Automated Glaucoma Screening Using Optic Disc-Cup Segmentation and Hybrid Machine Learning Models.

BioFusion Hackathon 2026
IEEE Engineering in Medicine and Biology Society
University of Sri Jayawardenapura

Team Name: Eye - Clinic
Team Members: R. P. Thilokya Angeesa, K.A. Nadil Hesara Kulathunge, U.D.K. Chamodh Nethsara
Affiliation: University of Moratuwa

Literature Review

Glaucoma is a leading cause of irreversible vision loss worldwide, motivating extensive research into automated detection using retinal fundus imaging. Early work such as Cheng et al. (2011) introduced the ORIGA dataset and relied on optic disc segmentation with handcrafted structural features [1]. While clinically interpretable, these methods suffered from limited generalization due to manual feature engineering.

The REFUGE challenge (2018) established standardized benchmarks for optic disc and cup segmentation and glaucoma classification, with U-Net–based deep learning models achieving significant performance improvements [2]. However, most approaches treated segmentation and classification as isolated tasks, with limited integration of anatomical measurements into diagnostic decisions. More recent datasets such as G1020 emphasized the need for larger and more diverse data sources to improve robustness [3].

Despite progress, many end-to-end models remain black-box systems with limited interpretability, while purely image-level classifiers ignore explicit modeling of optic disc and cup structures. Dataset bias and domain shift across imaging devices also persist. This work addresses these gaps by integrating anatomically guided segmentation with interpretable feature-based and image-based classification in a unified framework aligned with clinical reasoning.

Problem Identification

Glaucoma affects millions worldwide and is often asymptomatic in its early stages, leading to late diagnosis and irreversible vision loss. In many regions, particularly in developing countries, access to specialist ophthalmological screening is limited due to shortages of trained clinicians and the high cost of advanced diagnostic equipment.

Although fundus photography is widely available, reliable glaucoma detection remains challenging, as image interpretation requires expert knowledge and manual measurement of optic disc and cup boundaries is subjective and varies across clinicians.

This project addresses the need for an automated and interpretable glaucoma detection system that provides both quantitative structural measurements and probabilistic risk estimates. Such a system can assist clinicians, reduce diagnostic workload, and support large-scale, cost-effective screening for early glaucoma detection. Therefore, the primary objective of this work is to support early-stage screening rather than definitive clinical diagnosis

Dataset Justification

This study uses a combined dataset from ORIGA, REFUGE, and G1020, providing expert-annotated optic disc and cup segmentations along with glaucoma labels. Images were cropped around the optic disc region to reduce background noise and focus learning on clinically relevant anatomy.

Potential class imbalance and domain variation across datasets were addressed through stratified splitting and imbalance-aware training strategies. Due to differences in acquisition protocols and imaging devices across datasets, residual domain shift may remain despite unified preprocessing.

Methodology

Data processing pipeline

Retinal fundus images and annotations from ORIGA, REFUGE, and G1020 were resized to 512×512 pixels using bilinear interpolation for images and nearest-neighbor interpolation for masks. Strict image–mask pairing was enforced. REFUGE predefined splits were preserved, while ORIGA and G1020 were split using a reproducible 70/15/15 strategy.

Segmentation outputs were anatomically constrained to ensure the optic cup lay within the optic disc. From valid masks, clinically relevant features including vertical cup-to-disc ratio (vCDR), disc and cup areas, and vertical diameters were extracted, with failed segmentations filtered out.

Model Architecture

The proposed system follows a hybrid decision-level fusion architecture (Figure 1). Optic disc and optic cup segmentation is performed using a deep learning-based U-Net with a ResNet-34 encoder pretrained on ImageNet, producing pixel-level anatomical masks.

Structural features extracted from the segmentation outputs are processed by an XGBoost-based feature classifier, trained with stratified cross-validation and probability calibration, to estimate a structure-derived glaucoma probability P(A). In parallel, an EfficientNet-B0 model pretrained on ImageNet directly estimates an image-based glaucoma probability P(B) from fundus images. The final glaucoma diagnosis is obtained through a decision-level fusion strategy implemented using a meta-classifier that learns to combine both probabilistic outputs.

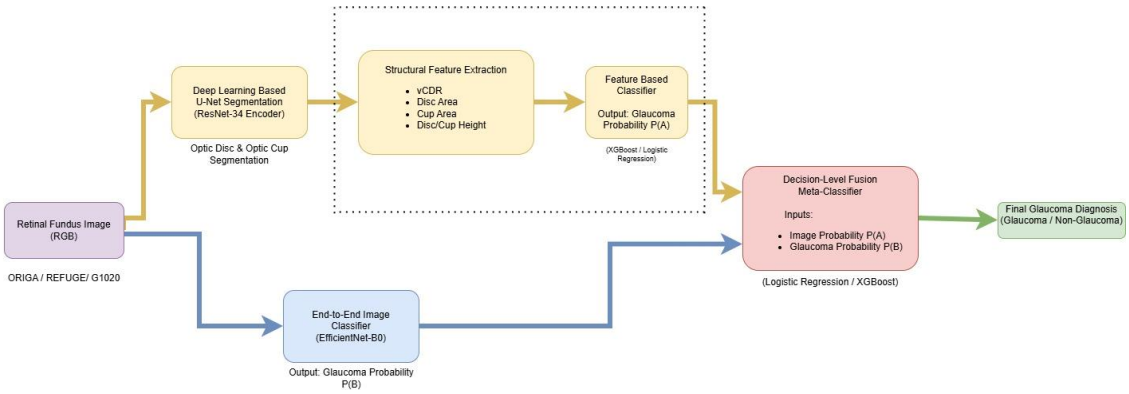

Figure 1: The hybrid decision-level fusion architecture of the system

Training Process

The segmentation network was trained using a combined Binary Cross-Entropy and Dice loss, optimized with AdamW at a learning rate of 1×10^{-3} for up to 20 epochs, selecting the model with the highest validation Dice score. The structure-based classifier was trained on segmentation-derived biomarkers, with class imbalance handled using weighted loss adjustment.

The image-based classifier was trained using Binary Cross Entropy with logits, data augmentation and weighted sampling to address class imbalance. Optimization was performed using AdamW with learning rate scheduling

Validation Strategy

Stratified data splitting was employed throughout to preserve class distribution and prevent data leakage. Segmentation performance was evaluated using the Dice coefficient, while classification performance was assessed using ROC-AUC, sensitivity and specificity.

For the fused classifier, probability calibration and threshold optimization using Youden's J statistic was explored, with the final operating point selected to prioritize sensitivity in accordance with screening objectives. Random seeds were fixed across experiments to ensure reproducibility and consistent evaluation. Given the screening-oriented objective, operating thresholds were selected to favor sensitivity over specificity.

Pretrained Model Usage & Adaptation

Rationale and Training Strategy

Pretrained models were employed to improve generalization and training stability given the limited size and variability of labeled medical imaging datasets. ImageNet-pretrained encoders provide robust low-level representations that transfer effectively to retinal fundus images when fine-tuned, reducing

overfitting and accelerating convergence. All pretrained networks were fine-tuned end-to-end to enable adaptation to domain-specific retinal characteristics.

Modifications

For segmentation, a U-Net architecture with a ResNet-34 encoder pretrained on ImageNet was employed, with the decoder adapted to output two segmentation channels corresponding to the optic disc and optic cup. For image-based classification, the pretrained EfficientNet-B0 model was modified by replacing its original classification head with a single node output layer to support binary glaucoma prediction. No additional layers were introduced.

Risk & Bias Discussion

Domain mismatch between natural images and retinal fundus images may introduce bias, particularly in higher-level features. This risk is mitigated through domain-specific fine-tuning, data augmentation, and the inclusion of a structure-based diagnostic pathway grounded in clinically interpretable features. False positives may occur in cases with physiologically large optic cups or myopic eyes, which share structural characteristics with glaucomatous optic neuropathy.

Results

The proposed glaucoma detection system was evaluated on a held-out test set using ROC-AUC as the primary performance metric. The structure-based classifier, using segmentation-derived features such as the cup-to-disc ratio, achieved a test ROC-AUC of 0.6945, reflecting moderate discriminative capability.

The EfficientNet-B0 classifier demonstrated improved performance of a test ROC-AUC of 0.7123. This depicts that deep image features provide additional discriminatory information beyond explicit anatomical measurements. The hybrid decision level fusion model achieved the best overall performance, with a test ROC-AUC of 0.7298. At the selected operating threshold, the fused model reached a high sensitivity (93.8%) with lower specificity (32.2%) that is consistent with a screening-focused design prioritizing detection of glaucomatous cases.

Overall, the results depicts that combining anatomical features with image-based predictions improves glaucoma detection compared to either approach alone.

Real world applications

The proposed system can be deployed as a decision-support tool alongside digital fundus cameras for automated glaucoma screening, particularly in rural and resource-limited settings. Primary users include ophthalmologists and optometrists, with secondary use by primary healthcare workers in community screening programs.

Fundus images captured during routine examinations are automatically analyzed, with risk scores supporting triage and remote specialist consultation. The system supports screening rather than replacing comprehensive clinical diagnosis and may be affected by image quality and dataset bias.

Marketing and impact strategy

The system can be adopted by hospitals, eye clinics, screening centers, and public health programs as a cost-effective decision-support tool for glaucoma screening. By providing automated risk estimation alongside interpretable structural indicators, it enables early detection, reduces clinician workload, and supports consistent large-scale screening using standard fundus images and open-source software, deployable in both cloud-based and local environments.

Future improvements

Future work will focus on improving robustness and clinical applicability. This includes expanding training with larger, more diverse multi-center datasets to reduce domain bias and improve generalization across imaging devices. Incorporating longitudinal patient data and additional clinical indicators such as intraocular pressure and visual field metrics, could further enhance diagnostic reliability. Model explainability can be strengthened by integrating attention-based visual explanations alongside structural biomarkers Finally, real-world validation through prospective clinical studies and deployment on edge or mobile platforms will be explored to support scalable screening in low-resource settings.

References

[1] J. Cheng et al., “Optic disc segmentation in retinal fundus images using structured learning, IEEE Transactions on Medical Imaging, vol. 30, no. 6, pp. 1226–1239, 2011

[2] Huazhu Fu, Fei Li, José Ignacio Orlando, Hrvoje Bogunović, Xu Sun, Jingan Liao, Yanwu Xu, Shaochong Zhang, Xiulan Zhang, "REFUGE: Retinal Fundus Glaucoma Challenge", IEEE Dataport, July 9, 2019, doi:10.21227/tz6e-r977

[3] M. N. Bajwa, G. A. P. Singh, W. Neumeier, M. I. Malik, A. Dengel, and S. Ahmed, “G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection,” arXiv:2006.09158 [eess.IV], 2020.