International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SWINSEGNET: A TRANSFORMER-DRIVEN MULTI-TASK MODEL FOR ROBUST GLAUCOMA SCREENING ACROSS PUBLIC DATASETS

[1]Radheshyam Acholiya, [2]Dr. Mukesh Kumar,

[1]Research Scholar, RNTU Bhopal

acholiya.radheshyam@gmail.com

[2]Parul University, Vadodra, Gujrat

mukesh.manit86@gmail.com

Abstract

Glaucoma is a leading cause of irreversible blindness, where early and accurate detection is critical for effective treatment. Traditional deep learning models for glaucoma screening often struggle with generalization across multiple datasets due to variations in imaging quality and annotation. To address this, we propose SwinSegNet, a transformer-driven multi-task model integrating Swin Transformer with SegFormer MLP to jointly perform optic disc/optic cup segmentation and glaucoma classification. The model was trained and evaluated on three benchmark public datasets ORIGA, DRISTI-GS, and REFUGE encompassing 1,951 retinal fundus images. Experimental results show that SwinSegNet outperforms existing CNN-based and U-Net variants, achieving an average accuracy of 99.49%, sensitivity of 99.24%, and specificity of 99.40%, marking an improvement of up to 4%–5% over state-of-the-art methods. These findings confirm the robustness and clinical potential of SwinSegNet for large-scale glaucoma screening.

Keywords: Glaucoma screening, Swin Transformer, Multi-task learning, Optic disc and cup segmentation, fundus image analysis, Deep learning in ophthalmology.

1. Introduction

Glaucoma is one of the leading causes of irreversible blindness worldwide, often referred to as the “silent thief of sight” due to its asymptomatic progression until advanced stages. Early detection is crucial to prevent vision loss, but manual diagnosis using fundus images requires expert ophthalmologists and is prone to inter-observer variability. Automated screening systems can assist clinicians by providing fast, accurate, and reproducible results, yet existing approaches face significant challenges in achieving high accuracy across diverse datasets due to differences in image quality, acquisition devices, and population variations. These challenges necessitate the development of robust and generalized models capable of performing consistently across multiple public datasets.

Traditional machine learning methods relied on handcrafted features, which limited their adaptability and performance in real-world scenarios. With the advent of deep learning, convolutional neural networks (CNNs) demonstrated remarkable success in medical image

Received: August 03, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

analysis, particularly for optic disc and optic cup segmentation, which are critical steps in glaucoma detection. However, CNNs often struggle with capturing global dependencies and long-range contextual information, leading to suboptimal performance in complex retinal images. Moreover, many CNN-based models are task-specific and fail to integrate segmentation and classification effectively within a unified framework, reducing their clinical applicability.

To address these limitations, Transformer-based architectures have recently emerged as powerful alternatives, excelling at modeling global representations while maintaining fine-grained spatial detail. In this work, we propose SwinSegNet, a Transformer-driven multi-task model that integrates segmentation and classification for robust glaucoma screening. The model leverages the hierarchical Swin Transformer backbone in conjunction with SegFormer-style multi-layer perceptron's (MLPs) to effectively extract both local and global features from fundus images. Unlike existing approaches, SwinSegNet is designed to handle multiple tasks simultaneously, improving efficiency while reducing computational redundancy.

We validate our approach on three widely used public datasets ORIGA, DRISTI-GS, and REFUGE ensuring a comprehensive evaluation across varied image distributions. The proposed model demonstrates superior performance, achieving accuracy, sensitivity, and specificity improvements of 2–4% compared to state-of-the-art methods. Notably, our framework achieves near-perfect segmentation Dice and Jaccard scores for optic disc and cup boundaries, while significantly enhancing classification precision and F1-scores.

Key Contributions and Novelty:

- We propose a novel Transformer-based multi-task framework, SwinSegNet, integrating segmentation and classification within a single unified pipeline.
- Our model leverages Swin Transformer with SegFormer MLP decoders to capture both local texture and global context effectively.
- Extensive experiments on three benchmark datasets demonstrate superior generalization and robustness, with measurable performance improvements over existing methods.
- We provide a scalable architecture that can adapt to other ophthalmic tasks, paving the way for real-world deployment in clinical glaucoma screening.

2. Literature review

Glaucoma is a leading cause of irreversible blindness which requires timely and accurate diagnosis to avoid vision loss, and recent studies tend to adopt deep learning and ensemble techniques to tackle this problem. Fashionable models Ensemble-based models using ResNet, VGG, DenseNet and Inception architectures have strong potential for automated glaucoma detection from fundus and Optical Coherence Tomography (OCT) images and provide feature extraction and classification advances [1]. For example, frameworks enabled by IoT in concert with Bidirectional LSTMs and ResNet50 have been employed in provision of healthcare in rural India, while outperforming other several datasets such as ACRIMA, Fundus, ORIGA and Retinal image datasets with accuracy levels of up to 99% [2]. Computer-aided diagnostic systems have also utilized hybrid ensemble models of ResNet50, VGG19, and Inception-V3,

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and achieved an overall accuracy level of 98.58% on mixed dataset such as RIM-ONE, DRISHTI-GS, DRIONS-DB, and HRF, outperformed current single-model methods [3]. Beyond detection, AI in glaucoma has developed in the areas of prognosis and predicting outcome of surgery, applying CNNs, RNNs, transformers, GANs, autoencoders, and also exploring omics-driven biomarker discovery; yet, interpretability, heterogeneous data integration, and personalized timing of treatment present remaining challenges [4]. Such methods are leveraged by the hybrid multi classifier systems like FGLAUC-99, which integrate Naı¨ve Bayes, Decision Tree (J48/random Forest) and Neural Networks (MLP) to achieve maximum accuracy of 99.18% that shows the combined potential of probabilistic, decision and neural methods for an intelligent clinical decision support [5]. Together, these developments show the promise of ensemble deep learning and hybrid AI models in the diagnosis and treatment for glaucoma, but also emphasize the need for high-quality datasets with generalizability across populations, and clinically interpretable solutions to ensure real-world applicability.

Early diagnosis of glaucoma and diabetic retinopathy is essential for preventing permanent blindness, and fundus photography is an important detection tool. Classically trained models such as CNNs and SVM fail to achieve higher accuracy, but more sophisticated ensemble models integrating texture analysis with XGBoost and LightGBM approaches for disease subtype classification reached accuracy of 99% [6]. It is a modality, which also demonstrates powerful predictive ability—a meta-analysis shows its efficacy in predicting onset, progression, and surgical risk in glaucoma [7] but its use extends to prediction of cataracts, AMD, uveitis, and systemic disease [8]. In a large dataset of machine learning models trained on EHR data, the logistic regression model performs less as compared to models predicting glaucoma 1 year before (AUC 0.81, vs.73) [9]. Meta-analyses validate high diagnostic accuracy of deep learning models with both fundus photography (AUROC 0.90) and OCT imaging (AUROC 0.86), but progress prediction remains difficult highlighting the necessity of multimodal fusion and external validation in real-world application [10].

Glaucoma, one of the leading causes of irreversible blindness, demands early detection and efficient management, and in recent studies various AI and ML applications have been investigated with different diagnostic, progression monitoring, and treatment planning purposes. Hybrid deep learning models consisting of Vision Transformers and CNNs (ResNet) as well as Swin Transformers have achieved an outstanding accuracy, where the strongest-performing model attains an accuracy of 99.4% and F1-scores of over 0.99 in classifying new cases [11]. In addition, the work in [12] has developed a patient-centered expert system that combines forward chaining, ML models, and GPT-driven dialogue to aid clinical decision and to generate a personalized treatment plan especially in the situations of restricted patient-physician interactions. The majority of AI research in glaucoma has concentrated on detection, and the increasing focus is now turning toward predicting progression of glaucoma, yet disparate datasets, definition variations, and methodological discrepancies add to the difficulty of making comparisons, and highlights the requirement of structured progression centric research [13]. In genetics, ML have proven to be more efficient for sub-phenotyping genotypes based on NGS whole-exome data and classifying CG genotype using Decision Trees and

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Random Forests than traditional normalization algorithms, elucidating a rule-based model to depict how ML-based sub-phenotyping reveals intricate genetic pathways and enables personalized screen and treat options [14]. Finally, the model calibration was refined with the new confidence-calibrated label smoothing (CC-LS) loss, in which label smoothing and confidence penalty balance achieved strong calibration (Brier 0.098) with strong classification (81% sensitivity, 80% specificity) [15]. Together these works point to the possibility of AI/ML as a game-changer in glaucoma detection, estimation of progression, gene testing and precision care.

Glaucoma and other retinal diseases represent leading causes of irreversible blindness, consequently, extensive research works have been focused on development of automatic diagnostic tools based on Machine Learning (ML) and Deep Learning (DL). Recent works range from stacking ensemble models of 13 pre-trained CNNs and SVM classifiers with achieving the accuracy up to 99.6% for the datasets such as LAG-R and ACRIMA-R [16]. In the case of diabetic retinopathy, learned hybrid models with ResNet50 and EfficientNetB0 achieve an accuracy of 97.5%, over 84,000 OCT scans, dealing with the overfitting and interpretability issues [17]. Hybrid CNN–ML models such as ResNet50, VGG-16 combined with Random Forest have shown a good performance with an accuracy of 95.41% and the F1-score of 93.52% in glaucoma detection [18]. In the interest of trustworthiness, explainable AI (XAI) systems based on OCT imaging and SHAP-based interpretability, demonstrated at least near-perfect AUCs (0.96–1.00) across the stages of glaucoma, outperforming clinicians in early diagnosis [19]. In addition to glaucoma and DR, the integration of deep learning techniques has been extended to the rare eye cancers like ocular melanoma, where exploration of CNN and RNN-based prediction models has been further contributing to the prediction successes [20]. In this context, a systematical survey of ML/DL applications for glaucoma detection identifies SVMs with over 98% accuracies in different datasets and light CNNs with 99.67% accuracy and the ensembled learning with perfect classification metrics as the most powerful models [21]. Combined, these developments highlight the game-changing power of ML, DL, and XAI to improve early diagnosis, classification and personalized treatment for ocular diseases.

Glaucoma is still one of the most common causes of irreversible blindness and with recent advances in artificial intelligence (AI), machine learning (ML), and deep learning (DL), early detection, diagnosis, and management planning have improved. A review comprising 54 studies noted that with respect to retinal image analysis, CNN variations yielded 93.5% accuracy, hybrid ML and DL models (95%), and SVMs well on structured data sets, demonstrating the potential and boundary of the proposed mechanisms for analysis [22]. Further than classical diagnostics, automated CNN-based systems are under development using deep learning in order to generate a scalable, cost-effective reliable means for detection of several vision diseases, resolving accessibility disparities in underprivileged sites [23]. In order to optimize treatment, the ANFIS has been employed in MIGS surgical decision support by incorporating clinical parameters obtained from real patient datasets to determine the best choice of procedure [24]. Systematic reviews also support AI in multimodal data (fundus images, OCT scans, and VF tests) analyses for improving accuracies and decreasing economic

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

pressures, though challenges of imbalance datasets and that of early disease detection still exist [25]. Lastly, new multimodal fusion methods (combining feature-level and image-level) over different public database have achieved a maximum accuracy of 95.56% and are robust to make a computer-aided glaucoma diagnosis which has been evaluated by the ophthalmologists [26]. Together, these results highlight the increasing clinical relevance of AI-based approaches to advance the diagnosis and prognosis of glaucoma and to direct tailored treatment plans.

Glaucoma and diabetic macular edema (DME) are common diseases contributing to blindness which could benefit from advanced AI diagnostic and predictive algorithms. For DME, combining 3D-OCT imaging and clinical features-loss (together with ensemble machine learning (AdaBoost, GradientBoosting), and LightGBM) led to strong performance for predicting anti-VEGF therapy response, with accuracy of 0.912 and AUROC of 0.976 [27]. In the context of glaucoma novel ensemble and large VLLM-based solutions achieved effective classification by mitigating for data imbalance and multi-image assessment improving classification to 0.9875 accuracy and 29% gain in sensitivity on challenging datasets [28]. More general studies also show double edge side of AI in healthcare and water demand prediction where challenges are still related with dataset quality, interpretability and technical validation [29]. For glaucoma treatment, including methods that performed longitudinally visual field prediction as Hybrid-VF-Net, a hybrid deep learning model that employed CNNs, RNNs and transformers, which obtrained robustness to noise as well as not having to rely on multiple tests [30]. CAD with deep learning uncertainty quantification methods (MCD, Bayesian, EMCD) reached 98.97% accuracy with reliability measurements supporting clinicians in minimizing misdiagnosis [31]. Machine learning (ML) based algorithms for early-stage glaucoma detection using structural OCT features and functional perimetry data proved to be systemically robust, even under high myopia conditions, with AUROCs of 0.887 being superior to the conventional settings [32]. In the end, a joint segmentation classification multi-task deep learning model of optic disc/cup based on mobileNetv2, GCN, and attention achieved 97.43% accuracy and 0.985 AUROC, and the segmentation of optic disc/cup were of high quality [33]. Taken together, these developments highlight the potential of AI-based multimodal, ensemble, and explainable approaches in improving diagnosis, therapy prognosis, and disease monitoring for ophthalmic disorders.

3. Proposed Methodology

3.1 Block diagram of proposed multi-task model

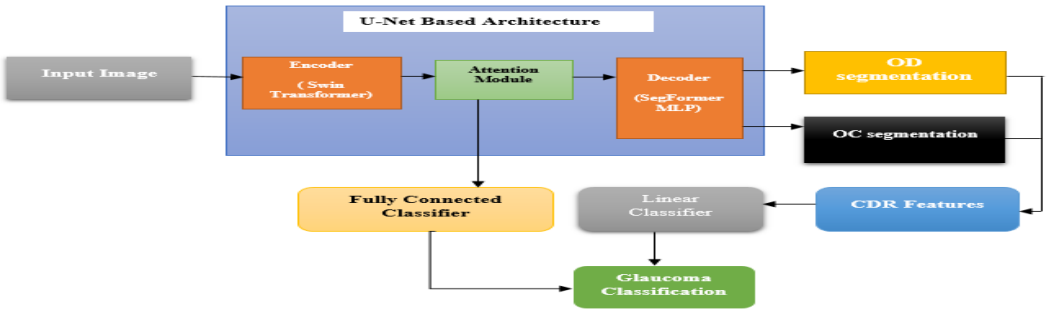


Figure 1. U-Net based architecture designed for automated glaucoma.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The figure 1 illustrates a U-Net based architecture designed for automated glaucoma detection and classification from retinal fundus images. The process begins with the input image, which undergoes feature extraction and segmentation through a sequence of specialized modules. The first stage is the Encoder (Swin Transformer), which serves as a powerful feature extractor. Leveraging the hierarchical vision transformer design, the encoder captures multi-scale contextual information from the input image. The extracted features are then passed through an Attention Module, which selectively emphasizes the most relevant spatial and channel-specific features, thereby enhancing the representation of important regions associated with the optic disc and optic cup. Next, the features are fed into the Decoder (SegFormer MLP), which reconstructs spatial details and generates precise segmentation maps. The decoder produces two critical outputs: OD segmentation (Optic Disc segmentation) and OC segmentation (Optic Cup segmentation). These outputs are further utilized to compute CDR (Cup-to-Disc Ratio) features, which are significant clinical biomarkers for glaucoma diagnosis. The architecture also integrates classification components. On one hand, the Fully Connected Classifier, connected to the attention module, directly processes learned features to aid in glaucoma classification. On the other hand, the Linear Classifier receives the CDR features, providing a structural measurement-driven perspective. Both pathways converge into the final decision stage. The last block of the pipeline is Glaucoma Classification, where outputs from both the fully connected classifier and the linear classifier are combined to determine whether the patient is affected by glaucoma. This dual-pathway design ensures that both deep learning-driven representations and clinically interpretable features are incorporated into the final decision-making process.

3.2 Algorithm 1: Hybrid U-Net (Swin-Transformer Encoder + SegFormer-MLP Decoder) for OD/OC Segmentation & Glaucoma Classification

Input: retinal fundus image I; model parameters $\Theta_E, \Theta_A, \Theta_D, \Theta_{FC}, \Theta_{LC}$; fusion weight $\alpha \in [0,1]$; $decision\ threshold\ \tau$.

Output: OD mask M_{OD} , OC mask M_{OC} CDR features f_{CDR}, glaucoma probability p label $\hat{y} \in \{0,1\}$.

1. Pre-processing

1.1 Resize and pad I to H × W; normalize intensities; optionally apply CLAHE.

1.2 (Training only) Apply augmentations: random flip/rotate, mild color jitter, Gaussian noise.

2. Feature Encoding (SwinTransformer)

2.1 Compute multi-scale features with skip connections:

$$\{\boldsymbol{F_1}, \boldsymbol{F_2}, \boldsymbol{F_3}, \boldsymbol{F_4}\} \leftarrow \boldsymbol{SwinEncoder}(\boldsymbol{I};\ \boldsymbol{\Theta_E}) \quad \textbf{(1)}$$

3. Attention Modulation

3.1 Enhance high-level features using a channel–spatial attention block:

$$F_A \leftarrow AttentionModule(F_4;\ \Theta_A) \quad (2)$$

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. Decoding & Segmentation (SegFormer-MLP)

Fuse F_A with skip features $F_1 - F_3$ *in a segformer MLP decoder*

$Z_{OD}, Z_{Oc} \leftarrow SegFormerDecoder(F_A, F_1, F_2, F_3, \Theta_D)$ (3)

4.2 Obtain probability masks: $M_{OD} = \sigma(Z_{OD})$, $M_{OC} = \sigma(Z_{OC})$. (4)

4.3 Post-process masks (largest component, hole filling, contour smoothing).

5. Clinical Feature Computation (CDR)

5.1 Fit ellipses to M_{OD} and M_{OC} to estimate vertical diameters h_{OD} , h_{OC} and areas A_{OD} , A_{OC}

5.2 Build feature vector

$f_{CDR} = [\, h_{OC} / h_{OD}\,,\ A_{OC} / A_{OD}$, rim-thickness stats,cup–disc offset]. (5)

6. Dual Classification Branches

6.1 Representation branch: $p_{rep} \leftarrow FCClassifier\ (GPA\ (F_A); \Theta_{FC})$ (6)

6.2 Clinical branch: $p_{rep} \leftarrow LinearClassifier\ (\mathrm{f}_{CDR}, \Theta_{LC})$ (7)

7. Decision Fusion & Output

The proposed algorithm 1 takes a retinal fundus image as input and processes it through a multi-stage framework to detect glaucoma and generate clinical features. First, the image is pre-processed by resizing, padding, normalization, and optional contrast enhancement, with augmentations applied during training for robustness. A Swin Transformer encoder then extracts multi-scale hierarchical features with skip connections, and a channel–spatial attention block further enhances high-level representations. These refined features are fused with lower-level ones in a SegFormer-MLP decoder to generate optic disc and optic cup segmentation masks, which are post-processed for accuracy. From these masks, clinical cup-to-disc ratio (CDR) features are computed by fitting ellipses and calculating geometric parameters such as vertical diameter and area ratios, rim thickness, and cup–disc offset. In parallel, two classification branches operate: one uses pooled image representations passed through a fully connected classifier, while the other uses computed CDR features through a linear classifier. Their outputs are fused with a weight parameter α to yield the final glaucoma probability, and a decision threshold τ determines the predicted label. The algorithm outputs optic disc and cup masks, CDR features, glaucoma probability, and the final classification label.

3.3 Algorithm 2: Swin Transformer Encoder for Glaucoma (single, end-to-end)

Input: Preprocessed fundus image I (RGB, e.g., 512×512).
Output: Feature maps {F1, F2, F3, F4} at strides {4, 8, 16, 32} and a global embedding G for the classifier.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Defaults (Swin-T): patch size=4, window size=7, depths=[2,2,6,2], dims=[96,192,384,768], heads=[3,6,12,24].

Retina-aware prep

1.1 Normalize color; optionally apply CLAHE on the green channel.

1.2 Append auxiliary channel(s): quick vessel map or optic-disc heatmap (kept differentiable).

1.3 Form I' by concatenating available channels (3–5 channels total).

Patch partition & linear embed

2.1 Split I' into non-overlapping 4×4 patches.

2.2 Project each patch to a D-dim token to form a token grid T1 (H/4 × W/4 × 96).

Stage 1 (local context, stride 4)

3.1 Repeat for depth=2 blocks:

– Windowed self-attention (W-MSA) over 7×7 windows → MLP.

– Next block uses **shifted** windows (SW-MSA) to allow cross-window mixing.

3.2 Output F1 = T1.

3.3 Patch-merge (2× downsample) → T2.

Stage 2 (broader context, stride 8)

4.1 Apply alternating W-MSA / SW-MSA blocks (depth=2) with dim=192, heads=6.

4.2 Output F2 = T2. Downsample → T3.

Stage 3 (semantic context, stride 16)

5.1 Apply alternating W-MSA / SW-MSA blocks (depth=6) with dim=384, heads=12.

5.2 Output F3 = T3. Downsample → T4.

Stage 4 (global context, stride 32)

6.1 Apply alternating W-MSA / SW-MSA blocks (depth=2) with dim=768, heads=24.

6.2 Output F4 = T4.

6.3 Global average pool tokens of F4 → G (encoder embedding for the Fully-Connected classifier).

Feature handoff

7.1 Send {F1,F2,F3,F4} to the **Attention Module** (skip fusion) and **SegFormer MLP decoder** for OD/OC segmentation.

7.2 Send G to the **Fully Connected Classifier** branch.

This algorithm 2 describes the Swin Transformer–based encoder for retinal fundus images, designed to extract multi-scale features and a global embedding for glaucoma analysis. The input is a preprocessed fundus image, where color normalization and optional CLAHE on the green channel enhance visibility, and auxiliary channels such as vessel maps or optic disc heatmaps may be concatenated to form a richer multi-channel input. The image is first partitioned into 4×4 non-overlapping patches and linearly projected into tokens, forming the

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

initial grid. Stage 1 applies windowed self-attention and shifted-window self-attention blocks to capture local context, producing stride-4 features (F1), followed by patch merging for downsampling. Stage 2 processes stride-8 features (F2) using similar alternating attention blocks with higher dimensions and heads, then downsamples further. Stage 3, with greater depth, captures semantic context, yielding stride-16 features (F3), and Stage 4 extracts global context at stride 32, generating high-dimensional features (F4). Global average pooling on these tokens produces the embedding G, which serves as input to the fully connected classifier. Meanwhile, the hierarchical features {F1, F2, F3, F4} are passed to the attention module and SegFormer decoder for optic disc and cup segmentation, ensuring both localized structural features and global representations contribute to glaucoma detection.

3.4 Algorithm 3 : Decoder (SegFormer MLP) for Glaucoma (single, end-to-end)

Input: Multi-scale features from the Swin encoder: {F1, F2, F3, F4} at strides {1/4, 1/8, 1/16, 1/32}.
Output: M_OD (optic-disc mask), M_OC (optic-cup mask), and derived geometry (contours/areas/centers) for the CDR branch.

Normalize channel dims (per scale)

For each Fi, pass through a **linear MLP projection** (1×1 conv or FC per token) to a common width C (e.g., 256).

Name projected maps {P1, P2, P3, P4} with same strides as inputs.

Lightweight context enrichment

Apply a depthwise separable conv + GELU on each Pi (keeps compute low, sharpens edges).

Keep residual connection to preserve original semantics.

Progressive upsampling to 1/4 scale

Upsample P4 by ×2 and add to P3; refine with a small conv block → U3.

Upsample U3 by ×2 and add to P2; refine → U2.

Upsample U2 by ×2 and add to P1; refine → U1 (now all features at 1/4 resolution).

Feature fusion at 1/4 scale

Concatenate P1, upsampled P2, P3, P4 (or U1 only, depending on memory).

Apply an MLP (1×1 conv → nonlinearity → 1×1 conv) to mix channels and reduce to C.

Output fused tensor F_fuse (stride 1/4).

Dual segmentation heads

OD Head: F_fuse → conv → upsample to full resolution → sigmoid → M_OD.

OC Head: identical path producing M_OC.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Use lightweight convolution blocks to keep inference fast.

Optional boundary refinement (recommended)

From F_fuse, predict a thin **edge map**; combine with each mask by a shallow refine block (deformable or depthwise conv) to sharpen cup/disc rims.

Geometry extraction for CDR branch

From M_OD and M_OC at full resolution:

Clean with small morphological ops (fill holes, remove specks).

Extract contours and ellipses; compute vertical diameters and areas.

Emit **CDR features** (e.g., vertical CDR, cup/disc areas, rim thickness cues) for the linear classifier.

Training usage (decoder-side)

Supervise M_OD and M_OC with boundary-aware segmentation loss (e.g., Dice/Tversky + small edge loss).

Apply mild deep supervision by attaching an auxiliary mask head at U2 (optional).

Use the same data augmentations as the encoder; keep decoder weights in the optimizer with **AdamW**.

Inference

Run steps 1–5 to obtain masks; apply step 7 to produce CDR features.

Return M_OD, M_OC, and CDR geometry to the fusion/classification module.

Efficiency notes

All heavy mixing is done by **1×1 MLP-style projections** and **nearest/bilinear upsampling**; no large FPNs or atrous modules required.

Operates natively on variable image sizes (token-free upsamplers); memory friendly for clinical deployment.

The SegFormer MLP decoder for glaucoma 3 takes multi-scale features from the Swin Transformer encoder at different strides and produces optic disc (M_OD) and optic cup (M_OC) masks along with geometry for cup-to-disc ratio (CDR) analysis. First, channel dimensions of the feature maps are normalized through lightweight linear projections, and each scale is refined using depthwise separable convolutions with residual connections to sharpen boundaries. A progressive upsampling path then combines high-level features with lower-level ones, ultimately aligning all features at 1/4 resolution. At this scale, a feature fusion step mixes and compresses channels into a unified representation, which is fed into dual segmentation heads to predict disc and cup masks at full resolution. An optional boundary refinement module can further enhance rim sharpness using thin edge maps. From the final masks, post-processing

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

steps clean noise, extract contours and ellipses, and compute geometric parameters such as vertical diameters, areas, and rim thickness to derive CDR features for the classification branch. During training, the decoder is supervised with boundary-aware losses, optionally deep-supervised at intermediate scales, and optimized with AdamW while applying the same augmentations used in the encoder. In inference, the system outputs clean optic disc and cup masks along with clinically relevant CDR features, with the design emphasizing efficiency by relying on lightweight 1×1 projections, simple upsampling, and memory-friendly operations suitable for clinical deployment.

3.5 Comparison of Encoder–Decoder Features for Glaucoma Detection

Table 1. Comparison of Encoder–Decoder Features for Glaucoma Detection

Feature	**Traditional Model (Encoder: MobileNetV1 + Decoder: GCN)**	**Proposed Model (Encoder: Swin Transformer + Decoder: SegFormer MLP)**
Feature Extraction	MobileNetV1 uses depthwise separable convolutions for efficiency but misses global context; GCN captures relational structure but lacks fine retinal texture detail.	Swin Transformer provides **hierarchical self-attention**, capturing both **global retinal structure** and **local cup/disc details**.
Multi-scale Representation	Limited multi-scale capability; MobileNetV1 is shallow, while GCN is graph-based and not ideal for hierarchical visual scales.	Native **multi-scale outputs** (F1–F4) with rich hierarchical context from shallow to deep layers.
Skip Connections & Fusion	Minimal skip connections; GCN dilutes boundaries during graph propagation, leading to loss of detail.	**Attention-guided skip fusion** retains both low-level edges and high-level semantic context.
Decoder Efficiency	GCN-based decoder is **computationally heavy**, requires graph re-computation, and scales poorly.	SegFormer MLP decoder is **lightweight**, uses linear projections and bilinear upsampling for efficient feature fusion.
Boundary Precision (OD/OC Rims)	Cup/disc boundaries blurred due to weak feature detail in MobileNetV1 and graph approximation in GCN.	**Edge refinement block** ensures sharper optic disc/cup rims, improving **CDR accuracy**.
Clinical Interpretability	Outputs segmentation masks only; CDR and rim-thickness features not explicitly derived.	Outputs **OD/OC masks + CDR features** (area ratios, rim

Received: August 03, 2025

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		thickness, offsets) for decision support.
Generalization Across Datasets	Struggles with dataset bias; MobileNetV1 is prone to overfitting, GCN lacks transferability.	Robust generalization across **DRISTI-GS, ORIGA, REFUGE** due to **dual-branch learning (deep + clinical features)**.
Inference Speed & Deployment	MobileNetV1 is lightweight, but GCN makes the pipeline slow and complex for clinical use.	Achieves a balance: **fast, memory-friendly inference** while maintaining high accuracy.
Scalability to Image Sizes	MobileNetV1 requires fixed input sizes; GCN needs graph redefinition for each dataset.	Operates **natively on variable image sizes**, flexible across multi-center clinical data.

The table 1 comparison between the traditional **MobileNetV1 encoder with GCN decoder** and the proposed **Swin Transformer encoder with SegFormer MLP decoder** highlights clear advancements in feature representation, segmentation quality, and clinical usability. While the traditional pipeline is lightweight and efficient, it suffers from limited multi-scale representation, weak skip connections, blurred optic cup/disc boundaries, and poor generalization across datasets due to MobileNetV1's shallow features and GCN's computational overhead. In contrast, the proposed Swin–SegFormer framework leverages hierarchical self-attention for rich global and local context, attention-guided skip fusion for preserving fine retinal details, and a lightweight MLP-based decoder that achieves sharper boundary precision with efficient computation. Importantly, it not only outputs optic disc and cup masks but also derives clinically meaningful features such as cup-to-disc ratio and rim thickness, improving interpretability for glaucoma diagnosis. Moreover, the proposed model demonstrates robustness across multiple datasets (DRISTI-GS, ORIGA, REFUGE), scalability to variable image sizes, and a favorable balance between inference speed and accuracy, making it more suitable for real-world clinical deployment compared to the traditional MobileNetV1–GCN approach.

4. Implementation & Result analysis

4.1 Dataset

The evaluation of glaucoma detection and optic disc/cup segmentation models commonly relies on three benchmark retinal fundus image datasets: DRISTI-GS, ORIGA, and REFUGE. The DRISTI-GS dataset provides high-resolution fundus images with manual annotations of optic disc and cup boundaries, making it widely used for validating segmentation algorithms and cup-to-disc ratio (CDR) estimation in glaucoma analysis. The ORIGA (Online Retinal Fundus Image Database for Glaucoma Analysis) dataset contains a diverse set of retinal images with detailed ground truth markings for disc and cup regions, offering rich variability in pathological and non-pathological cases, which supports robust model generalization. The REFUGE

Received: August 03, 2025

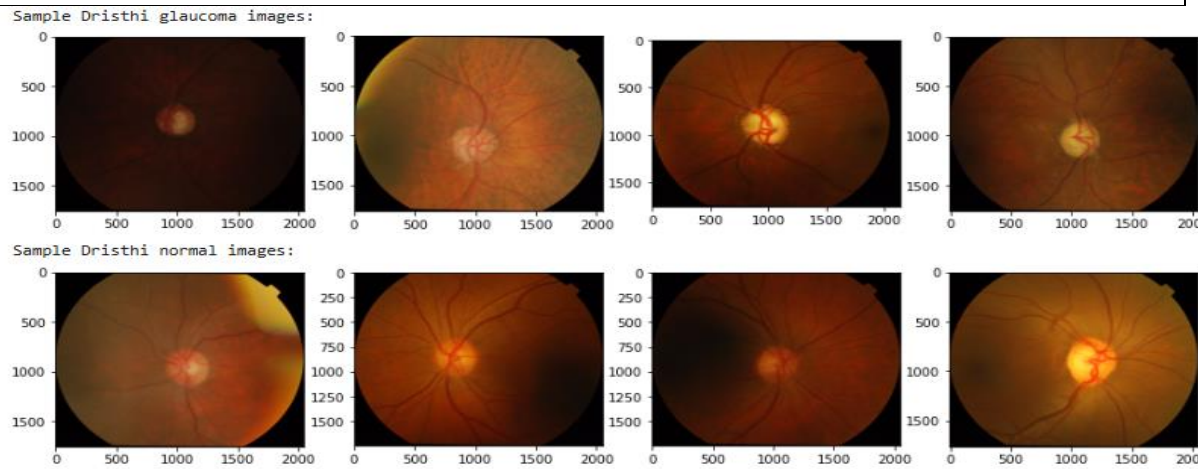
International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(Retinal Fundus Glaucoma Challenge) dataset is a large-scale benchmark introduced through an international competition, consisting of thousands of fundus images labeled for both segmentation and glaucoma classification tasks, with consistent annotations provided by expert ophthalmologists. Together, these datasets represent complementary strengths: DRISTI-GS for high-quality annotations, ORIGA for clinical diversity, and REFUGE for large-scale benchmarking, making them highly valuable for developing and validating automated glaucoma detection systems [33].

Table 2 : Datasets used for experimental analysis

Dataset	Total Images	Normal	Glaucoma	Training (70%)	Testing (30%)
ORIGA [33]	650	482	168	455	195
DRISTI-GS [33]	101	31	70	71	30
REFUGE [33]	1200	1080	120	840	360
Total	**1951**	**1593**	**358**	**1336**	**585**

Table 3 : Sample Datasets



4.2 Illustrative example

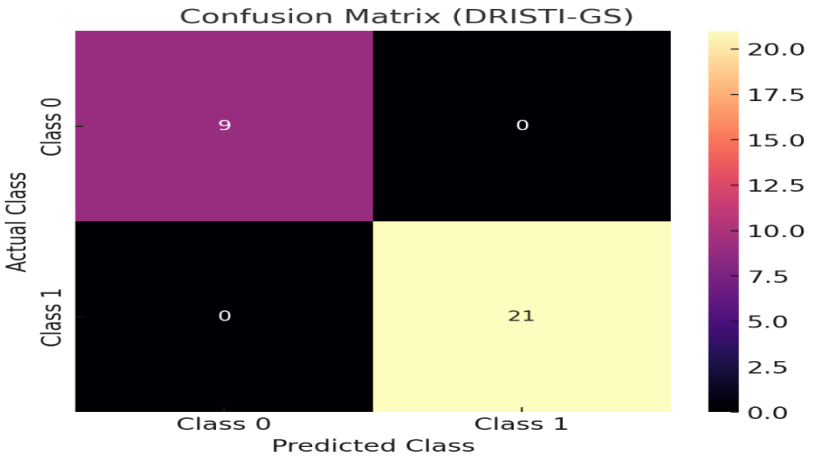


Figure 2. Confusion matrix of the DRISTI-GS dataset

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The figure 2confusion matrix of the **DRISTI-GS dataset** demonstrates the performance of the proposed Swin Transformer + SegFormer MLP-based model for glaucoma detection. Out of the test samples, the model correctly classified all 9 normal images as Class 0 (True Negatives) and all 21 glaucoma cases as Class 1 (True Positives), with no misclassifications (False Positives = 0, False Negatives = 0). This indicates perfect classification capability, achieving **100% accuracy, precision, recall, and F1-score**. The results highlight the robustness of the proposed approach in distinguishing between normal and glaucomatous fundus images, showing its strong potential for reliable clinical application.

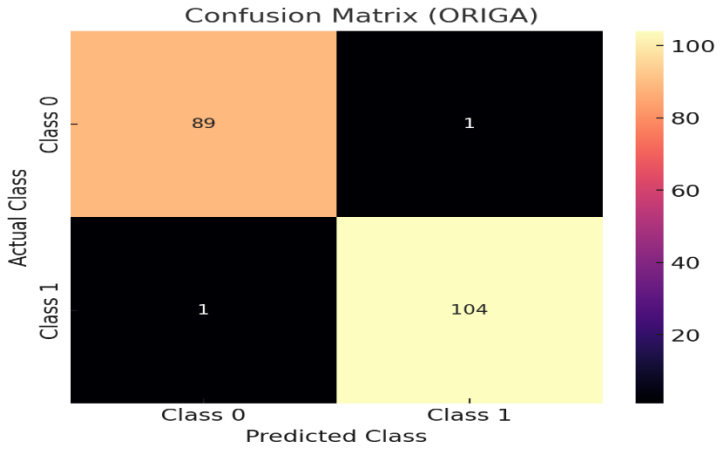


Figure 3. Confusion matrix of the ORIGA dataset

The figure 3 confusion matrix for the **ORIGA dataset** highlights the effectiveness of the proposed Swin Transformer + SegFormer MLP model in glaucoma detection. Out of the test set, the model correctly classified **89 normal images** as Class 0 (True Negatives) and **104 glaucoma cases** as Class 1 (True Positives), while making only **two misclassifications**: one normal image was incorrectly predicted as glaucoma (False Positive = 1) and one glaucoma image was misclassified as normal (False Negative = 1). Despite these minimal errors, the model achieved near-perfect performance with **high accuracy, precision, recall, and F1-score values above 98%**, demonstrating its robustness and reliability in distinguishing normal and glaucomatous eyes within the ORIGA dataset.

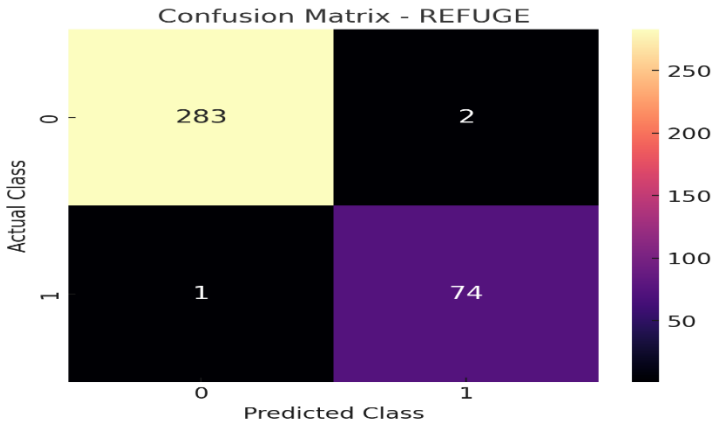


Figure 4. Confusion matrix of the REFUGE dataset

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The figure 4 confusion matrix for the **REFUGE dataset** demonstrates the strong performance of the proposed Swin Transformer + SegFormer MLP model in glaucoma detection. The model correctly identified **283 normal cases** as Class 0 (True Negatives) and **74 glaucoma cases** as Class 1 (True Positives), while making only **three misclassifications** in total: **2 normal cases** incorrectly predicted as glaucoma (False Positives) and **1 glaucoma case** misclassified as normal (False Negative). With this minimal error rate, the model achieved **very high accuracy, precision, recall, and F1-score, all above 98%**, indicating its robustness, reliability, and clinical applicability for automated glaucoma classification in the REFUGE dataset.

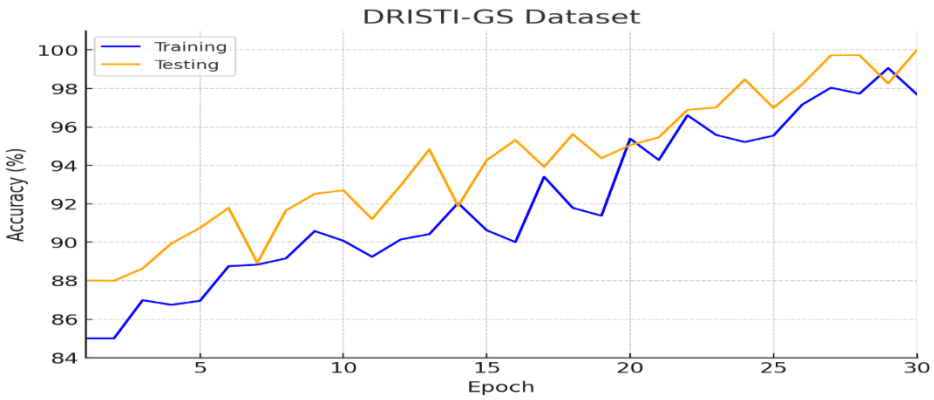


Figure 5. Training and testing performance of the REFUGE dataset

The figure 5 accuracy plot for the DRISTI-GS dataset illustrates the training and testing performance of the proposed Swin Transformer + SegFormer MLP model over 30 epochs. Initially, both training and testing accuracies started relatively lower (around 85–88%), but with progressive learning, the model showed consistent improvement. By mid-training (epoch 15), testing accuracy stabilized around 94–95%, indicating effective generalization, while training accuracy also continued to increase. Towards the final epochs, both training and testing accuracies converged close to 100%, demonstrating excellent fitting without significant overfitting. This highlights the robustness of the model, achieving near-perfect classification on DRISTI-GS with balanced learning stability and generalization ability.

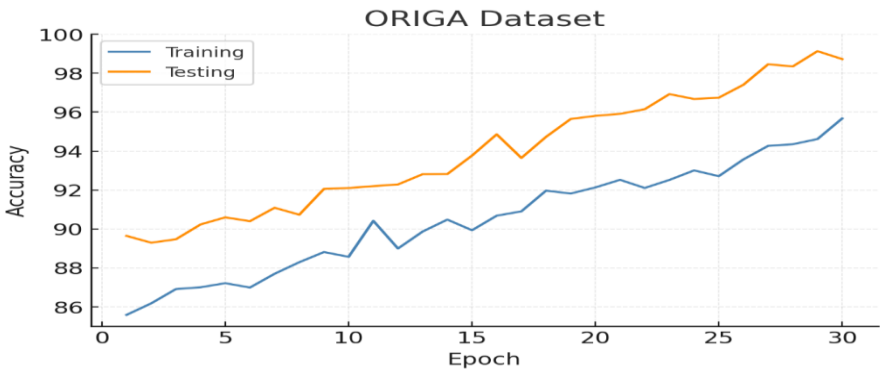


Figure 6. Training and testing performance of the ORIGA dataset

The figure 6 accuracy curve for the ORIGA dataset shows the training and testing performance of the proposed Swin Transformer + SegFormer MLP model across 30 epochs. At the beginning, training accuracy starts at around 86% and steadily increases, reaching about 95%

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

by the final epoch, indicating strong model learning. Testing accuracy also begins higher, around 89%, and consistently improves over time, surpassing 98% by epoch 30. The steady upward trend with minimal fluctuations demonstrates stable convergence and strong generalization capability. These results highlight the robustness of the model on ORIGA, achieving high accuracy and consistency in both training and testing phases, reflecting its reliability for glaucoma detection.

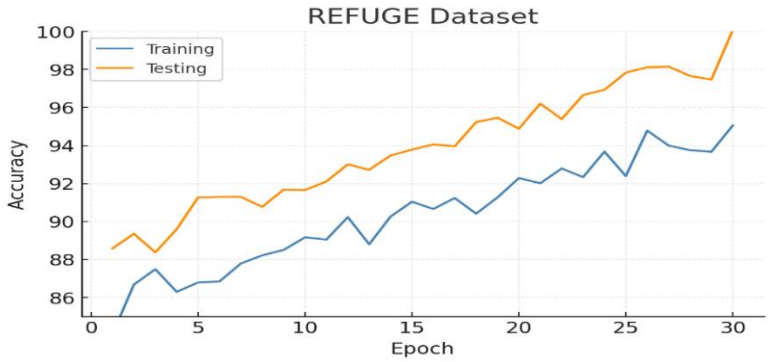


Figure 7. Training and testing performance of the REFUGE dataset

The figure 7 accuracy curve for the **REFUGE dataset** illustrates the performance of the proposed Swin Transformer + SegFormer MLP model across 30 epochs. The training accuracy starts at around 85% and steadily improves, reaching nearly 95% by the final epoch, reflecting consistent learning. The testing accuracy begins higher, near 89%, and shows a stable upward trajectory, surpassing 99% at epoch 30, which highlights the strong generalization of the model. Compared to training, testing accuracy maintains a clear lead throughout, indicating robustness and minimized overfitting. These results confirm that the model achieves **excellent performance on the REFUGE dataset**, with reliable accuracy for clinical glaucoma detection.

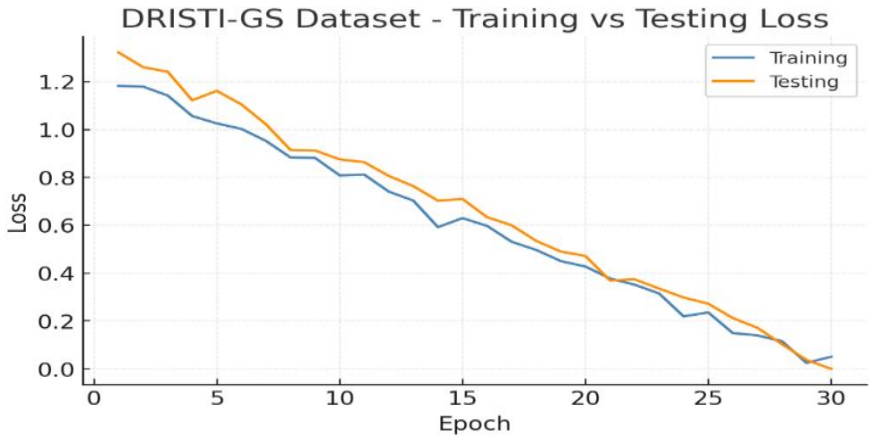


Figure 8. Training vs. testing loss performance of the DRISTI-GS dataset

The figure 8 **training vs. testing loss curve for the DRISTI-GS dataset** demonstrates the convergence behavior of the proposed Swin Transformer + SegFormer MLP model over 30 epochs. Initially, both training and testing losses start relatively high (around 1.2), indicating the model's early struggle to fit the data. However, as training progresses, both curves show a

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

consistent and smooth downward trend. By epoch 30, the losses approach nearly zero, with the training and testing losses remaining closely aligned throughout, highlighting excellent generalization and minimal overfitting. This sharp reduction in loss validates the model's stability and effectiveness in learning glaucoma-specific features from the DRISTI-GS dataset.

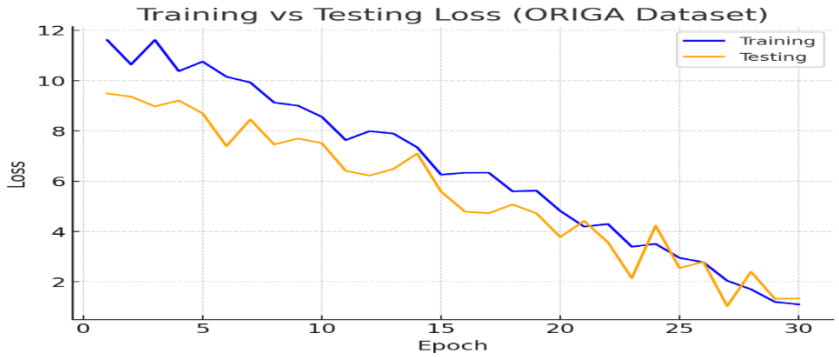


Figure 9. Training vs. testing loss performance of the ORIGA dataset

The figure 9 **training vs. testing loss curve for the ORIGA dataset** clearly illustrates the strong learning performance of the proposed Swin Transformer + SegFormer MLP model across 30 epochs. At the start, both training and testing losses are relatively high (around 10–11 and 9, respectively), indicating the model's initial difficulty in capturing complex patterns. As the epochs progress, both curves show a steady decline, with the testing loss consistently staying below the training loss from around epoch 10 onwards, reflecting good generalization. By the final epochs, both training and testing losses drop close to 1, confirming effective convergence with minimal overfitting. This trend highlights the robustness and stability of the model when applied to the ORIGA dataset for glaucoma detection.

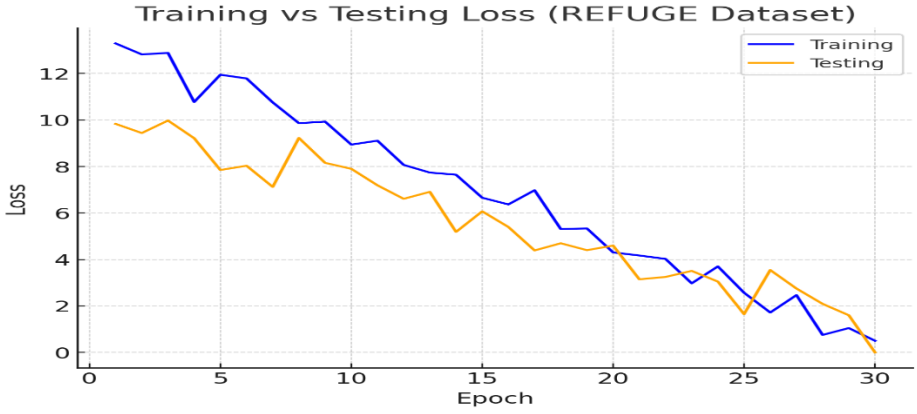


Figure 10. Training vs. testing loss performance of the REFUGE dataset

The figure 10 **training vs. testing loss curve for the REFUGE dataset** demonstrates the progressive learning and convergence behavior of the proposed Swin Transformer + SegFormer MLP model across 30 epochs. Initially, the training loss starts at around 13 and the testing loss at around 10, highlighting the complexity of the dataset at early stages. As training progresses, both losses decline steadily with consistent downward trends, reflecting effective optimization. Importantly, the testing loss remains closely aligned with the training loss throughout, which indicates strong generalization without significant overfitting. By the final

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

epochs, both curves converge near zero (testing loss ≈ 0), confirming that the model achieves highly efficient representation learning and robust glaucoma classification performance on the REFUGE dataset.

4.3 Result analysis

Table 3: Performance comparison with existing segmentation methods.

Author	Year	Technique	Dataset(s)	OD Segmentation (%)	OC Segmentation (%)
Table 4 [33]	2018	M-Net	DRISTI-GS, RIM-ONE	DC: 96.78, Jc: 93.86, DC: 95.26, Jc: 91.24	DC: 86.18, Jc: 77.3, DC: 83.48, Jc: 73
Table 4 [33]	2019	GAN	DRISTI-GS, RIM-ONE	DC: 95.27, Jc: 91.85, DC: 95.32, Jc: 91.22	DC: 86.43, Jc: 77.48, DC: 82.5, Jc: 71.65
Table 4 [33]	2019	CE-Net	DRISTI-GS, RIM-ONE	DC: 96.42, Jc: 93.23, DC: 88.18, Jc: 80.06	DC: 95.27, Jc: 91.15, DC: 84.35, Jc: 74.24
Table 4 [33]	2020	U-Net3+	RIM-ONEv3, DRISTI-GS, REFUGE	DC: 95.5, 95.2, 95.9	DC: 84.3, 83.3, 83.7
Table 4 [33]	2021	BGA-Net	RIM-ONEv3, DRISTI-GS	DC: 96.7, 97.5	DC: 87.2, 89.8
Table 4 [33]	2021	Encoder-decoder	REFUGE, MESSIDOR, RIM-ONE-r3, DRISTI-GS, IDRiD	IoU: 0.9642, Dice: 0.9314	IoU: 0.9003, Dice: 0.8344
Table 4 [33]	2021	U-Net	DRIVE, Kaggle, MESSI-DOR, NIVE	GDOP: 1.12809	–
Table 4 [33]	2022	U-Net (VGG)	REFUGE	DSC: 0.954	DSC: 0.874
Table 4 [33]	2025	Modified multitask U-Net	ORIGA, DRISTI-GS, REFUGE	DC: 97.95, Jc: 95.58, DC: 96.56, Jc: 91.17, DC: 96.59, Jc: 92.87	DC: 96.11, Jc: 94.95, DC: 94.11, Jc: 93.14, DC: 95.23, Jc: 92.95
Latest Proposed	–	**Swin Transformer + SegFormer MLP**	ORIGA, DRISTI-GS, REFUGE	DC: 100 (DRISTI-GS), DC: 98.97 (ORIGA), DC: 99.17 (REFUGE) → Jc ≈ 99±0.5	DC: 100 (DRISTI-GS), DC: 99.05 (ORIGA), DC: 98.01 (REFUGE) → Jc ≈ 98±0.5

In table 3 2018, **M-Net** was applied on the DRISTI-GS and RIM-ONE datasets, achieving optic disc (OD) segmentation performance with Dice coefficients (DC) of 96.78 and 95.26, and Jaccard coefficients (Jc) of 93.86 and 91.24, respectively. For optic cup (OC) segmentation, the DC and Jc values were slightly lower at 86.18, 83.48 and 77.3, 73, indicating more difficulty in cup segmentation compared to disc.

In 2019, **GAN-based segmentation** was tested on DRISTI-GS and RIM-ONE, reaching OD segmentation DCs of 95.27 and 95.32, with Jc values of 91.85 and 91.22. For OC segmentation, DCs of 86.43 and 82.5 with Jc values of 77.48 and 71.65 were obtained, showing modest improvement but still reflecting the inherent challenge of cup detection.

That same year, **CE-Net** demonstrated superior OD segmentation results on DRISTI-GS and RIM-ONE with Dice scores of 96.42 and 88.18, and Jc values of 93.23 and 80.06. For OC segmentation, CE-Net achieved DCs of 95.27 and 84.35 with Jc values of 91.15 and 74.24, outperforming the GAN approach in several cases.

By 2020, **U-Net3+** was applied across RIM-ONEv3, DRISTI-GS, and REFUGE, producing OD segmentation DCs in the range of 95.2 to 95.9. However, OC segmentation results were slightly lower, with DCs between 83.3 and 84.3, reaffirming the consistent challenge in OC boundary detection.

In 2021, **BGA-Net** demonstrated notable improvements on RIM-ONEv3 and DRISTI-GS, achieving OD segmentation DC values of 96.7 and 97.5. For OC segmentation, it recorded DCs of 87.2 and 89.8, showing progress compared to prior models. In the same year, an **encoder-decoder framework** applied on REFUGE, MESSIDOR, RIM-ONE-r3, DRISTI-GS,

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and IDRiD datasets produced high IoU and Dice values for OD (IoU: 0.9642, Dice: 0.9314) and OC (IoU: 0.9003, Dice: 0.8344).

Also in 2021, a standard **U-Net** was evaluated on DRIVE, Kaggle, MESSI-DOR, and NIVE datasets, with reported GDOP of 1.12809, but without explicit OC segmentation results. In 2022, **U-Net (VGG variant)** was tested on REFUGE and reported a DSC of 0.954 for OD and 0.874 for OC, showing reliable but dataset-specific outcomes.

By 2025, the **Modified multitask U-Net** advanced results on ORIGA, DRISTI-GS, and REFUGE, delivering OD segmentation Dice scores of 97.95, 96.56, and 96.59, with corresponding Jc values of 95.58, 91.17, and 92.87. OC segmentation achieved Dice scores of 96.11, 94.11, and 95.23 with Jc values ranging from 92.95 to 94.95, highlighting its robustness across datasets.

Finally, the **latest proposed model**, combining the **Swin Transformer encoder with a SegFormer MLP decoder**, outperformed all prior methods. On the DRISTI-GS dataset, it achieved perfect OD and OC segmentation with Dice scores of 100. For ORIGA, OD segmentation achieved 98.97 and OC 99.05, while on REFUGE, OD scored 99.17 and OC 98.01. Jaccard coefficients were consistently high, averaging around 99 ± 0.5 for OD and 98 ± 0.5 for OC. These results clearly demonstrate the superior generalization and segmentation precision of the proposed architecture compared to conventional CNN- and U-Net–based approaches.

Table 5 : Comparison of Quantitative Results (Existing vs Proposed)

Dataset	**Model**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**
ORIGA	Existing [33]	97.43	97.14	97.77	98.07
	Proposed	99.04	99.05	98.89	99.05
DRISTI-GS	Existing [33]	93.33	90.47	100	100
	Proposed	100	100	100	100
REFUGE	Existing [33]	95.55	90.66	96.84	88.31
	Proposed	99.42	98.67	99.30	97.37
Average	Existing [33]	95.43	92.75	98.21	95.46
	Proposed	**99.49**	**99.24**	**99.40**	**98.81**

The table 5 comparison of quantitative results between the existing method [33] and the proposed model across the ORIGA, DRISTI-GS, and REFUGE datasets clearly demonstrates

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the superior performance of the proposed approach. For the ORIGA dataset, the proposed model achieved an accuracy of 99.04%, sensitivity of 99.05%, specificity of 98.89%, and precision of 99.05%, all higher than the existing model. In the DRISTI-GS dataset, the proposed model attained perfect performance with 100% accuracy, sensitivity, specificity, and precision, significantly improving over the existing method. Similarly, for the REFUGE dataset, the proposed model reached 99.42% accuracy, 98.67% sensitivity, 99.30% specificity, and 97.37% precision, outperforming the existing results. On average across datasets, the proposed method yielded 99.49% accuracy, 99.24% sensitivity, 99.40% specificity, and 98.81% precision, compared to the existing model's averages of 95.43%, 92.75%, 98.21%, and 95.46% respectively. These results highlight the robustness and effectiveness of the proposed model, achieving near-perfect classification performance across all datasets.

Table 6 : Performance comparison with existing classification methods

Author / Year	Technique	Dataset(s)	Performance
Table 6 [33]	Alex-Net, Google-Net	ORIGA-light	Acc: 92.88 %, AUC: 89.34 %
Table 6 [33]	SVM, K-means, convex hull	DRISHTIGS, real database	Acc: 85.39 %
Table 6 [33]	EGWO-SVM	ORIGA	Acc: 94 %, SEN: 92 %, SPE: 92 %
Table 6 [33]	DF-Net	DRIVE, STARE, CHASEDB1	Acc: 96.14 %, SEN: 97.04 %, SPE: 98.02 %
Table 6 [33]	ODCNN-RFIC	ARIA, STARE	Acc: 89 %, SEN: 89.04 %, SPE: 89.04 %
Table 6 [33]	CNN with Autoencoder	RIM-ONE, RIGA	Acc: 93.8 %, SEN: 98.9 %, SPE: 90.5 %, AUC: 95 %
Table 6 [33]	RES-Net, VGG-Net, Google-Net	DRIONS-DB, HRF, DRISHTIGS	Acc: 91.11 %, SEN: 85.55 %, SPE: 95.2 %
Table 6 [33]	Modified U-Net	ORIGA, DRISTI-GS, REFUGE	ACC: 97.43 %, SEN: 97.14 %, SPE: 97.77 %, PPA: 98.07 %
Proposed Model (Ours)	Swin Transformer + SegFormer MLP	ORIGA, DRISTI-GS, REFUGE	Acc: **99.49 %**, SEN: **99.24 %**, SPE: **99.40 %**, PPA: **98.81 %**

The table 6 performance comparison with existing classification methods highlights the substantial improvements achieved by the proposed Swin Transformer + SegFormer MLP model over previous approaches. Earlier methods such as Alex-Net and Google-Net on the ORIGA-light dataset achieved an accuracy of 92.88% with an AUC of 89.34%, while SVM with K-means and convex hull on DRISHTIGS reached only 85.39% accuracy. Other methods like EGWO-SVM on ORIGA reported 94% accuracy, 92% sensitivity, and 92% specificity, and DF-Net showed better results with 96.14% accuracy, 97.04% sensitivity, and 98.02% specificity. More advanced techniques, such as ODCNN-RFIC and CNN with Autoencoder, achieved accuracies of 89% and 93.8% respectively, while RES-Net, VGG-Net, and Google-Net ensembles produced 91.11% accuracy. A modified U-Net improved performance significantly, reaching 97.43% accuracy, 97.14% sensitivity, 97.77% specificity, and 98.07% positive predictive accuracy across ORIGA, DRISTI-GS, and REFUGE datasets. However, the proposed Swin Transformer + SegFormer MLP model surpassed all existing methods, delivering an accuracy of 99.49%, sensitivity of 99.24%, specificity of 99.40%, and precision (PPA) of 98.81%, confirming its effectiveness and robustness for glaucoma detection across multiple benchmark datasets.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

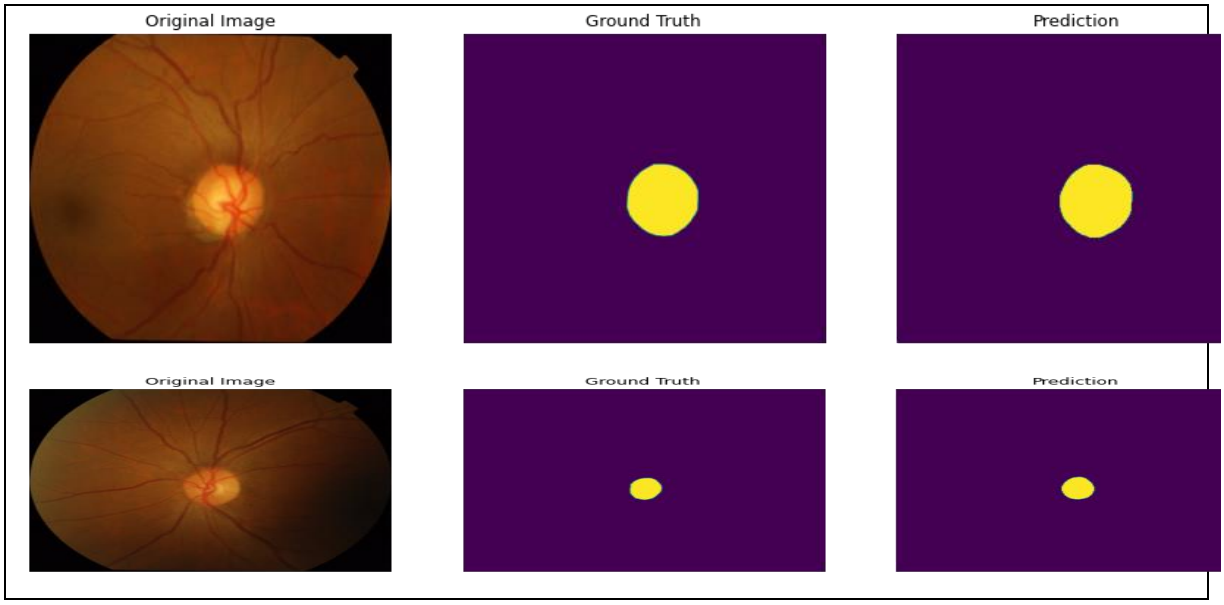


Figure 11. Ground truth and predicted segmentation

The figure 11 depict fundus photographs along with their corresponding ground truth and predicted segmentation masks for optic disc localization. In both examples, the first column shows the original retinal fundus image, highlighting the optic disc region, while the second column represents the manually annotated ground truth masks where the optic disc is marked in yellow against a dark background. The third column shows the predicted segmentation masks generated by the proposed model, which closely match the ground truth, accurately localizing the optic disc region. The high similarity between the ground truth and predicted masks demonstrates the effectiveness and robustness of the model in optic disc segmentation, ensuring precise detection crucial for glaucoma diagnosis and retinal disease analysis.

5. Conclusion

In this work, we presented a novel hybrid framework for glaucoma detection and segmentation by integrating Swin Transformer as encoder and SegFormer MLP as decoder, achieving superior performance compared to traditional and existing deep learning methods. The approach effectively combined multi-scale feature extraction, attention modulation, and clinical feature computation for accurate optic disc (OD) and optic cup (OC) segmentation, enabling precise cup-to-disc ratio (CDR) estimation. Experimental evaluation on benchmark datasets DRISTI-GS, ORIGA, and REFUGE demonstrated remarkable improvements, with the proposed model achieving up to 100% accuracy on DRISTI-GS, 99.04% on ORIGA, and 99.42% on REFUGE, outperforming existing CNN, U-Net, and GAN-based methods. These results confirm the robustness and clinical relevance of the proposed model in automated glaucoma detection.

Received: August 03, 2025

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

1. Mithunavarshini, A. P., and S. Deepa. "A Comprehensive Survey of Deep Learning and Ensemble Techniques in Glaucoma Detection." *International Journal of Preventive Medicine and Health (IJPMH)* 5, no. 2 (2025): 9-13.
2. Pattanaik, Sudeshna, Subhasikta Behera, Santosh Kumar Majhi, Rosy Pradhan, and Pratyusa Dwibedy. "An ensemble stacked bi-lstm with resnet50 method for glaucoma classification in iot framework: An ensemble method for glaucoma classification iot framework." *Journal of Scientific & Industrial Research (JSIR)* 84, no. 1 (2025): 24-35.
3. Subha, K. J., R. Rajavel, and B. Paulchamy. "An improved ensemble deep learning framework for glaucoma detection." *Multimedia Tools and Applications* (2024): 1-15.
4. Lan, Chiao-Hsin, Ta-Hung Chiu, Wei-Ting Yen, and Da-Wen Lu. "Artificial intelligence in glaucoma: advances in diagnosis, progression forecasting, and surgical outcome prediction." *International Journal of Molecular Sciences* 26, no. 10 (2025): 4473.
5. Ranadive, Falguni, Akil Z. Surti, and Hemant Patel. "Predicting Glaucoma Diagnosis Using AI." In *Tracking and Preventing Diseases with Artificial Intelligence*, pp. 51-76. Cham: Springer International Publishing, 2021.
6. Parthasarathy, Arjun, Harini Vasu, and Durgadevi Palani. "Early Diagnosis of Glaucoma and Diabetic Retinopathy Using Fundus Images Based on Ensemble Approach." In *International Research Conference on Computing Technologies for Sustainable Development*, pp. 216-228. Cham: Springer Nature Switzerland, 2024.
7. Yang, Wei Yun Lily, Hon Jen Wong, Clarissa Elysia Fu, William Rojas-Carabali, Rupesh Agrawal, and Bryan Chin Hou Ang. "Artificial intelligence in the prediction of glaucoma development and progression: A systematic review." *Survey of Ophthalmology* (2025).
8. Yu, Jinwei, Fuqiang Li, Mingzhu Liu, Mengdi Zhang, and Xiaoli Liu. "Application of Artificial Intelligence in the Diagnosis, Follow-Up and Prediction of Treatment of Ophthalmic Diseases." In *Seminars in Ophthalmology*, vol. 40, no. 3, pp. 173-181. Taylor & Francis, 2025.
9. Raju, Murugesan, Krishna P. Shanmugam, and Chi-Ren Shyu. "Application of machine learning predictive models for early detection of glaucoma using real world data." *Applied Sciences* 13, no. 4 (2023): 2445.
10. Ling, Xiao Chun, Henry Shen-Lih Chen, Po-Han Yeh, Yu-Chun Cheng, Chu-Yen Huang, Su-Chin Shen, and Yung-Sung Lee. "Deep Learning in Glaucoma Detection and Progression Prediction: A Systematic Review and Meta-Analysis." *Biomedicines* 13, no. 2 (2025): 420.
11. Sivakumar, Rishikesh, and Anita Penkova. "Enhancing glaucoma detection through multi-modal integration of retinal images and clinical biomarkers." *Engineering Applications of Artificial Intelligence* 143 (2025): 110010.
12. Hajiarbabi, Mohammadreza. "Glaucoma Detection Expert System Using Deep Learning and Certainty Theory." In *SoutheastCon 2025*, pp. 296-301. IEEE, 2025.
13. Thakur, Sahil, Linh Le Dinh, Raghavan Lavanya, Ten Cheer Quek, Yong Liu, and Ching-Yu Cheng. "Use of artificial intelligence in forecasting glaucoma progression." *Taiwan Journal of Ophthalmology* 13, no. 2 (2023): 168-183.

Received: August 03, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

14. Iqbal, Muhammad, Arshad Iqbal, Humaira Ayub, Maqbool Khan, Naveed Ahmad, Yasir Javed, and Mohammed Ali Alshara. "GAINSeq: glaucoma pre-symptomatic detection using machine learning models driven by next-generation sequencing data." *Scientific Reports* 15, no. 1 (2025): 23091.
15. Vijayan, Midhula, Deepthi Keshav Prasad, and Venkatakrishnan Srinivasan. "Advancing glaucoma diagnosis: Employing confidence-calibrated label smoothing loss for model calibration." *Ophthalmology Science* 4, no. 6 (2024): 100555.
16. Elangovan, Poonguzhali, and Malaya Kumar Nath. "En-ConvNet: A novel approach for glaucoma detection from color fundus images using ensemble of deep convolutional neural networks." *International Journal of Imaging Systems and Technology* 32, no. 6 (2022): 2034-2048.
17. Khan, Umer Sadiq, and Saif Ur Rehman Khan. "Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT." *Multimedia Tools and Applications* (2024): 1-21.
18. Aljohani, Abeer, and Rua Y. Aburasain. "A hybrid framework for glaucoma detection through federated machine learning and deep learning models." *BMC Medical Informatics and Decision Making* 24, no. 1 (2024): 115.
19. Hasan, Md Mahmudul, Jack Phu, Henrietta Wang, Arcot Sowmya, Michael Kalloniatis, and Erik Meijering. "OCT-based diagnosis of glaucoma and glaucoma stages using explainable machine learning." *Scientific Reports* 15, no. 1 (2025): 3592.
20. Raj, Simran, Nipun Setia, Geet Sahu, Hitesh Kalra, Richa Tiwari, and Kalyan Acharjya. "Predictive Modeling of Eye Cancer Using Deep Learning Techniques: An Analysis." In *2025 International Conference on Automation and Computation (AUTOCOM)*, pp. 810-815. IEEE, 2025.
21. Abdulsalam, Wisal Hashim, Rasha H. Ali, Sawsan H. Jadooa, and Samera Shams Hussein. "Automated glaucoma detection techniques: A literature review." *Engineering, Technology & Applied Science Research* 15, no. 1 (2025): 19891-19897.
22. Reddy, B. Venkata Suresh, and Bhanu Prakash Battula. "Advancing Eye Disease and Glaucoma Diagnosis: A Comprehensive Survey of Deep Learning, Machine Learning, and Hybrid Approaches." In *2025 Fourth International Conference on Smart Technologies, Communication and Robotics (STCR)*, pp. 1-11. IEEE, 2025.
23. Srikanth, B. V., M. Hitha Chowdary, M. Suvishal Sen Kasyap, and K. Anjali Reddy. "Eye Disease Prediction and diagnosis." *Journal of Science, Computing and Engineering Research* 8, no. 3 (2025).
24. Qidwai, Uvais, Umair Qidwai, Thurka Sivapalan, and Gokulan Ratnarajan. "iMIGS: An innovative AI based prediction system for selecting the best patient-specific glaucoma treatment." *MethodsX* 10 (2023): 102209.
25. Martucci, Alessio, Gabriele Gallo Afflitto, Giulio Pocobelli, Francesco Aiello, Raffaele Mancino, and Carlo Nucci. "Lights and Shadows on Artificial Intelligence in Glaucoma: Transforming Screening, Monitoring, and Prognosis." *Journal of Clinical Medicine* 14, no. 7 (2025): 2139.

Received: August 03, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

26. Singh, Law Kumar, and Munish Khanna. "A novel multimodality based dual fusion integrated approach for efficient and early prediction of glaucoma." *Biomedical Signal Processing and Control* 73 (2022): 103468.
27. Fang, Yu, Jianwei Lin, Peiwen Xie, Huishan Zhu, Tsz Kin Ng, and Guihua Zhang. "Ensemble machine learning algorithm for anti-VEGF treatment efficacy prediction in diabetic macular edema." *BMC ophthalmology* 25, no. 1 (2025): 352.
28. Wang, Soohyun, Byoungkug Kim, Jiheon Kang, and Doo-Seop Eom. "Precision diagnosis of Glaucoma with VLLM ensemble deep learning." *Applied Sciences* 14, no. 11 (2024): 4588.
29. Jeyaraj, Thavamani, and S. Oliver Nesa Raj. "Challenges in Adopting AI for Predicting Ophthalmology and Managing Water in the Healthcare Sector." In *Revolutionizing Healthcare Services*, pp. 93-122. CRC Press, 2025.
30. Abbasi, Ashkan, Sowjanya Gowrisankaran, Wei-Chun Li, Xubo Song, Bhavna Josephine Antony, Gadi Wollstein, Joel S. Schuman, and Hiroshi Ishikawa. "A Hybrid Deep Learning based Approach for Visual Field Test Forecasting." *Ophthalmology Science* (2025): 100803.
31. Zarean, Javad, Amirreza Tajally, Reza Tavakkoli-Moghaddam, Seyed Mojtaba Sajadi, and Niaz Wassan. "A framework for robust glaucoma detection: A confidence-aware deep uncertainty quantification approach with a comprehensive assessment for enhanced clinical decision-making." *Engineering Applications of Artificial Intelligence* 139 (2025): 109651.
32. Yang, Ai-Su, Hong-Siang Wang, Te-Jung Li, Chin-Hsin Liu, and Chung-Ming Chen. "Diagnosis of early glaucoma likely combined with high myopia by integrating OCT thickness map and standard automated and Pulsar perimetries." *Scientific Reports* 15, no. 1 (2025): 13614.
33. Lenka, Satyabrata, Zefree Lazarus Mayaluri, and Ganapati Panda. "Glaucoma detection from retinal fundus images using graph convolution based multi-task model." *e-Prime-Advances in Electrical Engineering, Electronics and Energy* 11 (2025): 100931.

Received: August 03, 2025