Published online: 8 April 2026

Abstract
Glaucoma is a leading cause of irreversible blindness and often advances without symptoms. This PRISMA/PICOS–guided systematic review synthesizes 100 studies (to January 2025) on artificial intelligence (AI) for glaucoma detection, staging, and monitoring using color fundus photography (CFP), optical coherence tomography (OCT), and OCT angiography (OCTA). Roughly 45% of studies use CFP, 55–60% use OCT, and 10–15% use OCTA; about one quarter fuse multiple imaging types with clinical data. Across methods—convolutional neural networks, Transformers, and hybrid fusion—internal test results are strong, with several systems reporting specialist-level sensitivity and specificity. Translation to practice remains limited by small, single-center cohorts, device/domain shifts that reduce generalization, a lack of prospective trials, and explainability approaches that are misaligned with clinical reasoning. We recommend multi-center, multi-ethnic, multi-vendor benchmarks, rigorous external and prospective validation, clinically meaningful and standardized explainability, and cost- and workflow-aware deployment. Addressing these gaps could enable earlier detection, more personalized monitoring, and real-world impact in reducing preventable blindness. Within this review, we investigate key challenges revolving around the use of AI models for glaucoma imaging across different learning paradigms, improving model efficiency, and coupling them with complementary techniques and multimodal data (CFP, OCT, and OCTA). We hope this review can give a comprehensive picture of AI-based methods for glaucoma to readers in the field of medical image analysis.

Keywords Glaucoma · Explainable AI (XAI) · Color fundus photography · Deep learning · Machine learning · Optical coherence tomography

Ahmed M. Abd El-Gawad and Sarah Hassan have contributed equally to this work.

Extended author information available on the last page of the article

1 Introduction

Glaucoma is a vision-threatening ocular morbidity that leads to progressive optic nerve damage with resultant changes in head and visual field secondary to the death of the retinal ganglion cells (Kang and Tanna 2021). Glaucoma is considered the leading cause of irreversible visual impairment. It currently affects 3.5% of people aged 40 to 80 years, with an increasing incidence as life expectancy increases. Approximately 112 million individuals are expected to have glaucoma by 2040 (Tham et al. 2014). Rapid diagnosis is crucial, as significant damage to ganglion cells is needed for the visual field defect to develop (Weinreb et al. 2014). Primary glaucoma has many risk factors, including old age, a family history of glaucoma, race, and ocular hypertension (Hollands et al. 2013). Advancing age may lead to aging-induced damage to retinal ganglion cells and the nerve fiber layer. Oxidative stress that occurs due to the accumulation of reactive oxygen species and defective antioxidants can cause further damage to the optic nerve and retinal ganglion cells (Freddo and Gong 2009; Sugar 1942). Many theories exist about ganglion cell damage in glaucoma. Mechanical theory hypothesizes that the damage occurs secondary to the direct mechanical effect of high intraocular pressure (IOP). The indirect ischemic theory postulates that high IOP leads to changes in retinal circulation that cause ischemia and damage to retinal ganglion cells. It was also proposed that glaucoma is caused by a neurodegenerative process related to a deficiency of neurotrophic factors, increased oxidative stress and mitochondrial dysfunction (Weinreb et al. 2014; But et al. 2016; Marquis and Whitson 2005; Wagner et al. 2022). The research paper is broken down into six main sections as shown in Fig. 1.

Glaucoma can be classified into primary and secondary according to etiology, and further classified into open and closed angle according to the status of the anterior chamber angle (Lee and Higginbotham 2005). Primary open-angle glaucoma represents the most prevalent

Fig. 1 Structural framework of the paper, outlining the organization of sections and the logical flow of the study

type of glaucoma, in which outflow resistance is present at the trabecular meshwork or Schlemm's canal. This type of glaucoma causes vision impairment with minimal symptoms (Kwon et al. 2009; Stamer et al. 2015). Secondary open-angle glaucoma is less frequent and includes pigmentary glaucoma (Simcoe et al. 2020), pseudoexfoliation glaucoma (Elhawy et al. 2012), uveitic and traumatic glaucomas, steroid-induced glaucoma, and glaucoma secondary to increased episcleral venous pressure (Sung and Barton 2004; Joshi et al. 2017). On the other hand, angle-closure glaucomas are characterized by more evident patient complaints. They occur due to iridocorneal angle narrowing and are less common than open-angle glaucoma but are responsible for 50% of glaucoma-related blindness (Zhang et al. 2021). Secondary angle-closure glaucomas include phacomorphic glaucoma, iridocorneal endothelium syndromes, large iris or ciliary body masses, or drug-induced angle closure (Masoomian et al. 2020; Denis 2007; Jain et al. 1983; Rodrigues et al. 2016).

Congenital glaucoma is considered a distinct entity that occurs when IOP increases early in life, causing ocular component stretching and late optic nerve damage. Surgical management is the mainstay of treatment for congenital glaucoma. Table 1 summarizes the main clinical characteristics and symptoms of different types and stages of glaucoma, including open-angle primary glaucoma (POAG), open-angle primary glaucoma (PACG), secondary glaucoma, and congenital glaucoma, highlighting differences in anatomical changes and presentation to the patient (Essuman and Beyuo 2024).

More than half of patients with glaucoma are unaware that they have the disease (Brusini et al. 2021). Diagnosis requires a multidisciplinary approach, starting from risk assessment and clinical examination to IOP measurement, OCT imaging of the retinal nerve fiber layer and optic nerve head, and visual field testing. Figure 2 displays an overview of the traditional steps of detection and diagnosis of glaucoma (Weinreb and Khaw 2004). Visual

Table 1 Clinical classification of glaucoma showing major types (POAG, PACG, secondary and congenital) with its stages, characteristic changes in the optic nerve, and typical symptoms, from early to advanced disease

Type	Stage	Clinical signs & characteristics	Symptoms
POAG	Early	Mild optic nerve cupping, slight RNFL thinning, minimal field loss	Often none; mild peripheral vision changes
	Advanced	Marked cupping, very thin RNFL, extensive visual field loss	Tunnel vision, possible central vision loss
PACG	Acute	Narrow/closed angle, corneal edema, high IOP	Eye pain, headache, halos
	Chronic	Gradual angle closure, optic nerve damage	Gradual vision loss
Secondary glaucoma	Variable	Damage due to trauma, uveitis, steroid use, etc	Depends on cause
Congenital glaucoma	Infantile/Juvenile	Enlarged cornea, corneal clouding, high IOP	Light sensitivity, tearing, enlarged eyes

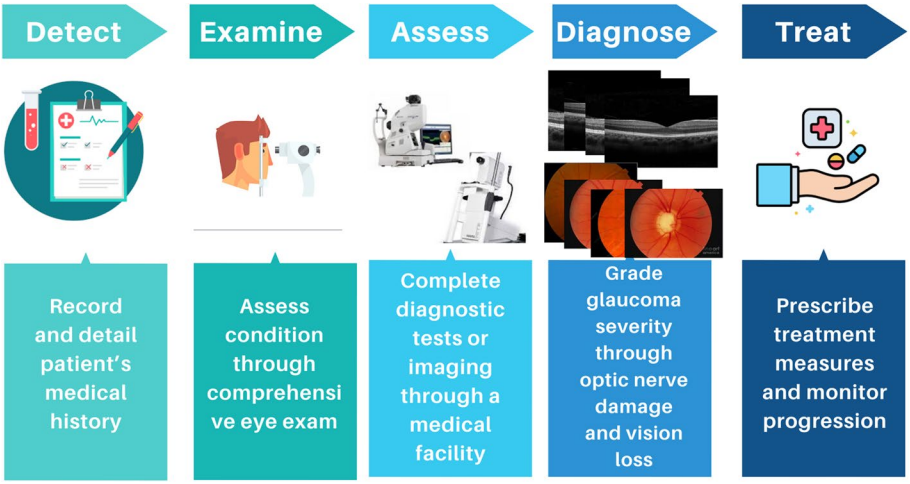


Fig. 2 Traditional clinical approach for diagnosing and grading glaucoma. The first step of the process is recording and detailing the patient's medical history, followed by a detailed eye examination to test the eye's health. These diagnostic tests, including images obtained with OCT and fundus cameras, are then conducted in a medical facility to generate objective measurements. Based on these results, clinicians make a call about the scope of glaucoma, optic nerve damage, or vision loss. The last step is to obtain a diagnosis, provide correct treatment measures, and monitor the disease development over time

field loss develops late, as central vision and Snellen visual acuity are preserved until the advanced stage. Correlation between OCT imaging and visual field changes can usually be observed. The severity of glaucoma can be categorized as mild, moderate, or severe on the basis of functional visual field abnormalities (Hood et al. 2019; Gedde et al. 2021).

Diagnosis of glaucoma relies on many clinical and investigational tools, with difficulty in detection especially early in the disease process (Susanna et al. 2015). Automated AI models could minimize subjective factors, as they help interpret retinal and optic nerve images. AI also has several other applications for glaucoma: it can enhance workflows in glaucoma clinics, saving time and increasing accuracy (Jan et al. 2024). Artificial intelligence can help diagnose glaucoma from optic disc photographs with high sensitivity and specificity, though physiological disc cupping can lead to false-positive results (Li et al. 2018).

Treatment of glaucoma depends on decreasing IOP. Medical treatments that decrease aqueous humor production or increase drainage have a major role in managing open-angle glaucoma and stabilizing angle-closure glaucoma before laser or surgical intervention. Laser treatments such as laser peripheral iridotomy, laser trabeculoplasty, laser iridoplasty, and diode laser therapy also play a role in glaucoma management. Surgical interventions include angle-based and filtering surgeries (Wagner et al. 2022; Rodgers et al. 2018).

Glaucoma evaluation begins with a thorough examination of head images of the optic nerve, which ophthalmologists carefully assess for signs of structural damage. However, the process can be slowed by challenges such as human error and the long time needed for accurate interpretation, which can delay early detection. Therefore, automated diagnostic strategies are important to overcome these limitations in order to have faster and more accurate diagnoses.

Unlike conventional image processing methodologies, which, although well-founded and easily understandable, generally require manual parameter tuning and have a hard time dealing with varying image quality. In terms of accuracy, they are generally outperformed by deep learning (DL) methods. As can be seen in Fig. 3, deep learning (DL) frameworks can be identified based on the model's capacity to derive hierarchical representation and abstract concepts from the data itself tations directly from raw retinal imagery, though practical deployments often integrate pre-process ing to improve input fidelity (Fig. 3a). To mitigate the data scarcity inherent in glaucoma studies, transfer learning (TL) exploits inductive biases from pretrained architectures, thereby accelerating convergence and reducing computational demands (Fig. 3b). Furthermore, federated learning (FL) facilitates collaborative, multi-institutional optimization, enhancing model generalizability while adhering to strict privacy constraints by retaining data on local nodes (Fig. 3c). We fundamentally distinguish feature extraction within this taxonomy: while conventional machine learning relies on hand-crafted descriptors, DL methodologies are defined by automated, end-to-end representation learning (Fig. 3d). These distinctions establish a rigorous conceptual framework for the AI methodologies examined herein.

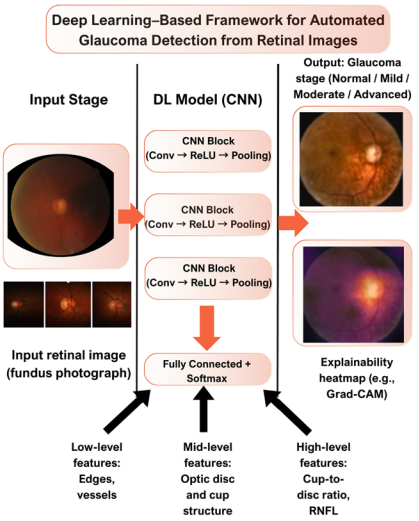
DL models also scale effectively, with performance improving as more data become available. Despite these benefits, DL models are sometimes criticized for their "black-box" nature and high computational requirements. Research has demonstrated their success in both binary classification (distinguishing glaucoma from normal) and multi-class grading of disease severity, using a range of public and private datasets.

ML methods provide more flexibility in selecting features and designing classifiers, offering interpretable decision-making for moderately sized datasets. However, their effectiveness depends heavily on careful manual feature engineering and pre-processing. Hybrid methods–combining traditional image processing with DL–aim to boost robustness and accuracy but may also increase implementation complexity and inherit drawbacks from both approaches.

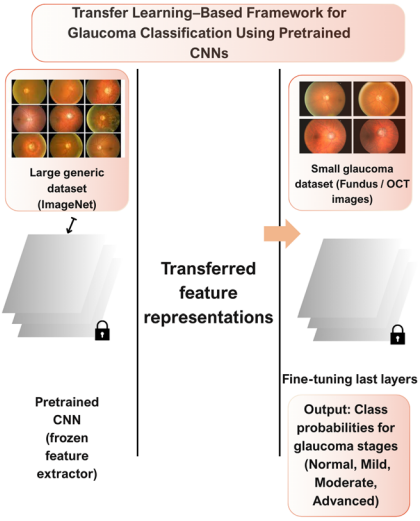
Furthermore, Fig. 4 in 3C and Fig. 4 in 3D illustrate the distribution of studies by year of publication. There has been a noticeable upward trend in the number of publications that address AI-based DR classification, achieving its peak in 2021 with 16 studies.

The distribution of the selected studies by publisher is shown in Fig. 4 in 3F and Fig. 4 in 3E. The most frequent publishers were Scientific Reports (26 studies), JAMA (12 studies), and IEEE (7 studies).

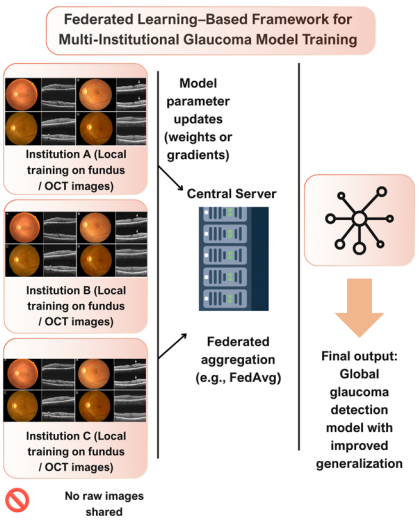
In light of the growing adoption of artificial intelligence across ophthalmic imaging, this review aims to examine how effectively AI-based methods can detect glaucoma, grade disease severity, and predict disease progression across commonly used imaging modalities, including fundus photography, optical coherence tomography (OCT), and optical coherence tomography angiography (OCTA). AI algorithms have proven to be highly accurate for the detection and diagnosis of glaucoma, and many systems are now achieving levels of accuracy that are close to and even higher than those of expert clinicians. Deep learning-based systems on fundus photography and OCT have also shown outstanding performance in discrimination between healthy and glaucomatous eyes and grading of disease severity. These models promise earlier diagnosis and better monitoring of progression, especially useful given that glaucoma progresses silently until late in the disease. However, most published evidence relies on small or homogeneous cohorts. In addition to questioning whether available results will be generalizable across heterogeneous populations, imaging machines



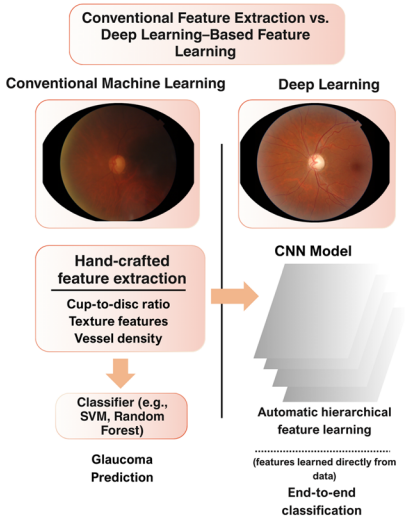
(a) Deep learning–based framework for glaucoma detection. The model learns hierarchical features directly from retinal images and performs end-to-end classification.



(b) Transfer learning–based framework using pretrained CNNs. A pretrained model is fine-tuned on a smaller glaucoma dataset to reduce training cost and improve convergence.

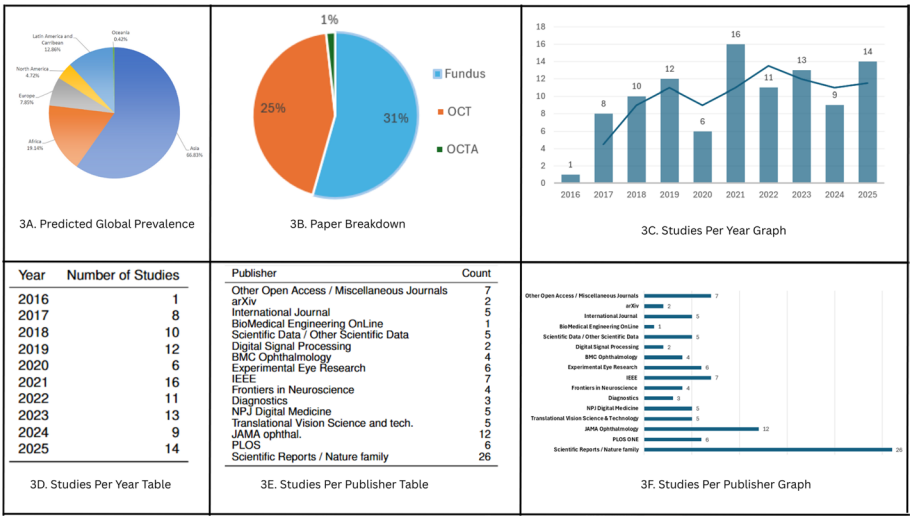


(c) Federated learning framework for multi-institutional glaucoma training. Local models are trained independently, while only model updates are shared with a central server to preserve privacy.



(d) Comparison between conventional machine learning and deep learning–based feature learning. Handcrafted features are replaced by automatically learned representations in deep models.

Fig. 3 Overview of artificial intelligence paradigms for glaucoma analysis: **a** deep learning, **b** transfer learning, **c** federated learning, and **d** conventional feature extraction versus deep feature learning



Year	Number of Studies
2016	1
2017	8
2018	10
2019	12
2020	6
2021	16
2022	11
2023	13
2024	9
2025	14

Publisher	Count
Other Open Access / Miscellaneous Journals	7
arXiv	2
International Journal	5
BioMedical Engineering OnLine	1
Scientific Data / Other Scientific Data	5
Digital Signal Processing	2
BMC Ophthalmology	4
Experimental Eye Research	6
IEEE	7
Frontiers in Neuroscience	4
Diagnostics	3
NPJ Digital Medicine	5
Translational Vision Science and tech.	5
JAMA ophthal.	12
PLOS	6
Scientific Reports / Nature family	26

Fig. 4 **3A** The predicted global prevalence of Glaucoma by 2040 (Zhang et al. 2021). **3B** The percentage of the paper is broken down into using different imaging tactics in Glaucoma. **3C** Glaucoma studies per year graph. **3D** Glaucoma studies per year table. **3E**. Glaucoma studies per publisher table. **3F** Glaucoma studies per publisher graph

and clinical facilities. Interpretability is also a problem, as much as most AI systems are black boxes, and clinicians may consequently lack full faith in the outcomes. All in all, AI undoubtedly has clear potential to support glaucoma detection and management, but greater validation in higher scales, greater transparency and integration into real practice workflows are needed before it can be safely implemented into practice.

2 Research questions

The following research questions were developed through a systematic, PRISMA-guided evaluation of the current landscape in AI-based glaucoma research. They aim to address recurring limitations found in recent studies, ranging from multimodal integration to the 'black box' interpretability problem.

- How can multimodal imaging approaches that combine OCT, OCTA, and fundus photography improve glaucoma detection and grading compared to single-modality strategies?
- How reliable are structural biomarkers derived from OCT (e.g., RNFL thickness) in predicting functional outcomes (visual field loss), and how can AI enhance this structure–function correlation?
- What strategies can overcome the limitations of small or single-center datasets highlighted in many of the reviewed studies?
- What is the comparative diagnostic value of fundus photography vs. OCT in early glaucoma detection?
- Can combining OCT, fundus imaging, and clinical data (IOP, VF, CCT) improve sensi-

tivity and specificity compared to uni-modal approaches?
- How do OCTA (angiography) or wide-field OCT expand our ability to detect subtle changes not visible in standard scans?
- What role do demographic and genetic factors (such as age, race, and family history) play in influencing glaucoma prevalence, and how can these be incorporated into predictive models?
- How could AI help reduce global disparities in blindness related to glaucoma, especially in countries with low awareness and coverage of treatment?
- How can the "black box" of DL models be addressed and reinforce their explainability to expand the trust of clinicians and acceptance?

.

3 Glaucoma statistics

3.1 Global glaucoma burden: an escalating global challenge

Glaucoma remains one of the biggest threats to vision across the world, stealing vision silently and irreversibly. The picture painted by these numbers is of grave concern. By 2040, millions more individuals will be affected by this situation, with Asia bearing the greatest strain at roughly 67% of worldwide cases, followed by Africa with about 19% of cases, while Latin America and the Caribbean regions account for approximately 13%. Developed regions are also not spared, although their percentage is lower and includes Europe at 7.85%, North America at 4.72%, and Oceania with less than 1% of cases. This fact highlights the importance of regular eye exams to allow early detection, especially in settings where the disease burden is significant.

3.2 Overall glaucoma prevalence

As illustrated by Figure 5 in 4A, the prevalence of glaucoma can vary significantly among countries. The high reported rate is 6.8% in Ghana, followed by 5% in Japan, 4.4% in Bahrain, and 3.9% in the USA. Other countries, such as Iran and the Netherlands, report a prevalence of roughly 1% each. Geographic differences in glaucoma occurrence are shown in the illustration. This variation may arise from changes in environmental exposures, genetic makeup, screening programs, or access to healthcare. More intensive screening and public health initiatives might be necessary in high-prevalence nations. (Khalili et al. 2022; Keel et al. 2018).

3.3 Glaucoma grading standards and clinical relevance

Precise severity stratification is paramount for both clinical management and rigorous evaluation. AI-based diagnostic systems. Glaucoma clinically manifests as progressive structural neurodeg- Anatomical generation of the optic nerve head and retinal nerve fiber layer, with concurrent functional Visual-field deficits. Strong staging enables optimum therapeutic decision-making and tailored surveillance Lancet intervals, prevention of irreversible visual

Fig. 5 **4A** Overall glaucoma prevalence by country (Khalili et al. 2022; Keel et al. 2018), **4B** distribution of glaucoma subtypes in the USA (POAG vs PACG) (Kwon et al. 2009; Masoomian et al. 2020; Khalili et al. 2022; Keel et al. 2018), **4C** gender-based glaucoma prevalence by country (Keel et al. 2018; Mannaf et al. 2024), **4D** age-stratified glaucoma prevalence across five countries (Khalili et al. 2022; Mannaf et al. 2024), **4E** awareness and treatment coverage for glaucoma across five countries (Khalili et al. 2022; Husain et al. 2024), **4F** glaucoma-related vision loss across eleven countries (Khalili et al. 2022; Dielemans et al. 1994; Ehrlich et al. 2024; Fujiwara et al. 2022)

impairment. A number of globally standardized grading frameworks form the basis for clinical practice, as outlined in the summary in Table 2. The Hodapp-Parrish-Anderson system stratifies disease severity based on funct Tional visual field loss, whereas the ISGEO classification synthesizes structural and functional crite it is done and standardized for diagnosis across populations. Further, Advanced Glaucoma Intervention Study (AGIS) score provides a quantitative measure for functional staging, which is widely used in longitudinal research. These classification systems drive clinical practice: early-stage disease will often be longitudinal observation and periodic rescreening, whereas advanced pathology constitutes a high This is a risk category that requires immediate therapeutic intervention to impede its progression. The ground truth, in computational ophthalmology, is standardized criteria such that truth annotation for model training and validation.

3.4 USA glaucoma subtypes – POAG vs PACG

The distribution of glaucoma subtypes is shown in Fig. 5 in 4B, POAG making up the vast majority (3%) and PACG accounting for only 0.3%. The figure confirms that POAG is the dominant form of glaucoma in Western populations like the U.S. PACG, though less common, remains clinically significant, especially in populations of Asian descent where its prevalence is higher (Kwon et al. 2009; Masoomian et al. 2020).

3.5 Gender-based prevalence

In Fig. 5 in section 4C, the prevalence of glaucoma is compared between males and females in various countries. In most nations, men have higher rates, especially in Ghana (7.7% male vs. 6.1% female) and Japan (5.8% male vs. 4.4% female). The gender disparity may reflect a combination of biological susceptibility, behavioral factors (e.g., access to care), and possible diagnostic bias. The figure suggests that males are at slightly higher risk in most countries, although the differences vary by region (Keel et al. 2018; Mannaf et al. 2024).

Table 2 Standard glaucoma grading systems

Grading system	Basis of grading	Stages/definition	References
Hodapp–Parrish–Anderson (HPA)	Visual field loss (SAP; Mean Deviation – MD)	Early: MD > -6 dB Moderate: MD -6 to -12 dB Severe: MD < -12 dB	Hodapp et al. (1993)
ISGEO classification	Structural (optic disc, RNFL) + Functional (VF)	Category 1: Structural damage + VF defect Category 2: Advanced structural damage without VF Category 3: Blindness with glaucoma features	Foster (2002)
AGIS score	Visual field defect scoring	0–5: Mild 6–11: Moderate 12–20: Severe	Investigators (1994)

SAP, standard automated perimetry; MD, mean deviation; RNFL, retinal nerve fiber layer; VF, visual field

3.6 Age stratified glaucoma prevalence

Illustrated by Fig. 5 in section 4D is the prevalence of glaucoma across different age groups in five countries: Australia, South Korea, Japan, Ghana, and the USA. The data show a clear trend of increasing glaucoma prevalence with age in all countries. Ghana exhibits the highest overall prevalence, particularly in the 80+ age group (14.6%), followed closely by Japan (13.1%) and Australia (12.9%). South Korea, on the other hand, continuously records the lowest rates in every age group, with the 80+ group reaching an all-time high of 10.7%. Prevalence rates in the United States are mild, steadily rising to 11% in the oldest age group. With significant variation by region that might be due to variations in genetics, availability to healthcare, or diagnostic procedures, these findings demonstrate a strong association between aging and glaucoma risk.

3.7 Awareness and treatment coverage

Critical insights can be found and stressed in Figure 5 in part 4E. Glaucoma awareness is heavily influenced by treatment coverage, but awareness alone is not always sufficient to guarantee care. Countries such as the USA, with a high awareness of 65%, also demonstrate high treatment rates at 60%, which suggests that with an effective healthcare system, the patients are being followed up with. For example, in India, Bahrain, and China, the treatment rates are lower than awareness, which can be caused by potential barriers. Such barriers can be from cost, limited access to services, or even inadequate healthcare infrastructure. What is most alarming is the gap in Ghana, where both awareness is at 18%, and treatment are at 10%. These results reinforce the need to incorporate public Education, health access, and policy issues regarding reducing the glaucoma burden worldwide.

3.8 Glaucoma related vision loss

The rates of glaucoma-related severe visual impairment and blindness are presented in Figure 5, section 4F. blindness across eleven countries. Ghana has the highest burden of blindness, where 25% of individuals suffer from severe visual impairment and 18% experience blindness. Bahrain, China, and Bangladesh also reported high rates of vision loss, which could suggest a lack of early detection and access to treatment. On the other hand, Iran, the Netherlands, and the USA report significantly lower values, and Iran shows the lowest at 5 and 2 percent, respectively. These findings highlight worldwide differences in glaucoma-related results and propose that targeted interventions and access to eye care services could help prevent vision loss, especially in low-resource regions.

3.9 Smoking status and glaucoma risk

This Table 3 explores the relationship between smoking status and glaucoma risk based on unadjusted odds ratios (OR) for males and females. Among males, ex-smokers and regular smokers show slightly lower odds of glaucoma compared to nonsmokers, although these associations are not statistically significant. Heavy male smokers exhibit a modest increase in risk (OR = 1.17), while the “other type” smoking group shows a higher OR of 1.60, again without statistical significance. For females, only the “other type” group has a reported OR,

Table 3 Unadjusted OR for disease risk according to smoking status and gender

Smoking status	Male OR (Unadjusted)	Female OR (Unadjusted)	Male (Unadjusted)	Female (Unadjusted)
Non-smoker	Reference	Reference	Reference	Reference
Ex-smoker	0.51 (0.21–1.23)	–	0.135	–
Regular smoker	0.61 (0.27–1.39)	–	0.245	–
Heavy smoker	1.17 (0.55–2.47)	–	0.671	–
Other type	1.60 (0.56–4.57)	6.67 (0.87–51.17)	0.38	0.068

The table shows comparisons across nonsmokers, ex-smokers, regular smokers, heavy smokers, and other smoking types, with results stratified for males and females. (Khalili et al. 2022)

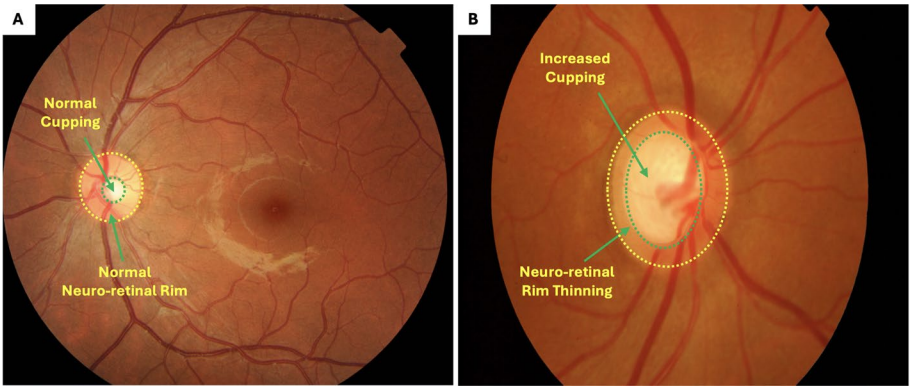


Fig. 6 Color fundus photographs illustrating the difference between a normal optic nerve head (**A**) and glaucomatous optic neuropathy (**B**). The glaucomatous disc shows increased cupping and thinning of the neuroretinal rim

which is markedly elevated at 6.67. Although the p-value (0.068) is just above the typical significance threshold, it suggests a potential association that may be relevant in larger studies. This data indicates possible sex-specific patterns in smoking-related glaucoma risk and highlights the need for more gender-inclusive research in glaucoma epidemiology.

4 Diagnostic imaging modalities of glaucoma

Diagnostic imaging modalities for glaucoma include CFP, OCT, and OCTA. CFP provides a magnified, flattened image of the optic nerve head (ONH) and surrounding retinal tissues using a specialized fundus camera. It is widely used to document optic disc appearance, peripapillary atrophy, and retinal nerve fiber layer (RNFL) defects. Historically, CFP has been considered the **gold standard** for documenting structural glaucomatous changes and monitoring progression. Fig. 6 illustrates fundus images showing features of glaucomatous optic neuropathy, including increased cup-to-disc ratio and thinning of the neuroretinal rim.

OCT is used to capture cross-sectional images of the retina and optic nerve, providing quantitative measurements of RNFL thickness, ganglion cell complex (GCC) thickness, and ONH morphology. These parameters are highly sensitive for detecting early glaucomatous damage, often before visual field loss becomes apparent, as shown in Fig. 7. OCT is **clinically accepted** and widely used in practice for glaucoma diagnosis and monitoring, complementing CFP.

OCTA provides high-resolution, depth-resolved images of the retinal and ONH microvasculature without the need for dye injection. By detecting motion contrast from red blood cells, OCTA visualizes vessel density and perfusion in the peripapillary and macular regions, as shown in Fig. 8. OCTA can also detect glaucomatous changes in microvasculature, such as reduced perip density, focal capillary dropout, and macular perfusion. Its higher resolution compared to other imaging modalities enables the recognition of subtle vascular changes; however, interpretation can be affected by motion artifacts and segmentation errors.OCTA is considered a clinically convenient imaging method, with imaging time being significantly shorter than OCT. plementary tool, providing additional vascular information that supports structural assessments from CFP and OCT.

In summary, the **traditional gold standard** for the assessment of the structure of the eye in glaucoma still remains CFP. Though OCT and OCTA are **clinically acceptable and com-**

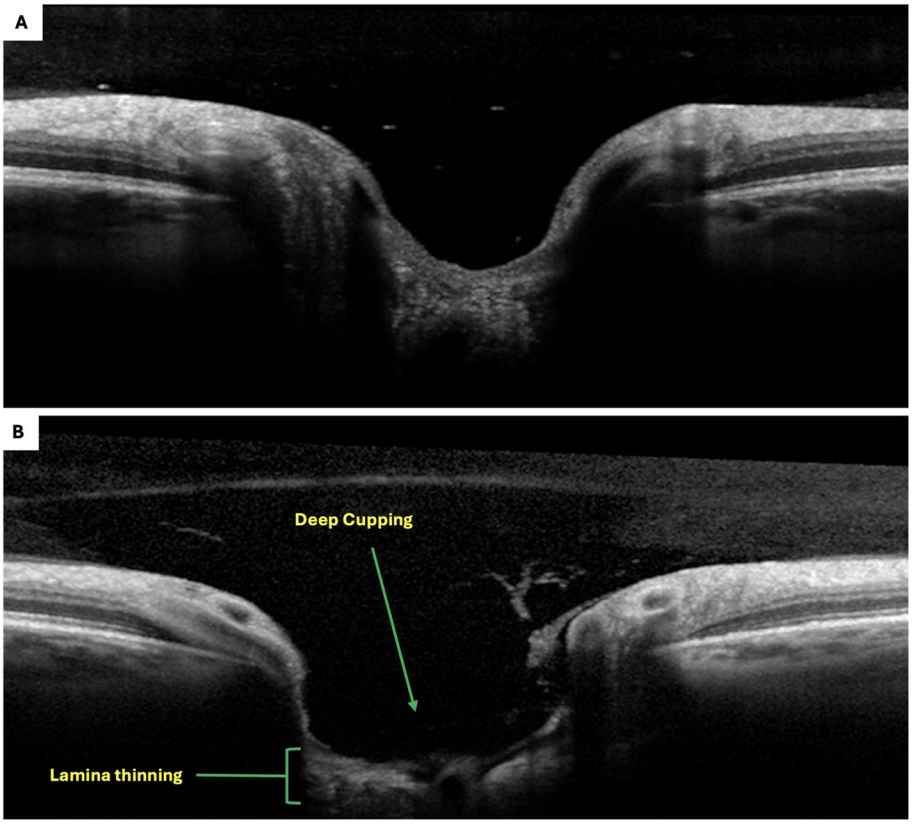


Fig. 7 OCT scans of the optic nerve head in a normal eye (**4A**) and a glaucomatous eye (**B**). The glaucomatous eye demonstrates excavation of the optic cup and thinning of the lamina cribrosa

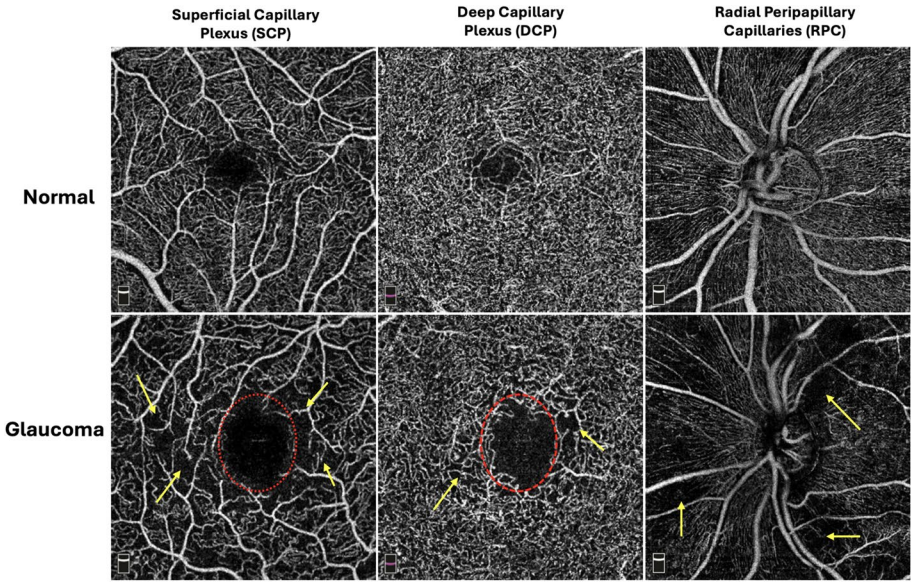


Fig. 8 OCTA scans comparing normal (top row) and glaucomatous (bottom row) eyes across three vascular layers: superficial capillary plexus (SCP), deep capillary plexus (DCP), and radial peripapillary capillaries (RPC). Glaucomatous eyes exhibit reduced vessel density (yellow arrows), particularly in the RPC layer surrounding the optic nerve head, and increased foveal avascular zone (red circles)

plementary modalities, enhancing early detection of anterior segment diseases remains an area of detection and in-depth evaluation of glaucomatous changes.

5 Materials and methods

5.1 Initial search

This review was overseen in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The PICOS (Population, Intervention, Comparison, Outcomes, and Study design) framework was used to define the eligibility criteria. Search terms included "Glaucoma," "Primary Open-Angle Glaucoma," "Primary Angle-Closure Glaucoma," "Optical Coherence Tomography," "Fundus Photography," "Artificial Intelligence (AI)," "Machine Learning," "Deep Learning," and "Computer Vision." Searches were performed across multiple databases such as IEEE Xplore, LWW Journals, IVOS, Tandfon Online, MDPI, Springer Nature, and Online Library up to the date of January 2025. All retrieved records were compiled in a Microsoft®Excel spreadsheet for further screening.

5.2 Preliminary screening

The review process began with an initial screen designed to efficiently filter the collection of articles. The first step was the immediate exclusion of any documents not published in

English or those only available as a standalone abstract. Critically, non-original research formats were also eliminated. This included setting aside case reports, opinion pieces, and pre-synthesized works such as meta-analyses. Once the remaining studies were confirmed to meet the basic structural requirements, the next hurdle was redundancy. This was tackled by developing a specialized MATLAB script (R2024b, The MathWorks Inc., Natick, MA, USA). The custom tool cross-referenced all Digital Object Identifiers (DOIs), detecting and eliminating duplicate articles automatically. In order to prepare for the final, comprehensive eligibility evaluation, the first author, publication year, journal, and DOI were carefully documented in the structured database that was created from the resulting, unique body of literature.

5.3 Eligibility assessment

All articles were evaluated according to the PICOS criteria outlined in Table 4. To be considered for inclusion, a study had to involve patients diagnosed with glaucoma at any stage or subtype and employ an artificial intelligence–based approach for detection, classification, or grading of the disease. Furthermore, the study was required to report at least one quantitative performance metric, such as accuracy, sensitivity, specificity, F1-score, Kappa score, or area under the curve (AUC). Studies that detailed only theoretical models without experimental validation, used animal research, or did not provide quantifiable data were disqualified.

Importantly, during the eligibility assessment, we explicitly evaluated whether each study incorporated explainable AI (XAI) methods. Studies employing XAI techniques were systematically identified, documented, and categorized alongside conventional AI approaches. This consideration was fully integrated into our PRISMA-based workflow, ensuring that the review captures both standard AI studies and those leveraging explainability, and that inclusion decisions consistently reflected the presence or absence of XAI.

The screening and data extraction processes were carried out independently by two reviewers, with any disagreements resolved through discussion with a third reviewer to ensure consistency and accuracy in the selection process.

Table 4 The inclusion criteria were utilized to choose publications for the glaucoma systematic review

Parameter	Description
Population	Patients with glaucoma (any stage or subtype)
Intervention	AI-based methods for glaucoma detection, classification, or grading
Comparison	Related AI-based studies with reported performance metrics
Outcomes	Accuracy, sensitivity, specificity, precision, F1-score, Kappa score, or AUC
Study design	All studies designs in English, excluding case reports, editorials, letters, meta-analyses, animal studies, and abstract-only papers

Studies using AI-based techniques for the detection, categorization, or grading of patients with glaucoma at any stage were included in the review. Eligible articles were written in English and reported performance indicators such as F1-score, AUC, sensitivity, specificity, and accuracy. Studies with only abstracts, editorials, letters, case reports, meta-analyses, and animal research were not included

A total of 100 studies were included in this review after excluding non-reputable articles through a systematic screening and eligibility assessment process. The selection process is illustrated in Fig. 9, which presents a PRISMA flowchart depicting each stage of identification, screening, eligibility evaluation, and final inclusion. From the initial 290 articles retrieved across multiple databases (IEEE Xplore, PubMed, Google Scholar, Science Direct, etc.), 229 articles remained after excluding those not published within the last decade. Of these, 229 articles were screened for eligibility based on titles, abstracts, and language, leading to the exclusion of 88 articles. The remaining 141 full-text articles were assessed using the PICOS criteria shown in Table 4, and 20 were excluded for not meeting these standards, leaving 121 articles for further evaluation and resulting in the final inclusion of 100 studies.

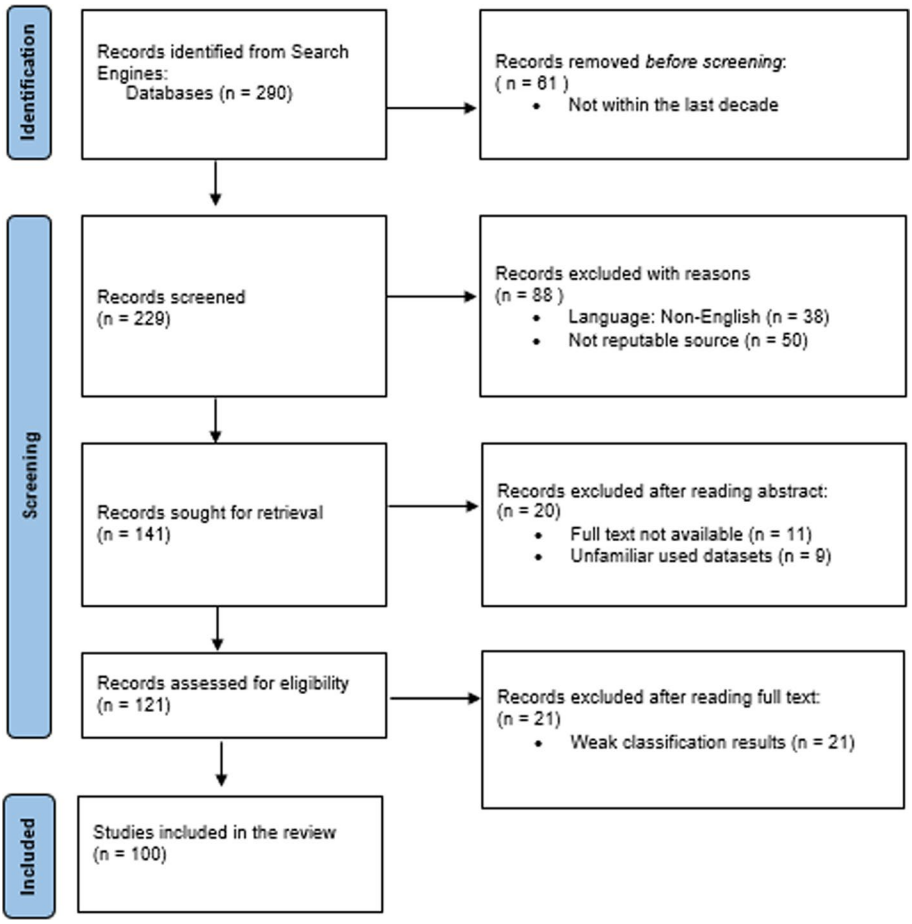


Fig. 9 PRISMA flowchart showing the study selection process. Out of 200 records retrieved, 81 were removed as duplicates or outdated. After screening 229 studies, 88 were excluded (non-English or non-reputable). From 141 abstracts, 20 were excluded, leaving 121 full texts assessed. Of these, 21 were excluded for weak results, and finally 100 studies were included in the review. *Definitions:Not reputable source:* Predatory journals or non–peer-reviewed sources. *Unfamiliar used datasets:* Rarely cited, poorly documented, or non-public datasets. *Weak classification results:* Performance metrics below standard benchmarks or insufficiently validated results

5.4 Glaucoma detection using OCT scans

As shown in Table 5, multiple studies have implemented binary and multi-class classification for glaucoma detection using OCT scans, achieving high diagnostic accuracy through DL and ML approaches (Kim et al. 2017). With the use of creative feature engineering like RNFL4, Kim et al.'s Random Forest model achieved outstanding results (98% accuracy, 0.979 AUC), setting significant initial standards for traditional machine learning techniques. Using C5.0 decision trees with distinct diagnostic criteria, interpretability is maintained in the interim. Later DL approaches, like (Russakoff et al. 2020)'s gNet3D 3D CNN (0.88 AUC), used natural OCT volumes but eliminated some interpretability. Medeiros et al. (2019) filled this gap by using fundus photos to estimate quantitative RNFL values (r = 0.832) while maintaining usefulness.

Valuable insights are revealed by the shift from single-modality to multimodal methods. Medeiros et al. (2021)'s comprehensive multimodal method (0.967 AUC), which comprised fundus images, clinical data, and macular OCT, was more effective than An et al. (2018)'s hybrid OCT + LSFG model (87.8% accuracy), which showed the value of integrating data types. Similarly, Xiao et al. (2021) demonstrated that when properly integrated, dual-modal combinations (OCT + VF, 0.950 AUC) can produce good outcomes. In contrast to Wu et al. (2021), which showed more consistent performance across disease stages (0.78–0.93 AUC) in a Taiwanese cohort, Li et al. (2023) reported performance drops in Caucasian versus Asian populations (0.83 vs 0.96 AUC) due to anatomical differences, indicating that disease advancement markers may be more generalizable than structural features.

Escamez et al. (2021)'s clinically actionable gradient boosting model (0.953 AUC, unambiguous RNFL thresholds) and Alqudah et al. (2022)'s more accurate but less interpretable DL model (98.6% accuracy) demonstrate the balance between interpretability and accuracy. Biarnés et al. (2023) showed that uncontrolled methods can yield new insights and attain accurate results (96.1%). Performance is also affected by dataset size: Omodaka et al. (2017) concentrated on disc classification (87.8% accuracy, 49 eyes), whereas Thakur and Juneja (2020) pointed out that even robust preliminary results (0.96 AUC) might not ensure external validity (0.78–0.95 AUC). In contrast to Mehta et al. (2021)'s larger UK Biobank dataset (2,476 eyes), Yoon et al. (2021)'s microbiome approach (96% accuracy) presented new markers, albeit in a restricted sample (n = 25). This suggests that innovative biomarker investigation holds promise.Technical sophistication is illustrated by Thompson et al. (2020)'s segmentation-free DL (0.96 AUC), while differences in performance across disease stages ((Wu et al. 2022)'s 0.9073–0.9841 AUC range) versus (Christopher et al. 2018)'s findings on detection versus progression (0.95 vs 0.74 AUC) suggest the need for context-specific solutions. The area is moving toward larger, diversified datasets, multimodal integration, explainable AI, and clinical workflow optimization due to recurring problems across studies, such as dataset constraints, cultural generalizability gaps, clinical application, and comprehension. Hasan et al. (2025) performed multi-class classification of glaucoma severity (Normal, Early, Moderate, Advanced) using a private OCT dataset of 602 scans from the Centre for Eye Health, UNSW, Sydney, Australia. Rasel et al. (2024) achieved high performance in multi-class glaucoma classification using OCT images (AUC = 0.960), though reliance on a single-center dataset may limit external validity. Both studies focus on glaucoma progression rather than diagnosis, but Mariottoni et al. (2023) provides

Table 5 Summary of key studies on glaucoma detection using OCT scans, highlighting dataset diversity, applied AI methods, and diagnostic outcomes (accuracy, sensitivity, specificity, AUC, F1-score)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Hasan et al. (2025)	Multi-class classification (Normal, Early, Moderate, Advanced Glaucoma)	Private OCT dataset (602 scans), Center for Eye Health,UNSW, Sydney, Australia	Accuracy: 75.2%, Sensitivity:79.1%, Specificity:71.4%, AUC: 0.816 (early), 0.844 (moderate), 0.932 (advanced), F1 Score:78.4%	Likely restricted to a specific population or imaging device, limiting cross-device and cross-population generalization
Rasel et al. (2024)	Private: 1,200 OCT images from Aravind Eye Hospital	Multi-class classification (Normal, Early, Moderate, Severe Glaucoma)	Accuracy: 90.1%; Sensitivity: 89.1%; Specificity: 91.3%; AUC: 0.950, F1-Score:-	Likely tested on a limited dataset, potentially from a single device or clinical site, reducing external validity
Omodaka et al. (2017)	163 eyes from 105 OAG patients (Japan); train: 114 eyes, test: 49 eyes	Neural Network on 91 quantified parameters (demographics, OCT, LSFG)	Accuracy: 87.8%, Sensitivity: –, Specificity: –, AUC:.91, F1-Score:75	Small sample; single ethnicity (Japanese); high myopia excluded; cross-sectional; LSFG parameters minimally contributory; rare disc types underpowered
Koornwinder et al. (2025)	1472 patients with glaucoma, used RNFL OCT scans	Glaucoma progression (not healthy vs glaucoma)	Accuracy: –, Sensitivity: –, Specificity: –, AUC: –, F1-Score: –	Model performance for very early or atypical glaucoma cases may not be fully established
Mariottoni et al. (2023)	14034 SD-OCT scans from 816 eyes, 462 individuals	Glaucoma progression	Accuracy:-, Sensitivity: 87.3%, Specificity: 86.4% AUC: 0.98, F1-Score: –	Requires further prospective validation to assess utility in real-world settings
Thompson et al. (2020)	1,154 eyes (635 patients: 612 glaucoma, 542 normal), Duke Glaucoma Repository	Segmentation-free deep learning on 20,806 SD-OCT B-scans	Accuracy:-, Sensitivity: 95%, Specificity: 81%, AUC: 0.96, F1-Score: 0.92	Single institution; cross-sectional; excludes high myopia; external validation needed; Grad-CAM resolution limited; cannot assess progression longitudinally
Russakoff et al. (2020)	Public: 2,805 OCT cubes of 1,095 eyes (Stanford); External:505 eyes (CUHK),336 eyes (India)	3D deep learning using full macular OCT cube scans; stratified analysis across myopia subgroups; cross-ethnic/geographic validation	Accuracy:91%, Sensitivity:89%, Specificity:96%, AUC:0.893, F1 score:-	The study emphasized algorithmic performance but did not evaluate workflow integration or clinical decision-making impact
Smith and Doe (2024)	Longitudinal: 99 eyes (70 glaucoma/suspects, 29 healthy); plus 1,687 OCT scans for short-term variability (Topcon + Heidelberg)	Introduced RNFL + GCL probability change maps (pc-maps) for detecting glaucomatous progression; agreement with visual field (VF) confirmed progression	Accuracy:91%, Sensitivity:89.1%, Specificity:91.3%, AUC:.960, F1 score:-	Study design appears retrospective with a limited sample size; prospective, longitudinal validation is needed

Table 5 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Kim et al. (2023)	Private: 2,607 OCT scans, University of Tokyo, Japan	Deep learning applied to vertical macular OCT scans in myopic eyes; compared against circumpapillary OCT; external validation included	Accuracy: 96.5%, Sensitivity: 96.4%, Specificity: 96.5%, AUC: 0.976, F1 Score:-	Likely based on a single-center dataset, reducing external validity across ethnic groups and imaging systems
Chiang et al. (2024)	Private (450) conducted at University of California, San Diego (UCSD)	Identifies macula + wide-field combination as most effective; high diagnostic performance using DL	Accuracy:- Sensitivity: 93.1%, Specificity: 90.0%, AUC: 0.962, F1 Score:-	Did not evaluate predictive performance for glaucoma progression over time
George et al. (2020)	Private OCT(3,782) Cleveland Clinic, Cole Eye Institute	Attention-guided 3D CNN applied to volumetric OCT; robustness improved vs baseline; also explored VFI regression	Accuracy:93.1%, sensitivity: -, specificity: -, AUC: 93.7%, F1 Score:92.088%	Lacks incorporation of complementary clinical data
Maetschke et al. (2019)	Private (Mayo Clinic) (265)	Deep learning directly on unsegmented OCT volumes, avoiding retinal layer segmentation; CAM highlights clinically meaningful ONH subregions	Accuracy:-, sensitivity: -, specificity: -, AUC: 0.94, F1 Score:0.979	The study does not evaluate glaucoma progression or temporal changes in OCT volumes
Pham et al. (2025)	Private – Wilmer Eye Institute, Johns Hopkins University 70,575 paired OCT/VFs (model) from 44,659 eyes; 4,044 eyes in the progression set	Robust AI model predicts VF worsening using OCT; validated on longitudinal data	Accuracy:-, sensitivity: 81.8%, specificity: -, AUC: 0.95, F1 Score:-	May be limited to a specific clinical center or population, reducing generalizability across ethnic groups and imaging systems
Asano et al. (2021)	648 eyes (358 subjects: 90 normal, 558 glaucoma) from multiple Japanese hospitals (University of Tokyo, Inouye Eye Hospital, JR Tokyo General Hospital, Hiroshima Memorial, Osaka University, Kyoto Prefectural University, Oike-Ganka Ikeda Clinic)	Using deep learning and transfer learning	Accuracy.:- Sensitivity: - Specificity: - AUC: -, F1 Score:-	Deep learning and transfer learning outputs may lack transparency for clinical decision-making, affecting adoption

Table 5 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Asaoka et al. (2019)	4316 OCT images, 1371 eyes with open angle glaucoma (OAG), and 193 normal eyes	Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular OCT images	Accuracy: 93.7%, Sensitivity: 82.5%, Specificity: 93.9%, AUC: -, F1 Score:-	The model is specialized for early-onset glaucoma, which may limit performance on typical adult-onset cases
Singh et al. (2021)	Private and public datasets (OCT only) 70 glaucoma + 70 healthy	Extracted 45 features, evaluated with multiple models (KNN)	Accuracy.:- Sensitivity: 100%, Specificity: 85.7%, AUC: 0.996, F1 Score:-	The AI system may provide limited explainability for ophthalmologists, affecting adoption in clinical workflows
Akter et al. (2023)	370 glaucomatous and 410 normal images	VGG16 - feature visualization techniques (heatmaps)	Accuracy.: 93%, Sensitivity: 94.5%, Specificity: 95.9%, AUC: 0.988, F1 Score:95%	Effectiveness in detecting early glaucoma or subtle structural changes may not be fully evaluated
Lee et al. (2020)	307 healthy images and 86 glaucoma images (SD-OCT)	DL classifier using bottleneck features from GCIPL & RNFL maps	Accuracy:-, sensitivity: 94.7%, specificity: 100%, AUC:.99, F1 Score:-	Likely limited to a single-center or homogeneous patient population, which may reduce generalizability across ethnicities and imaging devices
Ashtari-Majlan and Masip (2025)	Existing dataset of 3D OCT scans, 263 healthy scans and 847 open-angle glaucoma	Spatial-aware Transformer-GRU Framework using 3D OCT Imaging	Accuracy: 89.19%, sensitivity: 91.83%, specificity: 79.67%, AUC:94.2%, F1 Score:-	Practical performance in diverse clinical workflows or noisy real-world data may not be fully assessed
Ran et al. (2019)	4877 SDOCT collected from Chinese University of Hong Kong Eye Centre and the Hong Kong Eye Hospital	Compared 2D vs 3D deep learning systems for optic neuropathy detection	Accuracy:91%, sensitivity: 89%, Specificity:86%, AUC:0.893−0.897, F1 Score:-	The model's performance on external, real-world datasets is not assessed, which may limit clinical applicability
Muhammad et al. (2017)	57 glaucomatous eyes and 45 healthy eyes using OCT	Hybrid approach:(CNN + Random Forest)	Accuracy:63.7%–93.1%, sensitivity: -, specificity: -, AUC: -, F1 Score:-	Focuses on classification of glaucoma suspects but does not evaluate longitudinal progression or predictive capabilities
Xu et al. (2019)	Chinese-American subjects, 4036 AS-OCT images were collected from 791 subjects	Deep learning classifiers trained on anterior segment OCT	Accuracy:89%, sensitivity: 95.8%, specificity: 74.4%, AUC: 0.964, F1 Score:-	Deep learning classifiers may lack transparency, making clinical adoption more challenging
Wu et al. (2021)	470 eyes from 265 participants at Fu Jen Catholic University Hospital; Spectralis OCT, 114 parameters including cRNFL, BMO-MRW, macular thickness	Binary classification	Accuracy:-, sensitivity: -, specificity: -, AUC: 0.9459, F1 Score:-	Single-center, Taiwanese cohort; relatively small sample; OCT-device specific; strict exclusion criteria; age mismatch between groups; black-box nature of classifiers

Table 5 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Wu et al. (2022)	752 eyes (498 glaucoma, 254 normal) from Fu-Jen Catholic University Hospital, Taiwan	SVM using Spectralis OCT features (peripapillary + macular)	Accuracy:-, sensitivity: -, specificity: -, AUC: 0.82, F1 Score:-	Taiwanese-only cohort; age imbalance; excluded high myopia/peripapillary atrophy; cross-sectional case–control design; no longitudinal progression data; small sample size; device-specific OCT
Biarnés et al. (2023)	211 eyes total (80 healthy, 131 glaucoma) from Spain; external validation on 102 eyes	Unsupervised clustering (Model-Based Clustering, k-means, Hierarchical) on Cirrus SD-OCT (pRNFL + GCIPL)	Accuracy: 91.7, Sensitivity:-, Specificity: 94.3, AUC:-, F1 Score:-	Small sample size; case–control design; early/pre-perimetric glaucoma excluded; single ethnicity (Spanish); limited OCT features; age imbalance; device-specific parameters

The table also identifies major research gaps, including small or homogeneous samples, device dependency, and limited external validity across populations

quantitative performance metrics showing high sensitivity and AUC, whereas Koornwinder et al. (2025) reports no performance metrics, making comparison difficult.

Smith and Doe (2024) proposed RNFL + GCL probability change maps for glaucoma progression detection with agreement to visual fields, but the retrospective, small-sample design limits generalizability, highlighting the need for prospective validation. Chiang et al. (2024) demonstrated high diagnostic performance using combined macula and wide-field OCT (AUC = 0.962), though progression prediction was not assessed. George et al. (2020) applied attention-guided 3D CNNs on a larger volumetric OCT dataset (AUC = 93.8) and explored VFI regression, yet the absence of complementary clinical data may limit real-world applicability. Maetschke et al. (2019) achieved high diagnostic performance (AUC = 0.94) using unsegmented OCT with CAM interpretability, but did not assess progression. Pham et al. (2025) leveraged a large longitudinal dataset to predict visual field worsening (AUC = 0.95), though generalizability may be limited due to single-center data.

Asano et al. (2021) used a multi-center dataset with DL + transfer learning, though interpretability may limit clinical adoption, whereas Asaoka et al. (2019) achieved high accuracy for early-onset glaucoma using macular OCT, with limited generalizability to typical adult-onset cases. Singh et al. (2021) reported near-perfect sensitivity on a small dataset using feature-engineered classical ML, though clinical interpretability is limited. Akter et al. (2023) applied DL with heatmap visualization, enhancing interpretability, though early glaucoma performance remains uncertain. Lee et al. (2020) achieved very high sensitivity and perfect specificity using SD-OCT bottleneck features on a small, single-center dataset, whereas Ashtari-Majlan and Masip (2025) applied a larger 3D OCT dataset with a spatial-aware Transformer-GRU framework, trading slightly lower accuracy and specificity for volumetric modeling, with untested performance in diverse clinical settings. Ran et al. (2019) compared 2D versus 3D deep learning on a large SDOCT dataset, achieving high sensitivity (89%) but without external validation, whereas Muhammad et al. (2017) used a smaller hybrid CNN–Random Forest approach, showing variable accuracy (63.7–93.1%) without longitudinal assessment. Xu et al. (2019) applied DL to 4,036 anterior segment

OCT images from 791 Chinese-American subjects, achieving a high AUC of 0.964, though limited interpretability may hinder clinical adoption.

5.5 Glaucoma grading and classification fundus and OCT imaging

Table 7 presents several studies that implemented ML and DL for different gradings of glaucoma using OCT scans and fundus images. Akter et al. (2022) developed a deep learning system for glaucoma detection using OCT scans, achieving expert-level accuracy (AUC: 0.99, sensitivity: 98.7%, specificity: 98.5%). Their approach combines structural features (RNFL thickness, cup-to-disc ratio) and functional features (visual field metrics) with segmented ONH cup area analysis. The custom CNN outperformed standard models (ResNet18/VGG16) while using a simpler architecture. Although limited by manual segmentation and dataset size, the system shows strong potential for clinical screening. Integrated Grad-CAM visualizations enhance interpretability by highlighting diagnostically relevant ONH regions. Noury et al. (2022) developed a 3D convolutional neural network (CNN) trained on raw SD-OCT ONH cube scans to detect glaucoma using multimodal clinical labels, validated across ethnically diverse real-world datasets. The model achieved strong performance with AUCs ranging from 0.80 (Hong Kong) to 0.94 (India) and highlighted the lamina cribrosa as a key diagnostic region via saliency maps. A novel pipeline, DiagFind, further confirmed that cropped ONH regions retained diagnostic information, achieving an AUC of 0.77. Performance was comparable to a glaucoma specialist (AUC = 0.92 vs. 0.91) despite relying only on OCT volumes. Limitations include exclusion of glaucoma suspects, lack of detailed signal strength analysis, and challenges in generalizing across OCT devices and variable-quality scans. These findings suggest that the lamina cribrosa may provide critical but underutilized structural cues in AI-based glaucoma detection. Figure 10 shows an example of a computer-aided detection (CAD) system for glaucoma, demonstrating how automated algorithms analyze retinal images to extract features and assist in the classification of healthy versus glaucomatous eyes.

Hasan et al. (2025) proposed an explainable ML framework using OCT-derived spatial and frequency domain features to diagnose and stage glaucoma with high precision. Using SHAP and partial dependency analysis (PDA), they developed interpretable models that achieved AUCs of 0.97 (overall), 0.96 (early), 0.98 (moderate), and 1.00 (advanced glaucoma). Random Forest and SVM classifiers outperformed clinicians by up to 12.6% in early-stage detection. The models were deployed in web- and Excel-based clinical tools for real-time glaucoma likelihood scoring. Limitations include reliance on CIRRUS OCT data, with generalizability to other devices and broader populations yet to be validated. Russakoff et al. (2020) developed a 3D DL model (gNet3D) to detect referable glaucoma using full macular OCT cube scans from real-world clinical data. The model was trained on 2,805 scans from 1,095 eyes and externally validated on datasets from Hong Kong and India. Automated retinal layer segmentation improved spatial context and model performance, yielding an AUC of 0.88 (vs. 0.81 without pre-processing) and up to 0.95 across moderate myopia. gNet3D outperformed previous models but showed varying results across ethnic and geographic datasets. Limitations include absence of suspect cases in external sets, reliance on retrospective data, and exclusion of some clinical risk factors. Smith and Doe (2024) applied longitudinal OCT probability change maps to detect early or suspect glaucoma in 99 eyes (70 cases, 29 controls). RNFL change maps revealed significantly greater

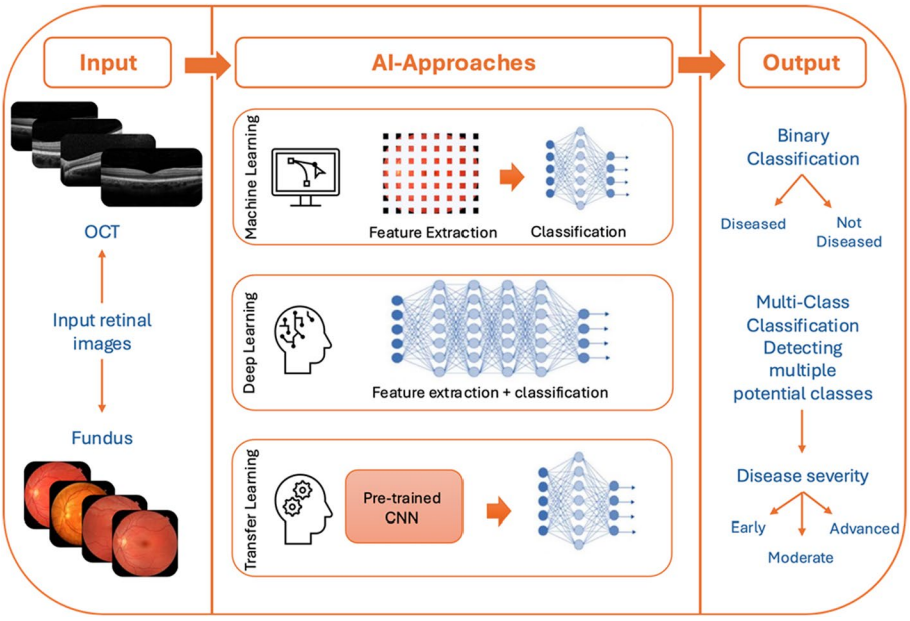


Fig. 10 Generalized workflow of CAD system for glaucoma detection and grading. Input retinal images (OCT and fundus) are processed through AI-based approaches–machine learning with feature extraction, deep learning with joint feature learning and classification, and transfer learning with pre-trained CNN models–leading to outputs such as binary classification (diseased vs. normal), multi-class classification, and severity staging (early, moderate, advanced)

abnormal regions in cases (mean 4.7%) compared with controls (mean 1.7%, $P < 0.001$). At 95% specificity, the method achieved 53% sensitivity for detecting early glaucomatous damage. Limitations include small sample size and exclusive focus on RNFL changes. Rasel et al. (2024) compared 2D and 3D CNNs for glaucoma detection using macular and optic nerve head OCT scans. The 2D-ResNet18 model (with pre-trained weights and XGBoost aggregation) outperformed 3D models, achieving AUCs of 0.960 (macular) and 0.943 (ONH). 3D CNNs suffered from overfitting due to limited data and lack of pre-trained models. Key limitations included low-resolution ONH volumes. Results favored 2D approaches for current applications, but future work with larger datasets could optimize 3D CNNs. Kim et al. (2023) developed a deep learning system using macular vertical OCT scans to detect glaucoma in myopic eyes, achieving an AUC of 0.981, outperforming circumpapillary scans (AUC: 0.914), especially in eyes with large PPA (AUC: 0.976 vs. 0.914). Including demographic data further boosted performance (AUC: 0.986). The study employed EfficientNet models on 2,358 OCT scans. Nevertheless, it was limited by a Korean-only cohort, possible augmentation biases, and imbalanced datasets (3:1 glaucoma/healthy ratio). TThese findings emphasize the necessity for varied confirmation to guarantee clinical relevance while highlighting the superiority of retinal vertical scans for high myopia. Saha et al. (2023) developed a lightweight CNN system for glaucoma detection using fundus photos, achieving 97.4% accuracy (AUC: 0.993) with MobileNetV3Small. Their simplified YOLO model efficiently localized the optic nerve head, requiring 12× less memory than ResNet50. The system demonstrated that resource-efficient models can match com-

plex architectures when analyzing specific retinal regions. Limitations include reliance on cropped images and dataset imbalances. This work supports potential integration into portable devices but requires broader validation. An AI system created by Son et al. (2023) uses retinal pictures to identify 15 anomalies and diagnose 8 eye conditions with expert-level accuracy (average AUROC 0.992). The AI's reasoning is visible because to their novel Counterfactual Attribution Ratio (CAR), which shows how particular results affect diagnoses. To ensure that the procedure easy to comprehend, the system use EfficientNet-B7 to discover features and then basic classifiers for the final diagnosis. Despite considerable accuracy variance, tests from several facilities performed reliably (AUROC 0.915–0.983). By interacting with the CAR system, physicians can examine and adjust diagnoses, increasing their trust in the AI's recommendations. Mariottoni et al. (2023) developed a deep learning model to detect glaucoma progression using spectral-domain OCT scans, achieving high diagnostic accuracy (AUC 0.938, 87.3% sensitivity, 86.4% specificity). The CNN-based system outperformed traditional trend analyses by 29.7% in sensitivity while providing localized progression heatmaps. Trained on 14,034 scans with expert consensus labels, the model's likelihood ratios significantly altered progression probabilities in 74% of cases. Though limited to structural data from a single center, its interpretable outputs enhance clinical utility, demonstrating AI's potential to augment glaucoma monitoring with actionable insights. Hussain et al. (2023) developed an AI system combining OCT scans, visual fields, and clinical data to predict glaucoma progression (AUC: 0.83). Their CNN-LSTM model, enhanced with GAN-generated synthetic OCT images, enabled prediction 6–9 months earlier, outperforming single-modality approaches. Key innovations included focal loss for class imbalance and segmentation-free analysis. While limited by a small sample size (86 patients), this multimodal framework shows promise for personalized glaucoma management. In order to identify glaucoma, Chiang et al. (2024) developed an AI system that analyzes wide-field OCT scans. It outperforms conventional single-region scans by automatically identifying important eye components and diagnosing the condition with 99% accuracy. The technique looks at the macula and optic nerve at the same time. Although encouraging, further extensive dataset testing is required. This strategy might enhance clinical glaucoma diagnosis. Thiéry et al. (2023) presented PointNet, a geometric deep learning technique that uses 3D point clouds from OCT images to diagnose glaucoma. Their approach outperformed RNFL thickness measurements (AUC: 0.80) and conventional 3D CNNs (AUC: 0.87–0.91) with an accuracy of 95% (AUC: 0.95). The method captured important structural changes while reducing the complexity of the data by turning scans into simplified point clouds (1,000 points per eye). Unverified posterior tissue borders and device-specific testing are among the drawbacks. With little input data, this approach could expedite the identification of glaucoma. Chiang et al. (2024) also developed a deep learning system to detect glaucoma in highly myopic eyes using fundus photographs. Their approach combined a ConvNeXt Base architecture with a Convolutional Block Attention Module (CBAM), achieving 82.16% accuracy (AUC: 0.894). The model outperformed traditional methods by focusing on optic disc and retinal nerve fiber layer features, even in eyes with severe myopia (-6D to -9D). While effective, the study was limited to Asian populations and excluded pathological myopia cases. This AI tool could assist non-specialists in early glaucoma screening but requires validation across diverse ethnic groups. George et al. (2020) developed an attention-guided 3D-CNN (AG-OCT) for glaucoma detection using OCT scans, achieving 93.8% AUC. Their model used three pathways–global, focused

(occluded), and local (cropped)–guided by grad-CAM heatmaps to highlight critical retinal structures. The approach outperformed traditional methods by 7% and identified unexpected biomarkers such as the RPE and photoreceptor layers. While effective, the model required high-resolution scans and excluded non-glaucomatous optic neuropathies. This framework shows promise for direct 3D scan analysis but needs validation on diverse datasets. A multi-task deep learning system for glaucoma classification was created by Kooner et al. (2022), which achieved 71.3% accuracy (F1: 0.624) in multi-class staging and 83.9% accuracy (F1: 0.838) in binary classification. Their model outperformed single-modality techniques by 9–12% by combining structural OCT variables with functional visual field data. Inter-class similarity in the early stages caused performance to drop in multi-class progression grading, despite its effectiveness in binary detection. The algorithm was very good at detecting glaucoma early on, but it had trouble telling mild instances from advanced ones. Hemelings et al. (2021) demonstrated that deep learning can detect glaucoma (AUC: 0.94 [95% CI: 0.92–0.96]) and estimate vertical cup-disc ratio (VCDR) (R^2: 0.77 [95% CI: 0.76–0.79]; MAE: 0.19) from fundus images, even when the optic nerve head (ONH) is obscured. Their models leveraged peripapillary retinal features, achieving significant performance (AUC: 0.88; R^2: 0.37) with 60% of the ONH masked. Saliency maps highlighted inferotemporal and superotemporal regions, correlating with glaucoma pathology. While robust, the study used fixed crop sizes and lacked myopia-specific validation. These findings underscore AI's ability to identify subtle, non-ONH biomarkers for glaucoma screening. Xiao et al. (2021) created a deep learning system that can anticipate surgical results and categorize the etiology of macular holes (MH) (idiopathic vs. secondary). Combining OCT pictures with clinical data, their model produced AUCs of 0.965 (etiology), 0.904 (all MH), and 0.947 (idiopathic MH). With 97 and 95% accuracy in etiology classification, respectively, it fared better than single-modality models.9% sensitivity for predicting idiopathic MH closure. A tiny secondary MH sample size was one of the constraints. Although it needs more widespread validation, this tool might enhance surgical planning. Li et al. (2019) developed an attention-based deep learning system for glaucoma detection using fundus images. Their AG-CNN model achieved 95.3% accuracy, 95.4% sensitivity, and 95.2% specificity (AUC: 0.967) on the LAG test set by incorporating ophthalmologist attention maps and multi-scale feature learning. The system outperformed existing methods by >5% accuracy and maintained strong performance on external validation (RIM-ONE: AUC 0.916). Key innovations included guided backpropagation for precise lesion localization and attention-guided feature learning. While demonstrating clinical potential, limitations included dependency on quality attention maps and moderate performance drops on external datasets. This approach shows promise for interpretable glaucoma screening but requires further validation. Pham et al. (2025) developed a machine learning model to predict glaucoma progression using OCT-derived visual field estimates. Their model achieved an AUC of 0.95 on the test set with 81.8% sensitivity, demonstrating strong diagnostic performance. By analyzing structural OCT data, the system could effectively identify eyes at risk of visual field deterioration. Although promising, accuracy was reduced in moderate-stage glaucoma cases. This approach shows potential for complementing traditional glaucoma monitoring but requires further validation in diverse clinical populations. Shoukat et al. (2023) created a deep learning system that uses retinal fundus images to automatically diagnose glaucoma. On the G1020 dataset, their ResNet-50-based model had an AUC of 0.97, sensitivity of 99.3%, and specificity of 96.52%. The method showed great accuracy (98.48%) in detecting initial

stages of glaucoma by using grayscale fundus images and data augmentation techniques to improve performance. Despite demonstrating strong performance on several datasets (RIM-ONE, ORIGA, and DRISHTI-GS), the model's specificity was lower on the ORIGA dataset (79.26%). This strategy offers a viable and effective substitute, but it still has to be tested in a variety of clinical settings. Huang et al. (2022) developed an advanced deep learning system for glaucoma detection using multi-modal ophthalmic data. Their probabilistic model demonstrated exceptional performance, achieving an AUC of 0.996 with 97.4% overall accuracy. The system showed balanced diagnostic capability with 97.6% sensitivity (correctly identifying glaucoma cases) and 97.2% specificity (correctly identifying normal eyes). These results represent a significant improvement over previous methods, particularly in the model's ability to maintain high accuracy across both disease detection and normal case identification. However, extensive evaluation across a wide range of patient populations is essential before the system can be implemented on a wide scale. Chincholi and Koestler (2024) developed transformer-based deep learning models (ViT and DETR) for glaucoma detection using fundus images, achieving AUCs of 0.95 (OHTS) and 0.92 (DIGS/ADAGES). Their system demonstrated balanced performance with 87–89% sensitivity and 80–82% specificity by analyzing cup-to-disc ratios (threshold >0.6) and neuroretinal rim thinning. While DETR showed marginally better accuracy (90.48% vs. ViT's 87.87%), both models surpassed traditional methods in localization precision. Limitations included potential dataset bias from clinic-sourced images. This approach shows promise for automated glaucoma screening but requires validation in diverse populations. A generalizable deep learning regression model (G-RISK) was created by Hemelings et al. (2023) for the purpose of screening fundus pictures for glaucoma. Their method performed exceptionally well on a variety of datasets, with population-level data showing AUC values of 0.976 (BMES) and 0.984 (GHS). The model outperformed the 85% sensitivity threshold suggested by Prevent Blindness America, with high sensitivities of 87.3% (BMES) and 90.3% (GHS). Eleven publicly available datasets were used to further validate the model's stability; the AUCs ranged from 0.854 to 0.988. G-RISK fared better than conventional binary classification techniques and even outperformed image-measured VCDR as a glaucoma risk proxy by utilizing a regression-based framework trained on VCDR estimates. These results highlight the model's potential for widespread clinical application; yet, additional thorough research conducted in real-world settings is necessary to validate its therapeutic efficacy. Lin et al. (2024) developed *Brighteye*, a Vision Transformer (ViT)-based framework for glaucoma screening using color fundus photographs. The system integrates YOLOv8 for optic disc detection and cropping, followed by a ViT model for glaucoma classification and glaucomatous feature analysis. *Brighteye* achieved a sensitivity of 85.7% at 95% specificity, surpassing the Prevent Blindness America criteria, and reduced the Hamming distance for feature classification from 0.2470 to 0.1250 through optimized pre-processing. Ranked 5th among 226 entries in the JustRAIGS challenge, the model demonstrated robustness despite being trained on a limited set of annotated data (80 images for YOLOv8). Key innovations include patch feature aggregation in ViT and alignment with clinical workflow (e.g., optic disc-centric ROI cropping). However, challenges remain in detecting subtle features such as shallow RNFL defects, as shown by misclassified cases. The code is publicly available, facilitating further research and deployment. Sharma et al. (2025) developed *AI-GS*, a hybrid multi-model artificial intelligence network for glaucoma screening using macula-centered fundus images. The system integrates six lightweight deep learning models

(total size: 110 MB) to perform simultaneous segmentation (optic cup/disc, fovea), classification (glaucoma/normal), and detection of structural biomarkers (disc hemorrhages, RNFL defects). *AI-GS* achieved a sensitivity of 93.5% at 95% specificity on internal testing and maintained 80.5% sensitivity at 91.1% specificity in real-world screening (Miyagi dataset), outperforming standalone binary classifiers. Huang et al. (2025) developed the Multi-modal Longitudinal Estimation Deep Learning (MLEDL) system, a deep learning framework that predicts current and future visual field (VF) loss in glaucoma using color fundus photographs (CFPs) and clinical narratives. The system integrates ResNet-50 with multimodal data fusion, achieving a pointwise mean absolute error (MAE) of 3.10 dB for future VF prediction. Clinical validation showed 83% agreement with ophthalmologist-graded VFs using the HPA method. Heatmaps revealed structure-function relationships between optic nerve damage and VF loss patterns, enhancing interpretability. Despite its strong performance ($R^2 = 0.839$ for longitudinal predictions), the model underestimates moderate VF defects (accuracy: 53-56%) and is affected by external dataset variability (MAE: 4.42 dB on PKU data). The framework's reliance on CFPs and clinical text, rather than OCT or perimetry, makes it viable for resource-limited settings. Code and data are available for research upon request. By integrating OCT and visual field data, Pascal et al. (2022) created a multimodal deep learning system for glaucoma detection. With an external AUC of 0.911 and an internal AUC of 0.996, the framework maintains high generalization by integrating structural OCT characteristics with functional visual field measurements using an improved neural network design. The method outperformed conventional diagnostic criteria for glaucoma screening, exhibiting clinically strong performance with 84.5% sensitivity and 95.6% specificity. Wang et al. (2010) proposed a deep semi-supervised learning (SSL) approach, *SRC-MT*, for glaucoma detection and severity grading using retinal fundus images. The method leverages a mean teacher framework with a DenseNet121 backbone and sample relation consistency regularization to utilize both labeled and unlabeled data. *SRC-MT* achieved an average AUC of 0.8944 on global field-of-view (FOV) regions and 0.8969 on local disc regions when trained with 600 labeled and 5,401 unlabeled images from their proprietary EyeW dataset. Performance improved with higher-quality unlabeled images (AUC: 0.8972 for "good" quality vs. Wang et al. (2025) 0.8908 for "usable" quality) and more labeled data (AUC: 0.9270 with full labels). The model showed superior performance on in-distribution data but experienced degradation on out-of-distribution datasets (AUC: 0.7342 on AIROGS, 0.5090 on EDDFS), highlighting challenges with distribution shift. Comparing global FOV to local disc areas and quantitatively assessing the effect of unlabeled data quality were two significant innovations. The dataset size (7,503 photos) and dependence on single-annotator labels were drawbacks, too. The study highlights the significance of data quality and distribution alignment while offering helpful recommendations for SSL deployment in clinical contexts. There was no mention of code availability. Yi et al. (2022) achieved high multiclass classification accuracy (0.975) across glaucoma stages but suffered from limited model transparency, whereas An et al. (2018) reached lower accuracy (87.8) using a small, single-center Japanese dataset with multi-modal OCT/LSFG features, highlighting constraints in generalizability, model depth, and clinical applicability. Zhou et al. (2025) demonstrated high AUC (0.97) and F1-score (95) on a new 2,000-image Multi-Glau dataset, but external validation is still pending, whereas Oh et al. (2021) achieved slightly lower AUC (0.8998) but strong sensitivity (95) and multi-ethnic, multi-center validation on >100k images, providing much greater external validity and clinical generalization. Yousefi et al. (2018)

focused on longitudinal progression detection with a smaller, primarily White cohort, achieving high specificity (95) but without reported accuracy or AUC, and their model relied on linear progression assumptions and minimum VF visits, limiting external generalizability. In contrast, Dixit et al. (2021) used a much larger, single-center dataset with hundreds of thousands of VF and clinical samples, reporting accuracy of 91–93 and AUC 0.89–0.93, but faced class imbalance, excluded severe cases, and still had limited generalizability beyond their center. Li et al. (2019) used a large, publicly available fundus image dataset with attention-based CNNs and interpretable heatmaps, achieving high accuracy (95.3) and AUC (0.967), but performance in multi-center or real-world clinical settings is still untested. In contrast, Maetschke et al. (2019) applied deep learning to a small private OCT dataset (265 scans), achieving slightly lower AUC (0.94) and focusing on clinically meaningful regions via CAM, yet did not assess progression or temporal changes, limiting longitudinal applicability. Kim et al. (2017) achieved very high accuracy (0.98) and AUC (0.979) using feature combinations from RNFL in a relatively small dataset, but it requires further validation on diverse populations to confirm robustness. Kashyap et al. (2022) applied an improved U-Net to 650 fundus images from a Singapore Malay cohort, with potential limitations in detecting subtle or early glaucoma. Wu et al. (2022) performed a large-scale meta-analysis of 197,174 fundus and 16,039 OCT images, showing strong AUC (0.92–0.96), yet did not address practical clinical deployment or workflow integration. Zhang et al. (2021) leveraged a very large, multi-center AIROGS fundus dataset with a transformer-based segmentation/classification model, achieving high sensitivity (95%) but moderate specificity (85.3%) and AUC (0.8998), though prevalence estimates may be imprecise for specific subgroups due to meta-analytic variation. In contrast, An et al. (2019) used a smaller, single-center Japanese cohort with perfect sensitivity (100) and good specificity (78.9%, AUC 0.94) for binary glaucoma detection, but the study's generalizability is limited by sample size, ethnic homogeneity, high myopia prevalence, and lack of longitudinal data. Xiong et al. (2022) developed a multimodal deep learning FusionNet combining visual fields and peripapillary OCT scans, achieving high sensitivity and specificity (95% each) with an AUC of 0.917. However, its performance was limited by a single-ethnicity (Chinese), hospital-based dataset, exclusion of comorbid retinal/optic nerve diseases, reduced accuracy on external testing, and potential variability across OCT devices. Alhadeff et al. (2017) analyzed correlations between fundus photo features and glaucomatous damage on OCT and visual fields in a small cross-sectional cohort (100 glaucoma/suspect eyes, 62 healthy), but the study was limited by sample size, inclusion of early glaucoma cases only, inter-grader variability, and very few disc hemorrhages. In contrast, Yousefi et al. (2018) applied machine learning to detect longitudinal visual field progression in 270 eyes over 9 years, achieving 72% sensitivity but was constrained by a predominantly White cohort, potential confounders, linear progression assumptions, and discrete labeling of progression. Asaoka et al. (2016) applied a deep feedforward neural network to standard automated perimetry for detecting preperimetric glaucoma, achieving 74.9% sensitivity and 79% specificity (AUC 0.926), but was limited by single-center data, no external validation, and exclusion of non-glaucomatous VF loss. In contrast, Martin et al. (2018) used Random Forests on 24-hour contact lens sensor (CLS) parameters for POAG detection, achieving only 61.1% accuracy, constrained by CLS artifacts, inclusion of treated patients, and limited generalizability, highlighting that perimetry-based neural networks outperformed sensor-based ML in early glaucoma detection.

5.6 Explainable artificial intelligence (XAI) in glaucoma detection and classification

To address the growing concern regarding the limited interpretability of AI-based glaucoma models, and in alignment with the emphasis on explainable artificial intelligence (XAI) in this review, we introduce a dedicated subsection summarizing current XAI approaches and how they are applied in glaucoma detection and classification. Although deep learning systems have achieved strong diagnostic performance, their clinical translation remains constrained by limited transparency, making interpretable and clinically aligned explanations essential for building trust and supporting real-world deployment.

Building on the conceptual framework presented in Fig. 11; Table 6 summarizes representative studies that explicitly integrate explainable artificial intelligence (XAI) techniques into glaucoma detection and classification. The table provides a structured overview of how different XAI approaches are implemented in practice, highlighting their interpretability mechanisms and clinical relevance. By consolidating empirical evidence across recent studies, this overview connects theoretical XAI concepts with real-world diagnostic applications, demonstrating how interpretability contributes to trustworthy and clinically meaningful AI systems in ophthalmology.

In summary, the integration of explainable artificial intelligence into glaucoma detection represents a shift toward more clinically trustworthy AI systems. The reviewed studies show that XAI techniques can link model predictions to meaningful ophthalmic structures, improving clinician confidence and interpretability. The diversity of current XAI approaches–ranging from visualization-based methods to quantitative feature analyses–also reflects the maturation of the field.

Importantly, interpretability is becoming an integral component of model design rather than an optional addition. This supports transparent decision-making, facilitates error analysis, and enhances monitoring of disease progression. Together, these advances bring XAI-enabled glaucoma models closer to safe and reliable clinical deployment.

5.7 Glaucoma fundus photography

Table 8 summarizes studies that applied machine learning and deep learning approaches to glaucoma detection and grading using fundus photographs. Lin et al. (2024) developed a system for glaucoma detection and glaucomatous feature classification using the private JustRAIGS dataset, which included over 101,000 fundus images. Their model achieved strong diagnostic performance with a sensitivity of 95%, specificity of 85.7%, and an AUC of 0.995. However, since the dataset was relatively homogeneous and limited to color fundus photographs, the model's generalizability across diverse populations and imaging devices remains uncertain. Wang et al. (2025) proposed a multi-class classification model for glaucoma severity (none, mild, moderate, severe) trained on 7,503 fundus images, of which 503 were labeled and 7,000 were unlabeled. The method achieved an AUC of 0.8944 and a sensitivity of 0.7765. While effective, the approach heavily relied on unlabeled data, raising concerns about data quality and representativeness that may influence model robustness. Bajwa et al. (2019) used a number of datasets, including ORIGA, HRF, DRIONS-DB, and DRIVE, to perform multi-class classification (glaucoma vs. normal vs. diabetic retinopathy). Although their model's main focus was on binary glaucoma vs normal identification, it did reach an AUC of 0.868. There was a significant clinical gap since early or

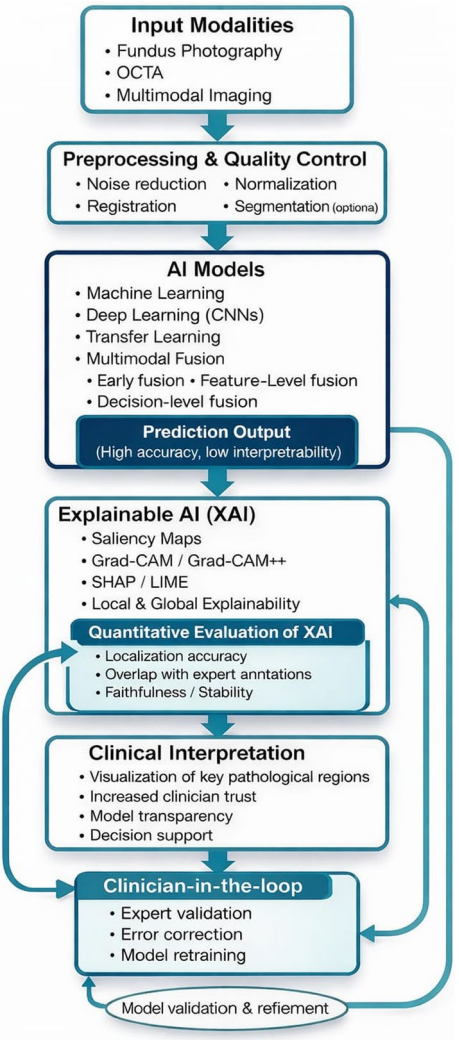


Fig. 11 Schematic overview of an explainable artificial intelligence (XAI)–driven framework for ophthalmic imaging–based disease detection and classification. The diagram illustrates the pipeline from input imaging modalities through AI prediction and integration of XAI techniques, including saliency maps, class activation mapping, and perturbation-based methods such as SHAP and LIME, with quantitative evaluation and clinician-in-the-loop validation

pre-perimetric glaucoma detection was not addressed. Ahn et al. (2018) used 1,542 fundus pictures to create a multi-class model that differentiates between normal, early, and advanced glaucoma. The system's AUC was 0.94 and its total accuracy was 87.9%. The model has trouble detecting glaucoma early, which is essential to stopping the disease's progression, despite its encouraging results.In order to predict glaucoma in 96 patients and 25 controls, Yoon et al. (2021) investigated oral microbiome profiling using machine learning (Random Forest and CPAR rule mining). The model achieved 96% accuracy, but the small Korean-only control group and reliance on 16 S rRNA sequencing limit broader applicability and cannot infer causality. Das et al. (2024) proposed AES-Net, an enhanced self-attention CNN for multi-stage glaucoma classification, tested across several datasets (ACRIMA, RIM-ONE, Drishti-GS1, HRF, SJCHOI86-HRF). The method achieved an accuracy of 92.67%, though dataset details were not fully provided, reducing reproducibility. Das and Nayak

Table 6 Overview of explainable artificial intelligence (XAI) methods used in glaucoma detection and classification

References	XAI method	XAI findings	Clinical relevance
Christopher et al. (2018)	Occlusion testing	Model focused on clinically relevant optic nerve head regions	Increased trust by aligning CNN attention with ophthalmic knowledge
Kim et al. (2017)	Decision tree (C5.0)	Extracted explicit diagnostic rules linking RNFL and visual field features	Provided transparent, clinician-interpretable decision logic
Akter et al. (2023)	Grad-CAM	Activation maps highlighted glaucoma-related OCT regions	Enabled visual verification of model focus areas
Ran et al. (2019)	Heatmap visualization	Highlighted regions matched clinician-recognized pathology	Supported interpretability of deep learning predictions
Koornwinder et al. (2025)	SHAP values	Identified key predictive clinical and imaging features	Quantified feature importance for clinical decision support
Noury et al. (2022)	CAM/Grad-CAM heatmaps	Qualitative identification of novel retinal diagnostic regions	Visual explanation of deep learning model attention patterns
Hasan et al. (2025)	SHAP-based explainable ML	Quantified OCT feature contributions for glaucoma diagnosis and staging	Transparent quantitative interpretation supporting clinical decisions
Son et al. (2023)	Concept-based interpretable DL	Explicit finding–disease relationship modeling	Interactive clinician-interpretable decision support
George et al. (2020)	Attention-based visualization	Highlighted volumetric regions linking structure and function	Improved interpretability of 3D OCT predictions
Huang et al. (2025)	Multimodal interpretable DL	Explained longitudinal visual field predictions from images and narratives	Transparent monitoring of glaucoma progression
Jalili et al. (2025)	Vision-language generative XAI (GPT-4V)	Generated interpretable textual explanations and feature identification	Clinician-readable diagnostic reasoning
Mehta et al. (2021)	Interpretable machine learning	Identified clinically meaningful feature contributions from multimodal data	Transparent decision support for glaucoma diagnosis
Escamez et al. (2021)	Interpretable ML classifier	Provided transparent rule-based explanations for early glaucoma diagnosis	Clinician-interpretable diagnostic decision support
Oh et al. (2021)	SHAP-based explainable machine learning	Quantified feature contributions and visualized diagnostic importance	Transparent clinician-interpretable glaucoma diagnosis

(2024) introduced GS-Net, a global self-attention guided CNN for multi-stage glaucoma classification. Their model achieved an accuracy of 84.91%, but lacked external validation and was limited to multi-class settings, leaving generalizability untested. Akter et al. (2025) developed an explainable system for glaucoma stage classification using visual field functional data and 476 balanced VF plots (early, moderate, advanced). Their model achieved 99.0% specificity and an F1-score of 96.8%. However, the dataset showed limited diversity and imbalance in advanced cases. Mouhafid et al. (2026) performed multi-class glaucoma diagnosis using 1,600 fundus images (Drishti-GS1, RIM-ONE, ACRIMA) with GlauVGG-Net and multi-transform dynamic augmentation. The system achieved 95.2% accuracy, 96% sensitivity, and 94.4% specificity. Despite strong performance, limited external testing constrains its clinical applicability. Zedan et al. (2025) introduced a dual-stage deep learning framework for glaucoma severity classification based on multiscale feature fusion, trained on 800 private OCT scans. Their model achieved 92.8% accuracy, 91.7% sensitivity, an

Table 7 Summary of key studies on glaucoma grading and classification using fundus and OCT imaging, highlighting dataset diversity, applied AI methods, and diagnostic outcomes (accuracy, specificity, sensitivity, AUC, F1-score)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Yi et al. (2022)	87 normal, 171 early, 79 intermediate and 165 terminal	Multi-class (Normal vs Early vs Intermediate vs Terminal)	Accuracy: 94.6%, Sensitivity: –, Specificity: –, AUC: 87.4%, F1-Score: 91.2%	The model's decision-making remained largely non-transparent, limiting clinical trust and adoption
An et al. (2018)	Private: 163 glaucomatous eyes from 105 patients; OCT (48 structural features), LSFG (36 blood-flow features), 7 demographic features; Training: 114 eyes, Test: 49 eyes	Multi-class classification (FI, SS, MY, GE)	Accuracy: 87.8%, Sensitivity: –, Specificity: 78.9%, AUC: 0.926, F1-Score: –	Small sample, single-center Japanese population; requires OCT/LSFG, limiting generalizability; deep learning not feasible; single-hidden-layer NN may underperform deeper models; clinical integration constrained by specialized equipment
Zhou et al. (2025)	2000 fundus images	Multi-Glau dataset	Accuracy: –, Sensitivity: –, Specificity: –, AUC:0.97, F1-Score: 95%	Dataset new, external testing pending
Oh et al. (2021)	Rotterdam EyePACS AIROGS dataset: 113,893 fundus images from 60,357 subjects, 500 sites, multi-ethnic	Referable vs. Non-referable glaucoma	Accuracy: 90.3%, Sensitivity: 95%, Specificity: 85.3%, AUC: 0.8998, F1-Score: –	Very strong – trained/validated on AIROGS (multi-center, multi-ethnic, >100k images), providing high external validity
Yousefi et al. (2018)	Private: Cross-sectional cohort: 2085 eyes (1214 subjects: 939 glaucoma, 1146 normal); Test-retest cohort: 133 eyes (71 glaucoma patients); Longitudinal cohort: 270 eyes (136 glaucoma patients, 17 VF tests over 9 years)	Longitudinal progression detection (Stable vs Progressing Glaucoma)	Accuracy: –, Sensitivity: 87%, Specificity: 95%, AUC: –, F1-Score: –	Cohort primarily White; minimum 5 VF visits required; linear progression assumption; excluded cataracts/secondary glaucoma but may still confound VF; external generalizability not tested
Dixit et al. (2021)	11,242 eyes (5,843 patients) from Johns Hopkins Wilmer Eye Institute; 672,123 VF tests + 350,437 clinical samples (IOP, CCT, CDR)	Longitudinal progression prediction (Stable vs Progressing Glaucoma)	Accuracy: 91-93%, Sensitivity: 74.5-95%, Specificity: 95.6%, AUC: 0.89-.93, F1-Score: –	Single-center, class imbalance (80–90% stable eyes), excluded severe glaucoma (MD < -15 dB), discordant ground truth among reference standards, limited generalizability

Table 7 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Akter et al. (2022)	Private OCT images (255 scans) collected at a single academic medical center (Jordan Univ. Hospital)	Multi-feature analysis combining clinical data (RNFL, CDR, MD, PSD) with OCT imaging; deep learning with Grad-CAM interpretability	Accuracy:98.6%, Sensitivity:98.7%, Specificity:98.5%, AUC:0.99, F1 Score:-	The study used only 255 OCT scans from a single academic center in Jordan, which raises concerns about external validity. Larger, multi-center, and ethnically diverse datasets are needed to confirm robustness
Noury et al. (2022)	Training: 1,564 scans (Stanford); Testing: 3,371 scans (Stanford, India, Nepal, Hong Kong)	3D CNN with cross-national validation; explainable AI via saliency maps; identified lamina cribrosa thinning as novel diagnostic biomarker	Accuracy:- Sensitivity:86%, Specificity:95%, AUC: Stanford 0.91, India 0.94, Nepal 0.87, Hong Kong 0.80; %, AUC: 0.91, F1 Score:Stanford 0.87, India 0.91, Nepal 0.80, Hong Kong 0.76	The study was cross-sectional and did not incorporate longitudinal data (progression detection) or other clinical modalities like visual fields
Hussain et al. (2023)	Private dataset (1,806 OCT images), Aier Eye Hospital Group, China	Lightweight CNN (MobileNetV2) with VGG19 ensemble; generative algorithm–guided preprocessing	Accuracy: 99.3%, Sensitivity: 99.5%, Specificity: 98.8%, AUC: 0.992, F1 Score:-	Study did not evaluate integration into routine practice, including interpretability, workflow compatibility, and clinician acceptance
Thiéry et al. (2023)	OCT 3D point cloud – Yonsei University College of Medicine, Seoul, Korea	Geometric Deep Learning applied to OCT 3D point clouds; interpretable, requires less data than 3D CNNs	Accuracy:- Sensitivity:- Specificity:- AUC: 0.95 ± 0.01, F1 Score:-	Emphasis on geometric modeling may limit applicability to conventional OCT or fundus image analyses without additional preprocessing
Chiang et al. (2024)	Fundus (3,088 images total) from high myopia glaucoma (1,540) and high myopia non-glaucoma (1,548)	Addresses diagnostic challenges in myopic eyes; Grad-CAM interpretable, good performance despite overlapping features	Accuracy:82.16%, sensitivity: 81.04%, specificity: 83.27%, AUC: 0.894, F1 Score: 81.92%	AI outputs may not yet be fully interpretable or seamlessly integrated into routine ophthalmic practice
Kooner et al. (2022)	OCTA – Private (735 subjects; 1,371 eyes) Seoul National University Bundang Hospital, Korea	Integration of OCT/OCTA features with ML models; >2,500 models evaluated (classical + deep learning); nested 5×10-fold cross-validation; externally validated	Accuracy:83.9%, sensitivity: 86-87%, specificity: 84.2%, AUC: -, F1 Score:62.4%	No evaluation of predictive capability for glaucoma progression over time

Table 7 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Li et al. (2019)	Public – LAG (Large-scale Attention-based Glaucoma) dataset 5,824 fundus images 2,392 glaucomatous 3,432 normal Collected from Chinese Medical Institutions with expert annotations	CNN with attention mechanism; first glaucoma dataset with attention region labels; interpretable heatmaps; benchmarked against existing methods	Accuracy: 95.3%, Sensitivity: 95.4%, Specificity: 95.2%, AUC: 0.967, F1 Score:-	Model performance in multi-center or real-world clinical settings remains untested
Maetschke et al. (2019)	Private (Mayo Clinic) (265)	Deep learning directly on unsegmented OCT volumes, avoiding retinal layer segmentation; CAM highlights clinically meaningful ONH subregions	Accuracy:-, sensitivity: -, specificity: -, AUC: 0.94, F1 Score:-	The study does not evaluate glaucoma progression or temporal changes in OCT volumes
Huang et al. (2022)	Private: 2,650 eyes (1,324 healthy, 1,326 glaucoma) From Samsung Medical Center, Korea	Using single RNFL thickness map with high diagnostic accuracy; suitable for clinical implementation	Accuracy: 97.4%, Sensitivity: 97.6%, Specificity: 97.2%, AUC: 0.996, F1 Score:-	While probabilistic outputs provide uncertainty estimates, the model's decision-making process may remain difficult for clinicians to interpret
Chincholi and Koestler (2024)	Public: 2,380 images from OHTS + 1,195 from DIGS/ADAGES (total 3,575 images)	Transformer-based deep learning model focusing on ONH photography; validated across two large public datasets	Accuracy:- Sensitivity: 87–89%, Specificity: 80–82%, AUC: 0.95, F1 Score:-	Transformer-based models are often evaluated on retrospective datasets; real-world prospective validation may be lacking
Kim et al. (2017)	100 cases test, 399 cases for training	Tested multiple combinations of features to best reflect the whole RNFL	Accuracy 98%, sensitivity 98.3%, specificity 97.5%, AUC 0.979, F1 Score:-	Additional validation on diverse, independent datasets is needed to confirm robustness and clinical applicability
Kashyap et al. (2022)	650 fundus images, Singapore malay eye study	Improved U-Net model with minor feature extraction	Accuracy.:98.82%, Sensitivity: 4.05%, Specificity: 0.8-4%, AUC: 0.983%, F1 Score:-	May have reduced sensitivity for detecting early or subtle glaucoma changes
Wu et al. (2022)	197,174 fundus and 16,039 OCT	Meta-analysis of ML models	Accuracy:-, sensitivity: -, specificity: -, AUC: 0.92–0.96, F1 Score:-	Does not directly address real-world deployment challenges or clinical workflow integration

Table 7 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Zhang et al. (2021)	AIROGS dataset: 113,893 fundus images (60,357 subjects) from 500 sites	Segmentation + transformer-based classification (Vdisk/Vcenter dual model); quality control for ungradable images	Accuracy:-, sensitivity:95%, specificity: 85.3%, AUC:.8998, F1 Score:-	Meta-analysis combines studies with varying ethnicities, age ranges, and diagnostic criteria, which may limit the precision of prevalence estimates for specific subgroups
An et al. (2019)	Private: 208 glaucomatous eyes (including 49 preperimetric), 149 healthy eyes; single-center Japanese cohort	Binary classification	Accuracy:-, sensitivity: 100%, specificity: 78.9%, AUC: 0.940, F1 Score:-	Single-center, Japanese cohort; higher myopia in glaucoma group; small sample size; cross-sectional design; no longitudinal progression data
Escamez et al. (2021)	UK Biobank: 863 healthy subjects (1,193 eyes), 771 glaucoma subjects (1,283 eyes), 55 progress-to-glaucoma subjects (98 eyes); cross-sectional, multimodal	Binary classification	Accuracy:75%, sensitivity: -, specificity: -, AUC: 0.95, F1 Score:-	Self-reported glaucoma diagnoses; limited CFP quality; OCT motion artifacts; healthy volunteer bias; exclusion of comorbid conditions; generalizability to real-world populations uncertain
Li et al. (2023)	Asian Dataset: 257 glaucoma patients (307 eyes) + 257 controls (307 eyes); Test: 183 glaucoma (206 eyes) + 173 controls (206 eyes); Caucasian Dataset: 57 glaucoma (87 eyes) + 81 controls (145 eyes); multi-ethnic cohorts, OCT-derived RNFL features	Binary classification	Accuracy:92%, sensitivity: 95%, specificity: 95%, AUC: 0.96, F1 Score:-	Small Caucasian sample; limited ethnic diversity; reliant on clinician-defined glaucoma; not tested on other ethnicities or OCT devices; external sharing restrictions
Thakur and Juneja (2020)	260 retinal fundus images (DRISHTI-GS: 101, RIM-One: 159) with optic disc/cup segmentation and CDR values	Binary classification	Accuracy:97.2%, sensitivity: 95%, specificity: 88.23%, AUC:-, F1 Score:-	Small dataset (260 images); requires optic disc/cup segmentation; traditional ML only (no deep learning); performance sensitive to image quality; computational cost of feature extraction and selection

Table 7 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Escamez et al. (2021)	175 eyes (90 early glaucoma, 85 healthy controls) with peripapillary RNFL thickness from spectral-domain OCT	Binary classification	Accuracy:89%, sensitivity: 89%, specificity: 88.5%, AUC: 0.953, F1 Score:-	Only early glaucoma cases (MD > -6 dB), Caucasian participants, excluded low-quality OCT scans; dependency on segmentation accuracy; weighted model trades sensitivity for specificity
Medeiros et al. (2019)	32,820 pairs of fundus photos + SD-OCT scans from 2,312 eyes (1,198 participants, Duke Glaucoma Repository)	SD-OCT-trained deep learning for RNFL thickness prediction	Accuracy:83.7%, sensitivity: 95%, specificity: 76%, AUC: 0.944, F1 Score:-	Retrospective, single-center; trained only on global RNFL thickness; no longitudinal validation; image quality not formally assessed; 30% RNFL variance unexplained
Xiong et al. (2022)	2,463 VF-OCT pairs from 1,083 patients (ZOC China training/ validation/ internal test; First Hospital Shijiazhuang external test)	FusionNet: Multimodal DL using visual fields + peripapillary circular OCT scans	Accuracy:-, sensitivity: 95%, specificity: 95%, AUC: 0.917, F1 Score:-	Single ethnicity (Chinese), hospital-based; excluded comorbid retinal/optic nerve diseases; performance dropped on external test; underpowered subgroup analyses; technical/device variability across OCT devices
Alhadeff et al. (2017)	100 glaucoma/ suspect eyes, 62 healthy eyes	Correlation of fundus photo features with glaucomatous damage detected on OCT (RGC+ and RNFL) and visual fields	Accuracy:-, sensitivity: -, specificity: -, AUC: -, Fl Score:-	Limited diversity, imbalance in advanced case
Yousefi et al. (2018)	270 eyes (136 glaucoma patients, longitudinal 17 visits over 9 years)	Detection of longitudinal visual field progression in glaucoma using ML	Accuracy:-, sensitivity: 72%, specificity: 95%, AUC: -, F1 Score:-	Primarily white cohort; cataract/ other confounders; linear trend assumption; discrete labels; progression only
Christopher et al. (2018)	179 glaucoma eyes (93 patients), 56 healthy eyes (28 participants)	RNFL features identified by unsupervised ML on OCT predict glaucoma progression	Accuracy:-, sensitivity: 88%, specificity: 95%, AUC: 0.94, F1 Score:-	Short follow-up (2 yrs); glaucoma group older; high glaucoma:healthy ratio; only RNFL layers
Aljohani and Aburasain (2024)	2,775 fundus images from ACRIMA, G1020, ORIGA, REFUGE datasets	Hybrid framework using federated ML and deep learning (CNN + RF + GLCM features) for glaucoma detection	Accuracy: 95.41%, sensitivity: 95.6%, specificity: 92%, AUC: -, F1: 93.52	Grayscale conversion may lose RGB info; computational cost; limited dataset diversity; some false negatives

Table 7 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Asaoka et al. (2016)	171 VFs from 53 PPG eyes, 108 VFs from 87 healthy eyes	Deep feedforward neural network (FNN) for detecting preperimetric glaucoma from standard automated perimetry	Accuracy:-, sensitivity: 74.9%, specificity: 79%, AUC: 0.92.6, F1 Score:-	Single-center, no external validation, excluded non-glaucomatous VF loss, synthetic oversampling (SMOTE)
Martin et al. (2018)	136 healthy eyes, 136 POAG eyes (age-matched)	Machine learning (Random Forest) using 24-hour contact lens sensor (CLS) parameters ± IOP for POAG detection	Accuracy:61.1%, sensitivity: -, specificity: -, AUC: 0.681, F1 Score:-	CLS alone modest AUC; treated POAG patients; CLS artifacts; limited to POAG; black-box associations

The table also reflects current research directions in AI-based glaucoma staging and classification

AUC of 0.97, and an F1-score of 0.9296. Nevertheless, reproducibility and generalization remain limited due to dataset constraints. Thanapaisal et al. (2025) applied machine learning to classify glaucoma severity using 6,671 fundus images following the Hodapp–Parrish–Anderson staging system. Their approach achieved an accuracy of 96.6%, sensitivity of 95.6%, specificity of 97.7%, and an AUC of 0.985, though device diversity was limited, restricting clinical scalability. Hemelings et al. (2021) used 23,930 fundus images from Xuzhou Medical University to address dual challenges including VCDR prediction and glaucoma versus healthy categorization. With an AUC of 0.94 and sensitivity of 98.0% and specificity of 91.0%, the model's predictive value is diminished by its lack of real-world validation. Xiao et al. (2021) examined a related ophthalmic task, concentrating on predicting surgical outcomes and classifying idiopathic macular holes (IMH) vs secondary macular holes (SMH), using a dataset of 1,082 preoperative OCT scans from four Chinese hospitals. Although the study focused on the diagnosis and prognosis of macular holes rather than glaucoma, the results are not immediately applicable to disorders of the optic nerve. Nevertheless, their approach obtained 95% accuracy and an AUC of 0.965. With outside validation on the PKU dataset, citeHasan2025 examined 1,598 ZJU patient records, including CFP, VF, IOP, central corneal thickness (CCT), and clinical notes. Although they claimed to have deployed and tested the technology in the actual world, their objective was to forecast longitudinal VF progression. In order to estimate retinal nerve fiber layer (RNFL) thickness, Yang et al. (2022) investigated the utilization of 557 fundus pictures from 303 glaucoma patients, obtaining sensitivity of 92.0% and specificity of 86.9%. However, Real-world integration and prospective validation in diverse clinical settings are likely lacking. Saha et al. (2023) trained a deep learning model on a private dataset of 6,671 fundus images, incorporating attention mechanisms to enhance interpretability. Their automated pipeline achieved an accuracy of 96.6%, sensitivity of 95.6%, specificity of 97.7%, and an AUC of 0.985. However, Relied solely on color fundus photographs, excluding OCT, perimetry, or clinical data that could improve diagnostic robustness Son et al. (2023) proposed a lightweight deep learning model combined with SHAP-based interpretability, using 2,400 fundus images collected at GMKMCH, Salem, India. The system achieved strong diagnostic performance

Table 8 Summary of key studies on glaucoma detection and severity classification using fundus photography, highlighting dataset characteristics, applied AI methodologies, diagnostic performance metrics (accuracy, sensitivity, specificity, AUC, F1-score), and research gaps that point to challenges in early-stage detection and classification

Study, Year	Dataset (images)	Methodology	Results	Research gap
Lin et al. (2024)	Private: JustRAIGS dataset (101,423 images) Training: 2,616 referable glaucoma + 4,000 NRG (non-referable)	Glaucoma detection and glaucomatous feature classification	Accuracy:-, Sensitivity: 95%, Specificity: 85.7%, AUC: 0.995, F1-Score:-	Likely trained and validated on a relatively small or homogeneous dataset of color fundus photographs, which may limit generalizability across populations and imaging devices
Wang et al. (2025)	7,503 total fundus images: 503 labeled from Wenzhou Eye Hospital (WMU-GS), 7,000 unlabeled from EyeQ dataset	Multi-class glaucoma severity (none, mild, moderate, severe)	Accuracy: –,Sensitivity: 0.7765,Specificity: –,AUC: 0.8944,F1-Score: –	The method relies on unlabeled data, but the quality and representativeness of that data may significantly impact performance– this risk isn't fully resolved
Bajwa et al. (2019)	ORIGA 482 healthy images and 168 glaucomatous images. HRF 15 health, 15 glaucoma, 15 DR. 50 healthy eyes collected from Ophthalmology Department of Feiz Hospital, Isfahan, Iran. The Drive dataset, DRIONS-DB, contains 110 images	multi-class classification (Glaucoma vs Normal vs DR)	Accuracy: 91.43%,Sensitivity:71.17%, Specificity: 85%,AUC: 0.874,F1-Score: 0.4616	The system focuses more on binary glaucoma vs. normal classification. Detecting early or pre-perimetric glaucoma remains a gap
Ahn et al. (2018)	786 normal, 467 advanced glaucoma, and 289 early glaucoma fundus images	Multi-class (Normal vs Early vs Advanced Glaucoma)	Accuracy: 87.9%,Sensitivity: 98.3%,Specificity:97.5%,AUC: 0.979,F1-Score: –	The model struggled with early glaucoma detection, which is clinically the most critical for preventing progression

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Yoon et al. (2021)	96 glaucoma patients (62 POAG, 34 SG) and 25 healthy controls	Oral microbiome profiling with ML (Random Forest, CPAR rule mining) to predict glaucoma	Accuracy: 96% Sensitivity: –Specificity: –AUC: –F1-Score: –	Small control group; Korean-only population; 16 S rRNA sequencing; cannot infer causality
Das et al. (2024)	enhanced self-attention CNN for multi-stage glaucoma classification ACRIMA, RIM-ONE, Drishti-GS1, HRF, SJCHOI86-HRF (counts not detailed)	AES-Net: Adapter	Accuracy: 86.20% Sensitivity: –Specificity: –AUC: 0.9493,F1-Score:85.47%	Dataset details not fully provided
Das and Nayak (2024)	Fundus datasets (not fully specified)	GS-Net: Global Self-Attention Guided CNN for multi-stage glaucoma classification	Accuracy: 84.91%, Sensitivity: –,Specificity: –,AUC: 0.9454,F1-Score: 84.55%	No external validation and Multi-class
Akter et al. (2025)	Excellent stage classification; VF functional data with explainability	476 balanced VF plots (72 early, 43 moderate, 21 advanced, severe merged)	Accuracy: –Sensitivity: –Specificity: 99.0% AUC: –F1-Score: 96.8%	Limited diversity, imbalance in advanced cases
Mouhafid et al. (2026)	1600 fundus images (Drishti-GS1, RIM-ONE, ACRIMA)	Multi-class glaucoma diagnosis using GlauVGG-Net with multi-transform dynamic augmentation	Accuracy: 95.2%,Sensitivity: 96%,Specificity: 94.4%,AUC: –,F1-Score: –	Limited external testing
Zedan et al. (2025)	800 OCT scans (private)	Dual-Stage Deep-Learning Method for Glaucoma Severity Classification based on Multiscale Feature Fusion	Accuracy: 92.822%,Sensitivity: 0.9174,Specificity: –,AUC: 0.97,F1-Score: 0.9296	Limited reproducibility
Thanapaisal et al. (2025)	6671 fundus images (Hodapp–Parrish–Anderson staging)	Machine learning technology in the classification of glaucoma severity using fundus photographs	Accuracy:96.6%,Sensitivity: 95.6%,Specificity: 97.7%,AUC:.985,F1-Score: –	Lacks device diversity

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Hemelings et al. (2021)	Dual tasks: (1) Glaucoma vs Healthy (classification), (2) VCDR prediction (regression)	Private – Affiliated Hospital of Xuzhou Medical University (23,930 fundus images)	Accuracy: –, Sensitivity: 98.0%, Specificity: 91.0%, AUC: 0.94, F1-Score: –	The approach was not evaluated in real-world clinical settings, limiting evidence for practical deployment
Xiao et al. (2021)	Private dataset – 1,082 pre-op OCT scans, 330 eyes from 315 patients; collected across 4 hospitals (ZOC, Zhujiang Hosp., Kunming Univ. Hosp., Guangdong Provincial Hosp.), China	(1) IMH vs SMH classification (aetiology), (2) Post-op outcome prediction (Closed vs Open MH, IMH status)	Accuracy: 95%, Sensitivity: 93.8%, Specificity: 87%, AUC: 0.965, F1-Score: –	The study addressed macular hole diagnosis and outcomes, not glaucoma, meaning findings are not directly transferable to optic nerve diseases
Huang et al. (2025)	1,598 patient records from ZJU (CFP + VF + IOP + CCT + notes); external test on PKU dataset	VF progression prediction (longitudinal)	Accuracy: –, Sensitivity: –, Specificity: –, AUC: –, F1-Score: –	Prospective, real-world deployment and testing of the system were not reported
Yang et al. (2022)	557 eyes from 303 patients with glaucoma, fundus photographs	Predicting RNFL thickness in glaucoma (screening approach)	Accuracy: –, Sensitivity: 92.0%, Specificity: 86.9%, AUC: –, F1-Score: –	Real-world integration and prospective validation in diverse clinical settings are likely lacking
Saha et al. (2023)	Private: 6,671 fundus images	Deep learning model trained on color fundus photographs; attention modules used for interpretability; automated pipeline for glaucoma screening	Accuracy: 96.6%; Sensitivity: 95.6%; Specificity: 97.7%; AUC: 0.985; F1 score:-	Relied solely on color fundus photographs, excluding OCT, perimetry, or clinical data that could improve diagnostic robustness
Son et al. (2023)	Private dataset (2,400 fundus images), collected at Government Mohan Kumaramangalam Medical College & Hospital (GMKMCH), Salem, India	Lightweight deep learning model with SHAP for interpretability; interactive diagnosis system	Accuracy:- Sensitivity: 92.5–100%, Specificity: 88.3–100%, AUC: 0.980, F1-score:-	Absence of large-scale, device-independent, prospective clinical validation limits real-world applicability

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Li et al. (2019)	Public – LAG (Large-scale Attention-based Glaucoma) dataset 5,824 fundus images 2,392 glaucomatous 3,432 normal Collected from Chinese Medical Institutions with expert annotations	CNN with attention mechanism; first glaucoma dataset with attention region labels; interpretable heatmaps; benchmarked against existing methods	Accuracy: 95.3%, Sensitivity: 95.4%, Specificity: 95.2%, AUC: 0.967, F1 Score:-	Model performance in multi-center or real-world clinical settings remains untested
Shoukat et al. (2023)	Public fundus G1020, RIM-ONE, ORIGA, and DRISHTI-GS datasets	trained across multiple public cohorts; uses ResNet-50 with grayscale channel + augmentation	Accuracy: 98.48%, Sensitivity: 99.30%, Specificity: 96.52%, AUC: 97.0%, F1 Score: 98%	The model may not have been tested on independent or multi-center datasets to confirm robustness
Hemelings et al. (2023)	149,455 fundus images from 13 sources (BMES, GHS, AIROGS, ORIGA, REFUGE1/2, LAG, ODIR, GAMMA, RIM-ONE, ACRIMA, PAPILA)	Regression DL model for glaucoma screening; standardized preprocessing; externally validated across multiple cohorts	Accuracy:- Sensitivities: 87.3%, specificity: 0.976, AUC: 0.984, F1 Score:-	Prospective, multi-center clinical validation may be insufficient
Sharma et al. (2025)	Internal: Multi-center dataset (N unspecified) External validation: AIROGS and JustRAIGS	Hybrid AI: combines segmentation (optic disc/cup/hemorrhages) with classification; lightweight model (110 MB); externally validated	Accuracy.:- Sensitivity: 93.5% - 95%, specificity:- AUC: 0.983, F1 Score:-	Multi-model approach may reduce clinical interpretability despite improved performance
Pascal et al. (2022)	Total: 9,034 fundus photos Training: 3,655 Internal test: 1,566 External (SEED): 3,813	Multi-task CNN framework trained to classify referable vs non-referable glaucoma using CFPs	Accuracy.:- Sensitivity: 84.5%, Specificity: 95.6%, AUC: 0.996, F1 Score:-	Primarily designed for detection, not for monitoring disease progression over time

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Christopher et al. (2018)	The total dataset consisted of 7,411 stereo pairs split into 14,822 individual ONH images collected from 4,363 eyes of 2,329 individuals	Using multiple CNNs (VGG16, Inception) and comparing with and without transfer learning for ONH-based GON detection	Accuracy:.- Sensitivity: 84%, Specificity: 83%, AUC:.83, F1 Score:-	Primarily evaluates detection of existing glaucomatous optic neuropathy, with limited insights into early-onset cases or disease progression
Lee et al. (2021)	1072 eyes, 827 glaucoma-suspect	Deep learning model predicted RNFL thickness from CFPs to track structural changes and potential for development	Accuracy.:96%, Sensitivity: - Specificity: - AUC: -, F1 Score:-	Likely based on images from a limited clinical population, potentially affecting generalizability across ethnicities and imaging systems
Diaz-Pinto et al. (2019)	Existing databases five public	CNNs with fine-tuning and deep tuning strategies	Accuracy.:- specificity: 0.8580, sensitivity: 0.9346, AUC: 0.9605, F1 Score:-	The model may be less sensitive to very early or subtle glaucomatous changes
Liu et al. (2019)	retinal fundus images obtained from the Chinese Glaucoma Study Alliance, the Handan Eye Study, and online databases. 241 images were selected 1364 fundus photos with glaucomatous indications and 1768 fundus photographs without glaucomatous features	Improved U-Net model with minor feature extraction	Accuracy:- Sensitivity: 96.2%, Specificity: 97.7% AUC: 0.996, F1 Score:-	Does not address glaucoma progression over time or provide predictive monitoring
Shibata et al. (2018)	Deep residual learning algorithm	Improved U-Net model with minor feature extraction	Accuracy.:94.2-96%, Sensitivity: - Specificity: 95%, AUC: -, F1 Score:-	Deep residual learning models may offer limited clinical explainability for ophthalmologists

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Medeiros et al. (2021)	334,466 pairs of color fundus photographs and OCT	Deep learning system trained to detect progressive optic nerve damage	Accuracy.:- Sensitivity: 97%, Specificity: 80%, AUC: 0.996, F1 Score:-	Deep learning predictions may lack detailed explanations for clinicians, limiting transparency in decision-making
Raghavendra et al. (2018)	589 normal and 837 glaucoma	Deep convolution neural network using digital fundus images	Accuracy: 98.13%, sensitivity: 98%, specificity: 98.3%, AUC:-, F1 Score:-	Likely limited to specific clinical centers or a small number of fundus images, reducing generalizability across ethnicities and imaging systems
Abbas (2017)	Combination of four databases(public and private), 600 normal and 600 glaucoma, fundus images	Multimodal deep learning system with dual feature extraction before classification	accuracy: 99%, sensitivity: 84.5%, specificity: 98.01%, AUC:-, F1 Score:-	Early deep learning approaches often provide limited insight into decision-making, which may hinder clinical adoption
Li et al. (2018)	48116 fundus photographs	Deep learning system trained and validated using ophthalmologist-labeled fundus photos	Accuracy:-, sensitivity: 95.6%, specificity: 92%, AUC: 0.986, F1 Score:-	The system primarily detects existing glaucoma but does not address prediction of disease progression
Rogers et al. (2019)	94 scanned stereoscopic photographic slides	AI system compared directly with ophthalmologists' assessments	Accuracy:83.4%, sensitivity: 80.9%, specificity: 86.2%, AUC: - F1 Score:-	While automated, the system may require high-quality stereoscopic images, limiting applicability in routine clinical settings with variable image quality

Table 8 (continued)

Study, Year	Dataset (images)	Methodology	Results	Research gap
Almazroa et al. (2017)	RIGA dataset: 750 color fundus images from 3 sources (MESSIDOR: 460, Bin Rushed: 195, Magrabi: 95); Training: 200 images; Testing: 550 images	HCDR vs VCDR	Accuracy: 74.2% sensitivity: -, specificity: -, AUC: -, F1 Score:-	Small dataset, heterogeneous image quality, segmentation errors affect accuracy, limited external validation, reliance on manual annotations reduces generalizability
Shah et al. (2019)	Drishti-GS1: 50 train, 51 test; RIM-ONE v3: 100 train, 59 test; preprocessing: CLAHE, augmentations; 80% train, 20% validation	High CDR vs Low CDR (indirect glaucoma screening)	Accuracy:-, sensitivity: -, specificity: -, AUC: -, F1 Score:0.96	Small dataset size, cup segmentation challenging due to subtle boundaries, no direct glaucoma classification, RIM-ONE v3 train/test split not fixed, clinical thresholds for glaucoma not defined, limited external validation
Shinde (2021)	666 retinal fundus images (RIM-ONE:169, DRISHTI-GS:101, DRIONS-DB:110, JSIEC:124, DRIVE:44, Private NIO:118); augmented to 2,264 images	Binary classification	Accuracy: 100%, Sensitivity:100%, Specificity: 100%, AUC: -, F1 Score:-	Requires labeled segmentation masks for U-Net; PPA may affect segmentation; computational cost; excluded additional biomarkers (e.g., RNFL thickness); potential overfitting due to augmentation
Civit-Masot et al. (2020)	252 retinal fundus images (RIM-One V3 + DRISHTI-GS)	Dual ML system combining disc/cup segmentation (U-Net) and classification (MobileNet-V2) for glaucoma detection	Accuracy:-, sensitivity: 82%, specificity: 88%, AUC: 0.91−0.93, F1 Score:-	Small dataset (252 images), dependency on segmentation quality, elliptical post-processing assumption, sensitivity-specificity trade-off

(sensitivity: 92.5–100%, specificity: 88.3–100%, AUC: 0.980), and was designed for interactive diagnosis. Despite Absence of large-scale, device-independent,prospective clinical validation limits real-world applicability. Li et al. (2019) introduced the Large-scale Attention-based Glaucoma (LAG) dataset, comprising 5,824 annotated fundus images (2,392 glaucomatous and 3,432 normal) collected from Chinese medical institutions. Their CNN with attention mechanisms provided interpretable heatmaps and achieved an accuracy of 95.3%, sensitivity of 95.4%, specificity of 95.2%, and an AUC of 0.967. While Model performance in multi-center or real-world clinical settings remains untested. Shoukat et al. (2023) leveraged multiple public datasets (G1020, RIM-ONE, ORIGA, DRISHTI-GS) to train a ResNet-50 model with grayscale augmentation. Their framework reported accuracy of 98.48%, sensitivity of 99.30%, specificity of 96.52%, AUC of 0.97, and an F1-score of 98%. Nevertheless, The model may not have been tested on independent or multi-center datasets to confirm robustness. Hemelings et al. (2023) developed a regression-based deep learning model for glaucoma screening using a large-scale, multi-source dataset of 149,455 fundus images pooled from 13 cohorts, including BMES, GHS, AIROGS, ORIGA, REFUGE1/2, LAG, ODIR, GAMMA, RIM-ONE, ACRIMA, and PAPILA. The model achieved sensitivity of 87.3%, specificity of 97.6%, and an AUC of 0.984. Although Prospective, multi-center clinical validation may be insufficient. Sharma et al. (2025) developed a hybrid AI framework that integrates segmentation of optic disc, cup, and hemorrhages with classification, resulting in a lightweight model of 110 MB. Trained on a multi-center dataset (N unspecified) and externally validated on AIROGS and JustRAIGS, the system achieved sensitivity between 93.5–95% with an AUC of 0.983. Multi-model approach may reduce clinical interpretability despite improved performance. Pascal et al. (2022) proposed a multi-task CNN framework trained on 9,034 fundus photographs (3,655 for training, 1,566 for internal testing, and 3,813 from the SEED dataset for external validation). The model focused on classifying referable versus non-referable glaucoma, achieving sensitivity of 84.5%, specificity of 95.6%, and an AUC of 0.996. Primarily designed for detection, not for monitoring disease progression over time. Christopher et al. (2018) evaluated 14,822 individual ONH images derived from 7,411 stereo pairs collected from 4,363 eyes across 2,329 individuals. Multiple CNNs, including VGG16 and Inception, were compared with and without transfer learning for GON detection, yielding an AUC of 0.83. Primarily evaluates detection of existing glaucomatous optic neuropathy, with limited insights into early-onset cases or disease progression. Lee et al. (2021) investigated 1,072 eyes (827 glaucoma-suspect) to predict RNFL thickness from CFPs using deep learning, aiming to track structural changes and identify risk for disease development. Although specific performance metrics were not provided, this work highlighted potential for non-invasive monitoring. Likely based on images from a limited clinical population, potentially affecting generalizability across ethnicities and imaging systems. Diaz-Pinto et al. (2019) employed CNNs with fine-tuning and deep-tuning strategies on five existing public fundus databases. Their approach reported sensitivity of 93.46%, specificity of 85.80%, and an AUC of 0.9605. The model may be less sensitive to very early or subtle glaucomatous changes. Liu et al. (2019) analyzed retinal fundus images from the Chinese Glaucoma Study Alliance, the Handan Eye Study, and online databases, ultimately selecting 241 images for model development. Their improved U-Net model achieved sensitivity of 96.2%, specificity of 97.7%, and an AUC of 0.996. Despite these strong results, it does not address glaucoma progression over time or provide predictive monitoring. Shibata et al. (2018) applied a deep residual learning algo-

rithm using fundus images, introducing improvements over U-Net–based feature extraction. However, deep residual learning models may offer limited clinical explainability for ophthalmologists. Medeiros et al. (2021) leveraged 334,466 pairs of color fundus photographs and OCT scans to train a deep learning system for detecting progressive optic nerve damage. Their model achieved sensitivity of 97%, specificity of 80%, and an AUC of 0.996. Though highly accurate, deep learning predictions may lack detailed explanations for clinicians, limiting transparency in decision-making. Raghavendra et al. (2018) developed a deep convolutional neural network trained on 589 normal and 837 glaucomatous fundus images. The model achieved accuracy of 98.13%, sensitivity of 98%, and specificity of 98.3%. Nevertheless, it was likely limited to specific clinical centers or a small number of fundus images, reducing generalizability across ethnicities and imaging systems. Abbas (2017) introduced a multimodal deep learning system that integrated four public and private databases (600 normal and 600 glaucoma fundus images) with dual feature extraction before classification. The model reported accuracy of 99%, sensitivity of 84.5%, and specificity of 98.01%. However, early deep learning approaches often provide limited insight into decision-making, which may hinder clinical adoption. Li et al. (2018) trained a deep learning system on 48,116 ophthalmologist-labeled fundus photographs. The model achieved sensitivity of 95.6%, specificity of 92%, and an AUC of 0.986. However, The system primarily detects existing glaucoma but does not address prediction of disease progression. Rogers et al. (2019) analyzed 94 scanned stereoscopic photographic slides to compare AI performance directly with ophthalmologists. The system achieved an accuracy of 83.4%, though sensitivity and specificity were not reported. While automated, the system may require high-quality stereoscopic images, limiting applicability in routine clinical settings with variable image quality. Almazroa et al. (2017) investigated the RIGA dataset, comprising 750 color fundus images from three sources (MESSIDOR, Bin Rushed, Magrabi). Their model, trained on 200 images and tested on 550, reported an accuracy of 74.2% for HCDR vs VCDR analysis. Though Small dataset, heterogeneous image quality, segmentation errors affect accuracy, limited external validation, reliance on manual annotations reduces generalizability. Shah et al. (2019) employed Drishti-GS1 (50 training, 51 test) and RIM-ONE v3 (100 training, 59 test) datasets, using CLAHE preprocessing and data augmentation. Their approach focused on high versus low cup-to-disc ratio (CDR) as an indirect glaucoma screening method, achieving an F1-score of 0.96.Though Small dataset size, cup segmentation challenging due to subtle boundaries, no direct glaucoma classification, RIM-ONE v3 train/test split not fixed, clinical thresholds for glaucoma not defined, limited external validation. Shinde (2021) applied U-Net to 666 fundus images (augmented to 2,264) from multiple sources (RIM-ONE, DRISHTI-GS, DRIONS-DB, JSIEC, DRIVE, and private NIO). The model reported perfect performance (accuracy, sensitivity, and specificity of 100%), though likely influenced by augmentation. Requires labeled segmentation masks for U-Net; PPA may affect segmentation; computational cost; excluded additional biomarkers (e.g., RNFL thickness); potential overfitting due to augmentation. Civit-Masot et al. (2020) designed a dual ML system combining U-Net for disc/cup segmentation with MobileNet-V2 for glaucoma classification, trained on 252 images (RIM-ONE V3 and DRISHTI-GS). The model achieved sensitivity of 82%, specificity of 88%, and an AUC of 0.91–0.93. Though promising, this study relied on a small dataset (252 images), with dependency on segmentation quality, an elliptical post-processing assumption, and a sensitivity–specificity trade-off.

6 Glaucoma detection and classification approaches

Early detection of glaucoma is critical for preventing irreversible vision loss; however, the disease often progresses asymptomatically until substantial structural and functional damage has occurred. In the literature, AI-based approaches for glaucoma imaging can be broadly categorized into two main tasks: (i) binary detection, which aims to distinguish glaucomatous from healthy eyes, and (ii) multi-class classification and staging, which focuses on grading disease severity across different stages. This section reviews and organizes existing methods according to these task formulations, while highlighting the imaging modalities, datasets, and learning paradigms used to address each problem.

6.1 Binary glaucoma detection

A comprehensive clinical examination by an ophthalmologist remains the gold standard for glaucoma diagnosis. This typically includes assessment of intraocular pressure, visual field testing, and detailed evaluation of optic nerve head morphology through fundus photography, OCT and OCTA imaging. Recent advances in artificial intelligence have shown promising results in augmenting these diagnostic methods, with several studies demonstrating high accuracy in glaucoma detection using OCT and fundus images.

Table 9 illustrates the types of imaging, clinical datasets, and summary information that highlight the diversity and scope of publicly available resources for glaucoma research. These datasets vary in imaging modalities, including fundus photography, OCT and visual field tests, and provide a range of clinical annotations such as optic disc/cup segmentation, RNFL thickness maps and severity grading.

Examples of publicly available binary datasets include: - GAMMA dataset (Fundus + OCT, multi-class grading available) - REFUGE challenge dataset (Binary classification) - Drishti-GS, RIGA, G1020, LAG, AIROGS, ACRIMA datasets (Binary classification) - Multichannel Glaucoma Benchmark Dataset, Glaucoma Screening Dataset, GHTnet (Binary + multi-class) - Additional datasets from literature as cited Huang et al. (2023); Kumar et al. (2023); Kovalyk et al. (2022); Almazroa et al. (2018); Jalili et al. (2025); Bajwa et al. (2020); Raja et al. (2020); Milad et al. (2025); Robinson et al. (2018); Chavan and Choubey (2025); Kiefer (2023); Hai (2024); Lim et al. (2022).

Multimodal imaging approaches that integrate OCT, OCTA, and fundus photography provide complementary structural and vascular information, thereby improving sensitivity and specificity of glaucoma detection compared to single-modality strategies. The structure-function relationship is improved by artificial intelligence (AI) techniques, specifically residual deep learning, attention-based CNNs, and transformer models, which capture intricate irregular patterns. Structural biomarkers, like RNFL thickness, are valuable, but they only partially correlate with functional outcomes, like visual field deficits.

6.2 Multi-class classification, staging, and explainability

In addition to binary detection, AI models have been applied to multi-class classification for glaucoma staging (early, moderate, advanced). Challenges such as reliance on small or single-center datasets remain, highlighting the importance of federated learning, transfer learning, and data augmentation strategies to improve generalizability across populations.

Table 9 Comprehensive summary of publicly available datasets for glaucoma detection and classification

References, Year	Imaging type	Clinical dataset	Number of images	Resource link
Huang et al. (2023)	Fundus and OCT- Multi-class	GAMMA (Glaucoma grAading Multi-Modality dAtaset) (Normal, Early, Moderate, Advanced)	2,262 OCT B-scans, 4,282 fundus images	https://doi.org/10.6084/m9.figshare.c.6406319.v1
Kumar et al. (2023)	Fundus	REFUGE (Retinal Fundus Glaucoma Challenge) (Healthy vs. Glaucoma)	1,200 fundus images	https://doi.org/10.6084/m9.figshare.20123135.v2
Kovalyk et al. (2022)	Fundus	Drishti-GS (Healthy vs. Glaucoma)	101 fundus images	https://doi.org/10.6084/m9.figshare.14798004.v1
Almazroa et al. (2018)	Fundus	RIGA (Retinal fundus Images for Glaucoma Analysis) (Healthy vs. Glaucoma)	750 fundus images	https://deepblue.lib.umich.edu/data/concern/data_sets/3b591905z
Jalili et al. (2025)	Fundus	ACRIMA, RIM-ONE, Drishti-GS, REFUGE, G1020 (Mostly binary datasets combined; some extended for multi-class)	4,716 total fundus images (across datasets)	https://doi.org/10.6084/m9.figshare.7613135.v1 - https://medimrg.webs.ull.es/
Bajwa et al. (2020)	Fundus	G1020 (Healthy vs. Glaucoma)	1,020 fundus images	https://www.kaggle.com/datasets
Raja et al. (2020)	Fundus	LAG (Labelled fundus images for Glaucoma) (Healthy vs. Glaucoma)	4,503 fundus images	https://data.mendeley.com/datasets/2rnnz5nz74/2
Milad et al. (2025)	Fundus	Private Clinical Dataset (Visual field staging: early, moderate, advanced)	40,000+ VF tests + matched disc photos	https://airogs.grand-challenge.org/?utm_source=chatgpt.com
Robinson et al. (2018)	Fundus and Visual Field	ACRIMA (Healthy vs. Glaucoma)	705 fundus images	https://figshare.com/s/c2d31f850af14c5b5232
Chavan and Choubey (2025)	Fundus	REFUGE (Binary)	1,200 fundus images	https://figshare.com/articles/figure/Refuge/26049574?utm_source=chatgpt.com%26file=47095942
Kiefer (2023)	Fundus + OCT	(Mixed modalities; includes staging levels) Multichannel Glaucoma Benchmark Dataset	4,096+ samples	

Table 9 (continued)

References, Year	Imaging type	Clinical dataset	Number of images	Resource link
Hai (2024)	Fundus	Glaucoma Screening Dataset (Healthy vs. Glaucoma)	650 fundus images	https://ieee-dataport.org/documents/glaucoma-screening-dataset
Lim et al. (2022)	GHTnet Dataset	1,074 fundus images (Healthy vs. Glaucoma)	Glaucoma Screening Dataset	https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-022-00933-z#MOESM1

The table details key resources across different imaging modalities, including fundus photography and OCT, with information on dataset size, image types, and clinical annotations ranging from binary classification (healthy vs. glaucoma) to multi-class grading (normal, early, moderate, advanced). Resource links are provided to facilitate accessibility and reproducibility

Comparisons between imaging modalities reveal that OCT often outperforms fundus photography for early glaucoma detection, yet combining both modalities with clinical data (including IOP, VF, and CCT) yields superior diagnostic accuracy. Furthermore, newer technologies such as OCTA and wide-field OCT expand detection capabilities by revealing microvascular and peripheral retinal changes invisible to standard scans.

Demographic and genetic factors, such as age, race, and family history, also play a significant role in glaucoma prevalence, and incorporating these into predictive models allows for more personalized risk assessment. Importantly, AI holds great potential to reduce global disparities in glaucoma-related blindness, particularly in low-resource settings. Nevertheless, the "black box" nature of DL models presents a barrier to clinical translation; techniques such as SHAP, Grad-CAM, and interpretable CNN variants are therefore critical for enhancing explainability and building clinician trust.

As illustrated in Fig. 12, diverse AI model families–including CNN, VGG, transformers, hybrid methods, and random forest classifiers–have been applied across both binary detection and multi-class staging, collectively advancing the field of glaucoma detection and aligning with the global health priorities and epidemiological insights presented in the subsequent sections of this paper.

7 Limitations and research gap

Although significant progress has been made in developing AI-based systems for glaucoma detection and staging, several key limitations and gaps still prevent widespread clinical adoption, as shown by Fig. 13. These issues appear consistently across the reviewed studies and are further reflected in the research gap analysis.

Limitations:

- Poor external validity results from reliance on tiny or homogeneous datasets, which often contain less than 500 photos.
- Global application is limited since populations in larger datasets (like LAG and JustRAIGS) are still limited to particular ethnic or regional groups.
- An excessive dependence on color fundus photos at the expense of other modalities like

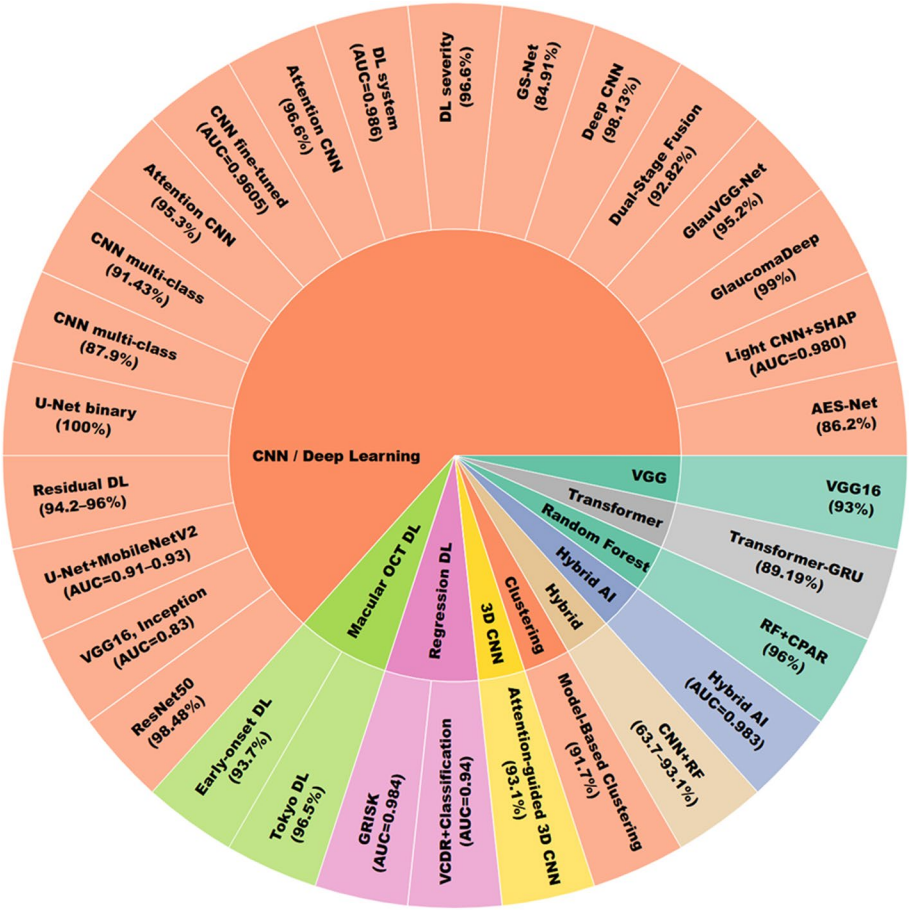


Fig. 12 Families of AI models used for glaucoma detection. The inner circle groups models into categories such as CNN/DL, VGG, Transformer, Random Forest, Hybrid AI, Clustering, and OCT-based methods. The outer circle gives examples of specific models within each family, like U-Net, ResNet, VGG16, Transformer-GRU, and Random Forest, with their reported performance (accuracy or AUC). This overview highlights the variety of approaches and their effectiveness in glaucoma detection

OCT, visual field data, or clinical characteristics.

- Dependency on optic disc/cup segmentation, which generates mistakes because of uneven human labels, elliptical fitting assumptions, and unclear boundaries.
- There are trade-offs between sensitivity and specificity due to the sensitivity of classification performance on segmentation quality.
- AI models' limited interpretability–early CNNs were opaque, and although attention mechanisms like SHAP enhanced explainability, they are still unable to match clinician thinking.
- Lack of validation in prospective, real-world clinical situations; the majority of research use internal cohorts or retrospective, carefully managed datasets.
- There was little thought given to implementation in low-resource environments, clinician usability, or integration into clinical workflow.

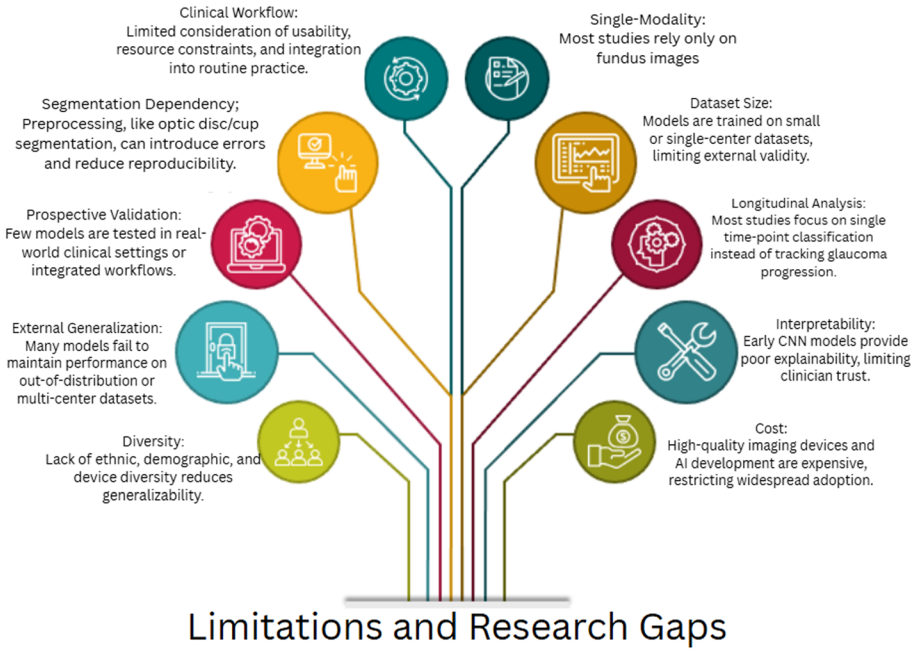


Fig. 13 Tree-structured framework showing the main research gaps in using explainable AI for glaucoma detection. The framework covers challenges across clinical workflow integration, reliance on segmentation, limited real-world validation, weak generalization to multi-center data, small and non-diverse datasets, dependence on single imaging types, and lack of longitudinal studies. It also highlights issues of interpretability, usability, and cost, which restrict large-scale clinical use

- Prioritize single-time-point classification above tracking glaucoma progression.
- Reduced ability to generalize across various imaging devices and acquisition techniques, which frequently leads to poorer performance on external datasets.

Despite progress, important gaps remain in the literature, especially in the research gaps stated below.

Research Gaps:

- Need for large, diverse, multi-center datasets representing different ethnicities, devices, and clinical environments.
- To improve diagnostic reliability, strong multimodal integration techniques that combine fundus, OCT, visual field, and clinical characteristics are being developed.
- Investigating longitudinal data to track the evolution of glaucoma predictably will allow for earlier interventions and more individualized treatment.
- Enhancement of interpretability techniques to better match model thinking with the judgment of ophthalmologists.
- Clinical trials are used for prospective validation to evaluate performance in practical settings.
- Establishment of standardized benchmarks and reporting practices for glaucoma AI systems to allow fair comparison across all studies.

- Addressing ethical and regulatory challenges, including transparency, bias mitigation, and approval pathways for clinical deployment.
- Ensuring patient privacy and secure handling of sensitive medical imaging data in AI research and deployment.

8 Conclusion

Reviewing evidence from 100 articles provided a comprehensive evaluation of artificial intelligence-based approaches for glaucoma detection, grading, and progression monitoring. OCTA, OCT, and fundus imaging were taken into consideration to evaluate whether they can outperform regular machine learning models by providing more accurate results and predictions. AI has shown a stronger and more stable potential in filling in the gaps in glaucoma research and unknown information. Despite the information gained from AI, it is still under development and requires further work to improve accuracy, as many studies rely on similar methods rather than exploring new and improved approaches.

Most deep learning models continue to function as "black boxes," which undermines clinician trust, complicates regulatory approval, and slows clinical integration. Moreover, current AI studies are limited by small dataset sizes, lack of external validation, and insufficient ethnic and demographic representativity, which may affect the generalizability of the results. Future research should prioritize larger, more diverse datasets and external validation to ensure that AI models are robust and equitable across populations. This type of data will help patients receive earlier detection and more personalized care.

The integration of AI into glaucoma care carries significant potential benefits. In resource-limited regions, lightweight AI models deployed via mobile or tele-ophthalmology platforms could facilitate population-level screening, thereby addressing the challenge of underdiagnosis. In specialized clinical environments, advanced OCT- and multimodal-based systems may enhance disease monitoring, support treatment planning, and enable earlier interventions that preserve vision and quality of life. Additionally, by assisting clinicians in making more proactive treatment decisions, AI-based progression prediction models may lessen the burden of irreversible vision loss.

In conclusion, the advancement of artificial intelligence is encouraging and moving in the correct direction for the future, even though considerable advancements are still required. AI techniques for glaucoma detection may become a permanent part of clinical practice. Millions of patients who are at risk of blindness could benefit from improved diagnostic accuracy, accessibility, and efficiency if explainable, multimodal, and clinically validated AI frameworks continue to evolve.

Author Contributions Conceptualization: A.E.-B., G.G., M.G. Methodology and study design (PRISMA/PICOS): M.E., A.M.A., S.H. Systematic search and screening: S.H., S.A.H., A.S., T.C. Data extraction and quality assessment: S.A.H., A.S., T.C., I.S., E.A.A., H.M.A., A.T.K. Analysis and synthesis: M.E., A.M.A., S.H., I.S. Clinical interpretation and validation: I.S., E.A.A., H.M.A. Figures (including PRISMA diagram) and tables: M.E., S.H., A.M.A., S.A.H., A.S. Writing—original draft: M.E., S.H. Writing—review & editing: A.M.A., S.A.H., A.S., T.C., I.S., E.A.A., M.G., G.G., H.M.A., A.T.K., A.E.-B. Supervision: A.E.-B., G.G., M.G. All authors reviewed and approved the final manuscript. Initials mapping: A.M.A. = Ahmed M. Abd El-Gawad; S.H. = Sarah Hassan; M.E. = Mohamed Elsharkawy; S.A.H. = Shahad Al Hamadani; A.S. = Aliyah Shivel; T.C. = Tracy Couch; I.S. = Ibrahim Saleh; E.A.A. = Eman A. Atallah; M.G. = Mohammed Ghazal; G.G. = Guruprasad Giridharan; H.M.A. = Hanan M. Amer; A.T.K. = Abeer Twakol Khalil; A.E.-B. = Ayman El-Baz.

Funding No funding was received for this work.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

Informed Consent Not applicable.

References

Abbas Q (2017) Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning. Int J Adv Comput Sci Appl 8(6):41–47

Ahn JM, Kim S, Ahn K-S, Cho S-H, Lee KB, Kim US (2018) A deep learning model for the detection of both advanced and early glaucoma using fundus photography. PLoS ONE 13(11):0207982

Akter N, Fletcher J, Perry S, Simunovic MP, Briggs N, Roy M (2022) Glaucoma diagnosis using multi-feature analysis and a deep learning technique. Sci Rep 12(1):8064

Akter N, Perry S, Fletcher J, Simunovic MP, Stapleton F, Roy M (2023) Glaucoma detection and feature visualization from oct images using deep learning. medRxiv

Akter N, Gordon J, Li S, Poon M, Perry S, Fletcher J, Chan T, White A, Roy M (2025) Glaucoma detection and staging from visual field images using machine learning techniques. PLoS ONE 20(1):0316919

Alhadeff PA, De Moraes CG, Chen M, Raza AS, Ritch R, Hood DC (2017) The association between clinical features seen on fundus photographs and glaucomatous damage detected on visual fields and optical coherence tomography scans. J Glaucoma 26(5):498–504

Aljohani A, Aburasain RY (2024) A hybrid framework for glaucoma detection through federated machine learning and deep learning models. BMC Med Inform Decis Mak 24(1):115

Almazroa A, Alodhayb S, Raahemifar K, Lakshminarayanan V (2017) An automatic image processing system for glaucoma screening. Int J Biomed Imaging 2017:1–19

Almazroa AA, Alodhayb S, Osman E, Ramadan E, Hummadi M, Dlaim M, Alkatee M, Raahemifar K, Lakshminarayanan V (2018) RRetinal fundus images for glaucoma analysis: the Riga dataset. In: Zhang J, Chen P-H (eds) Medical imaging 2018: imaging informatics for healthcare, research, and applications. SPIE. https://doi.org/10.1117/12.2293584

An G, Omodaka K, Tsuda S, Shiga Y, Takada N, Kikawa T, Nakazawa T, Yokota H, Akiba M (2018) Comparison of machine-learning classification models for glaucoma management. J Healthc Eng 2018:1–8

An G, Omodaka K, Hashimoto K, Tsuda S, Shiga Y, Takada N, Kikawa T, Yokota H, Akiba M, Nakazawa T (2019) Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images. J Healthc Eng 2019:1–9

Asano S, Asaoka R, Murata H, Hashimoto Y, Miki A, Mori K, Ikeda Y, Kanamoto T, Yamagami J, Inoue K (2021) Predicting the central 10 degrees visual field in glaucoma by applying a deep learning algorithm to optical coherence tomography images. Sci Rep 11(1):2214

Asaoka R, Murata H, Iwase A, Araie M (2016) Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology 123(9):1974–1980

Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, Kanamoto T, Ikeda Y, Mori K, Iwase A, Shoji N, Inoue K, Yamagami J, Araie M (2019) Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. Am J Ophthalmol 198:136–145

Ashtari-Majlan M, Masip D (2025) Spatial-aware transformer-gru framework for enhanced glaucoma diagnosis from 3d oct imaging. IEEE J Biomed Health Inform, 1–11

Bajwa MN, Malik MI, Siddiqui SA, Dengel A, Shafait F, Neumeier W, Ahmed S (2019) Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. BMC Med Inform Decis Mak 19(1):136

Bajwa MN, Singh GAP, Neumeier W, Malik MI, Dengel A, Ahmed S (2020) G1020: a benchmark retinal fundus image dataset for computer-aided glaucoma detection. arXiv. https://doi.org/10.48550/ARXIV.2006.09158

Biarnés M, Ventura-Abreu N, Rodríguez-Una I, Franquesa-Garcia F, Batlle-Ferrando S, Carrión-Donderis MT, Castro-Domínguez R, Millá E, Muniesa MJ, Pazos M (2023) Classifying glaucoma exclusively with oct: comparison of three clustering algorithms derived from machine learning. Eye 38(5):841–846

Brusini P, Salvetat ML, Zeppieri M (2021) How to measure intraocular pressure: an updated review of various tonometers. J Clin Med 10(17):3860

But NH, Ayub MH, Ali MH (2016) Challenges in the management of glaucoma in developing countries. Taiwan J Ophthalmol 6(3):119–122

Chavan S, Choubey N (2025) Glaucoma detection and severity classification based on Glaucoattent net framework. Int J Mach Learn Cybern 16(7–8):4849–4878. https://doi.org/10.1007/s13042-025-02547-7

Chiang CYN, Braeu FA, Chuangsuwanich T, Tan RKY, Chua J, Schmetterer L, Thiery AH, Buist ML, Girard MJA (2024) Are macula or optic nerve head structures better at diagnosing glaucoma? An answer using artificial intelligence and wide-field optical coherence tomography. Transl Vis Sci Technol 13(1):5

Chincholi F, Koestler H (2024) Transforming glaucoma diagnosis: transformers at the forefront. Front Artif Intell 7:1324109

Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, Girkin CA, Liebmann JM, Zangwill LM (2018) Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep 8(1):16685

Civit-Masot J, Dominguez-Morales MJ, Vicente-Diaz S, Civit A (2020) Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction. IEEE Access 8:127519–127529

Das D, Nayak DR (2024) Gs-net: global self-attention guided CNN for multi-stage glaucoma classification. arXiv preprint arXiv:2409.16082

Das D, Nayak DR, Pachori RB (2024) Aes-net: An adapter and enhanced self-attention guided network for multi-stage glaucoma classification using fundus images. Image Vis Comput 146:105042

Denis P (2007) Le glaucome du syndrome irido-cornéo-endothélial. J Fr Ophtalmol 30(2):189–195

Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A (2019) Cnns for automatic glaucoma assessment using fundus images: an extensive validation. Biomed Eng Online 18(1):29

Dielemans I, Vingerling JR, Algra D, Hofman A, Grobbee DE, Jong PTVM (1994) Risk factors for open-angle glaucoma in a population-based study: the Rotterdam study. Ophthalmology 101(11):1851–1855. https://doi.org/10.1016/s0161-6420(94)31090-6

Dixit A, Yohannan J, Boland MV (2021) Assessing glaucoma progression using machine learning trained on longitudinal visual field and clinical data. Ophthalmology 128(7):1016–1026

Ehrlich JR, Burke-Conte Z, Wittenborn JS, Saaddine J, Omura JD, Friedman DS, Flaxman AD, Rein DB (2024) Prevalence of glaucoma among us adults in 2022. JAMA Ophthalmol 142(11):1046

Elhawy E, Kamthan G, Dong CQ, Danias J (2012) Pseudoexfoliation syndrome, a systemic disorder with ocular manifestations. Hum Genomics 6(1):22

Escamez CSF, Martinez SP, Fernandez NT (2021) High interpretable machine learning classifier for early glaucoma diagnosis. Int J Ophthalmol 14(3):393–398

Essuman VA, Beyuo VM (2024) Congenital glaucoma: the 'not-so-silent' thief of sight in children. Community Eye Health 36(121):10–12

Foster PJ (2002) The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol 86(2):238–242. https://doi.org/10.1136/bjo.86.2.238

Freddo TF, Gong H (2009) Etiology of intraocular pressure elevation in primary open angle glaucoma. Optometric Glaucoma Soc e-J 4(1):0709

Fujiwara K, Yasuda M, Hata J, Nakano S, Hashimoto S, Ueda E, Nakamura S, Murakami Y, Nakamuro T, Iwase A, Araie M, Tawara A, Kubota T, Yoshitomi T, Ninomiya T, Sonoda K-H (2022) Prevalence of glaucoma and its systemic risk factors in a general Japanese population: the Hisayama study. Transl Vis Sci Technol 11(11):11

Gedde SJ, Vinod K, Wright MM, Muir KW, Lind JT, Chen PP, Li T, Mansberger SL (2021) Primary open-angle glaucoma preferred practice pattern®. Ophthalmology 128(1):71–150

George Y, Antony BJ, Ishikawa H, Wollstein G, Schuman JS, Garnavi R (2020) Attention-guided 3d-cnn framework for glaucoma detection and structural-functional association using volumetric images. IEEE J Biomed Health Inform 24(12):3421–3430
Hai ZH (2024) Glaucoma screening dataset. IEEE DataPort. https://doi.org/10.21227/B2R5-5Z41
Hasan MM, Phu J, Wang H, Sowmya A, Kalloniatis M, Meijering E (2025) Oct-based diagnosis of glaucoma and glaucoma stages using explainable machine learning. Sci Rep 15(1):3592
Hemelings R, Elen B, Barbosa-Breda J, Blaschko MB, De Boever P, Stalmans I (2021) Deep learning on fundus images detects glaucoma beyond the optic disc. Sci Rep 11(1):20313
Hemelings R, Elen B, Schuster AK, Blaschko MB, Barbosa-Breda J, Hujanen P, Junglas A, Nickels S, White A, Pfeiffer N, Mitchell P, De Boever P, Tuulonen A, Stalmans I (2023) A generalizable deep learning regression model for automated glaucoma screening from fundus images. Npj Digit Med 6:112
Hodapp E, Parrish RK, Anderson DR (1993) Clinical decisions in glaucoma. Mosby, St. Louis
Hollands H, Johnson D, Hollands S, Simel DL, Jinapriya D, Sharma S (2013) Do findings on routine examination identify patients at risk for primary open-angle glaucoma?: The rational clinical examination systematic review. JAMA 309(19):2035
Hood DC, Tsamis E, Bommakanti NK, Joiner DB, Al-Aswad LA, Blumberg DM, Cioffi GA, Liebmann JM, De Moraes CG (2019) Structure-function agreement is better than commonly thought in eyes with early glaucoma. Invest Opthalmol Vis Sci 60(13):4241
Huang X, Sun J, Gupta K, Montesano G, Crabb DP, Garway-Heath DF, Brusini P, Lanzetta P, Oddone F, Turpin A, McKendrick AM, Johnson CA, Yousefi S (2022) Detecting glaucoma from multi-modal data using probabilistic deep learning. Front Med 9:923096
Huang X, Kong X, Shen Z, Ouyang J, Li Y, Jin K, Ye J (2023) Grape: A multi-modal dataset of longitudinal follow-up visual field and fundus images for glaucoma management. Sci Data. https://doi.org/10.1038/s41597-023-02424-4
Huang X, Kong X, Yan Y, Gao Z, Li Z, Zhang C, Jin K, Ye J (2025) Interpretable longitudinal glaucoma visual field estimation deep learning system from fundus images and clinical narratives. Npj Digit Med 8(1):389
Husain KA, Alaali H, Alarayedh GG (2024) Prevalence and characteristics of glaucoma among patients presenting to ophthalmology clinics in a tertiary hospital in the kingdom of bahrain. Cureus. https://doi.org/10.7759/cureus.54129
Hussain S, Chua J, Wong D, Lo J, Kadziauskiene A, Asoklis R, Barbastathis G, Schmetterer L, Yong L (2023) Predicting glaucoma progression using deep learning framework guided by generative algorithm. Sci Rep 13(1):19960
Investigators AGIS (1994) Advanced glaucoma intervention study 2: visual field test scoring and reliability. Ophthalmology 101(8):1445–1455. https://doi.org/10.1016/S0161-6420(94)31171-7
Jain IS, Gupta A, Dogra MR, Gangwar DN, Dhir SP (1983) Phacomorphic glaucoma-management and visual prognosis. Indian J Ophthalmol 31(5):648–653
Jalili J, Jiravarnsirikul A, Bowd C, Chuter B, Belghith A, Goldbaum MH, Baxter SL, Weinreb RN, Zangwill LM, Christopher M (2025) Glaucoma detection and feature identification via gpt-4v fundus image analysis. Ophthalmol Sci 5(2):100667. https://doi.org/10.1016/j.xops.2024.100667
Jan C, He M, Vingrys A, Zhu Z, Stafford RS (2024) Diagnosing glaucoma in primary eye care and the role of artificial intelligence applications for reducing the prevalence of undetected glaucoma in australia. Eye 38(11):2003–2013
Joshi K, Gupta P, Chhabra S, Bhagat P, Bharti S, Minz RW (2017) Immune complex-positive glomerulonephritis in antineutrophil cytoplasmic antibody-positive patients. J Postgrad Med Educ Res 51(4):166–169
Kang JM, Tanna AP (2021) Glaucoma. Med Clin North Am 105(3):493–510
Kashyap R, Nair R, Gangadharan SMP, Botto-Tobar M, Farooq S, Rizwan A (2022) Glaucoma detection and classification using improved u-net deep learning model. Healthcare 10(12):2497
Keel S, Xie J, Foreman J, Lee PY, Alwan M, Fahy ET, Wijngaarden P, Fan Gaskin JC, Ang GS, Crowston JG, Taylor HR, Dirani M (2018) Prevalence of glaucoma in the Australian national eye health survey. Br J Ophthalmol 103(2):191–195. https://doi.org/10.1136/bjophthalmol-2017-311786
Khalili AF, Razzaghi S, Motlagh BF, Faramarzi E, Zeinalzadeh AH (2022) Prevalence of primary open-angle glaucoma and its relationship with smoking in the population of the azar cohort: a cross-sectional study. Middle East Afr J Ophthalmol 29(3):109–115
Kiefer R (2023) SMDG, A standardized fundus glaucoma dataset. https://doi.org/10.34740/KAGGLE/DS/2329670
Kim SJ, Cho KJ, Oh S (2017) Development of machine learning models for diagnosis of glaucoma. PLoS ONE 12(5):0177726
Kim J-A, Yoon H, Lee D, Kim M, Choi J, Lee EJ, Kim T-W (2023) Development of a deep learning system to detect glaucoma using macular vertical optical coherence tomography scans of myopic eyes. Sci Rep 13(1):8040

Kooner KS, Angirekula A, Treacher AH, Al-Humimat G, Marzban MF, Chen A, Pradhan R, Tunga N, Wang C, Ahuja P, Zuberi H, Montillo AA (2022) Glaucoma diagnosis through the integration of optical coherence tomography/angiography and machine learning diagnostic models. Clin Ophthalmol 16:2685–2697

Koornwinder A, Zhang Y, Ravindranath R, Chang RT, Bernstein IA, Wang SY (2025) Multimodal artificial intelligence models predicting glaucoma progression using electronic health records and retinal nerve fiber layer scans. Transl Vis Sci Technol 14(3):27

Kovalyk O, Morales-Sánchez J, Verdú-Monedero R, Sellés-Navarro I, Palazón-Cabanes A, Sancho-Gómez J-L (2022) Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci Data. https://doi.org/10.1038/s41597-022-01388-1

Kumar JRH, Seelamantula CS, Gagan JH, Kamath YS, Kuzhuppilly NIR, Vivekanand U, Gupta P, Patil S (2023) Chákṣu: a glaucoma specific fundus image database. Sci Data. https://doi.org/10.1038/s41597-023-01943-4

Kwon YH, Fingert JH, Kuehn MH, Alward WLM (2009) Primary open-angle glaucoma. N Engl J Med 360(11):1113–1124

Lee DA, Higginbotham EJ (2005) Glaucoma and its treatment: a review. Am J Health Syst Pharm 62(7):691–699

Lee J, Kim YK, Park KH, Jeoung JW (2020) Diagnosing glaucoma with spectral-domain optical coherence tomography using deep learning classifier. J Glaucoma 29(4):287–294

Lee T, Jammal AA, Mariottoni EB, Medeiros FA (2021) Predicting glaucoma development with longitudinal deep learning predictions from fundus photographs. Am J Ophthalmol 225:86–94

Li Z, He Y, Keel S, Meng W, Chang RT, He M (2018) Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 125(8):1199–1206

Li L, Xu M, Wang X, Jiang L, Liu H (2019) Attention based glaucoma detection: A large-scale database and cnn model. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 10571–10580

Li C, Chua J, Schwarzhans F, Husain R, Girard MJA, Majithia S, Tham Y-C, Cheng C-Y, Aung T, Fischer G, Vass C, Bujor I, Kwoh CK, Popa-Cherecheanu A, Schmetterer L, Wong D (2023) Assessing the external validity of machine learning-based detection of glaucoma. Sci Rep 13(1):558

Lim WS, Ho H-Y, Ho H-C, Chen Y-W, Lee C-K, Chen P-J, Lai F, Jang J-SR, Ko M-L (2022) Use of multimodal dataset in ai for detecting glaucoma based on fundus photographs assessed with oct: focus group study on high prevalence of myopia. BMC Med Imaging. https://doi.org/10.1186/s12880-022-00933-z

Lin H, Apostolidis C, Katsaggelos AK (2024) Brighteye: Glaucoma screening with color fundus photographs based on vision transformer. In: 2024 IEEE international symposium on biomedical imaging (ISBI). IEEE, pp 1–4

Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, Li S, Wang H, Mou D, Pang R, Yang D, Zangwill LM, Moghimi S, Hou H, Bowd C, Jiang L, Chen Y, Hu M, Xu Y, Kang H, Ji X, Chang R, Tham C, Cheung C, Ting DSW, Wong TY, Wang Z, Weinreb RN, Xu M, Wang N (2019) Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol 137(12):1353

Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R (2019) A feature agnostic approach for glaucoma detection in oct volumes. PLoS ONE 14(7):e0219126

Mannaf SMA, Islam MS, Islam MN, Rahman MM, Parvin S, Rahman S, Sarker BKD (2024) Population-based survey of the prevalence and types of glaucoma in Bangladesh. BMJ Open Ophthalmol 9(1):001609. https://doi.org/10.1136/bmjophth-2023-001609

Mariottoni EB, Datta S, Shigueoka LS, Jammal AA, Tavares IM, Henao R, Carin L, Medeiros FA (2023) Deep learning–assisted detection of glaucoma progression in spectral-domain oct. Ophthalmol Glaucoma 6(3):228–238

Marquis RE, Whitson JT (2005) Management of glaucoma: focus on pharmacological therapy. Drugs Aging 22(1):1–21

Martin KR, Mansouri K, Weinreb RN, Wasilewicz R, Gisler C, Hennebert J, Genoud D, Shaarawy T, Erb C, Pfeiffer N, Trope GE, Medeiros FA, Barkana Y, Liu JHK, Ritch R, Mermoud A, Jinapriya D, Birt C, Ahmed II, Kranemann C, Höh P, Lachenmayr B, Astakhov Y, Chen E, Duch S, Marchini G, Gandolfi S, Rekas M, Kuroyedov A, Cernak A, Polo V, Belda J, Grisanti S, Baudouin C, Nordmann J-P, De Moraes CG, Segal Z, Lusky M, Morori-Katz H, Geffen N, Kurtz S, Liu J, Budenz DL, Knight OJ, Mwanza JC, Viera A, Castanera F, Che-Hamzah J (2018) Use of machine learning on contact lens sensor–derived parameters for the diagnosis of primary open-angle glaucoma. Am J Ophthalmol 194:46–53

Masoomian B, Saatchi M, Ghassemi F, Riazi-Esfahani H, Vahedian Z (2020) Angle closure glaucoma secondary to enlarged soemmering ring that is clinically similar to iris tumour. Int Med Case Rep J 13:327–330

Medeiros FA, Jammal AA, Thompson AC (2019) From machine to machine: an oct-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. Ophthalmology 126(4):513–521
Medeiros FA, Jammal AA, Mariottoni EB (2021) Detection of progressive glaucomatous optic nerve damage on fundus photographs with deep learning. Ophthalmology 128(3):383–392
Mehta P, Petersen CA, Wen JC, Banitt MR, Chen PP, Bojikian KD, Egan C, Lee S-I, Balazinska M, Lee AY, Rokem A (2021) Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images. Am J Ophthalmol 231:154–169
Milad D, Antaki F, Mikhail D, Farah A, El-Khoury J, Touma S, Durr GM, Nayman T, Playout C, Keane PA, Duval R (2025) Code-free deep learning glaucoma detection on color fundus images. Ophthalmol Sci 5(4):100721. https://doi.org/10.1016/j.xops.2025.100721
Mouhafid M, Zhou Y, Shan C, Xiao Z, Benzorgat N (2026) Multi-class glaucoma diagnosis using Glauvgg-net and multi-transform dynamic data augmentation. Digit Signal Process 168:105503
Muhammad H, Fuchs TJ, De Cuir N, De Moraes CG, Blumberg DM, Liebmann JM, Ritch R, Hood DC (2017) Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. J Glaucoma 26(12):1086–1094
Noury E, Mannil SS, Chang RT, Ran AR, Cheung CY, Thapa SS, Rao HL, Dasari S, Riyazuddin M, Chang D, Nagaraj S, Tham CC, Zadeh R (2022) Deep learning for glaucoma detection and identification of novel diagnostic areas in diverse real-world datasets. Transl Vis Sci Technol 11(5):11
Oh S, Park Y, Cho KJ, Kim SJ (2021) Explainable machine learning model for glaucoma diagnosis and its interpretation. Diagnostics 11(3):510
Omodaka K, An G, Tsuda S, Shiga Y, Takada N, Kikawa T, Takahashi H, Yokota H, Akiba M, Nakazawa T (2017) Classification of optic disc shape in glaucoma using machine learning based on quantified ocular parameters. PLoS ONE 12(12):0190012
Pascal L, Perdomo OJ, Bost X, Huet B, Otálora S, Zuluaga MA (2022) Multi-task deep learning for glaucoma detection from color fundus images. Sci Rep 12(1):12361
Pham AT, Bradley C, Hou K, Herbert P, Yohannan J (2025) Detecting glaucoma worsening using optical coherence tomography derived visual field estimates. Sci Rep 15(1):5013
Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR (2018) Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf Sci 441:41–49
Raja H, Akram MU, Khawaja SG, Arslan M, Ramzan A, Nazir N (2020) Data on oct and fundus images for the detection of glaucoma. Data Brief 29:105342. https://doi.org/10.1016/j.dib.2020.105342
Ran AR, Cheung CY, Wang X, Chen H, Luo L-Y, Chan PP, Wong MOM, Chang RT, Mannil SS, Young AL, Yung H-W, Pang CP, Heng P-A, Tham CC (2019) Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. Lancet Digit Health 1(4):172–182
Rasel RK, Wu F, Chiariglione M, Choi SS, Doble N, Gao XR (2024) Assessing the efficacy of 2d and 3d cnn algorithms in oct-based glaucoma detection. Sci Rep 14(1):11758
Robinson DC, Hand JA, Madsen MB, McKelvey KR (2018) The dat project, an open and decentralized research data tool. Sci Data. https://doi.org/10.1038/sdata.2018.221
Rodgers CD, Meyer AM, Rosenberg NC, Lukowski ZL, Schaefer JL, Martorana GM, Levine MA, Meyers CA, Sherwood MB (2018) The impact of conjunctival flap method and drainage cannula diameter on bleb survival in the rabbit model. PLoS ONE 13(5):0196968
Rodrigues GB, Abe RY, Zangalli C, Sodre SL, Donini FA, Costa DC, Leite A, Felix JP, Torigoe M, Diniz-Filho A, Almeida HG (2016) Neovascular glaucoma: a review. Int J Retina Vitreous 2(1):26
Rogers TW, Jaccard N, Carbonaro F, Lemij HG, Vermeer KA, Reus NJ, Trikha S (2019) Evaluation of an ai system for the automated detection of glaucoma from stereoscopic optic disc photographs: the european optic disc assessment study. Eye 33(11):1791–1797
Russakoff DB, Mannil SS, Oakley JD, Ran AR, Cheung CY, Dasari S, Riyazzuddin M, Nagaraj S, Rao HL, Chang D, Chang RT (2020) A 3d deep learning system for detecting referable glaucoma using full oct macular cube scans. Transl Vis Sci Technol 9(2):12
Saha S, Vignarajan J, Frost S (2023) A fast and fully automated system for glaucoma detection using color fundus photographs. Sci Rep 13(1):18408
Shah S, Kasukurthi N, Pande H (2019) Dynamic region proposal networks for semantic segmentation in automated glaucoma screening. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE, pp 578–582
Sharma P, Takahashi N, Ninomiya T, Sato M, Miya T, Tsuda S, Nakazawa T (2025) A hybrid multi model artificial intelligence approach for glaucoma screening using fundus images. Npj Digit Med 8(1):130
Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, Asaoka R (2018) Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep 8(1):14665

Shinde R (2021) Glaucoma detection in retinal fundus images using u-net and supervised machine learning algorithms. Intell-Based Med 5:100038
Shoukat A, Akbar S, Hassan SA, Iqbal S, Mehmood A, Ilyas QM (2023) Automatic diagnosis of glaucoma from retinal images using deep learning approach. Diagnostics 13(10):1738
Simcoe MJ, Weisschuh N, Wissinger B, Hysi PG, Hammond CJ (2020) Genetic heritability of pigmentary glaucoma and associations with other eye phenotypes. JAMA Ophthalmol 138(3):294
Singh LK, Pooja Garg H, Khanna M (2021) An artificial intelligence-based smart system for early glaucoma recognition using oct images. Int J e-Health Med Commun 12(4):32–59
Smith J, Doe J (2024) Oct probability change maps for detection of early glaucomatous progression. J Ophthalmol 15(3):123–135
Son J, Shin JY, Kong ST, Park J, Kwon G, Kim HD, Park KH, Jung K-H, Park SJ (2023) An interpretable and interactive deep learning algorithm for a clinically applicable retinal fundus diagnosis system by modelling finding-disease relationship. Sci Rep 13(1):5934
Stamer WD, Braakman ST, Zhou EH, Ethier CR, Fredberg JJ, Overby DR, Johnson M (2015) Biomechanics of Schlemm's canal endothelium and intraocular pressure reduction. Prog Retin Eye Res 44:86–98
Sugar HS (1942) Place of hemorrhagic glaucoma in etiologic classification of glaucoma. Arch Ophthalmol 28(4):587–598
Sung VCT, Barton K (2004) Management of inflammatory glaucomas. Curr Opin Ophthalmol 15(2):136–140
Susanna R, De Moraes CG, Cioffi GA, Ritch R (2015) Why do people (still) go blind from glaucoma? Transl Vis Sci Technol 4(2):1
Thakur N, Juneja M (2020) Classification of glaucoma using hybrid features with machine learning approaches. Biomed Signal Process Control 62:102137
Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y (2014) Global prevalence of glaucoma and projections of glaucoma burden through 2040. Ophthalmology 121(11):2081–2090
Thanapaisal S, Uttakit P, Ittharat W, Suvannachart P, Supasai P, Polpinit P, Sirikarn P, Hanpinitsak P (2025) Machine learning technology in the classification of glaucoma severity using fundus photographs. Sci Rep 15(1):26151
Thiéry AH, Braeu F, Tun TA, Aung T, Girard MJA (2023) Medical application of geometric deep learning for the diagnosis of glaucoma. Transl Vis Sci Technol 12(2):23
Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA (2020) Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. JAMA Ophthalmol 138(4):333
Wagner IV, Stewart MW, Dorairaj SK (2022) Updates on the diagnosis and management of glaucoma. Mayo Clinic Proc Innov Qual Outcomes 6(6):618–635
Wang YX, Xu L, Yang H, Jonas JB (2010) Prevalence of glaucoma in north China: The Beijing eye study. Am J Ophthalmol 150(6):917–924
Wang L, Zhang X, Li Z, Yu S, Wu Y, Zhang S, Jiang G, Tian B, Mei C, Pu J, Liang Y, Yi Q, Wu W (2025) A deep semi-supervised learning approach to the detection of glaucoma on out-of-distribution retinal fundus image datasets. BMC Ophthalmol 25(1):326
Weinreb RN, Khaw PT (2004) Primary open-angle glaucoma. Lancet 363(9422):1711–1720
Weinreb RN, Aung T, Medeiros FA (2014) The pathophysiology and treatment of glaucoma: a review. JAMA 311(18):1901
Wu C-W, Shen H-L, Lu C-J, Chen S-H, Chen H-Y (2021) Comparison of different machine learning classifiers for glaucoma diagnosis based on spectralis oct. Diagnostics 11(9):1718
Wu J-H, Nishida T, Weinreb RN, Lin J-W (2022a) Performances of machine learning in detecting glaucoma using fundus and retinal optical coherence tomography images: A meta-analysis. Am J Ophthalmol 237:1–12
Wu C-W, Chen H-Y, Chen J-Y, Lee C-H (2022b) Glaucoma detection using support vector machine method based on spectralis oct. Diagnostics 12(2):391
Xiao Y, Hu Y, Quan W, Yang Y, Lai W, Wang X, Zhang X, Zhang B, Wu Y, Wu Q, Liu B, Zeng X, Lin Z, Fang Y, Hu Y, Feng S, Yuan L, Cai H, Li T, Lin H, Yu H (2021) Development and validation of a deep learning system to classify aetiology and predict anatomical outcomes of macular hole. Br J Ophthalmol 107(1):109–115
Xiong J, Li F, Song D, Tang G, He J, Gao K, Zhang H, Cheng W, Song Y, Lin F, Hu K, Wang P, Olivia Li J-P, Aung T, Qiao Y, Zhang X, Ting D (2022) Multimodal machine learning using visual fields and peripapillary circular oct scans in detection of glaucomatous optic neuropathy. Ophthalmology 129(2):171–180
Xu BY, Chiang M, Chaudhary S, Kulkarni S, Pardeshi AA, Varma R (2019) Deep learning classifiers for automated detection of gonioscopic angle closure based on anterior segment oct images. Am J Ophthalmol 208:273–280
Yang H, Ahn Y, Askaruly S, You JS, Kim SW, Jung W (2022) Deep learning-based glaucoma screening using regional rnfl thickness in fundus photography. Diagnostics 12(11):2894

Yi S, Zhang G, Qian C, Lu Y, Zhong H, He J (2022) A multimodal classification architecture for the severity diagnosis of glaucoma based on deep learning. Front Neurosci 16:939472
Yoon BW, Lim S-H, Shin JH, Lee J-W, Lee Y, Seo JH (2021) Analysis of oral microbiome in glaucoma patients using machine learning prediction models. J Oral Microbiol 13(1):1962125
Yousefi S, Kiwaki T, Zheng Y, Sugiura H, Asaoka R, Murata H, Lemij H, Yamanishi K (2018) Detection of longitudinal visual field progression in glaucoma using machine learning. Am J Ophthalmol 193:71–79
Zedan MJM, Abdani SR, Badawi S, Al-Bashayreh M, Zulkifley MA (2025) Dual-stage deep-learning method for glaucoma severity classification based on multiscale feature fusion. Exp Eye Res 259:110567
Zhang N, Wang J, Li Y, Jiang B (2021) Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. Sci Rep 11(1):13762
Zhou Y, Nie H, Gong X, Dai M, Guo Z, Deng X, Li M, Liu Y, Sun L, Tang X, Zhou L, Tang Z, Xia Z, Feng L, Zhang W, Yi Q, Xia X, Xie B, Song W (2025) A three-tier ai solution for equitable glaucoma diagnosis across China's hierarchical healthcare system. Npj Digit Med 8(1):400

Authors and Affiliations

Ahmed M. Abd El-Gawad[1] · Sarah Hassan[2] · Mohamed Elsharkawy[2] · Shahad Al Hamadani[2] · Aliyah Shivel[3] · Tracy Couch[4] · Ibrahim Saleh[5] · Eman A. Atallah[6] · Mohammed Ghazal[7] · Guruprasad Giridharan[2] · Hanan M. Amer[8] · Abeer Twakol Khalil[8] · Ayman El-Baz[2]

✉ Ayman El-Baz
aselba01@louisville.edu

[1] Department of Electronics and Communications, Alamein International University, El Alamein, Egypt

[2] Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA

[3] Department of Chemistry and Forensic Science, Eastern Kentucky University, Kentucky, USA

[4] Department of Biomedical Engineering, Washington University in St. Louis, Missouri, USA

[5] Department of Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, USA

[6] Mansoura Ophthalmic Center, Mansoura University, Mansoura, Egypt

[7] Electrical and Computer Engineering Department, Abu Dhabi University, Abu Dhabi 59911, UAE

[8] Department of Electronics and Communications Engineering, Mansoura University, Mansoura, Egypt