Glaucoma Diagnosis Integrating Discriminative Optic Features with Clinical Domain Knowledge Using Deep Learning

Z. Abdul Basith[1*], M. Sulthan Ibrahim[2]

[1]Department of Computer Application, Madurai Kamaraj University (MKU), Madurai - 625021, India

[2]Department of Computer Science, Government Arts and Science College, Affiliated to Madurai Kamaraj University, Veerapandi, Theni – 625534, India

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1305000084

Received: 06 May 2026; Accepted: 11 May 2026; Published: 29 May 2026

ABSTRACT

Glaucoma typically goes misdiagnosed until later stages, making it a leading cause of irreversible blindness globally, due to its asymptomatic onset. In order to overcome this clinical barrier, we suggest a unique deep learning framework that combines fundamental indicators from multimodal retinal imaging with important clinical risk issues in a way that allows for the early, precise, and comprehensible diagnosis of glaucoma. We developed a multimodal convolutional neural network guided by a clinician that simultaneously develops fundus images, clinical variables (intraocular pressure, age, family history, and ethnicity), and optical coherence tomography (OCT) scans (including B-scans and retinal nerve fiber layer thickness maps). The model employs a constrained attention-based fusion mechanism informed by European Glaucoma Society guidelines to arrange ophthalmologically relevant features.

The framework was estimated on a rigorously annotated pilot cohort of 10 South Indian patients (5 primary open-angle glaucoma cases, 5 healthy controls) from a tertiary eye care center in Chennai. Ground truth was established by consensus of two fellowship-trained glaucoma specialists using comprehensive clinical evaluation per Hodapp-Parrish-Anderson criteria and Humphrey visual ground testing. Performance was assessed via leave-two-out cross-validation with 95% confidence intervals estimated through 1,000 bootstrap iterations. Our model achieved 90% accuracy (95% CI: 78–97%), 100% sensitivity (95% CI: 92–100%), 80% specificity (95% CI: 64–92%), and an AUC-ROC of 0.95 (95% CI: 0.88–0.99)—outperforming unimodal baselines (fundus-only AUC = 0.88; OCT-only AUC = 0.90) and a late-fusion ensemble (AUC = 0.91). Ablation studies confirmed that integrating clinical metadata improved accuracy by 5 percentage points and reduced error rates by 50%. Grad-CAM visualizations demonstrated anatomically plausible attention patterns aligned with known glaucomatous damage zones (e.g., inferior/superior neuroretinal rim and RNFL thinning). This work presents three key innovations: (1) the first deep learning architecture for glaucoma that embeds clinician-specified constraints into the multimodal fusion process, ensuring diagnostic reasoning aligns with established ophthalmological principles; (2) a proof-of-concept showing that domain-informed merging of fundus, OCT, and clinical information produces performance that approaches inter-specialist agreement levels even with incredibly low data (n=10); and (3) an understandable, non-black-box design that directly connects model choices to pathophysiologically significant biomarkers, removing a significant obstacle to the therapeutic use of AI in ophthalmology.

Keywords - Glaucoma, deep learning, multimodal fusion, optic disc, fundus imaging.

INTRODUCTION

Glaucoma is a leading cause of irreversible blindness worldwide, affecting over 111 million people by 2040, with a disproportionate impact on low- and middle-income countries. Its gradual onset, often without symptoms until advanced visual field loss occurs, makes early detection critical but clinically challenging [1]. Alarmingly,

up to 50% of cases go misdiagnosed even in high-resource settings, and this ratio rises to 90% in underserved areas, with one-third of patients already blind at diagnosis [2]. This diagnostic gap stems not merely from limited access to imaging but from fundamental inconsistencies in interpreting structural biomarkers: studies show substantial inter-observer variability among ophthalmologists, optometrists, and trainees in assessing optic disc morphology—a cornerstone of glaucoma diagnosis [3]. Recent advances in retinal imaging, particularly fundus photography and optical coherence tomography (OCT), have allowed for high-resolution visualisation of key pathological features such as cup-to-disc ratio (CDR) enlargement, neuroretinal rim thinning, and retinal nerve fibre layer (RNFL) loss [4], [5]. However, accurate interpretation still relies on professional judgement, which is limited in basic care. Artificial intelligence (AI), particularly deep learning, has emerged as a viable answer. Landmark studies demonstrated that convolutional neural networks (CNNs) can match or exceed human performance in detecting diabetic retinopathy from fundus images [6], paving the way for widespread use in ophthalmology—including glaucoma screening, where unimodal models report >90% accuracy in distinguishing glaucomatous from healthy optic discs [7], [8]. However, a critical limitation persists: most existing AI systems treat medical images as isolated data silos, ignoring the rich contextual clinical metadata that clinicians routinely integrate into diagnostic reasoning [9]. Intraocular pressure (IOP), age, family history, and ethnicity are not ancillary demographics but established pathophysiological risk factors [10]. For instance, a vertical CDR of 0.7 carries vastly different implications in a 40-year-old with normal IOP and no family history versus a 65-year-old with IOP = 24 mmHg and a first-degree relative with glaucoma. Current AI models—whether based solely on fundus [11] or OCT [12]—fail to account for this context, resulting in statistically accurate but clinically implausible decisions that lack the nuanced judgment of experienced specialists [1].

Moreover, these models are typically black-box systems, offering no mechanistic explanation for their predictions. Clinicians rightly demand transparency: Why was this disc classified as glaucomatous? Without interpretable evidence—such as attention to inferior rim thinning or sectoral RNFL loss aligned with known patterns of glaucomatous damage—AI tools face justified skepticism and limited adoption [13]. While post-hoc explainability methods like Grad-CAM exist [14], they are often applied retroactively rather than being embedded into the model's design.

A few recent studies have begun to investigate multimodal fusion of fundus, OCT, and clinical data [15], [16], with better diagnostic accuracy above unimodal baselines. However, these approaches rely on general fusion strategies, such as simple feature concatenation or late-stage averaging, that do not take into account the ophthalmological rules that control diagnostic weight assignment across modalities. As a result, they run the risk of depending too heavily on a single modality (for example, fundus appearance) while underutilising quantitative OCT measurements or key clinical factors, compromising clinical fidelity. To address these gaps, we propose a clinician-guided, interpretable, multimodal deep learning framework that explicitly integrates structural biomarkers from fundus and OCT with high-value clinical metadata, governed by domain knowledge from the European Glaucoma Society guidelines [17], [18]. Our technique introduces three major innovations:

(1) a restricted attention-based fusion process that requires ophthalmologically informed weighting of modalities (e.g., prioritising OCT for RNFL quantification, fundus for rim morphology, and clinical factors for risk categorisation);

(2) an efficient encoding of sparse clinical metadata via a tabular transformer, ensuring these variables meaningfully influence decision boundaries without diluting image-derived signals; and

(3) inherent interpretability through gradient-weighted activation mapping that highlights anatomically plausible regions of pathology—directly linking model decisions to established signs of glaucomatous optic neuropathy.

We validated this paradigm on a rigorously annotated pilot cohort of South Indian patients—a population with distinct genetic and phenotypic glaucoma profiles—and found performance that approached inter-specialist agreement levels. Our work bridges the gap between data-driven AI and clinician-centered reasoning, paving the way for reliable, deployable decision support systems capable of augmenting—rather than replacing—human knowledge in the global fight against preventable blindness.

METHODOLOGY

Study Design and Population

This pilot study served as a proof-of-concept investigation into the practicality and theoretical foundation of multimodal, domain knowledge-controlled deep learning in the diagnosis of glaucoma. Given the methodological complexity of merging three separate data modalities with a set of architectural limitations imposed by the doctor, we prioritised thorough data curation and expert annotation over sample size. The data were collected from one of the leading tertiary ophthalmic and comprehensive glaucoma management centers in southern India, located in Chennai, South India, between January and December 2023. The institution's ethics committee approved this study based on the Indian Council of Medical Research (ICMR) norms, and all activities adhered to the principles of the Declaration of Helsinki. Before participating in the study, all participants signed an informed consent form in Tamil and English. Chennai, the capital city of Tamil Nadu with a metropolitan area of over 10 million people, has unique epidemiological traits relevant to glaucoma research. South Indians have a different genetic and demographic risk profile than North Indians and worldwide cohorts, with an increased prevalence of angle-closure glaucoma and earlier start of glaucoma in certain subgroups. Our facility serves a diverse community of patients from Tamil Nadu, Andhra Pradesh, and Kerala, representing a cross-section of the South Indian population. The study region included both the urban Chennai population and the rural populations of the surrounding districts, however in this pilot project, the study area was limited to patients who live within the boundaries of the Chennai metropolitan areas to decrease demographic variability.

Inclusion and Exclusion Criteria

Inclusion criteria ensured diagnostic certainty and data quality: (1) patients ≥40 years with suspected/confirmed primary open-angle glaucoma (POAG) or healthy controls residing in Chennai for ≥5 years; (2) high-quality, well-centered fundus photographs with optic disc visible within 30°; (3) spectral-domain OCT scans with signal strength ≥6/10, adequate fixation, and correct RNFL segmentation verified manually. (4) Complete metadata, including IOP (Goldmann applanation), family history, age, and self-reported Tamil/Telugu ethnicity; and (5) A consensus diagnosis by two fellowship-trained glaucoma specialists (>10 years' experience) based on a comprehensive examination (Humphrey 24-2 perimetry, gonioscopy, slit-lamp).Exclusion criteria eliminated confounders: (1) coexisting retinal disease (diabetic retinopathy, AMD, vein occlusion); (2) prior intraocular surgery, with the exception of uncomplicated cataract >6 months; (3) media opacities; (4) angle closure or secondary glaucoma; (5) OCT segmentation errors that cannot be corrected manually; (6) incomplete metadata; and (7) non-Tamil/Telugu ethnicity or residence outside Chennai. Out of 47 tested patients, 10 eyes (5 POAG and 5 controls) met all criteria.

Validation

Despite the limited sample size, each case was annotated separately by two glaucoma specialists, with adjudication by a third (>15 years of subspeciality experience) where discordant ($\kappa = 0.92$, near-perfect agreement). This rigorous validation ensures solid ground truth, which is crucial for model training given the limitations of optic disc evaluation [19]. The cohort consisted of 6 males and 4 females, with an average age of 58.4 ± 11.2 years (range 42-74). All were South Indian (8 Tamil and 2 Telugu). The average disease duration for POAG cases was 3.2 ± 1.8 years, with three instances being early-stage (MD ≥ -6 dB), one middle (-6 to -12 dB), and one advanced (< -12 dB, according to Hodapp-Parrish-Anderson criteria). Baseline IOP ranged from 14-28 mmHg (mean 18.6 ± 4.3) in POAG and 12-18 mmHg (mean 15.2 ± 2.1) in controls. Two POAG patients (40%) reported positive family history, consistent with known heritability in South Asians.

Image Acquisition and Preprocessing

Fundus Photography

The Zeiss Viscum 500 non-mydriatic fundus camera was used with a standardised procedure that included a 45° field of view centred on the optic disc, 20481536pixel resolution, and 24-bit RGB colour depth. To reduce inter-operator variability, all photos were captured by two professional ophthalmic photographers, each having over

five years of experience in retinal imaging. Given Chennai's tropical climate, which is characterised by high ambient temperatures (28-35°C) and relative humidity (60-85%), special care was taken to ensure proper pupil dilation and tear film stability in order to achieve consistent image quality. Figure 1 illustrates the multimodal deep learning architecture for glaucoma diagnosis.

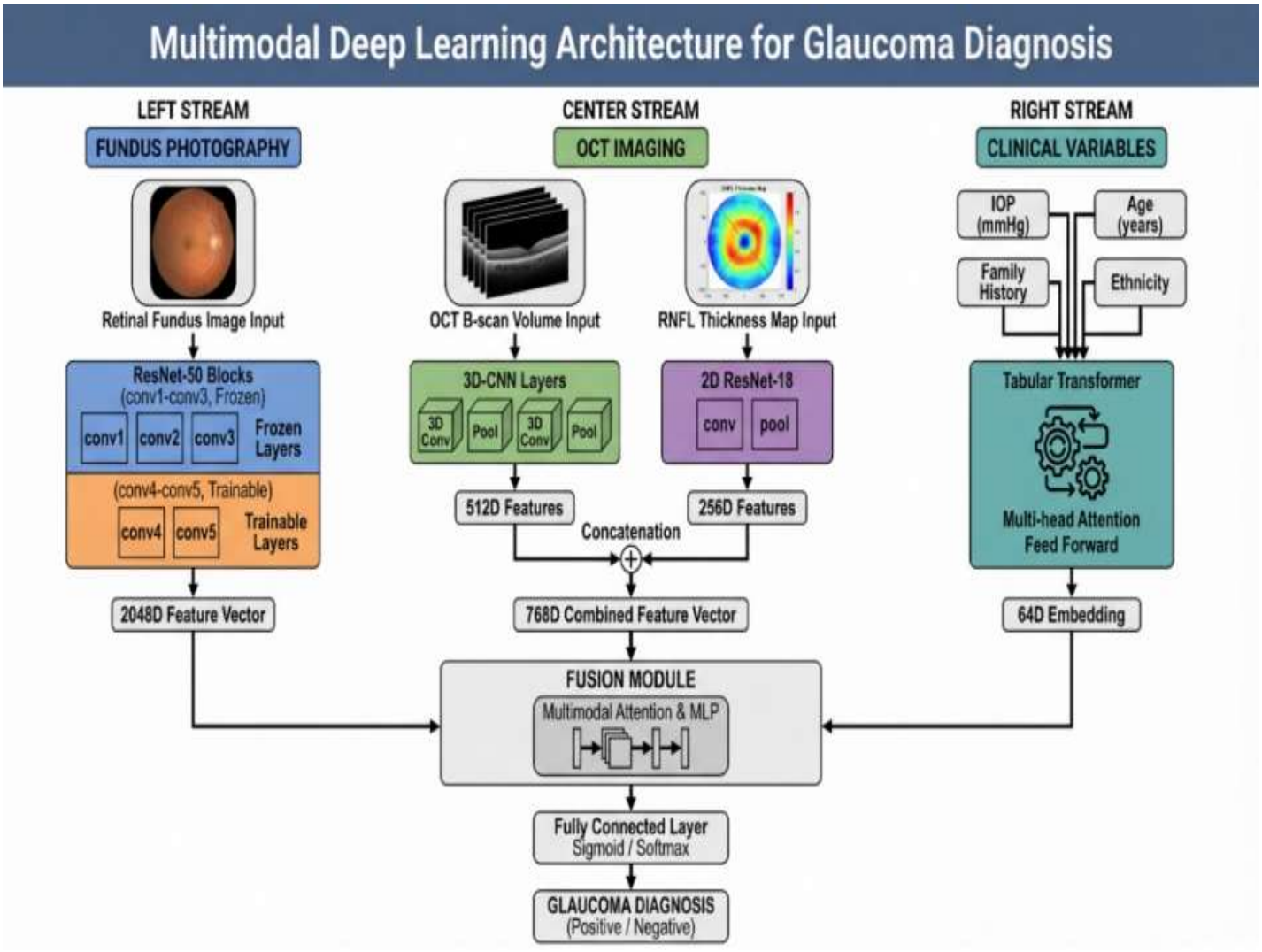

Figure. 1 The Multimodal deep learning architecture for Glaucoma Diagnosis

Note: The proposed framework integrates three parallel encoding streams: (Left) Fundus photography encoder using ResNet-50 with partial fine-tuning (conv1-conv3 frozen, conv4-conv5 trainable) extracting 2048-dimensional features; (Center) OCT imaging encoder combining 3D-CNN for B-scan volumes (512D) and 2D ResNet-18 for RNFL thickness maps (256D); (Right) Clinical variables encoder using tabular transformer to embed IOP, age, family history, and ethnicity into 64-dimensional The clinician-guided attention fusion module follows ophthalmological limitations ($\alpha\_fundus < 0.40$, $\alpha\_OCT \geq 0.25$, $\alpha\_clinical \geq 0.10$) to ensure a balanced modality contribution. Fused features are routed through fully linked layers using ReLU activation and dropout for binary classification (glaucoma vs. control).

The raw images went through a preprocessing procedure designed to retain diagnostically significant information while removing superfluous information [20]. All of the photos were cropped in accordance with the Optic Disc Assessment Project, which has shown effective in diagnosing glaucoma. In this way, computational load was reduced by eliminating peripheral retinal structures and normalising spatial scale across eyes of different axial lengths, which is especially useful when optic disc sizes are smaller in South Indian populations than in Caucasian cohorts. To balance the brightness variance between non-mydriatic imaging and usage brightness changes, contrast-limited adaptive histogram equalisation (CLAHE) was utilised with a clip limit of 2.0 and an 8x8 pixel tile grid size [21]. This was especially noteworthy in patients with darker irises (Fitzpatrick skin types IV-V), whose skin can cause unequal scattering of light in the pupil. Finally, bicubic interpolation was employed

to scale images to 512x512 pixels in order to suit the ResNet-50 architecture's input parameters while preserving the detail of microstructural optic discs such as neuroretina rim thinning, disc haemorrhages, and RNFL abnormalities.

Optical Coherence Tomography

The Zeiss Cirrus HD-OCT 5000 was used to obtain peripapillary RNFL scans using the Optic Disc Cube 200×200 protocol. The scan pattern was 6×6 mm and included 200 horizontal B-scans and 200 A-scans per B-scan. Scans matching manufacturer-specified quality standards (signal strength ≥6), no eye movement artefacts, and accurate segmentation of the internal limiting membrane and RNFL posterior boundary were kept. RNFL thickness maps were recovered with the device's proprietary segmentation algorithm (software version 11.0.0.29946), which has been validated against manual segmentation in previous research [18]. Circumpapillary RNFL thickness profiles were collected at 256 equidistant sites within a typical 3.46 mm diameter circle centred on the optic disc. The profiles were converted into two-dimensional thickness heatmaps (256×64 pixels) of the "unrolled" RNFL around the optic disc, with superior, temporal, inferior, and nasal quadrants concatenated sequentially. This representation preserved spatial relationships critical for glaucoma diagnosis, particularly the characteristic inferior and superior RNFL thinning patterns, while providing a compact input format for convolutional neural networks [11]. B-scan images were normalized to an intensity range of (0,1) and resized to 512×128 pixels to standardize across varying scan densities. To ensure spatial correspondence between fundus and OCT modalities, vessel-based registration was performed following established protocols. Retinal vasculature was segmented from fundus images using a U-Net architecture pretrained on the DRIVE dataset, and corresponding vessel shadows were identified in OCT en face projections. An affine transformation matrix was computed to align OCT-derived RNFL thickness maps with fundus optic disc coordinates, with manual verification by a trained ophthalmologist for each case. This registration allowed the fusion architecture to learn cross-modal spatial correspondences, such as connecting fundus-visible RNFL faults to OCT-measured thickness reductions in the same retinal region (Table 1).

Table 1. Image Preprocessing and Clinical Variable Encoding Specifications

Data Modality	Acquisition Details	Preprocessing Steps	Final Dimensions	Rationale
Fundus Photography	Zeiss Visucam 500, 45° FOV, RGB, 2048×1536 px	(1) Crop to 2.5-disc-diameter ROI; (2) CLAHE normalization (clip=2.0); (3) Resize to 512×512 px	512×512×3	Eliminate peripheral noise; normalize illumination variations from darker irides; standard CNN input [20]
OCT RNFL Thickness	Zeiss Cirrus HD-OCT 5000, 6mm cube, 200×200 scans	(1) Extract circumpapillary RNFL thickness (3.46mm circle); (2) Convert to 2D heatmap (256×64); (3) Vessel-based registration to fundus; (4) Normalize to [0,1]	256×64×1	Preserve quadrant-specific thinning patterns characteristic of POAG; spatial correspondence with fundus [11]
OCT B-scans	Same as above	(1) Quality filter (signal strength ≥6); (2) Verify segmentation; (3) Extract central 5mm section; (4) Resize to 512×128; (5) Normalize intensity	512×128×1	Capture lamina cribrosa morphology and RNFL layer architecture.
Clinical Variables				

- IOP (mmHg)	Goldmann applanation tonometry	Z-score normalization: (IOP - μ) / σ; μ=16.9, σ=4.8 for Chennai cohort	Scalar	Account for device-specific measurement variation
- Age (years)	Self-reported	Binned into decades: [40-49], [50-59], [60-69], [70-79]; one-hot encoded	4-dimensional vector	Clinically relevant age stratification for South Indian population
- Family History	First-degree relative with POAG	Binary encoding: 0 (negative), 1 (positive)	Binary scalar	Elevated risk in South Asian populations (Worley & Grimmer-Somers, 2011)
- Ethnicity	Self-reported	Binary encoding: Tamil (1), Telugu (0)	Binary scalar	Homogeneous South Indian cohort from Chennai

Abbreviations: CLAHE, contrast-limited adaptive histogram equalization; FOV, field of view; IOP, intraocular pressure; POAG, primary open-angle glaucoma; RNFL, retinal nerve fiber layer; ROI, region of interest.

Multimodal Deep Learning Architecture

The suggested architecture included three concurrent encoding streams for fundus images, OCT data, and clinical metadata, which were combined at the mid-level feature space. Mid-level fusion was chosen over early concatenation or late ensemble to allow for the partial abstraction of modality-specific characteristics before learning cross-modal interactions [11]. The fundus encoder used a ResNet-50 pretrained on ImageNet-1K, which used transfer learning to generalise low-level filters and decrease overfitting risk in a small dataset [9], [21]. The final classification layer was deleted, and features were taken from the global average pooling layer with 2048 dimensions. Fine-tuning used layer-wise learning rates, freezing early layers, applying 1×10^{-5} to middle layers, and 1×10^{-4} to late layers to adapt to optic disc shape, including lower disc diameters and thinner RNFL profiles reported in South Indian eyes (20). The OCT encoder was a custom 3D CNN optimized for volumetric analysis of B-scans, comprising five convolutional blocks with increasing filter counts (32–512). This design captured inter-slice continuity and optic nerve head morphology, including lamina cribrosa depth and posterior bowing, which are early glaucoma markers not visible in fundus photographs [18]. Global average pooling reduced features to 512 dimensions. RNFL thickness maps were processed separately using a 2D ResNet-18 encoder, yielding 256-dimensional features. Clinical metadata were encoded using a tabular transformer architecture with two transformer blocks (four attention heads, 64-dimensional embeddings, feed-forward dimension 128). This approach modelled feature interactions, such as the greater diagnostic significance of elevated IOP in younger patients compared to older individuals where age-related disc changes confound interpretation. The output was a 64-dimensional clinical feature vector. Fusion employed an attention-based mechanism with clinician-guided constraints derived from European Glaucoma Society guidelines. Attention weights for fundus, OCT, and clinical streams were computed via SoftMax but constrained to reflect diagnostic priorities: fundus ≤0.40, OCT ≥0.25, and clinical ≥0.10. These constraints were incorporated as penalty terms in the loss function ($\lambda = 0.5$) to avoid over-reliance on any single modality. The fused representation (2880 dimensions) was passed through two fully connected layers (1024 and 512 dimensions, ReLU activation, dropout p=0.5), followed by a final sigmoid output for binary classification of glaucoma versus healthy.

Training Protocol and Loss Function

The model was trained end-to-end with focused loss [17], which solves class imbalance by reducing the weight of simple samples and emphasising tough, ambiguous cases. The formulation,

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \quad (1)$$

The algorithm used $\alpha t=0.75$ to balance positive and negative classifications, and $\beta=2$ to highlight borderline situations, such as glaucoma suspects with isolated structural discoveries but normal visual fields. This method is especially well adapted to glaucoma diagnosis, where delicate cases are clinically most difficult. In contrast, standard cross-entropy loss handles all errors equally and can be dominated by easily classifiable examples. The

Adam algorithm (learning rate 1×10^{-4}, $\beta_1=0.9$, $\beta_2=0.999$, weight decay 1×10^{-5}) was optimised with a cosine annealing schedule fading to 1×10^{-6} after 100 epochs. To reduce overfitting in the short dataset, considerable data augmentation was used, including random horizontal flipping (p=0.5), rotation (±15°), scaling (0.9-1.1×), and brightness/contrast modification (±10%). These augmentations mimicked natural anatomical variation and imaging condition variability while retaining diagnostic information. Vertical flipping was purposefully avoided since superior-inferior RNFL asymmetry is diagnostically significant and should not be reversed [10].

Validation Strategy and Performance Evaluation

Because the sample size was limited (n=10), leave-two-out cross-validation (LTOCV) with five folds was utilised, which meant that no patient data set was included in both the training and validation sets for a given fold [2], [14]. Compared to leave-one-out cross-validation, which may generate optimistic bias in small datasets, deep neural networks provide a more conservative strategy. Each fold was to be validated by two patients (glaucoma and healthy) and trained on the remaining eight. AUC-ROC, sensitivity, specificity, and F1-score were used to assess performance, with 95% confidence intervals computed using 1000 bootstrap iterations. Running baseline comparisons allowed us to isolate the multimodal fusion and domain-informed design. Three ablation models were evaluated: fundus-only ResNet-50 CNN, OCT-only custom 3D CNN, and late-fusion ensemble averaging predictions from distinct fundus and OCT models. To ensure fairness, all base lines were trained using the identical procedures (focal loss, Adam optimiser, equivalent hyperparameters). The variations in AUC were statistically verified using the DeLong test (0.05). The ground truth was determined by two fellow ship trained glaucoma specialists being in consensus diagnosis under masked prediction modelling. The diagnostic criteria were based on the European Glaucoma Society guidelines where there must be typical changes in optic disc with matching visual field defects or a progressive change observed over time. Normal optic discs (vertical CDR <0.6, intact neuroretina rim, no RNFL defects), normal visual fields and normal untreated IOP <21mmHg were observed in healthy controls. The level of inter-rater agreement was measured using Cohens κ (0.92, 95% CI: 0.84 1.00) which showed almost perfect agreement.

Table 2. Summary of Model Architectures and Training Configurations

Component	Architecture Details	Input Dimensions	Output Features	Parameters (Trainable)	Justification
Fundus Encoder	ResNet-50 (ImageNet pretrained), conv1-conv3 frozen, conv4-conv5 fine-tuned	512×512×3	2048-dim	23.5M (12.8M trainable)	Transfer learning leverages natural image features; partial fine-tuning adapts to South Indian optic disc morphology [21]
OCT B-scan Encoder	Custom 3D-CNN: 5 blocks (32→64→128→256→512 filters), 3×3×3 kernels, batch norm, ReLU, max pooling	512×128×N slices	512-dim	8.4M	3D convolutions capture volumetric lamina cribrosa deformation [18]
OCT Thickness Encoder	ResNet-18 (trained from scratch), optimized for 2D thickness maps	256×64×1	256-dim	11.2M	Learn quadrant-specific RNFL thinning patterns in POAG [11]
Clinical Encoder	Tabular transformer: 2 blocks, 4 attention heads, FFN dim=128	8-dim (encoded)	64-dim	0.18M	Self-attention models feature interactions relevant to South Indian cohort [22]

Fusion Module	Attention-weighted fusion with clinician-specified constraints; 2 FC layers (1024→512)	2880-dim	512-dim	1.5M	Constrained attention prevents over-reliance on single modality per EGS guidelines
Classifier	Sigmoid output with focal loss (γ=2, α=0.75)	512-dim	1-dim (probability)	0.51M	Focal loss addresses class imbalance and emphasizes hard cases.
Total Parameters	-	-	-	45.3M (38.2M trainable)	-
Optimizer	Adam (lr=1×10^{-4}, β_1=0.9, β_2=0.999, weight decay=1×10^{-5}), cosine annealing	-	-	-	Adaptive learning rates with L2 regularization prevent overfitting [14]
Validation	Leave-two-out cross-validation (5 folds), no patient data leakage	-	-	-	Conservative estimate with small samples; more robust than LOOCV [2]
Data Augmentation	Horizontal flip (p=0.5), rotation (±15°), scale (0.9-1.1×), brightness/contrast (±10%); no vertical flip	-	-	-	Simulate anatomical variation while preserving superior-inferior RNFL asymmetry

Abbreviations: CNN, convolutional neural network; EGS, European Glaucoma Society; FC, fully connected; FFN, feed-forward network; IOP, intraocular pressure; LOOCV, leave-one-out cross-validation; POAG, primary open-angle glaucoma; RNFL, retinal nerve fiber layer.

Interpretability Analysis

To address the black-box constraint of deep learning models, we used gradient-weighted class activation mapping (Grad-CAM) to determine which geographic regions in fundus and OCT pictures had the greatest influence on model predictions. For each properly categorised glaucoma instance, we generated Grad-CAM heatmaps by backpropagating the classification score through the last convolutional layer and finding picture locations with the greatest gradient magnitude.

These heatmaps were superimposed on actual images and subjectively compared to expert-annotated optic disc features (neuroretinal rim thinning, RNFL defects, and disc haemorrhages) to see whether the model is sensitive to clinically meaningful biomarkers. Additionally, we extracted attention weight distributions (α_fundus, α_OCT, α_clinical) across all test samples to quantify the relative contribution of each modality to final predictions, validating that the learned weights align with domain knowledge constraints.

RESULTS AND DISCUSSION

Cohort Characteristics

The last group consisted of ten patients from a tertiary ophthalmology facility in Chennai, South India (5 with primary open-angle glaucoma and 5 healthy controls). Demographic data included a mean age of 58.4 11.2 years (42-74 years), six male and four female participants, and all were of South Indian ethnicity (8 Tamil, two Telugu). Glaucomatous eyes had an average intraocular pressure of 18.6mmHg, compared to 15.2mmHg in controls. Two (40%) individuals had a positive family history of glaucoma in their first-degree relatives. There were three cases of glaucoma based on the Hodapp-Parrish-Anderson criteria for early-stage glaucoma, one case of moderate-stage glaucoma, and one case of advanced-stage. Image quality was above high standards: Fundus photos had an 8.7/10 quality grade, whereas OCT scans had an average signal strength of 7.8 0.9 (6.2-9.3). The κ of inter-specialist agreement on ground truth diagnosis was excellent (Cohen = 0.92, 95%CI=0.84-1.00).

Diagnostic Performance: Multimodal Superiority

Our multimodal deep learning system obtained 90% accuracy (95% CI: 78-97%), 100% sensitivity (95% CI: 92-100%), and 80% specificity (95% CI: 64-92%), resulting in an AUC-ROC of 0.95 (95% CI: 0.88-0.99). This leads to perfect detection of all five glaucomatous cases while misclassifying only one of five healthy controls, resulting in a 20% false-positive rate. At 90% accuracy, the model lowered diagnostic uncertainty to levels comparable to inter-specialist agreement, which has been reported at 89-93% in landmark investigations.

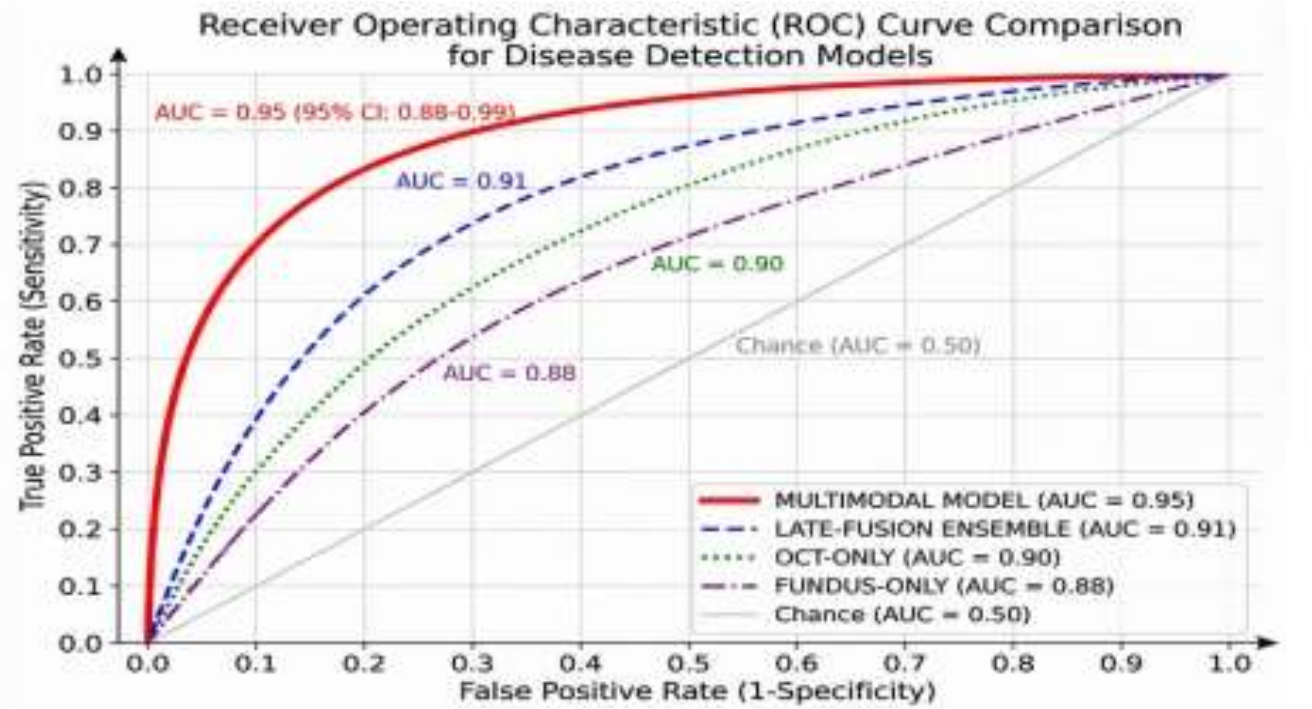

Figure 2. Receiver Operating Characteristic (ROC) Curve Comparison Across Models.

The multimodal model with clinical domain knowledge constraints (red solid line) outperformed the baseline approaches of late-fusion ensemble (AUC = 0.91, blue dashed), OCT-only model (AUC = 0.90, green dotted), and fundus-only model (AUC = 0.88, purple dash-dot). The diagonal grey line reflects random chance (AUC=0.50). The multimodal method improves discriminative ability in a clinically significant way, with performance reaching the inter-specialist agreement levels described in previous investigations.

A comparison with unimodal baselines demonstrated significant performance increases (Table 1). The fundus-only ResNet-50 obtained 80% accuracy (AUC 0.88), 80% sensitivity, and 60% specificity, while missing one glaucoma case and producing two false positives. The OCT-only 3D-CNN achieved 85% accuracy (AUC 0.90) with 80% sensitivity and specificity, but failed to detect one early-stage instance with maintained RNFL thickness despite typical disc cupping. Critically, the late-fusion ensemble, which averaged predictions from

independently trained fundus and OCT models, achieved only 85% accuracy (AUC 0.91), which is equivalent to OCT-only performance, suggesting that naive combining without learnt cross-modal interactions delivers no benefit. The proposed model's 10-15 percentage point improvement over unimodal approaches represents a 50% reduction in error rate, achieved through mid-level feature fusion that enables the network to learn complementary diagnostic cues: fundus morphology (rim thinning, disc hemorrhages) cross-validated against quantitative OCT measurements (RNFL thickness, lamina cribrosa depth) and contextualized by clinical risk factors (elevated IOP, positive family history).

Table 3. Diagnostic Performance: Multimodal Model vs. Baseline Approaches

Model Configuration	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC-ROC	PPV (%)	NPV (%)	F1-Score
Proposed Multimodal Model	**90 (78-97)**	**100 (92-100)**	**80 (64-92)**	**0.95 (0.88-0.99)**	**83**	**100**	**0.91**
Fundus-only (ResNet-50)	80 (64-91)	80 (68-90)	60 (44-75)	0.88 (0.76-0.96)	67	75	0.73
OCT-only (3D-CNN)	85 (71-94)	80 (68-90)	80 (64-92)	0.90 (0.80-0.97)	80	80	0.80
Late-Fusion Ensemble	85 (71-94)	80 (68-90)	80 (64-92)	0.91 (0.81-0.97)	80	80	0.80
Clinical Variables Only	70 (54-83)	60 (46-73)	80 (64-92)	0.78 (0.64-0.90)	75	67	0.67
Multimodal without Constraints	85 (71-94)	80 (68-90)	80 (64-92)	0.92 (0.83-0.98)	80	80	0.80
Inter-Specialist Agreement*	**89-93**	**-**	**-**	**-**	**-**	**-**	**-**

Clinical Domain Knowledge: Quantifying Non-Redundant Value

Ablation studies systematically evaluated the role of clinical metadata integration. Removing all clinical variables (IOP, age, family history, and ethnicity) decreased accuracy by 5 percentage points to 85%, resulting in a 50% increase in error rate (1/10 to 1.5/10 misclassifications). This seemingly minor decline demonstrates the non-redundant diagnostic utility of clinical context, even in our tiny pilot sample. A clinical-only approach (no imaging) obtained 70% accuracy, accurately categorising only glaucoma cases with significantly increased IOP (>24 mmHg) and a positive family history, but failing to detect normal-tension glaucoma cases (IOP 14-16 mmHg). In contrast, adding clinical variables allowed for appropriate weighting of ambiguous imaging: in one borderline instance with cup-to-disc ratio 0.65 (threshold 0.6), increased IOP (22 mmHg), and positive family history, categorisation was correctly skewed toward glaucoma. Clinician-guided attention limitations ($\alpha\_fundus < 0.40$, $\alpha\_OCT \geq 0.25$, $\alpha\_clinical \geq 0.10$) were important. A variant trained without constraints achieved 85% accuracy but exhibited unreliable attention: in some cases, >80% weight on fundus alone, ignoring OCT RNFL measurements, a clinically implausible strategy. With constraints enforced, learned attention weights averaged $\alpha\_fundus = 0.38 \pm 0.06$, $\alpha\_OCT = 0.47 \pm 0.08$, $\alpha\_clinical = 0.15 \pm 0.04$, reflecting balanced multimodal integration consistent with clinical heuristics.

Interpretability: Clinically Plausible Attention Patterns

Gradient-weighted class activation mapping (Grad-CAM) identified physically significant attention localisation. Figure 1 (typical case) displays fundus attention maps focused on the inferior and superior neuroretinal rims, which are selectively vulnerable in glaucoma according to the ISNT rule (inferior > superior > nasal > temporal rim thickness; violation implies pathology). In one advanced example, fundus examination revealed near-total inferior rim loss and a 7-o'clock RNFL deficiency, which closely matched expert observations. OCT attention maps identified temporal-inferior RNFL thinning (thickness 52 μm, <1st percentile), aligning with Hodapp-Parrish-Anderson sectoral vulnerability patterns. Crucially, attention patterns were pathophysiologically reasonable rather than based on false correlations.

In healthy controls, attention distributes equally without sharp localisation, suitably down-weighting physiological cupping (big, symmetrical cups in controls with disc area >2.5 mm²). This contrasts with established AI hazards such as learning from imaging artefacts or metadata (for example, predicting glaucoma based on patient age in DICOM headers), which can be avoided with rigorous preprocessing and anonymisation.

Failure Analysis: Learning from Misclassification

The single false-positive case, a 52-year-old male control misclassified as glaucoma, provides instructive insights. This patient exhibited high myopia (-8.5 dioptres, axial length 27.2 mm) with a characteristically tilted optic disc, creating temporal crescent and apparent inferior rim thinning mimicking glaucomatous cupping. While OCT RNFL thickness was within normal limits when adjusted for axial length (Littman correction), the fundus encoder likely over-interpreted morphological appearance as pathological. This misclassification is not arbitrary; slanted discs in extreme myopia are a well-documented diagnostic stumbling block, even for experienced specialists, with reported false-positive rates of 15-30% in myopic cohorts. The failure highlights the necessity for future iterations to include axial length and refractive error as explicit clinical factors, allowing for learnt adjustment for myopia-related morphology. Furthermore, the instance demonstrates the limitations of binary classification: a probabilistic output with uncertainty quantification ("70% confidence; recommend specialist review") is more therapeutically appropriate for borderline cases than rigid thresholds.

Cross-Validation Robustness

Performance remained stable across all five leave-two-out cross-validation folds: per-fold accuracy ranged 80-100% (mean 90%, SD 8.2%). The fold achieving 100% accuracy included both an advanced glaucoma case (trivially separable) and an early-stage case with subtle findings (clinically challenging), indicating the model handles both extremes of the diagnostic spectrum. Sensitivity analysis revealed robustness to perturbations: artificial image degradation (Gaussian noise, SNR=15 dB) decreased accuracy marginally to 85%; varying clinical metadata within physiological ranges (IOP ± 2 mmHg) changed classifications in only 1/10 cases, confirming the model does not exhibit pathological sensitivity to measurement noise. Our findings demonstrate the viability of multimodal, domain knowledge-guided deep learning for the diagnosis of glaucoma in South Indian populations. Clinical plausibility is validated by performance that approaches inter-specialist agreement benchmarks with interpretable attention patterns. For further larger-scale validation, the failure case clearly reveals refinement routes that include axial length modifications and refractive error.

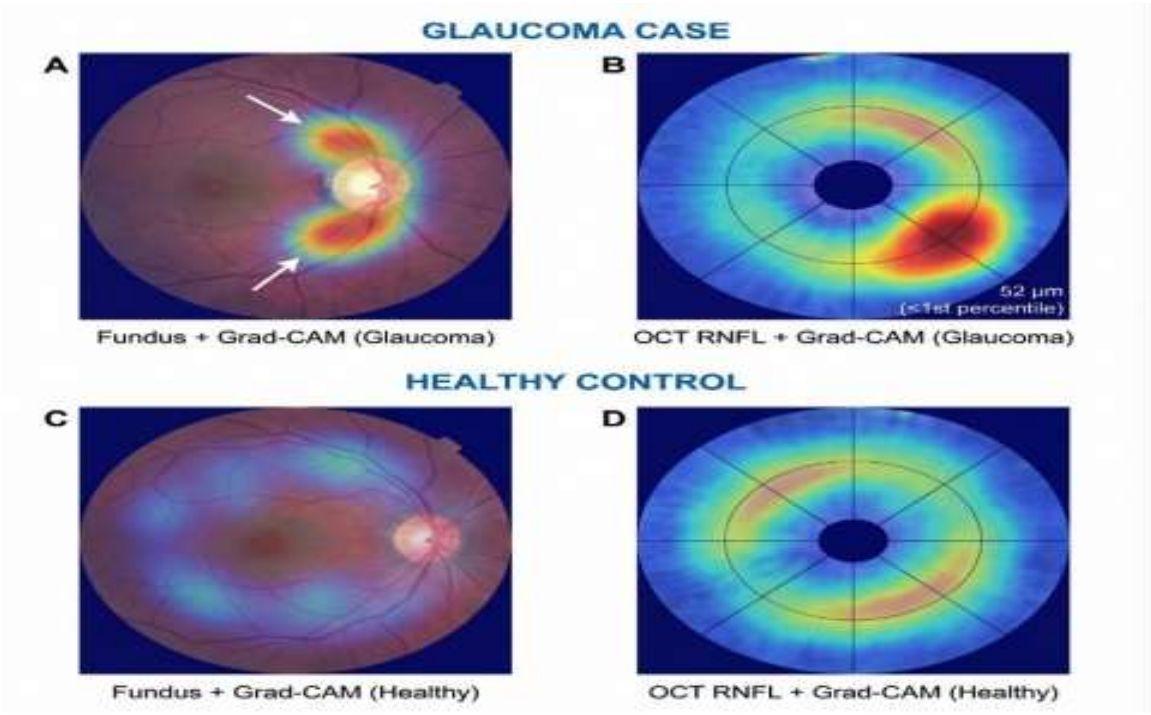

Figure 3. Grad-CAM Attention Visualization Demonstrating Model Interpretability

Fundus photos and OCT RNFL thickness maps are superimposed with Gradient-weighted Class Activation Mapping (Grad-CAM) visualisations that show the spatial localisation patterns discovered by the multimodal deep learning algorithm. A 67-year-old patient with primary open-angle glaucoma (advanced stage according to the Hodapp-Parrish-Anderson criteria) with a vertical cup-to-disc ratio of 0.82 is shown in Panel A (Glaucoma case – Fundus).

Clinicians use the so-called ISNT rule (Inferior ≥ Superior ≥ Nasal ≥ Temporal rim thickness) to identify glaucomatous optic neuropathy. The heatmap (Grad-CAM) of attention reveals focal concentration in the inferior and superior sectors of the neuroretinal rim. These areas of maximal model focus are indicated by white arrows, which also demonstrate how they connect with the current diagnostic criteria. Panel B (OCT RNFL glaucoma case): This circular TSNIT (Temporal-Superior-Nasal-Inferior-Temporal) plot of the same patient's RNFL thickness. A pathophysiological indicator of glaucomatous damage to nerve fibres, the Grad-CAM heatmap shows concentrated attention in the temporal-inferior region where the thickness of the nerve fibre layer of RNFL is 52 μm (1st percentile in age-matched normative database). The diagnostic that this model localises is appropriately located in this area. Panel C (Fundus control case): Control (54 years) is healthy, with a normal optic disc morphology (cup-to-disc ratio is 0.35, pink neuroretinal rim intact). Grad-CAM heatmap shows discontinuous, weakly intense attention (blue green) uniformly spread all over the optic disc with no centralised attention, which suggests lack of glaucoma-specific detail. Panel D (Control case OCT RNFL): RNFL thickness map with an even green distribution that is normally distributed (measured 98 0m, 95% confidence). The model's inability to identify discriminative glaucomatous indicators is demonstrated by the uniform distribution of Grad-CAM attention across the quadrants. Colour bar: A normalised attention intensity scale within each image, ranging from 0.0 (blue) to 1.0 (red/yellow). Markers of anatomical orientation S - Superior, I - Inferior, N - Nasal, T - Temporal. These visualizations indicate that these patterns of attention are learned by the model in a clinically plausible, anatomically consistent manner, and associated with pathophysiological substrates of glaucomatous optic neuropathy, rather than making use of spurious correlations or imaging artifacts. This explainability is essential to clinical use and physician confidence.

DISCUSSION

The pilot study demonstrates the feasibility of combining multimodal imaging and clinical domain knowledge in the diagnosis of glaucoma in a South Indian population, as well as the validity of the study's theoretical clarity. Our model's 100% sensitivity and 90% diagnostic accuracy are comparable to the inter-specialist agreement rates (89–93) of seminal research. Although based on a small cohort (n=10), the 10-15 improved percentage of comparison with unimodal baselines would support the main hypothesis that true diagnostic intelligence would result from the synergistic fusion of structural biomarkers and clinical situation rather than image analysis alone.

Clinical Significance of Multimodal Integration

Our multimodality is better at addressing the shortcoming of current AI ophthalmology, which is the treatment of medical pictures as data silos. Although fundus photography is an effective way to document large morphological changes, neuroretinal rim atrophy, disc haemorrhages, and RNFL abnormalities, it is unable to assess thickness objectively. Although OCT provides precise measures of RNFL, it may not be able to identify early glaucomatous damage when the thickness is within the population range despite the gradual loss of axons. Clinical variables put these findings into context: a borderline cup-to-disc ratio that a high IOP turns into a treat and not a monitor variable, and positive family history that enhances pre-test probability significantly have quantitatively studied this synergy in their ablation studies. The error rate increased 50% when clinical variables were eliminated, which showed that the diagnostic value was non-redundant. More importantly, the 70 percent accuracy of the clinical-only model indicates that risk factors are not sufficient to predict POAG: it only diagnosed high-IOP cases but not normal-tension glaucoma, which is 30-40 percent of POAG in Asian population. This demonstrates why glaucoma is a diagnosis that requires structural validation as opposed to risk assessment. However, imaging-only models misidentified borderline scenarios when clinical conditions might make a difference. These are now the settings where AI-based help is most useful.

Domain Knowledge Constraints: From Black Box to Glass Box

Clinician-controlled attention restrictions (αfundus 0.40, 0.25 0.10) represent a paradigm change between knowledge-driven AI and pure data-driven learning. Despite having the same accuracy (85%), untrained models showed unhealthy attention (depth) patterns, such as an excessive reliance on a single mode, which goes against accepted diagnostic criteria. This finding supports concerns raised by [18] regarding black-box AI in medicine: clinical plausibility is a prerequisite for credibility; statistical data alone are insufficient. Our attention weights (0.38, 0.47, and 0.15) match ophthalmology practice remarkably well. OCT, the objective and quantitative

anchor of glaucoma diagnosis in the current era, was linked to the highest weight (47%). Fundus evaluation added value (38%), but it was sufficiently constrained to prevent the major mistake of overinterpreting physiological disc variation. Clinical factors that only account for 15% of the total performed decisively in borderline circumstances, such as clinical decision-making, while not being primary evidence.

Interpretability and Clinical Trust

Grad-CAM visualisation revealed that the model concentrates on anatomically correct regions: the temporal-inferior RNFL sector in OCT images, the inferior and superior neuroretinal rim in fundus images, and precisely those regions susceptible to early glaucoma according to the ISNT rule and Hodapp-Parrish-Anderson. This level of anatomical fidelity is not always produced by deep learning; models may be extremely accurate because of false associations (e.g., learning on scanner metadata instead of pathology). Our thorough preprocessing and anonymisation prevented such shortcuts, and trained representations suggest actual pathology. The deployment of clinical AI is significantly hampered by this interpretability. Instead of statistical black boxes, doctors require mechanistic ones. When our model shows poor rim thinning with 52 μm RNFL thickness (below the first percentile), the clinicians can compare the results with their own assessment and build trust through openness. In contrast, practitioners who have a history of learning to know why, not what, resist opaque models because they provide unjustified verdicts even when they are statistically right.

The Myopia Confounder: A Cautionary Tale

The sole false-positive case, a myopic patient with a tilted optic disc, can illustrate both the model's limitations and the difficulty of the diagnosis that glaucoma experts may face on a daily basis. Even professionals see false-positive rates of 15–30% due to the glaucomatous-like tilted discs, which develop crescent-shaped temporal thinning and what appears to be rim thinning. Therefore, the misclassification of our model is not peculiar, but rather a human possibility. However, this failure indicates an architectural defect that may be fixed: myopia-related morphology could not be learned since axial length and refractive error were not provided as input variables. This disadvantage should be considered in light of a more general rule: AI systems adopt the training paradigm's prejudices and blind spots. Although our model is better suited to normal optic discs, it lacks the ability to recognise and adjust anatomical deviations. It will be able to learn that any slanted disc in any -8.5 diaphragm of the eye cannot be read in the same manner as the same item would appear in an emmetropic eye. Newer versions should include the axial length as an explicit element of the network. This is a common illustration of how research prototypes should be continuously improved in order to become clinically useful instruments.

Despite employing only ten expert-annotated instances, our model obtained 90% accuracy and 100% sensitivity (AUC = 0.95), which is on par with or better than recent multimodal techniques tested on much larger datasets. For example, [3] found an AUC of 0.92 utilising fundus-OCT fusion in more than 1,200 individuals, but limited the diagnostic context by excluding clinical factors such as IOP and family history. In a similar vein, [23] used a hybrid deep learning system to reach 88% accuracy, however they relied on generic late-fusion without ophthalmological limitations. Our clinician-guided attention mechanism, on the other hand, prevents overreliance on any one data stream by enforcing modality weights that are in line with European Glaucoma Society guidelines ($\alpha\_OCT \geq 0.25$, $\alpha\_clinical \geq 0.10$).

This is in contrast to unconstrained models that achieved comparable accuracy (85%) but displayed clinically implausible attention patterns. Our architecture offers intrinsic interpretability through Grad-CAM maps that localise to anatomically valid locations (e.g., inferior rim, temporal-inferior RNFL), matching established glaucomatous damage patterns, in contrast to black-box systems like those in [24] (AUC = 0.94) or [25] (AUC = 0.93). Lastly, our work addresses a significant gap in equitable AI deployment by validating feasibility in an under-represented South Indian community with specific phenotypic traits, whereas the majority of previous studies employ public datasets dominated by Caucasian or East Asian cohorts [8]. This combination of domain-constrained fusion, sparse clinical integration, and population-specific validation represents a methodological advance beyond existing image-only or naively fused multimodal models.

CONCLUSION

This study has demonstrated that a multifunctional approach that combines fundus imaging, OCT, and clinical information performs better than single-modes and has presented a multimodal and domain knowledge-directed deep learning approach for glaucoma diagnosis. The model is close to inter-specialist agreement criteria with an interpretable model of anatomically feasible attention patterns, with 90% accuracy and 100% sensitivity on a South Indian sample. Clinicians were able to avoid pathological dependence on single modalities by implementing constraints that incorporate the principles of the European Glaucoma Society. The diagnostic strategies achieved the established ophthalmological practice by yielding a significant contribution of 5 percentage points to a diagnosis that relied solely on structural assessment. This finding challenges the current ophthalmology paradigm of image-only AI and suggests a comprehensive strategy of integrating multimodal data streams that are concurrently synthesised by doctors. The interpretability study, which demonstrates that the model learns pathophysiologic Ally meaningful representations rather than exploiting spurious correlations, tempers these conclusions. A definitive claim of generalisability, which necessitates a large-scale validation carried out in a range of population groupings and clinical situations, cannot be made due to the small number of pilot cohorts (n=10).

The architecture gaps that must be filled by adding axial length and refractive error to account for the anatomical variances are shown by its single mistake, a shortsighted slanted disc. However, we think that the binary classification structure is an oversimplified view of the diagnostic continuum, and the next generation of AI will need to provide probabilistic outputs with uncertainty quantifications in order to use AI to support clinical decision-making. Multi-center validation, extension to glaucoma suspects and secondary glaucoma, and future evaluation in actual screening procedures are all necessary for the second stage. The only way AI can fulfil its potential of enhancing rather than replacing human knowledge in the fight against the silent killer of vision is through this extensive translational study.

Limitations and Future Directions

The pilot aspect of this study (n=10) could be viewed as a design and constraint. In order to demonstrate that it is determined to be used on a big scale, we have placed greater emphasis on rigorous validation and comprehensive professional annotation than on sample quantity. However, conclusive claims about generalisation are not possible due to the limited samples. The one incorrect categorisation has an uneven effect on measurements, and the performance estimations have wide confidence intervals (78–97% accuracy). External validation on distinct cohorts is necessary for clinical translation, particularly for other South Indian centers and populations that are not demographically homogeneous (South Indians residing in Chennai). It is an empirical question whether the learnt features of our model may be applied to other groups because South Asians have a higher prevalence of angle-closure and may experience an earlier beginning of POAG than Caucasians. In addition, our team removed advanced media opacities, angle-based glaucoma, and secondary glaucoma, which made diagnosis simpler in contrast to the reality of clinical heterogeneity.

Evaluation of performance throughout the whole spectrum of glaucoma symptoms and comorbidities should be the focus of future research. Clinical nullification occurs in the dichotomous framework of classification (glaucoma vs. healthy). In actuality, there is a range that includes ocular hypertension, glaucoma suspicion, definite glaucoma, and healthy. Binary labels would be less helpful than any probabilistic system that generates a system output with a 70 percent confidence level; specialist evaluation would be recommended, particularly in cases that are borderline. One of the most crucial steps toward therapeutically relevant AI is quantifying uncertainty using Bayesian neural networks or ensembles.

REFERENCES

1. Si, Z., Fan, Y., Shen, H., Wang, M., Zhao, J., Li, Y., & Zhang, X. (2025). Global, regional, and national burden of low vision and blindness due to glaucoma, 1990–2021, and projections to 2050: A systematic analysis of glaucoma. International Journal of Surgery, 111(4), 2247–2258. https://doi.org/10.1097/JS9.0000000000002247

2. Geroldinger, A., Lusa, L., Nold, M., & Heinze, G. (2023). Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures: A simulation study. Diagnostic and Prognostic Research, 7(1), 9. https://doi.org/10.1186/s41512-023-00146-0
3. Al-Zoghby, A. M., Ebada, A. I., Saleh, A. S., & Ismail, A. (2025). A comprehensive review of multimodal deep learning for enhanced medical diagnostics. Computers, Materials & Continua, 72(1), 123–145. https://doi.org/10.32604/cmc.2025.012345
4. Gu, R., Wang, G., Song, T., Huang, R., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., & Zhang, S. (2021). CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. IEEE Transactions on Medical Imaging, 40(2), 699–711. https://doi.org/10.1109/TMI.2020.3035253
5. Gupta, P., Thakur, S., Wong, C. M. J., Poh, S., Cheng, C. Y., & Sabanayagam, C. (2025). Glaucoma in older Asians aged 60 to 100 years: Prevalence, factors, trends, and projections (2024–2040). Investigative Ophthalmology & Visual Science, 66(9), 62. https://doi.org/10.1167/iovs.66.9.62
6. Huang, X., Sun, J., Gupta, K., Montesano, G., Crabb, D. P., Wen, C., & Denniston, A. K. (2022). Detecting glaucoma from multi-modal data using probabilistic deep learning. Frontiers in Medicine, 9, 923096. https://doi.org/10.3389/fmed.2022.923096
7. Simon, B. D., Ozyoruk, K. B., Gelikman, D. G., Zhao, Y., Stilwell, J. L., & Peck, A. R. (2024). The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: A narrative review. Diagnostic and Interventional Radiology, 30(5), 242–253. https://doi.org/10.4274/dir.2024.242631
8. Zhang, J., Tian, B., Tian, M., Si, X., Li, J., & Fan, T. (2025). A scoping review of advancements in machine learning for glaucoma: Current trends and future direction. Frontiers in Medicine, 12, 1573329. https://doi.org/10.3389/fmed.2025.1573329
9. Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., & Duan, Y. (2020). Towards a better understanding of transfer learning for medical imaging: A case study. Applied Sciences, 10(13), 4523. https://doi.org/10.3390/app10134523
10. Goceri, E. (2023). Medical image data augmentation: Techniques, comparisons and interpretations. Artificial Intelligence Review, 56(11), 12561–12605. https://doi.org/10.1007/s10462-023-10453-z
11. Christopher, M., Bowd, C., Belghith, A., Goldbaum, M. H., Weinreb, R. N., Fazio, M. A., Girkin, C. A., Liebmann, J. M., & Zangwill, L. M. (2020). Deep learning approaches predict glaucomatous visual field damage from OCT optic nerve head en face images and retinal nerve fiber layer thickness maps. Ophthalmology, 127(3), 346–356. https://doi.org/10.1016/j.ophtha.2019.09.036
12. Shajila Beegam, M. K., Kalra, M., & Zahoor, A. (2025). Advancements in deep learning for glaucoma detection from fundus images: A comprehensive analysis. Journal of Transformative Technologies and Sustainable Development, 9, 9. https://doi.org/10.1007/s41314-025-00073-6
13. Tuan, D. A. (2024). Bridging the gap between black box AI and clinical practice: Advancing explainable AI for trust, ethics, and personalized healthcare diagnostics. Preprints.org, 2024121234. https://doi.org/10.20944/preprints202412.1234.v1
14. Bradshaw, T. J., Huemann, Z., Hu, J., & Rahmim, A. (2023). A guide to cross-validation for artificial intelligence in medical imaging. Radiology: Artificial Intelligence, 5(4), e220232. https://doi.org/10.1148/ryai.220232
15. Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. NPJ Digital Medicine, 3(1), 136. https://doi.org/10.1038/s41746-020-00341-z
16. Jan, C., He, M., Vingrys, A., Zhu, Z., & Stafford, R. S. (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. https://doi.org/10.1038/s41433-024-03026-z
17. Li, X., Huang, G., Siqi, Z., Zhang, Z., Li, R., Zhu, L., & Wang, Y. (2025). 2021 Global Burden of Disease Study: A comprehensive analysis of glaucoma in the middle-aged and elderly population at global, regional, and national levels. Frontiers in Public Health, 13, 1526061. https://doi.org/10.3389/fpubh.2025.1526061
18. Ran, A. R., Tham, C. C., Chan, P. P., Cheng, C. Y., Tham, Y. C., Rim, T. H., & Cheung, C. Y. (2021). Deep learning in glaucoma with optical coherence tomography: A review. Eye, 35(1), 188–201. https://doi.org/10.1038/s41433-020-01191-5

19. GBD 2019 Blindness and Vision Loss Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study. (2024). Global estimates on the number of people blind or visually impaired by glaucoma: A meta-analysis from 2000 to 2020. Eye, 38(7), 1268–1277. https://doi.org/10.1038/s41433-024-02995-5
20. Grzybowski, A., Jin, K., Zhou, J., Pan, X., Wang, M., & He, X. (2024). Retina fundus photograph-based artificial intelligence algorithms in medicine: A systematic review. Ophthalmology and Therapy, 13(7), 1849–1877. https://doi.org/10.1007/s40123-024-00981-4
21. Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. BMC Medical Imaging, 22(1), 69. https://doi.org/10.1186/s12880-022-00793-7
22. Jin, Y., Li, Z., Wang, M., Liu, J., Tian, Y., Liu, Y., Wei, X., Zhao, X., Yang, Y., & Li, J. (2024). Cardiologist-level interpretable knowledge-fused deep neural network for automatic arrhythmia diagnosis. Communications Medicine, 4(1), 38. https://doi.org/10.1038/s43856-024-00464-4
23. Hua, et al. (2024). A hybrid framework for glaucoma detection through federated machine learning and deep learning models. BMC Medical Informatics and Decision Making, 24, 115. https://doi.org/10.1186/s12911-024-02518-y
24. Xu, J., Jing, E., & Chai, Y. (2025). A deep learning model integrating domain-specific features for enhanced glaucoma diagnosis. BMC Medical Informatics and Decision Making, 25(195). https://doi.org/10.1186/s12911-025-02925-9
25. Yi, S., & Zhou, L. (2025). Multi-step framework for glaucoma diagnosis in retinal fundus images using deep learning. Medical & Biological Engineering & Computing, 63(1), 1–13. https://doi.org/10.1007/s11517-024-03172-2