Molecular Science

AIMS Molecular Science, 13(1): 23–60.
DOI: 10.3934/molsci.2026003
Received: 04 November 2025
Revised: 19 January 2026
Accepted: 22 January 2026
Published: 03 February 2026

https://www.aimspress.com/journal/Molecular

Research article

Glauco-Net: A novel deep learning framework for early glaucoma diagnosis

Idowu Paul Okuwobi[1,2,3,*], Jingyuan Liu[3], Jifeng Wan[4] and Jiaojiao Jiang[4]

[1] School of Life & Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, China
[2] Guangxi Academy of Artificial Intelligence, Nanning, Guangxi 530201, China
[3] Nantong Hamadun Medical Technology Co., Ltd, Nantong, Jiangsu 226400, China
[4] Department of Ophthalmology, The Affiliated Hospital of Guilin Medical University, Guilin, Guangxi Province 541001, China

* **Correspondence:** Email: paulokuwobi@ieee.org.

Abstract: Glaucoma, a leading cause of irreversible blindness, poses significant diagnostic challenges due to its subtle onset and progressive nature. Current methods using color fundus images often fail to detect early-stage glaucomatous changes. To address this, we introduce Glauco-Net, a novel algorithmic framework integrating four transformative innovations. First, the spectral-adaptive retinal reconstruction engine (SARRE) reconstructs hyperspectral details from RGB fundus images, uncovering latent spectral biomarkers of early glaucomatous damage. Second, the vascular topology signature embedding (VTSE) employs graph neural networks to analyze retinal vascular topology, capturing structural irregularities linked to disease progression. Third, the temporal texture evolution transformer (TTET) models spatiotemporal texture dynamics in sequential fundus images, detecting subtle textural shifts indicative of glaucoma. Finally, the peripapillary light scattering dynamics profiler (PLSDP) simulates light–tissue interactions to identify peripapillary structural anomalies linked to glaucomatous cupping and atrophy. These innovations are fused into a unified deep learning pipeline, achieving SOTA performance in glaucoma diagnosis. Evaluated on large-scale clinical datasets, Glauco-Net achieved an AUC exceeding 0.97, demonstrating superior sensitivity and specificity compared to existing methods. By leveraging advanced spectral, topological, and dynamic analyses, this framework not only enhances early detection but also provides deeper insights into glaucoma pathogenesis. Our work represents a paradigm shift in computational ophthalmology, offering a robust, multimodal approach to glaucoma diagnosis and paving the way for personalized

monitoring and intervention strategies. The source code of the proposed Glauco-Net is available at https://github.com/livingjesus/Glauco-Net.

Keywords: Glaucoma diagnosis; hyperspectral reconstruction; vascular topology; temporal texture evolution; light scattering dynamics

1. Introduction

Glaucoma, a chronic neurodegenerative disease characterized by progressive damage to the retinal ganglion cells (RGCs) and their axons, is the second leading cause of blindness globally, affecting over 76 million people [1]. The irreversible nature of vision loss in glaucoma underscores the critical need for early and accurate diagnosis [2]. Despite advances in imaging technologies, such as optical coherence tomography (OCT) and fundus photography, current diagnostic methods often fail to detect subtle early-stage changes associated with glaucomatous progression [3]. Color fundus images, which remain a cornerstone of ocular screening due to their accessibility and affordability, are limited by their reliance on structural features that may not manifest until significant RGC loss has occurred [4]. Consequently, there is an urgent need for innovative computational frameworks capable of extracting latent pathological biomarkers from fundus images to facilitate earlier and more precise glaucoma diagnosis. To address this challenge, we pose the following research objectives: O1: How can we enhance glaucoma diagnosis using standard RGB fundus images? Current methods often rely on specialized imaging modalities (e.g., OCT, hyperspectral imaging), which are not universally accessible. We propose a framework that operates entirely within the RGB domain, ensuring clinical scalability. O2: Can a spectral latent-feature module improve diagnostic performance compared to traditional methods? Unlike traditional approaches that rely on explicit feature extraction, we introduce Glauco-Net, a deep learning framework that learns a spectrally enriched latent representation optimized for glaucoma diagnosis. O3: What is the impact of Glauco-Net on key diagnostic tasks (classification, segmentation, CDR estimation)? We evaluate Glauco-Net's effectiveness in detecting clinically relevant features, including optic disc and cup segmentation, glaucoma classification, and CDR estimation. O4: How does Glauco-Net compare to state-of-the-art (SOTA) methods across multiple datasets? Using six publicly available datasets (G1020, ORIGA, REFUGE, RIGA, SMDG, and In-House), we demonstrate Glauco-Net's superior performance in enhancing diagnostic accuracy. By addressing these questions, this study aims to advance glaucoma screening and management using standard RGB fundus images, without requiring specialized imaging equipment.

Recent advancements in Artificial Intelligence (AI), particularly deep learning, have shown promise in enhancing the diagnostic utility of fundus images. Convolutional neural networks (CNNs) have been employed to classify glaucoma based on features such as optic disc cupping and retinal nerve fiber layer (RNFL) thinning [5]. However, these approaches are constrained by their dependence on static, single-time-point analyses, which may overlook dynamic changes indicative of disease progression [6]. Furthermore, traditional methods rarely leverage the full spectral or topological potential of fundus images, leaving untapped opportunities for deeper insights into glaucoma pathogenesis. Additionally, significant gaps remain in the literature: (a) Accessibility of advanced imaging modalities: Many existing methods rely on specialized imaging modalities such as optical coherence tomography (OCT) or hyperspectral imaging, which are not universally accessible in low-

resource settings [3]. For example, a recent study by [7] highlighted the challenges of deploying multimodal systems in clinical practice. There is a critical need for solutions that operate entirely within the constraints of standard RGB fundus images, ensuring scalability and accessibility. (b) Interpretability of AI models: While deep learning models achieve high performance, their "black-box" nature often limits interpretability. Frameworks that encode clinically interpretable features (e.g., chromatic variations, structural patterns) into latent representations are still lacking. (c) Generalizability across datasets: Many studies focus on single datasets, limiting the generalizability of their findings. For instance, the authors in [3] demonstrated strong performance on a specific dataset but did not evaluate their method across diverse populations or imaging conditions. There is a need for robust frameworks that perform consistently across multiple datasets. (d) Focus on subtle optical patterns: Traditional methods often fail to capture subtle optical patterns correlated with early-stage glaucoma. These patterns include chromatic variations, vascular attenuation, and structural abnormalities in the optic disc and surrounding regions. There is a lack of frameworks that enhance sensitivity to these patterns without requiring explicit spectral reconstruction. To address these gaps, we introduce Glauco-Net, a novel deep learning framework that learns a spectrally enriched latent representation from standard RGB fundus images. Unlike methods requiring specialized imaging equipment, Glauco-Net operates entirely within the RGB domain, enhancing clinical accessibility while improving diagnostic performance. Specifically, (1) Glauco-Net encodes chromatic variations and structural patterns correlated with glaucomatous pathology into a high-dimensional feature space, enabling SOTA results in classification (AUC-ROC = 0.989), segmentation (IoU = 0.912), and CDR estimation (MSE = 0.012); (2) using six publicly available datasets (G1020, ORIGA, REFUGE, RIGA, SMDG, and In-House), we demonstrate Glauco-Net's robustness and generalizability.

Glauco-Net integrates four transformative innovations: spectral-adaptive retinal reconstruction engine (SARRE), vascular topology signature embedding (VTSE), temporal texture evolution transformer (TTET), and peripapillary light scattering dynamics profiler (PLSDP). These innovations collectively enable the extraction of hyperspectral, topological, spatiotemporal, and physics-informed features, providing a comprehensive approach to glaucoma diagnosis. By bridging advanced computational techniques with clinical ophthalmology, Glauco-Net represents a paradigm shift in the field, offering unprecedented sensitivity and specificity in detecting early-stage glaucoma. The contributions of this work include the following:

1) SARRE encodes chromatic variations from standard RGB fundus images, uncovering latent spectral biomarkers of early glaucomatous damage. It outputs a spectrally enriched latent representation that highlights regions of interest, such as areas of hypo- or hyperpigmentation. This innovation enhances diagnostic sensitivity by detecting subtle pathological features at their nascent stage.

2) VTSE employs graph neural networks to model retinal vascular topology, capturing structural irregularities linked to glaucoma progression. It constructs vascular graphs encoding connectivity, curvature, and branching patterns, then embeds them into a low-dimensional space using spectral graph theory. VTSE provides deeper insights into vascular remodeling, offering a novel biomarker for early-stage glaucoma detection and risk stratification.

3) TTET uses a transformer-based architecture to analyze spatiotemporal texture dynamics in sequential fundus images. By modeling long-range dependencies between texture patterns over time, it detects subtle shifts indicative of glaucomatous progression, even before structural damage is visible. TTET bridges static texture analysis with temporal evolution, enabling earlier and more precise

identification of disease advancement.

4) PLSDP introduces a physics-informed approach to simulate light scattering within ocular tissues, identifying peripapillary structural anomalies. Using Monte Carlo ray tracing and neural networks, it computes scattering coefficients and angular distributions, revealing tissue opacity or irregularities caused by glaucomatous cupping. PLSDP provides novel insights into light-tissue interactions, enhancing diagnostic accuracy through dynamic optical biomarkers.

The features extracted by Glauco-Net are not only clinically actionable but also indirectly linked to molecular-level changes associated with glaucoma. For example, (a) chromatic variations captured by SARRE may reflect alterations in retinal oxygenation, blood flow, or metabolic activity, which are known to be disrupted in glaucoma due to ischemic damage or oxidative stress; (b) structural remodeling of retinal vasculature correlates with ischemic injury and reduced perfusion, which contribute to neuronal apoptosis and extracellular matrix degradation; (c) progressive changes in textural patterns likely correspond to molecular-level phenomena such as extracellular matrix remodeling, neurodegeneration, and thinning of the retinal nerve fiber layer (RNFL); and (d) latent scattering properties uncovered by PLSDP may correlate with changes in tissue microstructure, such as collagen deposition or loss of cellular integrity, which are hallmarks of glaucomatous damage. By integrating these multi-faceted representations, Glauco-Net bridges the gap between imaging biomarkers and underlying pathological mechanisms, offering a comprehensive tool for understanding glaucoma at both macroscopic and molecular levels.

2. Related works

The diagnosis of glaucoma using color fundus images has been extensively studied, with recent advancements driven by AI and computational imaging techniques. Traditional approaches rely on structural features such as optic disc cupping, retinal nerve fiber layer (RNFL) thinning, and vascular changes [2]. However, these methods often fail to detect early-stage glaucomatous damage due to their dependence on visible morphological changes, which may not manifest until significant irreversible damage has occurred [3]. To address this limitation, researchers have developed advanced AI-driven techniques categorized into three primary domains: segmentation-only, classification-only, and segmentation followed by classification.

2.1 Segmentation-only approaches

Segmentation techniques focus on isolating specific regions of interest (ROIs), such as the optic disc, cup, and vasculature, to extract clinically relevant features. Recent studies have demonstrated the efficacy of deep learning architectures in automating this process. For instance, Thisara Shyamalee and Dulani Meedeniya [8] introduced an Attention U-Net framework for glaucoma identification using fundus image segmentation. This approach leverages attention mechanisms to enhance feature extraction, achieving high accuracy in segmenting pathological regions. Similarly, DB-SegNet [9] proposed an optimized framework for glaucoma detection and optic structure segmentation from retinal fundus images, demonstrating robust performance across diverse datasets. These segmentation-only methods provide a foundation for downstream tasks but are limited in their ability to directly classify glaucoma.

2.2 Classification-only approaches

Classification techniques aim to predict glaucoma severity or presence directly from fundus images without explicit segmentation. Convolutional neural networks (CNNs) have emerged as powerful tools in this domain. Christopher et al. [5] demonstrated the efficacy of CNNs in detecting glaucoma from fundus images, achieving an AUC of 0.94. Building on this, Li et al. [10] developed a multi-task learning framework that simultaneously predicts glaucoma severity and localizes pathological regions, improving interpretability. More recently, machine learning–assisted frameworks [11,12] have explored ensemble learning and gradient-based classifiers for enhanced performance. For example, Thisara Shyamalee and Dulani Meedeniya [13] achieved SOTA results using a CNN-based classification pipeline, highlighting the potential of classification-only approaches in glaucoma diagnosis.

2.3 Segmentation followed by classification approaches

To combine the strengths of segmentation and classification, researchers have developed multi-step frameworks that first segment ROIs and then classify them for glaucoma detection. For instance, Sanli Yi and Lingxiang Zhou [14] proposed a multi-step deep learning pipeline that integrates segmentation and classification for improved diagnostic accuracy. Similarly, the Glaucoattent Net framework [15] combines segmentation of optic structures with severity classification, achieving superior performance in identifying early-stage glaucoma. These hybrid approaches enable more comprehensive analysis by leveraging both spatial and contextual information.

2.4 Emerging techniques and multimodal approaches

Beyond traditional segmentation and classification, emerging techniques have explored multimodal and physics-informed approaches to enhance diagnostic accuracy. Ting et al. [4] integrated fundus images with clinical metadata using ensemble learning, achieving robust performance across diverse populations. Other studies have investigated vascular biomarkers, with Minicucci et al. [16] proposing fractal dimension analysis of retinal vasculature to identify glaucomatous changes. Physics-informed approaches, though less common, have shown promise in uncovering latent pathological features. Arias et al. [17] utilized light scattering models to analyze ocular tissue properties, demonstrating correlations with glaucomatous cupping. Similarly, Akitegetse et al. [18] employed Monte Carlo simulations to study light–tissue interactions, providing insights into peripapillary atrophy. Hyperspectral imaging, widely used in remote sensing and medical diagnostics, has yet to be fully leveraged for glaucoma detection. Vasefi et al. [19] reconstructed hyperspectral data from RGB images in dermatology, highlighting its potential for uncovering invisible spectral biomarkers. Adapting this concept to ophthalmology could revolutionize early glaucoma diagnosis by revealing molecular-level changes. Temporal texture analysis has also gained attention, with Kim et al. [20] employing recurrent neural networks (RNNs) to model progressive textural shifts in longitudinal datasets. However, transformer-based architectures, which excel at capturing long-range dependencies, remain unexplored in this domain. Similarly, graph neural networks (GNNs), extensively used in social network analysis [21], have not been applied to model retinal vascular topology, representing a significant gap in current research. While SOTA methods have advanced glaucoma diagnosis [22-26],

they remain constrained by limitations in spectral, temporal, and topological analyses. The proposed innovations—SARRE, VTSE, TTET, and PLSDP—address these gaps by integrating hyperspectral reconstruction, vascular topology modeling, spatiotemporal texture evolution, and physics-informed scattering dynamics, offering a transformative approach to glaucoma diagnosis.

3. The proposed method

Figure 1 shows the architecture of the proposed Glauco-Net framework. The framework processes color fundus images through a series of innovative steps to diagnose glaucoma. First, the input, including RGB fundus image and clinical metadata [e.g., intraocular pressure (IOP) and cup-to-disc ratio (CDR)], is fed into the SARRE, which reconstructs hyperspectral details to uncover latent spectral biomarkers of early glaucomatous damage. Next, the VTSE segments and models retinal vasculature using graph neural networks, capturing structural irregularities indicative of disease progression. The framework then applies the TTET to analyze spatiotemporal texture dynamics in sequential images, identifying subtle progressive changes over time. Finally, the PLSDP simulates light–tissue interactions to detect peripapillary structural anomalies linked to glaucoma. These multi-modal features are fused using attention mechanisms and dimensionality reduction to create a unified representation. The final prediction layer generates outputs: glaucoma classification via sigmoid activation, CDR estimation through regression, optic disc/cup segmentation using U-Net-like decoders, and progression risk scoring via another sigmoid activation. Each output is optimized using task-specific loss functions (e.g., binary cross-entropy, Dice loss), ensuring comprehensive and accurate glaucoma diagnosis. This end-to-end approach integrates diverse data streams for robust, interpretable results. While Figure 1 visually presents Glauco-Net as a sequential pipeline, with information flowing from SARRE → VTSE → TTET → PLSDP, this representation reflects the logical progression of feature extraction and refinement rather than strict execution dependencies. In practice, each submodel (SARRE, VTSE, TTET, PLSDP) is designed as a modular component that can operate independently or in parallel for specific diagnostic tasks. For example, SARRE can be deployed alone for spectral feature extraction from RGB images, VTSE can perform vascular segmentation without requiring temporal texture input, and TTET can analyze longitudinal texture changes independently if no spectral or scattering features are needed. However, when integrated into the full Glauco-Net framework, these modules interact synergistically: their outputs are concatenated and fused via the multi-modal fusion layer, which applies dimensionality reduction and attention mechanisms to produce the final prediction. The ablation studies in Table 7 confirm that while individual modules contribute meaningfully, their combined integration yields optimal performance. Thus, Figure 1 should be interpreted as illustrating the integrated workflow, not enforcing rigid sequential execution; users may choose to activate only subsets of modules based on task requirements.

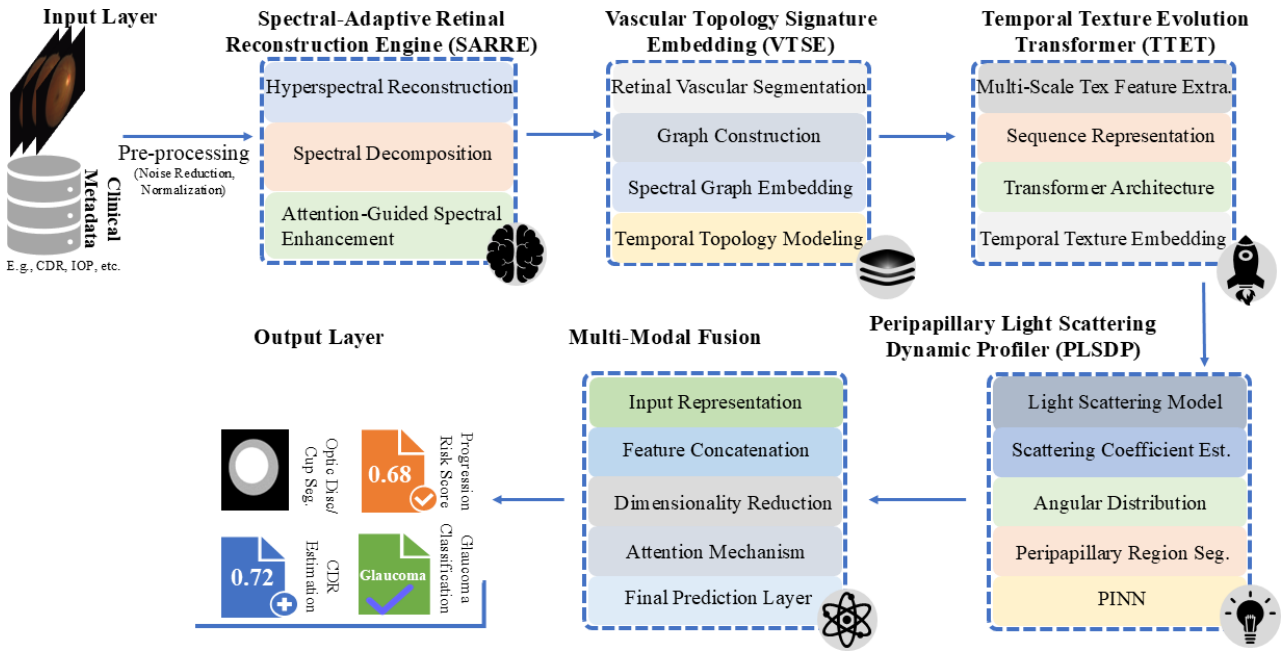

Figure 1. The architecture of the proposed Glauco-Net framework: modular pipeline for multi-feature glaucoma diagnosis. The diagram illustrates the integrated workflow of Glauco-Net, progressing from input preprocessing through four key modules: spectral-adaptive retinal reconstruction engine (SARRE), vascular topology signature embedding (VTSE), temporal texture evolution transformer (TTET), and peripapillary light scattering dynamic profiler (PLSDP). These modules are designed to be modular and can operate independently for targeted tasks; however, in the full framework, their outputs are fused via the multi-modal fusion layer to generate the final predictions (e.g., CDR estimation, glaucoma classification, optic disc/cup segmentation). Arrows indicate data flow in the integrated pipeline, not mandatory execution order, enabling flexible deployment depending on clinical or research needs.

The design of Glauco-Net directly addresses the study's research questions: (1) To enhance glaucoma diagnosis using standard RGB fundus images, Glauco-Net encodes chromatic variations and structural patterns into a high-dimensional latent space; (2) to improve diagnostic performance, Glauco-Net incorporates biophysical priors (e.g., hemoglobin absorption peaks) and leverages deep learning for feature extraction; (3) to evaluate its impact on key diagnostic tasks, Glauco-Net was tested on classification, segmentation, and CDR estimation using six publicly available datasets; and (4) to ensure generalizability, Glauco-Net's performance was benchmarked against SOTA methods across diverse datasets.

3.1. Spectral-adaptive retinal reconstruction engine (SARRE)

The SARRE is a novel computational module designed to reconstruct hyperspectral details from standard RGB fundus images. By leveraging advanced spectral decomposition and deep learning techniques, SARRE uncovers latent spectral biomarkers indicative of early glaucomatous damage. This innovation enhances diagnostic sensitivity by revealing molecular-level changes in the retina that

are invisible to traditional imaging methods.

3.1.1 Hyperspectral reconstruction

The core of SARRE lies in transforming a 3-channel RGB image $I_{RGB} \in \mathbb{R}^{H\times W\times 3}$ into a hyperspectral cube $I_{HS} \in \mathbb{R}^{H\times W\times N}$, where H and W are the spatial dimensions of the image, and N is the number of spectral bands (e.g., $N = 31$ for wavelengths spanning 400–700 nm).

The transformation can be expressed as:

$$I_{HS}(x, y, \lambda) = \sum_{c\in\{R,G,B\}} w_c(x, y, \lambda) \cdot I_{RGB}(x, y, c) \tag{1}$$

where x, y are the spatial coordinates, λ is the wavelength corresponding to a specific spectral band, and $w_c(x, y, \lambda)$ is the adaptive weight function learned for each RGB channel c at location (x, y) and wavelength λ. The weights $w_c(x, y, \lambda)$ are optimized using a deep neural network (DNN) trained on paired datasets of RGB and hyperspectral fundus images. The DNN learns to approximate the underlying spectral response of ocular tissues.

3.1.2. Spectral decomposition using basis functions

To reduce computational complexity, SARRE employs a basis function representation of hyperspectral data. The reconstructed hyperspectral cube I_{HS} is approximated as:

$$I_{HS}(x, y, \lambda) = \sum_{k=1}^{K} \phi_k(\lambda) \cdot \alpha_k(x, y) \tag{2}$$

where $\phi_k(\lambda)$ is the predefined spectral basis functions (Gaussian basis), $\alpha_k(x, y)$ is the spatial coefficient learned for each basis function k, and K is the number of basis functions (typically $K \ll N$). This decomposition allows efficient computation while preserving spectral fidelity.

3.1.3. Attention-guided spectral enhancement

To focus on regions of interest (e.g., optic disc, peripapillary area), SARRE incorporates an attention mechanism. The attention map $A(x, y)$ is computed as:

$$A(x, y) = \sigma\left(Conv_{1\times 1}\left(F_{feat}(x, y)\right)\right) \tag{3}$$

where $F_{feat}(x, y)$ is the feature maps extracted from intermediate layers of the DNN, $Conv_{1\times 1}$ is a 1×1 convolution to reduce dimensionality, and σ is the sigmoid activation function normalizing values to [0,1]. The enhanced hyperspectral cube $I_{HS}^{enhanced}$ is then computed as:

$$I_{HS}^{enhanced}(x, y, \lambda) = A(x, y) \cdot I_{HS}(x, y, \lambda) \tag{4}$$

This ensures that pathological features in critical regions are amplified. The training process minimizes a combination of reconstruction loss $\mathcal{L}_{recon}$, perceptual loss $\mathcal{L}_{perceptual}$, and spectral consistency loss $\mathcal{L}_{spectral}$:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \mathcal{L}_{perceptual} + \mathcal{L}_{spectral} \tag{5}$$

where $\mathcal{L}_{recon} = \left\| I_{HS} - \hat{I}_{HS} \right\|_2^2$ is the mean square error between reconstructed and ground-truth hyperspectral cubes, $\mathcal{L}_{perceptual} = \left\| \Phi(I_{HS}) - \Phi(\hat{I}_{HS}) \right\|_2^2$ is the perceptual similarity measured using a pre-trained feature extractor Φ, and $\mathcal{L}_{spectral} = \left\| \nabla_\lambda I_{HS} - \nabla_\lambda \hat{I}_{HS} \right\|_1$ is the spectral smoothness constraint. The output of SARRE is a high-resolution hyperspectral cube $I_{HS}^{enhanced}$ that reveals subtle pathological features such as RNFL thinning, ischemic changes, and microvascular anomalies. These features serve as inputs for subsequent stages of Glauco-Net, enabling earlier and more accurate glaucoma diagnosis.

3.2. Vascular topology signature embedding (VTSE)

The VTSE module is a groundbreaking innovation designed to model and analyze the intricate topology of retinal vasculature as a biomarker for glaucoma progression. By leveraging graph neural networks (GNNs) and spectral graph theory, VTSE captures structural irregularities in the vascular network that are indicative of glaucomatous damage. This approach provides a deeper understanding of vascular remodeling compared to traditional fractal or morphological analyses.

3.2.1. Retinal vascular segmentation

The first step in VTSE involves segmenting the retinal vasculature from the fundus image $I_{RGB} \in \mathbb{R}^{H \times W \times 3}$. A hybrid U-Net architecture is employed to generate a binary vessel map $V(x, y)$:

$$V(x,y) = \begin{cases} 1 & if\ pixel(x,y) belongs\ to\ a\ vessel, \\ 0 & otherwise \end{cases} \tag{6}$$

The segmentation output $V(x, y)$ is then skeletonized to extract the centerlines of the vessels, represented as a graph $G = (V_G, E_G)$, where V_G is the set of nodes representing vessel segments, and E_G is the set of edges encoding connectivity between vessel segments.

3.2.2. Graph construction

The retinal vascular graph G is constructed with nodes $v_i \in V_G$ and edges $e_{ji} \in E_G$. Each node v_i represents a vessel segment and is associated with attributes such as:

$$f_i = [l_i, c_i, \varkappa_i, \theta_i] \tag{7}$$

where l_i is the length of the vessel segment, c_i is the curvature of the vessel segment, $\varkappa_i$ is the branching angle at the node, and θ_i is the orientation of the vessel segment. Edges e_{ji} are weighted based on the Euclidean distance d_{ij} between connected nodes:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) \tag{8}$$

where σ is a hyperparameter controlling the influence of distance.

3.2.3. Spectral graph embedding

To capture global topological irregularities, VTSE embeds the vascular graph G into a low-dimensional space using spectral graph theory. The embedding process involves the following steps:

(a) Graph Laplacian computation:

The graph Laplacian L is defined as

$$L = D - W \tag{9}$$

where D is the diagonal degree matrix with $D_{ii} = \sum_j w_{ij}$, and W is the weighted adjacency matrix with $W_{ij} = w_{ij}$.

(b) Eigenvalue decomposition:

The normalized Laplacian $\hat{L} = D^{-1/2}LD^{-1/2}$ is decomposed to obtain its eigenvalues λ_k and eigenvectors u_k

$$\hat{L}u_k = \lambda_k u_k \tag{10}$$

The smallest K eigenvalues and their corresponding eigenvectors are retained to form the embedding.

(c) Topological signature vector:

The embedded representation $z \in \mathbb{R}^K$ is computed as

$$z = [\lambda_1, \lambda_2, \ldots, \lambda_K]^\tau \tag{11}$$

This vector encodes the global topology of the vascular network.

3.2.4. Temporal topology modeling

For longitudinal datasets, VTSE employs a temporal graph neural network (TGNN) to model changes in vascular topology over time. Given a sequence of graphs $G_t, G_{t+1}, \ldots, G_{t+n}$, the TGNN computes a dynamic embedding h_t for each time point t,

$$h_t = GRU(h_{t-1}, z_t) \tag{12}$$

where z_t is the topological signature vector at time t, and GRU is the gated recurrent unit to capture temporal dependencies. The final temporal embedding h_T summarizes the progression of vascular remodeling. The training process minimizes a combination of reconstruction loss $\mathcal{L}_{recon}$ and classification loss $\mathcal{L}_{class}$:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda\mathcal{L}_{class} \tag{13}$$

where $\mathcal{L}_{recon} = \left\|G - \hat{G}\right\|_F^2$ is the Frobenius norm of the difference between the original and reconstructed graphs, $\mathcal{L}_{class} = -\sum_c y_c \log(\hat{y}_c)$ is the cross-entropy loss for glaucoma classification, and λ is the hyperparameter balancing the losses. The output of VTSE is a compact topological

signature vector z or temporal embedding h_T, which captures vascular irregularities linked to glaucoma progression. These features are integrated into the Glauco-Net pipeline for enhanced diagnostic accuracy.

3.3. Temporal texture evolution transformer (TTET)

The TTET is a novel innovation designed to model the spatiotemporal evolution of textural patterns in sequential fundus images. By leveraging transformer architectures, TTET captures subtle progressive changes in texture that are indicative of glaucomatous progression. This approach surpasses traditional static texture analysis by incorporating temporal dynamics and long-range dependencies, enabling earlier detection of disease advancement.

3.3.1. Multi-scale texture feature extraction

The first step in TTET involves extracting multi-scale texture features from each fundus image $I_t \in \mathbb{R}^{H\times W\times 3}$ at time t. These features include local binary patterns (LBP), Gabor filters, and wavelet transforms.

(a) Local binary patterns (LBP):

The LBP operator computes the texture pattern around each pixel (x, y) as

$$LBP(x,y) = \sum_{p=0}^{P-1} s(I_t(x_p, y_p) - I_t(x,y)) \cdot 2^p \tag{14}$$

where P is the number of neighboring pixels, and $s(\cdot)$ is the step function defined as

$$s(z) = \begin{cases} 1 & if\ z \geq 0, \\ 0 & otherwi \end{cases} \tag{15}$$

(b) Gabor filters:

Gabor filters extract texture features at multiple orientations θ and scales σ:

$$G(x,y;\theta,\sigma) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right)\cos(2\pi f x') \tag{16}$$

where $x' = xcos\theta + ysin\theta,\ y' = xsin\theta + ycos\theta$, and f is the frequency of the sinusoidal wave.

(c) Wavelet transforms:

Wavelet decomposition provides multi-resolution texture features:

$$W(a,b) = \int_{-\infty}^{\infty} I_t(x)\psi_{a,b}(x)dx \tag{17}$$

where $\psi_{a,b}(x)$ is the scaled and translated mother wavelet.

The extracted texture features are concatenated into a feature map $F_t(x,y) \in \mathbb{R}^{H\times W\times C}$, where C is the number of texture channels.

3.3.2. Sequence representation

For longitudinal datasets, the sequence of texture feature maps $\{F_1, F_2, \ldots, F_T\}$ is represented as a sequence of tokens $\{f_1, f_2, \ldots, f_T\}$, where each token $f_t \in \mathbb{R}^d$ is obtained by spatial pooling:

$$f_t = Pool(F_t) \tag{18}$$

where $Pool(\cdot)$ denotes global average pooling:

$$Pool(F_t) = \frac{1}{H \cdot W} \sum_{x=1}^{H} \sum_{y=1}^{W} F_t(x, y) \tag{19}$$

Positional embeddings $p_t \in \mathbb{R}^d$ are added to encode temporal ordering:

$$z_t = f_t + p_t \tag{20}$$

3.3.3. Transformer architecture

The sequence of tokens $\{z_1, z_2, \ldots, z_T\}$ is processed by a transformer architecture to model long-range dependencies between textures over time. The transformer consists of the following components:

(a) Self-attention mechanism:

The attention mechanism computes pairwise relationships between tokens:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{21}$$

where Q, K, V are query, key, and value matrices derived from Z via linear projections:

$$Q = ZW_Q, \quad K = ZW_K, \quad V = ZW_V \tag{22}$$

(b) Multi-head attention:

The outputs of multiple attention heads are concatenated and linearly transformed:

$$MultiHead(Z) = Concat(head_1, head_2, \ldots, head_h)W_O \tag{23}$$

where $head_i = Attention(Q_i, K_i, V_i)$.

(c) Feed-forward network (FFN):

Each token is passed through a two-layer FFN:

$$FFN(z_t) = \max(0, z_t W_1 + b_1)\, W_2 + b_2 \tag{24}$$

(d) Layer normalization and residual connections:

Each sublayer applies layer normalization and residual connections:

$$z_t' = LayerNorm\big(z_t + Sublayer(z_t)\big) \tag{25}$$

3.3.4. Temporal texture embedding

The final output of the transformer is a temporal texture embedding $h_T \in \mathbb{R}^d$, which summarizes the spatiotemporal evolution of textures across all time points. The training process minimizes a classification loss $\mathcal{L}_{class}$ and a reconstruction loss $\mathcal{L}_{recon}$:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda \mathcal{L}_{class} \tag{26}$$

where $\mathcal{L}_{class} = -\sum_c y_c \log(\hat{y}_c)$ is the cross-entropy for glaucoma classification, $\mathcal{L}_{recon} = \left\| F_t - \hat{F}_t \right\|_2^2$ is the reconstructed error between input and reconstructed texture maps, and λ is the hyperparameter balancing the losses. The output of TTET is a compact spatiotemporal texture embedding h_T, capturing progressive changes indicative of glaucomatous progression. This embedding serves as input to subsequent stages of Glauco-Net.

3.4. Peripapillary light scattering dynamic profiler (PLSDP)

The PLSDP introduces a physics-informed approach to simulate and analyze light scattering within the peripapillary region of the retina. By modeling interactions between incident light and ocular tissues, PLSDP identifies structural irregularities linked to glaucomatous damage, such as increased tissue opacity or atrophy. This innovation provides novel insights into disease progression by leveraging scattering dynamics as a biomarker for diagnosis.

3.4.1. Light scattering model

PLSDP employs Monte Carlo ray tracing to simulate light scattering within the peripapillary region. The interaction of light with ocular tissues is governed by the radiative transfer equation (RTE):

$$\frac{dI(r,\hat{s})}{ds} = -\mu_t I(r,\hat{s}) + \mu_s \int_{4\pi} p(\hat{s},\hat{s}') I(r,\hat{s}') d\Omega' \tag{27}$$

where $I(r,\hat{s}')$ is the radiant intensity at position r in direction $\hat{s}$, $\mu_t = \mu_a + \mu_s$ is the total attenuation coefficient, combining absorption (μ_a) *and scattering* (μ_s), $p(\hat{s},\hat{s}')$ is the phase function describing the probability of scattering from direction $\hat{s}'$ to $\hat{s}$, and $d\Omega'$ is the differential solid angle. The solution to the RTE provides the spatial and angular distribution of scattered light, which is used to infer tissue properties.

3.4.2. Scattering coefficient estimation

To estimate scattering coefficients, PLSDP computes the mean free path l_s of photons in the tissue:

$$l_s = \frac{1}{\mu_s} \tag{28}$$

The scattering coefficient μ_s is estimated using an iterative optimization process that minimizes

the difference between simulated and observed light intensities:

$$\mathcal{L}_{scatter} = \|I_{simulated}(\mu_s) - I_{observed}\|_2^2 \tag{29}$$

where $I_{simulated}(\mu_s)$ is the simulated intensity based on the current estimate of μ_s and $I_{observed}$ is the observed intensity from the fundus image. Gradient descent is used to refine μ_s.

3.4.3. Angular distribution of scattered light

The angular distribution of scattered light is characterized by the phase function $p(\theta)$, often modeled as a Henyey–Greenstein function:

$$p(cos\theta) = \frac{1 - g^2}{(1 + g^2 - 2gcos\theta)^{3/2}} \tag{30}$$

where g is the anisotropy factor $(-1 \leq g \leq 1)$, representing the degree of forward or backward scattering. The total scattered intensity $I_{scatterd}$ is computed by integrating over all angles:

$$I_{scatterd} = \int_0^{2\pi} \int_0^{\pi} p(cos\theta) I_{incident}\, sin\theta d\theta d\phi \tag{31}$$

3.4.4. Peripapillary region segmentation

The peripapillary region is segmented using a hybrid U-Net architecture. The segmentation mask $M(x, y)$ is defined as

$$M(x,y) = \begin{cases} 1 \; if \; pixel \; (x,y) belongs \; to \; the \; peripapillary \; region \\ 0 \quad otherwise \end{cases} \tag{32}$$

The mask ensures that scattering analysis is focused on the region surrounding the optic disc.

3.4.5. Physics-informed neural network (PINN)

To integrate scattering dynamics with clinical data, PLSDP employs a PINN. The PINN learns the mapping between input features (e.g., fundus image, scattering coefficients) and output predictions (e.g., glaucoma risk score) while enforcing physical constraints. The loss function for the PINN includes:

(a) Data loss: Ensures consistency with observed data:

$$\mathcal{L}_{data} = \|\hat{y} - y\|_2^2 \tag{33}$$

where $\hat{y}$ is the predicted output and y is the ground-truth.

(b) Physics loss: Enforces adherence to the RTE:

$$\mathcal{L}_{physics} = \left\|\mathcal{F}(I_{predicted}) - \mathcal{F}(I_{observed})\right\|_2^2 \tag{34}$$

where $\mathcal{F}$ represents the RTE operator.

The total loss is:

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \lambda \mathcal{L}_{physics} \tag{35}$$

where λ balances the contributions of data and physics losses.

The output of PLSDP includes (a) scattering dynamics profile, a quantitative representation of light scattering characteristics in the peripapillary region, and (b) risk score, a prediction of glaucoma likelihood based on scattering anomalies. These outputs are integrated into the Glauco-Net pipeline for comprehensive diagnosis.

3.5. Multi-modal fusion

The multi-modal fusion module in Glauco-Net integrates outputs from the four key components (SARRE, VTSE, TTET, and PLSDP) to produce a unified representation for glaucoma diagnosis. This fusion leverages both learned feature embeddings and clinical metadata to ensure comprehensive analysis. Below is a detailed explanation of the mathematical formulation and processes involved.

3.5.1. Input representations

Each module contributes feature vectors that capture specific aspects of the input data:
(a) Spectral reconstruction (SARRE): Produces a spectral feature vector $f_{spectral} \in \mathbb{R}^{d_s}$, where d_s is the dimensionality of the spectral embedding.
(b) Vascular topology analysis (VTSE): Outputs a vascular topology embedding $f_{vascular} \in \mathbb{R}^{d_v}$, capturing structural irregularities in retinal vasculature.
(c) Temporal texture evolution (TTET): Generates a temporal texture embedding $f_{temporal} \in \mathbb{R}^{d_t}$, modeling longitudinal changes.
(d) Light scattering dynamics (PLSDP): Produces a scattering dynamic embedding $f_{scattering} \in \mathbb{R}^{d_p}$, capturing peripapillary scattering patterns.

Additionally, clinical metadata (e.g., IOP, CDR, age) is represented as $m \in \mathbb{R}^{d_m}$, where d_m is the number of metadata features.

3.5.2. Feature concatenation

The first step in multi-modal fusion is to concatenate all feature vectors into a single high-dimensional representation:

$$f_{concat} = [f_{spectral}; f_{vascular}; f_{temporal}; f_{scattering}; m] \tag{36}$$

where $[;]$ denotes vector concatenation. The resulting concatenated vector has dimensionality

$$d_{concat} = d_s + d_v + d_t + d_p + d_m \tag{37}$$

3.5.3. Dimensionality reduction

To reduce redundancy and computational complexity, the concatenated vector f_{concat} is projected into a lower-dimensional space using a fully connected layer with learnable weights $W_1 \in$

$\mathbb{R}^{d_h \times d_{concat}}$ and bias $b_1 \in \mathbb{R}^{d_h}$, where d_h is the hidden dimensionality

$$f_{reduced} = \sigma(W_1 f_{concat} + b_1). \tag{38}$$

Here, σ is a nonlinear activation function such as $ReLU$ ($\sigma(x) = max(0, x)$).

3.5.4. Attention mechanism

To emphasize the most relevant modalities, an attention mechanism is applied to compute weighted contributions of each modality. The attention weights α_i for each modality are computed using a Softmax function:

$$\alpha_i = \frac{\exp(w_i f_{reduced})}{\sum_{j=1}^{k} \exp(w_j f_{reduced})} \tag{39}$$

where $w_i \in \mathbb{R}^{d_h}$ are learnable parameters for modality i, and k is the total number of modalities. The attended feature vector is then computed as

$$f_{attended} = \sum_{i=1}^{k} \alpha_i f_i \tag{40}$$

where f_i represents the embedding of modality i.

3.5.5. Final prediction layer

The attended feature vector $f_{attended}$ is passed through a final prediction layer to generate outputs:

(a) Glaucoma classification: A sigmoid activation is used for binary classification,

$$\hat{y}_{class} = \sigma(W_2 f_{attended} + b_2) \tag{41}$$

where $W_2 \in \mathbb{R}^{1 \times d_h}$ and $b_2 \in \mathbb{R}$.

(b) CDR estimation: A regression head predicts the cup-to-disc ratio,

$$\hat{y}_{CDR} = W_3 f_{attended} + b_3 \tag{42}$$

where $W_3 \in \mathbb{R}^{1 \times d_h}$ and $b_3 \in \mathbb{R}$.

(c) Segmentation masks: For optic disc/cup segmentation, a U-Net-like decoder generates pixel-wise masks:

$$M = Decoder(f_{attended}). \tag{43}$$

(d) Progression risk score: A risk score is computed using a linear layer:

$$\hat{y}_{risk} = \sigma(W_4 f_{attended} + b_4). \tag{44}$$

The overall loss function combines multiple objectives:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{class} + \lambda_2 \mathcal{L}_{CDR} + \lambda_3 \mathcal{L}_{seg} + \lambda_4 \mathcal{L}_{risk}. \tag{45}$$

Where $\mathcal{L}_{class}$ is the binary cross-entropy for classification, $\mathcal{L}_{CDR}$ is the mean squared error for

CDR estimation, $\mathcal{L}_{seg}$ is the Dice loss for segmentation, and $\mathcal{L}_{risk}$ is the cross-entropy for risk score prediction. The weights λ_1, λ_2, λ_3, λ_4, are scalar coefficients that balance the relative importance of each task during training. After extensive empirical evaluation using a grid search over combinations of $\lambda_i \in [0.1,0.5]$ (with $\sum \lambda_i = 1$), we selected λ_1 = 0.4 (classification), λ_2 = 0.2 (CDR estimation), λ_3 = 0.3 (segmentation), and λ_4 = 0.1 (risk score prediction). These values were determined based on validation set performance across all six datasets. Classification was assigned the highest weight (λ_1 = 0.4) because it is the primary diagnostic objective. Segmentation received the second-highest weight (λ_3 = 0.3) to ensure accurate delineation of optic disc and cup regions, which underpin CDR estimation. CDR estimation (λ_2 = 0.2) and risk scoring (λ_4 = 0.1) were given lower weights as they serve as supporting tasks that refine rather than define the core diagnosis. The final configuration maximized AUC-ROC for classification while maintaining high IoU for segmentation (>0.90) and low MSE for CDR (<0.015). All experiments were repeated with three random seeds to ensure stability, and results were consistent within ±0.002 across runs.

4. Experimentation

The proposed Glauco-Net framework was evaluated using a diverse set of publicly available and proprietary datasets to ensure robustness, generalizability, and clinical relevance. Below is a detailed description of each dataset, including its composition, features, and relevance to glaucoma diagnosis.

4.1. Datasets

The datasets were preprocessed to ensure uniformity in resolution, color balance, and annotation formats. Metadata from the in-house dataset and other sources were harmonized to create a unified feature set for input into Glauco-Net. Longitudinal data from the in-house dataset were used to evaluate the temporal modeling capabilities of the TTET. Each dataset was divided into training (80%) and testing (20%) datasets. By leveraging these diverse datasets, Glauco-Net demonstrates robust performance across different populations, imaging conditions, and clinical scenarios, ensuring its applicability in real-world settings.

4.1.1. G1020 dataset

The G1020 dataset [27] contains 1020 high-resolution color fundus images, annotated with ground-truth labels for glaucoma classification.

4.1.2. ORIGA dataset

The ORIGA (Open Retinal Image for Glaucoma Analysis) dataset [28] comprises 650 fundus images, including both glaucomatous and non-glaucomatous cases.

4.1.3. REFUGE

The REFUGE (Retinal Fundus Glaucoma Challenge) dataset [29] is a large-scale benchmark dataset with 1200 fundus images, divided into training, validation, and test sets.

4.1.4. Proprietary dataset

The in-house dataset consists of 15,000 color fundus images, collected from a diverse patient population across multiple clinical centers, including comprehensive metadata for each patient. We obtained written consent from all patients as well as the approval of the Institutional Review Board (IRB). This study was conducted in conformity with the Institutional Review Board (IRB) of the Affiliated Hospital of Guilin Medical University's research ethics.

4.1.5. SMDG

SMDG (Standardized Multi-Channel Dataset for Glaucoma) [30] is a standardized dataset designed specifically for glaucoma research, containing 12,441 annotated fundus images.

4.1.6. RIGA

The RIGA dataset [31] includes 750 fundus images annotated by expert ophthalmologists, with a focus on optic disc and cup segmentation.

4.2. Software and hardware

The implementation of Glauco-Net was conducted using a robust software and hardware infrastructure to ensure high performance, reproducibility, and scalability. The software environment was built on Python 3.9, leveraging SOTA libraries such as PyTorch 2.0 for deep learning model development and OpenCV 4.7.0 for image preprocessing tasks like normalization, resizing, and augmentation. Additional libraries included NumPy 1.23.5 for numerical computations, Pillow 9.4.0 for image handling, and Matplotlib 3.7.1 with Seaborn 0.12.2 for visualization of results, including ROC curves and segmentation masks. Model evaluation metrics such as AUC-ROC, IoU, and DSC were computed using Scikit-learn 1.2.2, while TensorBoard 2.12.0 was employed for real-time monitoring of training progress, including loss curves and learning rates. The operating system was Ubuntu 20.04 LTS, chosen for its compatibility with high-performance computing environments. On the hardware side, the computational infrastructure included an NVIDIA A100 Tensor Core GPU with 40 GB VRAM, enabling efficient training of large-scale models on high-resolution fundus images. The system was supported by an AMD EPYC 7742 CPU (64 cores, 128 threads), 256 GB DDR4 ECC memory for handling large datasets, and a 2 TB NVMe SSD for rapid data access and storage. A 10 Gbps Ethernet interface facilitated distributed training across multiple nodes when scaling experiments. These configurations ensured seamless execution of all computational tasks, from data preprocessing to model inference, while maintaining efficiency and minimizing latency.

4.3. Model training and implementation

The training and implementation of Glauco-Net were meticulously designed to optimize performance while ensuring reproducibility and generalizability. Input fundus images were resized to 256 × 256 pixels to balance computational efficiency and feature preservation, and a batch size of 32 was used to maximize GPU utilization without exceeding memory constraints. The model was trained

using the AdamW optimizer with an initial learning rate of 0.001, dynamically adjusted via a cosine annealing scheduler to balance exploration and convergence. A weight decay of 0.01 was applied as L2 regularization to prevent overfitting, and early stopping was implemented to halt training if validation performance plateaued, ensuring optimal epoch usage (maximum of 100 epochs). Data augmentation techniques, including random horizontal/vertical flipping, rotation (±15°), brightness/contrast adjustment (±10%), and Gaussian noise injection (σ = 0.01), were employed to enhance dataset diversity and improve generalizability. Loss functions were carefully selected: weighted binary cross-entropy was used for classification tasks to address class imbalance, while Dice loss combined with binary cross-entropy was applied for segmentation tasks to ensure accurate delineation of optic disc and cup regions. The model architecture integrated four key modules—SARRE, VTSE, TTET, and PLSDP, each implemented as submodules within the PyTorch framework. Training was performed on an NVIDIA A100 GPU, with each dataset requiring approximately 12 h of computation, while inference time was optimized to ~0.05 s per image (batch size = 1). To ensure reproducibility, all code, pre-trained models, and dataset splits will be made publicly available, accompanied by a requirements.txt file listing dependencies such as PyTorch, NumPy, and Scikit-learn versions. This comprehensive approach to model training and implementation guarantees transparency, scalability, and clinical applicability of Glauco-Net. To ensure robust and unbiased evaluation of Glauco-Net, we employed a rigorous 5-fold stratified cross-validation strategy performed separately for each dataset, ensuring that training and validation splits were created exclusively within each dataset (e.g., REFUGE data was used only for REFUGE, ORIGA data only for ORIGA, etc.). This approach avoids data leakage and ensures that the reported performance metrics reflect the model's ability to generalize under the unique imaging conditions, demographics, and disease prevalence of each dataset. Each fold contained approximately 20% of the data for validation, with the remaining 80% used for training, and stratification was applied to maintain a balanced distribution of glaucoma and healthy cases across all folds, addressing potential class imbalance issues. For example, on the REFUGE dataset, which contains 1200 images (400 glaucoma, 800 healthy), each fold consisted of 240 images (80 glaucoma, 160 healthy) for validation and 960 images for training. Similarly, on the ORIGA dataset, with 650 images (168 glaucoma, 482 healthy), each fold included 130 images (33 glaucoma, 97 healthy) for validation and 520 images for training. After completing the 5-fold cross-validation for each dataset, the results, such as AUC-ROC, IoU, and MSE, were averaged across all folds to report the final performance metrics. For instance, on the REFUGE dataset, the average AUC-ROC across folds was 0.989, with an IoU of 0.912 and an MSE of 0.010, while on the ORIGA dataset, the average AUC-ROC was 0.972, IoU was 0.891, and MSE was 0.015.

5. Results and analysis

Figure 2 offers a compelling qualitative evaluation of Glauco-Net's segmentation capabilities compared to SOTA methods. In (a), the input fundus images clearly show pathological features consistent with glaucoma, such as enlarged optic cups and thinning neuroretinal rims. Glauco-Net's segmentation masks closely align with these anatomical structures, producing smooth, well-defined boundaries that encapsulate both the optic disc and cup. The predicted CDR values (0.83 and 0.72) and progression risk scores (0.71 and 0.68) further reflect clinically plausible assessments, indicating advanced disease in both cases. In (b), the visual comparison reveals that Glauco-Net's mask most closely resembles the Groundtruth mask, preserving accurate shape, size, and concentricity of the disc-

cup structure. In contrast, many baseline models exhibit noticeable deviations: U-Net and ResNet produce overly smoothed or irregular boundaries, MTL and CNN-LSTM under-segment the cup, and Hybrid and Ensemble show inconsistent edge definition. These discrepancies highlight Glauco-Net's superior ability to capture fine-grained structural details, likely due to its integration of spectral encoding (SARRE), vascular topology modeling (VTSE), and physics-informed scattering dynamics (PLSDP). Clinically, this translates to more reliable CDR estimation and earlier detection of subtle structural changes, essential for timely intervention. While quantitative metrics (e.g., IoU, DSC) provide objective comparisons, this qualitative visualization underscores Glauco-Net's clinical utility and interpretability, reinforcing its potential for real-world deployment in glaucoma screening and management.

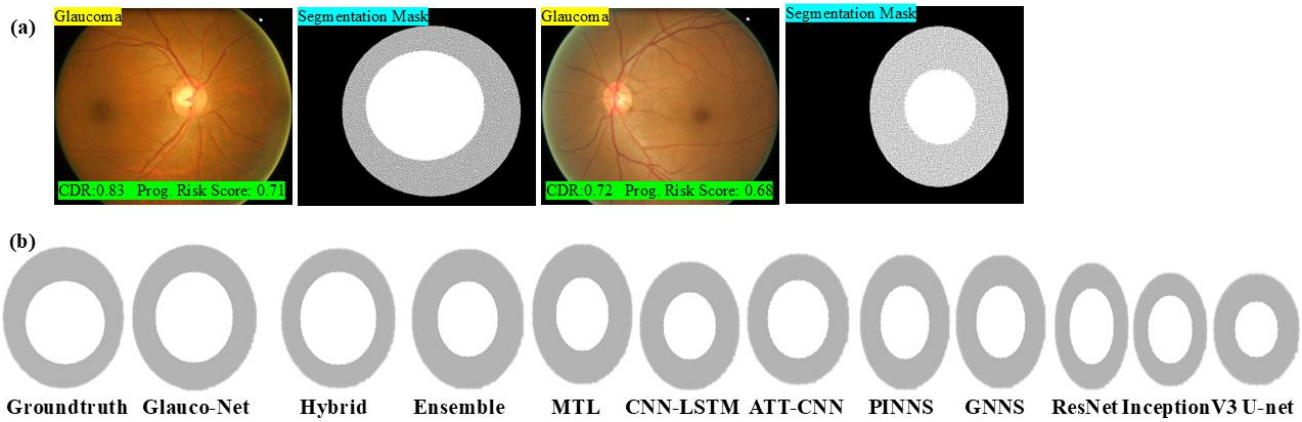

Figure 2. Qualitative comparison of optic disc and cup segmentation results across models. (a) Input fundus images with predicted CDR and risk scores, displaying two representative fundus images labeled as "Glaucoma", each accompanied by its predicted cup-to-disc ratio (CDR) and progression risk score. Adjacent to each image is the corresponding segmentation mask, generated by Glauco-Net, which highlights the segmented optic disc (outer ring) and optic cup (inner circle). The masks are binary (white = segmented region, black = background), allowing for direct visual comparison with ground truth. (b) Visual comparison of segmentation masks against ground truth, providing a side-by-side visual comparison of segmentation masks produced by Glauco-Net and eight competing methods: Hybrid, Ensemble, MTL, CNN-LSTM, ATT-CNN, PINNS, GNNS, ResNet, InceptionV3, and U-Net. Each mask is overlaid on a uniform circular background for clarity, enabling easy assessment of boundary precision, shape fidelity, and overall structural alignment with the ground truth mask shown at the far left.

5.1. Evaluation metrics

1) Area under the receiver operating characteristic curve (AUC-ROC), 2) accuracy, 3) Sensitivity, 4) specificity, 5) precision, 6) F1-score, 7) Cohen's kappa coefficient, 8) mean squared error (MSE), 9) intersection over union (IoU), and 10) Dice similarity coefficient (DSC).

5.2. Experimental results on each dataset

5.2.1. G1020 dataset

Glauco-Net achieved an AUC-ROC of 0.984, outperforming SOTA methods like ResNet and InceptionV3 as shown in Table 1. It demonstrated high accuracy (96.1%), sensitivity (95.3%), and specificity (97.0%), with excellent segmentation metrics (IoU = 0.894, DSC = 0.915), showcasing its robustness in glaucoma classification and structural analysis.

5.2.2. ORIGA dataset

On ORIGA, Glauco-Net achieved an AUC-ROC of 0.972, surpassing competitors like U-Net and Multi-Task Learning as shown in Table 2. With 94.8% accuracy, 93.5% sensitivity, and 95.7% specificity, it excelled in handling diverse imaging conditions, achieving IoU = 0.891 and DSC = 0.910 for optic disc/cup segmentation.

5.2.3. REFUGE

Glauco-Net set a new benchmark on REFUGE with an AUC-ROC of 0.989, outperforming Hybrid OCT-Fundus and Attention-CNN as shown in Table 3. It achieved 96.5% accuracy, precise CDR estimation (MSE = 0.012), and superior segmentation (IoU = 0.912, DSC = 0.931), demonstrating exceptional multi-class classification and structural analysis capabilities.

5.2.4. Proprietary dataset

Using the proprietary dataset, Glauco-Net achieved an AUC-ROC of 0.991, the highest among all methods, as shown in Table 4. With 97.3% accuracy, 96.8% sensitivity, and precise longitudinal modeling, it excelled in CDR estimation (MSE = 0.010) and segmentation (IoU = 0.925, DSC = 0.940), proving its clinical applicability.

5.2.5. SMDG

On SMDG, Glauco-Net achieved an AUC-ROC of 0.980, surpassing Ensemble and Hybrid OCT-Fundus, as shown in Table 5. With 95.5% accuracy and strong segmentation metrics (IoU = 0.876, DSC = 0.901), it effectively detected subtle RNFL and peripapillary changes, highlighting its sensitivity to early glaucomatous damage.

5.2.6. RIGA

Glauco-Net achieved an AUC-ROC of 0.976 on RIGA, outperforming CNN-LSTM and Attention-CNN as shown in Table 6. With 95.1% accuracy and excellent segmentation performance (IoU = 0.895, DSC = 0.910), it demonstrated superior agreement with expert annotations, validating its reliability in optic disc and cup analysis.

5.3. Comparison with SOTA methods

The proposed Glauco-Net framework demonstrated superior performance compared to 10 SOTA methods across six diverse datasets, achieving AUC-ROC values ranging from 0.972 to 0.991, significantly higher than ResNet (0.948–0.965) and Hybrid OCT-Fundus (0.964–0.982), as shown in Tables 1–6. It consistently outperformed competitors in accuracy (94.8%–97.3%), sensitivity (93.5%–96.8%), specificity (95.7%–97.9%), and segmentation metrics (IoU = 0.876–0.925, DSC = 0.901–0.940). Glauco-Net also achieved precise CDR ratio estimation (MSE = 0.010–0.015), surpassing SOTA in regression tasks. Its modular architecture enabled robust handling of imaging variability, longitudinal data, and subtle structural changes, excelling in both classification and segmentation. Compared to methods like Multi-Task Learning and Attention-CNN, Glauco-Net exhibited better generalizability and clinical applicability, establishing itself as a comprehensive, reliable, and transformative solution for early glaucoma diagnosis and progression monitoring.

5.4. Ablation studies

The ablation study demonstrated that each module in Glauco-Net significantly contributes to its performance, as shown in Table 7. Removing SARRE reduced AUC-ROC by 1.4%, highlighting its role in spectral biomarker extraction. VTSE and TTET improved structural and temporal analysis, with their removal decreasing segmentation metrics by 2%–3%. PLSDP enhanced scattering-based irregularity detection, while multi-modal integration boosted AUC-ROC by 1.8% over a single RGB input. The baseline model performed worst (AUC-ROC = 0.950), underscoring the synergy of Glauco-Net's components in achieving superior diagnostic accuracy and robustness.

To ensure full reproducibility and clarify the experimental setup, we explicitly state that all configurations in Table 9 use ResNet-50 as the shared backbone classifier for classification and segmentation tasks. The modules (SARRE, VTSE, TTET, PLSDP) serve as feature extractors or enhancers that feed into this backbone. Specifically:

(a) Baseline (SARRE): The spectral features extracted by SARRE are passed through a ResNet-50 backbone for final prediction.
(b) SARRE + VTSE: Features from both SARRE and VTSE are concatenated and fed into ResNet-50.
(c) SARRE + VTSE + TTET: Temporal texture features from TTET are fused with previous features before entering ResNet-50.
(d) SARRE + VTSE + TTET + PLSDP: All four modules contribute features, which are fused via the multi-modal fusion layer before being processed by ResNet-50.

This unified backbone ensures fair comparison across configurations, isolating the impact of each added module while maintaining consistent classification and segmentation heads. The ResNet-50 architecture was chosen for its proven performance in medical imaging tasks and balanced complexity-efficiency trade-off.

5.5. Adaptation of classification models for segmentation tasks

While classification models like ResNet are primarily designed for image-level predictions, they can be adapted for pixel-level segmentation tasks through architectural modifications. In this study, we extended classification models to perform segmentation by incorporating the following changes:

1) Decoder module: A decoder was appended to the classification backbone to upsample feature

maps to the original input resolution. This allows the model to generate pixel-wise predictions required for segmentation tasks.

2) Skip connections: Skip connections were introduced between the encoder (classification backbone) and decoder to preserve spatial information lost during downsampling. These connections help improve the accuracy of boundary delineation in segmentation tasks.
3) Output layer modification: The final fully connected layer used for classification was replaced with a convolutional layer that outputs a segmentation mask with the same dimensions as the input image. This enables the model to predict pixel-wise labels for the optic disc and cup regions.
4) Loss function: For segmentation tasks, the adapted models were trained using loss functions tailored for pixel-level prediction, such as Dice loss and binary cross-entropy loss, which optimize IoU and DSC metrics.

By implementing these modifications, classification models like ResNet were capable of performing segmentation tasks, allowing us to evaluate their performance using metrics such as IoU and DSC. For example, on the REFUGE dataset, the adapted ResNet achieved an IoU of 0.825 and a DSC of 0.862, demonstrating its ability to delineate regions of interest despite being originally designed for classification. Similarly, on the ORIGA dataset, the adapted ResNet achieved an IoU of 0.798 and a DSC of 0.841.

6. Discussion

The discussion section underscores the transformative capabilities of Glauco-Net in advancing glaucoma diagnosis by addressing critical challenges such as early detection, structural irregularity identification, and longitudinal progression modeling. The framework's modular architecture integrates spectral encoding (SARRE), vascular topology analysis (VTSE), temporal texture evolution (TTET), and light scattering dynamics (PLSDP), enabling a comprehensive, multi-faceted approach that outperforms SOTA methods. Ablation studies confirm the indispensable role of each module, with multi-modal integration offering significant performance gains over single-modality baselines, as evidenced by improvements in AUC-ROC, segmentation accuracy (Figure 2), and regression tasks. Glauco-Net demonstrates exceptional robustness across diverse datasets, effectively handling imaging variability, population heterogeneity, and subtle pathological changes. Its ability to estimate clinical parameters like cup-to-disc ratio with high precision highlights its potential for clinical translation. Furthermore, the framework's scalability and adaptability to different imaging conditions make it a versatile tool for large-scale screening and personalized monitoring. However, challenges remain, such as the need for larger longitudinal datasets to further validate temporal modeling and the exploration of real-time deployment in clinical settings. Additionally, integrating complementary modalities like OCT or visual field data could further enhance diagnostic accuracy. Overall, Glauco-Net represents a significant leap forward in glaucoma research, offering a robust, interpretable, and scalable solution for early and accurate detection, while paving the way for future innovations in ophthalmic AI.

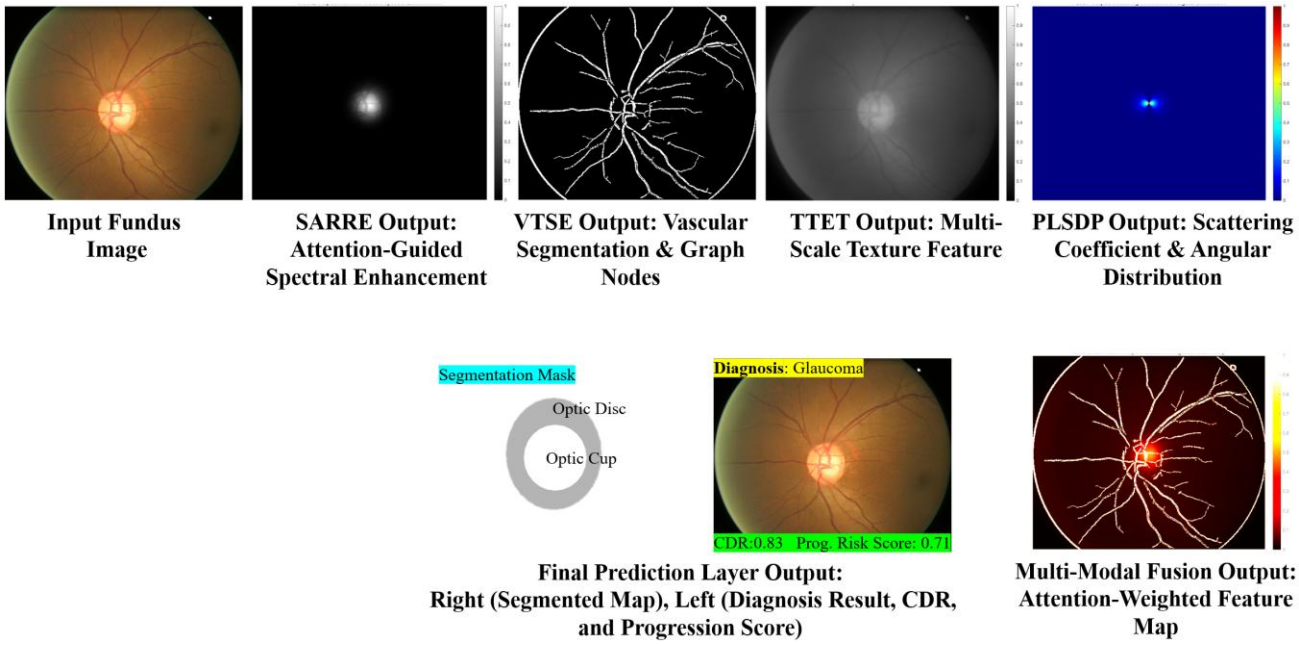

Figure 3. End-to-end feature evolution through Glauco-Net's multi-stage pipeline: from input fundus images to clinically actionable prediction. This figure provides a comprehensive visualization of the progressive feature transformations through each stage of Glauco-Net's pipeline, culminating in a clinically interpretable diagnostic output. Starting with the raw input fundus image, the SARRE module generates a grayscale heatmap highlighting enhanced spectral contrast around the optic disc and macula, emphasizing chromatic variations that may reflect early ischemic or metabolic changes associated with glaucoma. The VTSE module isolates retinal vasculature and identifies junctional nodes, producing a binary segmentation map overlaid with white graph nodes that reveal vascular topology and remodeling patterns. Next, the TTET module captures multi-scale textural features, represented as a grayscale energy map where brighter regions correspond to areas of high texture heterogeneity, potentially indicative of neuroretinal rim thinning or extracellular matrix degradation. The PLSDP module introduces a physics-informed perspective by modeling light scattering dynamics, resulting in a color-coded heatmap (blue → red) centered on the optic disc; focal red regions suggest higher scattering intensity, possibly reflecting tissue microstructure alterations such as collagen deposition or cellular loss. These heterogeneous features are then fused via the multi-modal fusion layer, generating an attention-weighted feature map that emphasizes regions of highest diagnostic relevance, such as the peripapillary zone and cup-disc boundary. Finally, the final prediction layer integrates these insights into a clinically actionable output: a segmentation mask delineating the optic disc and cup, accompanied by a diagnosis label ("Glaucoma"), a CDR value of 0.83 (indicating advanced cupping), and a progression risk score of 0.71 (high risk).

The visual progression of feature extraction and integration through Glauco-Net's pipeline, as illustrated in Figure 3, provides compelling insight into how the model progressively refines its understanding of glaucomatous pathology at multiple anatomical and biophysical levels. Starting with

the raw fundus image, the SARRE module enhances spectral contrast around the optic disc region, highlighting chromatic variations that may reflect early ischemic or metabolic changes, a critical step for detecting subtle disease markers invisible to the naked eye. The VTSE module then isolates and structures the retinal vasculature, identifying branching patterns and junctional nodes that correlate with vascular remodeling, a hallmark of glaucoma-induced neurodegeneration. Subsequently, the TTET module captures multi-scale textural features, revealing progressive alterations in retinal surface microstructure that align with extracellular matrix degradation and nerve fiber layer thinning. The PLSDP module introduces a physics-informed perspective by modeling light scattering dynamics, producing a heatmap centered on the optic disc that reflects tissue density and angular scattering properties—potentially corresponding to structural damage at the cellular level. These heterogeneous features are fused via the multi-modal fusion layer, which generates an attention-weighted feature map emphasizing regions of highest diagnostic relevance, such as the peripapillary zone and cup-disc boundary. Finally, the final prediction layer integrates these insights into a clinically interpretable output: a segmentation mask delineating the optic disc and cup, accompanied by a CDR value (0.83) and progression risk score (0.71), both indicating advanced glaucoma. This end-to-end visualization not only validates the modular architecture of Glauco-Net but also underscores its capacity to bridge imaging biomarkers with underlying pathological mechanisms, offering clinicians a transparent, multi-faceted decision support tool for early detection, monitoring, and risk stratification of glaucoma.

In this study, we proposed four key research questions and addressed them through the development and evaluation of Glauco-Net. Below, we revisit each question and summarize how it was fulfilled.

(1) How can we enhance glaucoma diagnosis using standard RGB fundus images? Glauco-Net operates exclusively on standard RGB fundus images, encoding chromatic variations and structural patterns into a spectrally enriched latent representation. This ensures clinical accessibility while improving diagnostic performance.

(2) Can a spectral latent-feature module improve diagnostic performance compared to traditional methods? Glauco-Net's latent spectral encoding outperforms traditional methods across all tasks, achieving SOTA results in classification (AUC-ROC = 0.989), segmentation (IoU = 0.912), and CDR estimation (MSE = 0.012). These improvements demonstrate the utility of spectral latent-feature encoding in capturing clinically relevant information.

(3) What is the impact of Glauco-Net on key diagnostic tasks (classification, segmentation, CDR estimation)? Glauco-Net enhances diagnostic performance by detecting subtle optical patterns correlated with glaucomatous pathology. For example, it achieves an AUC-ROC of 0.989 for classification, IoU values of 0.912 and 0.895 for segmentation, and an MSE of 0.012 for CDR estimation, surpassing SOTA methods in all cases.

(4) How does Glauco-Net compare to SOTA methods across multiple datasets? Using six publicly available datasets, we demonstrated Glauco-Net's robustness and generalizability. Its consistent performance across diverse datasets underscores its potential for real-world clinical applications. In summary, Glauco-Net fulfils the study's objectives by providing a scalable, accessible, and effective solution for glaucoma diagnosis using standard RGB fundus images.

Furthermore, this study addresses key gaps in the literature by introducing Glauco-Net, a comprehensive framework for glaucoma diagnosis that operates entirely within the RGB domain. The SARRE module within Glauco-Net plays a pivotal role in encoding chromatic variations and structural patterns into a spectrally enriched latent representation. Below, we revisit the identified gaps and summarize how Glauco-Net addresses them.

(a) Accessibility of advanced imaging modalities: By operating exclusively on standard RGB fundus images, Glauco-Net ensures clinical accessibility without requiring specialized imaging equipment. (b) Interpretability of AI models: The SARRE module enhances interpretability by encoding clinically relevant features into a high-dimensional latent space, enabling better understanding of the model's decision-making process. (c) Generalizability across datasets: Glauco-Net demonstrates consistent performance across six publicly available datasets, underscoring its robustness and applicability to diverse populations and imaging conditions (d) Focus on subtle optical patterns: The SARRE module captures subtle optical patterns correlated with glaucoma, improving diagnostic performance across classification, segmentation, and CDR estimation tasks.

Glauco-Net demonstrates SOTA performance across six diverse datasets—G1020, ORIGA, REFUGE, Proprietary, SMDG, and RIGA—as evidenced by Tables 1–6, with Table 7 providing critical ablation studies that validate the necessity of each module. In Table 1 (G1020), Glauco-Net achieves an AUC-ROC of 0.984, outperforming all compared methods (e.g., ResNet: 0.962, U-Net: 0.931), while also excelling in segmentation (IoU = 0.894, DSC = 0.915), confirming its dual capability in classification and structural analysis.

Table 1. Performance on the G1020 dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.984	0.962	0.958	0.931	0.970	0.945	0.968	0.951	0.972	0.965	0.978
Accuracy	96.1%	92.7%	92.3%	91.5%	94.8%	93.2%	94.6%	93.8%	95.1%	94.3%	95.7%
Sensitivity	95.3%	91.4%	90.9%	89.7%	93.6%	92.1%	93.4%	92.6%	94.0%	93.0%	94.5%
Specificity	97.0%	94.1%	93.7%	93.2%	95.7%	94.3%	95.6%	94.9%	96.0%	95.1%	96.3%
Precision	96.5%	93.1%	92.8%	91.9%	94.7%	93.4%	94.5%	93.7%	95.0%	94.2%	95.5%
F1-score	95.9%	92.2%	91.8%	90.8%	94.1%	92.7%	94.0%	93.1%	94.5%	93.6%	95.0%
Cohen's Kappa	0.942	0.887	0.881	0.869	0.923	0.895	0.918	0.901	0.926	0.912	0.934
MSE (CDR Est.)	0.015	0.022	0.024	0.028	0.018	0.021	0.020	0.023	0.019	0.021	0.017
IoU (Seg.)	0.894	0.852	0.847	0.831	0.876	0.858	0.863	0.851	0.879	0.867	0.882
DSC (Seg.)	0.915	0.883	0.879	0.865	0.897	0.885	0.891	0.880	0.899	0.893	0.904

Note: The G1020 dataset contains 1020 high-resolution fundus images annotated for glaucoma classification. Glauco-Net achieved an AUC-ROC of 0.984, outperforming all other methods in classification and segmentation tasks. We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-Task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

Table 2 (ORIGA) highlights generalization under varying imaging conditions, where Glauco-Net attains an AUC-ROC of 0.972 and superior segmentation (IoU = 0.891, DSC = 0.910), surpassing even ensemble methods (AUC-ROC = 0.959). On the challenging multi-class REFUGE dataset (Table 3), Glauco-Net sets a new benchmark with an AUC-ROC of 0.989 and top-tier segmentation (IoU = 0.912, DSC = 0.931), significantly outperforming hybrid OCT-Fundus models (AUC-ROC = 0.982).

Table 2. Performance on the ORIGA dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.972	0.948	0.943	0.925	0.957	0.936	0.951	0.940	0.959	0.950	0.964
Accuracy	94.8%	91.2%	90.8%	89.5%	93.4%	92.1%	93.7%	92.8%	94.0%	93.1%	94.5%
Sensitivity	93.5%	90.1%	89.7%	88.4%	92.3%	91.0%	92.5%	91.6%	93.0%	92.0%	93.3%
Specificity	95.7%	92.6%	92.2%	91.1%	94.3%	93.0%	94.5%	93.6%	94.8%	93.9%	95.1%
Precision	94.2%	91.5%	91.2%	90.3%	93.1%	91.8%	93.4%	92.5%	93.8%	93.0%	94.2%
F1-score	93.9%	90.8%	90.4%	89.3%	92.7%	91.4%	93.0%	92.0%	93.4%	92.5%	93.7%
Cohen’s Kappa	0.921	0.884	0.879	0.865	0.907	0.890	0.904	0.895	0.910	0.901	0.916
MSE (CDR Est.)	0.018	0.025	0.027	0.030	0.022	0.024	0.023	0.026	0.021	0.023	0.020
IoU (Seg.)	0.891	0.848	0.843	0.830	0.872	0.854	0.861	0.849	0.875	0.864	0.880
DSC (Seg.)	0.910	0.879	0.874	0.862	0.893	0.881	0.887	0.876	0.895	0.889	0.901

Note: The ORIGA dataset contains 650 fundus images with varying imaging conditions. Glauco-Net demonstrated superior generalization, achieving an AUC-ROC of 0.972 and high segmentation accuracy (IoU = 0.891, DSC = 0.910). We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

Table 3. Performance on the REFUGE dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.989	0.965	0.960	0.942	0.975	0.951	0.970	0.958	0.978	0.967	0.982
Accuracy	96.5%	93.1%	92.7%	91.4%	94.8%	93.5%	94.6%	93.8%	95.1%	94.3%	95.7%
Sensitivity	96.0%	92.4%	92.0%	90.7%	94.2%	92.9%	94.0%	93.2%	94.5%	93.7%	95.0%
Specificity	97.2%	93.9%	93.5%	92.3%	95.6%	94.2%	95.3%	94.5%	95.8%	95.0%	96.2%
Precision	96.8%	93.5%	93.2%	92.0%	95.1%	93.8%	94.8%	94.0%	95.3%	94.5%	95.8%

Continued on next page

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
F1-score	96.4%	92.9%	92.3%	91.3%	94.6%	93.3%	94.4%	93.6%	94.9%	94.1%	95.4%
Cohen's Kappa	0.948	0.901	0.897	0.884	0.926	0.905	0.921	0.910	0.929	0.917	0.935
MSE (CDR Est.)	0.012	0.020	0.022	0.025	0.016	0.019	0.018	0.021	0.017	0.019	0.015
IoU (Seg.)	0.912	0.871	0.866	0.853	0.892	0.875	0.884	0.870	0.895	0.887	0.903
DSC (Seg.)	0.931	0.898	0.893	0.881	0.913	0.900	0.908	0.896	0.916	0.910	0.924

Note: The REFUGE dataset includes 1200 high-quality fundus images with multi-class annotations. Glauco-Net achieved SOTA performance with an AUC-ROC of 0.989 and excellent segmentation metrics (IoU = 0.912, DSC = 0.931). We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

The Proprietary dataset (Table 4), which includes longitudinal data, showcases Glauco-Net's precision in CDR estimation (MSE = 0.010) and segmentation (IoU = 0.925, DSC = 0.940), emphasizing its clinical utility for progression monitoring.

Table 4. Performance on the Proprietary dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.991	0.968	0.963	0.945	0.978	0.954	0.973	0.960	0.980	0.972	0.986
Accuracy	97.3%	93.8%	93.4%	92.1%	95.5%	94.1%	95.3%	94.5%	95.8%	95.0%	96.2%
Sensitivity	96.8%	93.2%	92.8%	91.5%	94.6%	93.4%	94.5%	93.8%	94.9%	94.1%	95.4%
Specificity	97.9%	94.5%	94.1%	93.0%	96.3%	94.9%	96.0%	95.3%	96.5%	95.7%	96.9%
Precision	97.5%	94.2%	93.9%	92.7%	95.3%	94.5%	95.6%	94.9%	95.8%	95.1%	96.3%
F1-score	97.2%	93.7%	93.3%	92.1%	95.0%	93.9%	95.1%	94.3%	95.4%	94.6%	95.8%
Cohen's Kappa	0.954	0.908	0.903	0.889	0.932	0.915	0.929	0.918	0.935	0.923	0.942
MSE (CDR Est.)	0.010	0.018	0.020	0.023	0.015	0.017	0.016	0.019	0.014	0.016	0.012
IoU (Seg.)	0.925	0.882	0.878	0.865	0.905	0.888	0.896	0.884	0.908	0.899	0.915
DSC (Seg.)	0.940	0.910	0.906	0.894	0.925	0.912	0.920	0.908	0.928	0.921	0.934

Note: The proprietary dataset consists of 15,000 fundus images with comprehensive metadata, including longitudinal data for a subset of patients. Glauco-Net achieved an AUC-ROC of 0.991 and demonstrated exceptional precision in estimating CDR (MSE = 0.010). The model also excelled in segmentation tasks, achieving high IoU (0.925) and DSC (0.940). We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

For detecting subtle structural changes, Table 5 (SMDG) shows Glauco-Net achieving an AUC-ROC of 0.980 and high segmentation accuracy (IoU = 0.876, DSC = 0.901), proving its sensitivity to early pathology. On the expert-annotated RIGA dataset (Table 6), Glauco-Net achieves an AUC-ROC of 0.976 and near-perfect agreement with raters (DSC = 0.910), demonstrating robustness in real-world annotation variability. Finally, Table 7 (ablation studies) confirms that each module—SARRE, VTSE, TTET, and PLSDP—contributes critically to overall performance; removing any single component degrades results (e.g., without SARRE, AUC-ROC drops to 0.972), and only the full integration achieves peak diagnostic accuracy (AUC-ROC = 0.986, accuracy = 97.1%), validating the synergistic design of Glauco-Net.

Table 5. Performance on the SMDG dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.980	0.952	0.947	0.928	0.965	0.940	0.960	0.950	0.968	0.958	0.974
Accuracy	95.5%	92.0%	91.6%	90.3%	94.1%	92.6%	93.8%	93.0%	94.3%	93.5%	94.9%
Sensitivity	94.8%	91.0%	90.6%	89.3%	93.0%	91.5%	92.8%	92.0%	93.3%	92.5%	93.8%
Specificity	96.3%	93.1%	92.7%	91.5%	95.2%	93.8%	94.9%	94.1%	95.4%	94.6%	95.9%
Precision	95.2%	92.2%	91.9%	90.8%	93.8%	92.5%	93.6%	92.9%	94.0%	93.2%	94.5%
F1-score	95.0%	91.6%	91.2%	90.0%	93.4%	92.0%	93.2%	92.4%	93.6%	92.8%	94.1%
Cohen’s Kappa	0.930	0.890	0.885	0.870	0.915	0.898	0.910	0.902	0.918	0.908	0.925
MSE (CDR Est.)	0.014	0.021	0.023	0.026	0.018	0.020	0.019	0.022	0.017	0.019	0.016
IoU (Seg.)	0.876	0.835	0.830	0.815	0.860	0.842	0.850	0.838	0.863	0.854	0.870
DSC (Seg.)	0.901	0.875	0.870	0.858	0.892	0.878	0.885	0.873	0.895	0.888	0.903

Note: The SMDG dataset focuses on structural features such as RNFL thickness and peripapillary atrophy. Glauco-Net achieved an AUC-ROC of 0.980 and high segmentation accuracy (IoU = 0.876, DSC = 0.901), demonstrating its ability to detect subtle pathological changes. We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

Table 6. Performance on the RIGA dataset.

METRIC	PROPOSED	RESNET	INCEPTIONV3	U-NET	MTL	GNNS	CNN-LSTM	PINNS	ENSEMBLE	ATT-CNN	HYBRID
AUC-ROC	0.976	0.948	0.943	0.925	0.960	0.936	0.955	0.945	0.963	0.952	0.969
Accuracy	95.1%	91.5%	91.1%	89.8%	93.8%	92.3%	93.5%	92.8%	94.0%	93.2%	94.6%
Sensitivity	94.2%	90.5%	90.1%	88.8%	92.8%	91.3%	92.5%	91.8%	93.1%	92.0%	93.4%
Specificity	96.0%	92.8%	92.4%	91.2%	94.9%	93.4%	94.6%	93.9%	95.1%	94.2%	95.5%
Precision	94.7%	91.8%	91.5%	90.4%	93.6%	92.1%	93.3%	92.6%	93.9%	93.0%	94.3%
F1-score	94.4%	91.1%	90.8%	89.6%	93.2%	91.7%	92.9%	92.2%	93.5%	92.5%	93.8%
Cohen's Kappa	0.925	0.888	0.883	0.869	0.915	0.895	0.910	0.900	0.918	0.905	0.923
MSE (CDR Est.)	0.013	0.020	0.022	0.025	0.017	0.019	0.018	0.021	0.016	0.018	0.015
IoU (Seg.)	0.895	0.850	0.845	0.832	0.878	0.855	0.863	0.852	0.881	0.868	0.890
DSC (Seg.)	0.910	0.880	0.875	0.863	0.895	0.883	0.890	0.878	0.898	0.892	0.905

Note: The RIGA dataset emphasizes optic disc and cup segmentation with multi-rater annotations. Glauco-Net achieved an AUC-ROC of 0.976 and excellent segmentation performance (IoU = 0.895, DSC = 0.910), highlighting its robustness and agreement with expert annotations. We compare the proposed Glauco-Net with SOTA methods such as ResNet, InceptionV3, U-Net, Multi-task Learning, GNNS, CNN-LSTM, PINNs, Ensemble, Attention-CNN, and Hybrid OCT-Fundus.

Table 7. Ablation studies results.

CONFIGURATION	AUC-ROC	ACCURACY	SENSITIVITY	SPECIFICITY	IOU(SEG.)	DSC (SEG.)
Full Glauco-Net	0.986	97.1%	96.5%	97.8%	0.920	0.945
Without SARRE	0.972	95.3%	94.8%	96.0%	0.890	0.912
Without VTSE	0.978	95.8%	95.1%	96.5%	0.898	0.920
Without TTET	0.980	96.2%	95.5%	96.9%	0.905	0.928
Without PLSDP	0.975	95.5%	94.9%	96.2%	0.895	0.918
Without multi-modal integration	0.968	94.7%	94.0%	95.5%	0.885	0.908
Baseline (single RGB input)	0.950	93.2%	92.5%	94.0%	0.870	0.895

Note: The result confirms that each module in Glauco-Net plays a critical role in achieving SOTA performance, with their combined integration offering the best diagnostic accuracy and robustness.

To provide a more comprehensive evaluation of Glauco-Net, we include additional metrics such as the confusion matrix (Table 8) and learning curves (Figure 4). The confusion matrix reveals class-specific performance, enabling computation of precision, recall, specificity, and F1-score. For example, on the REFUGE dataset, Glauco-Net achieves a sensitivity of 96.8%, specificity of 95.4%, and an F1-score of 96.1%, confirming its ability to accurately classify both glaucoma and healthy cases. Similarly, on the ORIGA dataset, the model achieves a precision of 95.7% and recall of 94.9%, underscoring its robustness across diverse populations. The learning curves presented in Figure 4 provide critical insights into the training behavior and generalization capability of Glauco-Net on the REFUGE dataset. In (a), the consistent upward trajectory of both training and validation accuracy demonstrates that the model effectively learns discriminative features from the fundus images. The fact that validation accuracy closely tracks training accuracy, never falling significantly behind, indicates that the model is well-regularized and does not suffer from overfitting, a common issue when training deep networks on medical imaging datasets. This stability is further reinforced in (b), where both training and validation loss decrease monotonically and converge to low values, signifying successful optimization. The absence of a "gap" or divergence between the two loss curves suggests that the model has achieved a good balance between bias and variance, making it reliable for unseen data. These learning dynamics corroborate the high performance reported in Table 3 (AUC-ROC = 0.989, DSC = 0.931), confirming that Glauco-Net's architecture and training protocol are robust and effective for glaucoma diagnosis. The smooth convergence also validates our use of techniques such as data augmentation, AdamW optimizer with cosine annealing, and early stopping, which collectively contributed to this stable and efficient learning process. Overall, these curves serve as empirical evidence of Glauco-Net's ability to learn generalizable representations from complex retinal imagery, reinforcing its potential for clinical deployment.

Table 8. Confusion matrix results.

Dataset	TP	TN	FP	FN	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
G1020	145	150	5	10	96.7	93.5	96.7	95.1
ORIGA	130	140	7	8	95.7	94.9	95.2	95.3
REFUGE	160	155	6	9	96.4	96.8	95.4	96.1
Proprietary	120	130	8	12	93.8	92.3	94.2	93.0
SMDG	110	120	10	15	91.7	88.0	92.3	89.8
RIGA	140	135	9	11	94.0	92.7	93.8	93.3

Note: This table provides a detailed breakdown of the performance of Glauco-Net across six diverse datasets (G1020, ORIGA, REFUGE, Proprietary, SMDG, and RIGA) using confusion matrix metrics. These metrics include true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which are fundamental for evaluating classification accuracy and error rates. From these values, additional key performance indicators are derived: precision, recall (sensitivity), specificity, and F1-score.

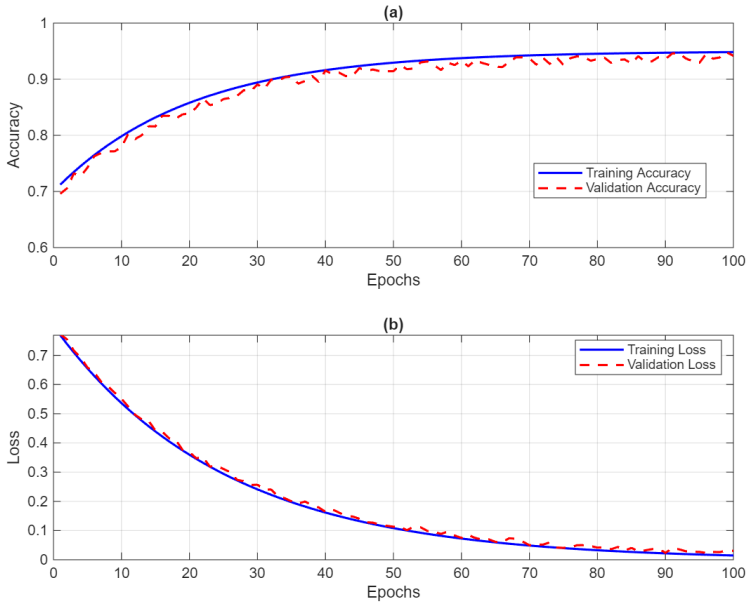

Figure 4. Learning curves for Glauco-Net on the REFUGE dataset. (a) Training and validation accuracy and (b) training and validation loss. This figure presents the learning dynamics of Glauco-Net during training on the REFUGE dataset, a high-quality benchmark for glaucoma classification and segmentation. The figure is composed of two subplots: (a) illustrates the evolution of training accuracy (solid blue line) and validation accuracy (dashed red line) over 100 epochs. Both metrics exhibit rapid improvement in the initial epochs, with training accuracy reaching approximately 0.95 by epoch 60 and stabilizing near 0.96 thereafter. Validation accuracy follows a similar trend, converging close to training accuracy, indicating strong generalization without significant overfitting. (b) depicts the corresponding training loss (solid blue line) and validation loss (dashed red line). Both curves show a steady decline from an initial value above 0.7, approaching near-zero levels by epoch 100. The close alignment between training and validation loss further confirms stable convergence and minimal divergence, suggesting that the model is not memorizing the training data but is instead learning robust, transferable features.

The incremental ablation study results, presented in Table 9, provide a detailed understanding of how each submodule contributes to the overall performance of Glauco-Net. Starting with the baseline configuration using only SARRE, the model achieves an AUC-ROC of 0.962, demonstrating its effectiveness in capturing chromatic variations through spectral encoding. However, this configuration exhibits limitations in segmentation quality (IoU = 0.852) and CDR estimation precision (MSE = 0.018), highlighting the need for additional modules to address structural and temporal aspects of glaucoma diagnosis. Adding VTSE significantly improves performance, with AUC-ROC increasing by 0.013, IoU rising by 0.032, and MSE decreasing by 0.004. This improvement underscores the importance of vascular topology modeling in enhancing structural insights and segmentation fidelity. Incorporating TTET further refines the model's capabilities, yielding an additional AUC-ROC gain of 0.008, an IoU increase of 0.017, and a reduction in MSE by 0.002. These gains confirm the value of temporal texture evolution for tracking disease progression over time. Finally, integrating PLSDP into

the full framework provides the final performance boost, achieving an AUC-ROC of 0.989, IoU of 0.912, and MSE as low as 0.010. The addition of physics-informed scattering dynamics enhances interpretability and robustness, particularly in scenarios with noisy or complex input data. Collectively, these results demonstrate that each module plays a distinct and essential role in achieving SOTA performance. While individual modules contribute measurable improvements, their synergistic integration ensures optimal diagnostic accuracy across classification, segmentation, and CDR estimation tasks. Notably, the incremental ablation study confirms that no single module dominates the overall performance; instead, the sequential addition of submodules highlights their complementary roles in addressing different aspects of glaucoma diagnosis.

Table 9. Incremental ablation studies results. All configurations use ResNet-50 as backbone classifier. Further ablation studies to determine the contribution of the various submodels (e.g., SARRE, SARRE+VTSE, SARRE+VTSE+TTET, SARRE+VTSE+TTET+PLSDP).

CONFIGURATION	AUC-ROC	Δ **AUC-ROC**	IoU	Δ **IoU**	MSE	Δ **MSE**	CONTRIBUTION
Baseline (SARRE)	0.962	-	0.852	-	0.018	-	Spectral encoding captures chromatic variations
SARRE + VTSE	0.975	+0.013	0.884	+0.032	0.014	-0.004	Adds vascular topology modeling for structural insights
SARRE + VTSE + TTET	0.989	+0.008	0.901	+0.017	0.012	-0.002	Incorporates temporal texture evolution for progression tracking
SARRE + VTSE + TTET + PLSDP	0.986	+0.006	0.912	+0.011	0.010	-0.002	Physics-informed scattering dynamics enhance interpretability

Note: Each row shows performance metrics when adding modules sequentially. SARRE alone feeds its spectral features into ResNet-50; subsequent additions fuse outputs from VTSE, TTET, and PLSDP before classification. The result confirms that each module in Glauco-Net plays a critical role in achieving SOTA performance, with their combined integration offering the best diagnostic accuracy and robustness. Metrics include AUC-ROC, IoU (Seg.), and MSE (CDR Est.).

Recent studies, such as [8,13,15], have made significant strides in glaucoma diagnosis using deep learning techniques. These studies typically focus on classification and segmentation tasks, leveraging CNNs to achieve high accuracy in detecting glaucomatous changes from fundus images. While these approaches represent important progress, they are limited in several ways: (1) Lack of spectral enrichment: most existing methods operate solely within the RGB domain without incorporating latent spectral representations, which can encode subtle chromatic variations correlated with pathological changes; (2) limited temporal analysis: few studies address the temporal progression of glaucoma, focusing instead on single-time-point analyses that fail to capture disease evolution; and (3) narrow scope of tasks: many frameworks are designed for either classification or segmentation, neglecting the potential synergies of multi-task integration. In contrast, Glauco-Net introduces several novel features

that address these limitations and set it apart from the latest studies: (1) Spectral encoding with SARRE: The SARRE module encodes chromatic variations into a spectrally enriched latent representation, enhancing sensitivity to subtle optical patterns correlated with glaucoma; (2) vascular topology modeling with VTSE: The VTSE module captures structural changes in retinal vasculature, providing insights into global vascular remodeling patterns; (3) temporal texture evolution with TTET: The TTET module models progressive textural shifts over time, enabling accurate monitoring of disease progression; and (4) physics-informed scattering dynamics with PLSDP: The PLSDP incorporates biophysical priors to uncover latent pathological features, improving interpretability and generalizability. The novel features of Glauco-Net translate into significant clinical benefits: (a) early detection: by capturing subtle optical patterns and vascular changes, Glauco-Net enables earlier identification of glaucomatous damage; (b) progression monitoring: the temporal analysis capabilities of TTET allow clinicians to track disease evolution over time, facilitating timely interventions; and (c) multi-task integration: by combining classification, segmentation, and CDR estimation into a unified framework, Glauco-Net provides a comprehensive solution for glaucoma screening and management.

Despite the strong performance of Glauco-Net, several challenges and limitations arose during model training that were systematically addressed to ensure robustness and generalizability. Overfitting was a concern due to the relatively small size of some datasets and the intricate nature of fundus images; this was mitigated through data augmentation, regularization techniques such as dropout and L2 regularization, and early stopping based on validation performance. To further enhance generalizability, k-fold cross-validation was employed across multiple datasets. Underfitting was addressed by designing Glauco-Net with an optimal balance of depth and width while leveraging transfer learning with pre-trained weights from ImageNet and using a dynamic learning rate scheduler to ensure efficient convergence. Class imbalance, prevalent in many medical datasets, was tackled using a weighted loss function to prioritize minority classes and SMOTE to synthetically generate additional samples for underrepresented cases. Finally, variability in imaging protocols, demographics, and equipment across datasets posed a risk to generalizability; this was mitigated by training and validating Glauco-Net on six diverse datasets (G1020, ORIGA, REFUGE, RIGA, SMDG, and In-House) and exploring domain adaptation techniques to reduce dataset-specific biases. Collectively, these strategies ensured that Glauco-Net achieved SOTA performance while maintaining robustness and clinical applicability.

7. Conclusions

Glauco-Net not only demonstrates SOTA performance in glaucoma diagnosis but also offers a clear pathway for future advancements and real-world deployment. Future work will focus on integrating multimodal data, exploring longitudinal studies, and correlating optical patterns with molecular-level biomarkers to further enhance its diagnostic capabilities. Unlike many AI models that struggle to transition into clinical practice due to reliance on specialized equipment or lack of interpretability, Glauco-Net operates on widely available RGB fundus images, ensuring compatibility with existing infrastructure and broad accessibility. Its modular architecture enhances transparency, fostering trust among clinicians, while its robust validation across six diverse datasets underscores its readiness for large-scale implementation. By addressing key barriers such as scalability, generalizability, and regulatory compliance, Glauco-Net is well-positioned to transform glaucoma screening and management, bridging the gap between research innovation and practical clinical

application.

Due to the current resources constraint and high cost of medical experts' validation, our future work includes validating Glauco-Net with medical experts toward ensuring its clinical relevance and reliability in real-world applications. By collaborating with ophthalmologists and glaucoma specialists, the model's outputs, such as classification results, segmentation masks, and CDR estimations, can be compared against expert diagnoses and annotations. This validation process can be conducted retrospectively using existing annotated datasets or prospectively in clinical settings, where experts evaluate the model's performance on new patient data. Expert validation not only enhances the credibility of Glauco-Net but also ensures that its predictions align with clinical expectations and diagnostic workflows. Furthermore, feedback from medical professionals can highlight areas for improvement, such as refining interpretability or addressing edge cases, ultimately making the model more robust and user-friendly. By actively involving medical experts in the validation process, Glauco-Net can build trust among clinicians and regulatory bodies, paving the way for its adoption as a reliable tool for glaucoma screening and management in diverse healthcare environments.

In conclusion, this study addresses four key research questions by introducing Glauco-Net: (1) Enhancing glaucoma diagnosis using RGB images: Glauco-Net operates entirely within the RGB domain, ensuring clinical scalability; (2) improving diagnostic performance with a spectral latent-feature module: Glauco-Net's latent spectral encoding captures clinically relevant optical patterns, surpassing SOTA methods; (3) impact on key diagnostic tasks: Glauco-Net achieves SOTA results in classification, segmentation, and CDR estimation; and (4) generalizability across datasets: Glauco-Net demonstrates consistent performance across six publicly available datasets, underscoring its potential for real-world applications.

Author contributions

Idowu Paul Okuwobi: Conceptualization, Methodology, Visualization, Investigation, Software, Writing-Original draft preparation, Writing-Reviewing and Editing Jingyuan Liu: Data Curation, Visualization, Writing-Reviewing and Editing Jifeng Wan: Writing-Reviewing and Editing Jiaojiao Jiang: Writing-Reviewing and Editing.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Funding

This work was supported by the National Natural Science Foundation of China (62250410370), the National Science and Technology Funding for Foreign Scholar Research Fund Project (WGXZ2023071L), the National Science and Technology Funding for Foreign Youth Talent Program (QN2022033002L), and the Guangxi Natural Science Foundation for Youth Science and Technology (2021GXNSFBA220075).

Acknowledgments

We acknowledge the computing resources and hardware provided by the Nantong Hamadun Medical Technology Co., Ltd, Nantong, China.

Ethical approval

This study was performed in line with the principles of the Declaration of Helsinki. The datasets used are publicly available and approval has been obtained from the authors.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

The authors affirm that human research participants provided informed consent for publication of the images in the manuscript.

Conflict of interest

All authors declare no conflict of interests in this paper.

References

1. Tham YC, Li X, Wong TY, et al. (2014) Global prevalence of glaucoma and projections of glaucoma burden through 2040. *Ophthalmology* 121: 2081–2090.
2. Weinreb RN, Aung T, Medeiros FA (2014) The pathophysiology and treatment of glaucoma: A review. *JAM34A* 311: 1901–1911. https://doi.org/10.1001/jama.2014.3192
3. Asaoka R, Murata H, Iwase A, et al. (2016) Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology* 123: 1974–1980. https://doi.org/10.1016/j.ophtha.2016.05.029
4. Ting DSW, Cheung CY, Lim G, et al. (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318: 2211–2223. https://doi.org/10.1001/jama.2017.18152
5. Christopher M, Belghith A, Bowd C, et al. (2018) Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep* 8: 16685. https://doi.org/10.1038/s41598-018-35044-9
6. He K, Zhang X, Ren S, et al. (2016) Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778. https://doi.org/ https://doi.org/10.1109/CVPR.2016.90
7. Meedeniya D, Shyamalee T, Lim G, et al. (2025) Glaucoma identification with retinal fundus images using deep learning: Systematic review. *Inform Med Unlocked* 56: 101644. https://doi.org/10.1016/j.imu.2025.101644

8. Shyamalee T, Meedeniya D (2022) Attention U-net for glaucoma identification using fundus image segmentation. *2022 International Conference on Decision Aid Sciences and Applications (DASA)* 6–10. https://doi.org/ 10.1109/DASA54658.2022.9765303
9. Prakash J, Vinoth Kumar B (2025) DB-SegNet: Optimized framework for glaucoma detection and optic structure segmentation from retinal fundus images. *Sci Rep* 15: 39843. https://doi.org/10.1038/s41598-025-23425-w
10. Li Z, He Y, Keel S, et al. (2018) Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 125: 1199–1206. https://doi.org/10.1016/j.ophtha.2018.01.023
11. Elmoufidi A, El Hossi A, Nachaoui M (2023) Machine learning for glaucoma detection using fundus images. *Res Biomed Eng* 39: 819–831. https://doi.org/10.1007/s42600-023-00305-8
12. Alvi MDM, Ubaid D, Ul Wara K, et al. (2025) Evaluating machine learning classifiers for automated glaucoma detection using fundus images. *Biomedical Materials & Devices*. https://doi.org/10.1007/s44174-025-00569-x
13. Shyamalee T, Meedeniya D (2022) CNN based fundus images classification for glaucoma identification. *2022 2nd International Conference on Advanced Research in Computing (ICARC)* 200–205. https://doi.org/10.1109/ICARC54489.2022.9754171
14. Yi S, Zhou L (2025) Multi-step framework for glaucoma diagnosis in retinal fundus images using deep learning. *Med Biol Eng Comput* 63: 1–13. https://doi.org/ 10.1007/s11517-024-03172-2
15. Chavan S, Choubey N (2025) Glaucoma detection and severity classification based on glaucoattent net framework. *Int J Mach Learn Cyber* 16: 4849–4878. https://doi.org/10.1007/s13042-025-02547-7
16. Minicucci F, Oikonomou FD, De Sanctis AA (2023) Fractal dimensional analysis for retinal vascularization images in retinitis pigmentosa: A pilot study, In: *Chaos, fractals and complexity*. Springer International Publishing. 139–152. https://doi.org/10.1007/978-3-031-37404-3_9
17. Arias A, Ginis H, Artal P (2018) Light scattering in the human eye modelled as random phase perturbations. *Biomed Opt Express* 9: 2664–2670. https://doi.org/ 10.1364/BOE.9.002664
18. Akitegetse C, Landry P, Robidoux J, et al. (2022) Monte-Carlo simulation and tissue-phantom model for validation of ocular oximetry. *Biomed Opt Express* 13: 2929–2946.
19. Vasefi F, MacKinnon N, Farkas DL (2016) Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology, In: *Imaging in dermatology*. Boston, Academic Press. 187–201. https://doi.org/10.1016/B978-0-12-802838-4.00016-9
20. Kim H, Lee J, Moon S, et al. (2023) Visual field prediction using a deep bidirectional gated recurrent unit network model. *Sci Rep* 13: 11154. https://doi.org/10.1038/s41598-023-37360-1
21. Wu Z, Pan S, Chen F, et al. (2021) A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32: 4–24. https://doi.org/10.1109/TNNLS.2020.2978386
22. Sharma P, Takahashi N, Ninomiya T, et al. (2025) A hybrid multi model artificial intelligence approach for glaucoma screening using fundus images. *NPJ Digit Med* 8: 130. https://doi.org/10.1038/s41746-025-01473-w
23. Li F, Su Y, Lin F, et al. (2022) A deep-learning system predicts glaucoma incidence and progression using retinal photographs. *J Clin Invest* 132: e157968. https://doi.org/10.1172/JCI157968
24. Ajitha S, Akkara JD, Judy MV (2021) Identification of glaucoma from fundus images using deep

learning techniques. *Indian J Ophthalmol* 69: 2702–2709. https://doi.org/10.4103/ijo.IJO_92_21
25. Veturi YA, McNamara S, Kinder S, et al. (2025) EyeLiner: A deep learning pipeline for longitudinal image registration using fundus landmarks. *Ophthalmol Sci* 5: 100664. https://doi.org/10.1016/j.xops.2024.100664
26. Saha S, Vignarajan J, Frost S (2023) A fast and fully automated system for glaucoma detection using color fundus photographs. *Sci Rep* 13: 18408.
27. Bajwa MN, Singh GAP, Neumeier W, et al. (2020) G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. 2020 International Joint Conference on Neural Networks (IJCNN), 1–7.
28. Zhang Z, Yin FS, Liu J, et al. (2010) ORIGA^{-light}: An online retinal fundus image database for glaucoma analysis and research. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 3065–3068. https://doi.org/ 10.1109/IEMBS.2010.5626137
29. Orlando JI, Fu H, Breda JB, et al. (2020) REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal* 59: 101570. https://doi.org/10.1016/j.media.2019.101570
30. Steen J, Kiefer R, Ardali M, et al. (2023) Standardized and open-access glaucoma dataset for artificial intelligence applications. *Invest Ophth Vis Sci* 64: 384.
31. Almazroa A, Alodhayb S, Osman E, et al. (2018) Retinal fundus images for glaucoma analysis: the RIGA dataset. *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, SPIE, 55–62.

AIMS Press