Interpretable and Low-Compute Multimodal Glaucoma Screening Using Optic Disc Biomarkers and Clinical Data

Amira SOLTANI
Esprit School Of Business
National Higher Engineering School of Tunis,
LR: LATICE
amira_soltani@yahoo.fr

Imed JABRI
National Higher Engineering School of Tunis,
LR: LATICE
imedjabri@yahoo.com

Abstract— **Early glaucoma screening is challenging because optic nerve head changes may be subtle in fundus images. This paper proposes an interpretable multimodal framework that combines optic disc biomarkers with clinical data for glaucoma detection. Cup-to-disc ratio and optic disc diameter are extracted from retinal images and fused with demographic variables before classification using AdaBoost, Support Vector Machine, and Logistic Regression. On a dataset of 204 fundus images, AdaBoost achieved the best performance, with 0.98 accuracy, 0.97 precision, 0.99 F1-score, and 0.99 AUC. The results show that optic disc biomarkers are the main predictors of glaucoma, while clinical variables provide complementary information.**

Keywords— ***Glaucoma; fundus images; optic disc morphology; cup-to-disc ratio; multimodal classification; interpretability; AdaBoost.***

I. INTRODUCTION

Glaucoma is one of the leading causes of irreversible blindness worldwide, and its early diagnosis remains difficult because the disease often progresses silently in its initial stages. In clinical practice, screening mainly relies on the visual assessment of structural changes in the optic nerve head from retinal fundus images, particularly the cup-to-disc ratio (CDR), which is a key indicator of glaucomatous damage [1,2, 3,4]. However, these structural changes may be subtle, making diagnosis subjective, operator-dependent, and time-consuming.

To address these limitations, many computer-aided glaucoma detection systems have been proposed. Early approaches mainly relied on optic disc and cup segmentation and on handcrafted structural features extracted from fundus images [5,6, 7, 8, 9]. Later studies introduced machine learning models based on image-derived features or clinical records, while more recent work has increasingly focused on deep learning methods [10, 11, 12, 13]. Although deep learning has shown strong performance, its clinical adoption remains limited by the need for large annotated datasets and by the lack of interpretability of many black-box models.

Another limitation of many existing studies is their strong dependence on image information alone, even though glaucoma risk is also influenced by demographic and clinical factors such as age [3], [4]. This highlights the need for screening approaches that are not only accurate, but also transparent, computationally efficient, and suitable for relatively small clinical datasets.

In this context, this work proposes an interpretable multimodal framework for glaucoma screening that combines structural optic disc biomarkers with clinical variables. The method extracts meaningful indicators such as cup-to-disc ratio and optic disc diameter from retinal images and fuses them with demographic information including age, sex, and origin. These multimodal features are then used within an ensemble learning framework for glaucoma classification, with the objective of preserving both predictive performance and clinical interpretability [10, 13].

The remainder of this paper is organized as follows. Section II briefly reviews previous work on glaucoma detection and machine learning-based screening methods. Section III presents the proposed methodology, from image preprocessing and biomarker extraction to multimodal feature integration and classification. Section IV reports the experimental results and discusses their clinical relevance. Section V concludes the paper and outlines possible directions for future research.

II. RELATED WORK

Automated glaucoma screening has long relied on the analysis of retinal fundus images, particularly through the structural assessment of the optic nerve head. Early computer-aided approaches focused on explicit anatomical measurements, such as optic disc and optic cup boundaries, from which clinically meaningful indicators like the cup-to-disc ratio were derived. These methods were attractive because they were transparent and directly aligned with ophthalmic reasoning, even though their performance depended strongly on the quality of segmentation and handcrafted feature extraction [6,7, 8, 9].

As the field evolved, several studies extended image-based analysis by incorporating machine learning models trained on structural descriptors and patient-related variables. In this direction, Kovalyk-Borodyak *et al.* proposed a multimodal framework using binocular fundus information and clinical data from the PAPILA dataset, showing that the integration of both-eye information and clinical context can improve automated glaucoma diagnosis. Their study also highlighted the role of explainability tools such as SHAP in understanding model behavior, which is particularly relevant in medical applications where transparency matters as much as predictive performance [14].

More recently, deep learning has become the dominant paradigm in glaucoma detection. A 2025 systematic review and meta-analysis by Ling *et al.* synthesized 48 studies and reported strong overall performance for deep learning systems using fundus photography and OCT. For fundus-based diagnosis, the pooled sensitivity and specificity reached 0.92 and 0.93, respectively, with an AUROC of 0.90. However, the same review also emphasized that internal validation results were generally better than external validation results, pointing to persistent concerns regarding real-world generalization and clinical transferability [15].

This observation is consistent with broader recent reviews. Beegam, Kalra, and Zahoor noted in 2025 that despite the impressive progress of deep learning for fundus-based

glaucoma detection, major challenges remain, including limited generalizability, insufficient explainability, data variability, and ethical concerns related to trustworthy AI deployment [16]. Similarly, Ashtari-Majlan, Dehshibi, and Masip provided a 2024 survey of deep-learning-based glaucoma diagnosis across fundus, OCT, and visual field imaging, showing that the literature is moving toward more sophisticated architectures while still facing unresolved issues related to dataset diversity, reproducibility, and clinical translation [17].

A particularly notable development in recent work is the shift toward lightweight yet clinically oriented AI systems. Sharma *et al.* introduced the AI-GS framework in *npj Digital Medicine*, combining multiple lightweight deep learning models for fundus-image-based glaucoma screening. Rather than relying on a single black-box classifier, their system integrates segmentation, structural sign detection, and classification, while explicitly combining deep learning predictions with conventional glaucomatous optic-disc feature analysis. This trend is important because it shows that even within modern deep learning pipelines, structural reasoning remains clinically valuable [18].

Another major direction is the rise of foundation and self-supervised models. Chuter *et al.* evaluated RET Found, a foundation artificial intelligence model, for glaucoma detection from color fundus photographs using a diverse clinical dataset. Their study assessed performance across race, age, glaucoma severity, training cycles, and sample size, illustrating how current research increasingly examines not only overall accuracy, but also subgroup behavior and training-data efficiency [19].

In parallel, multimodal learning is gaining increasing importance in glaucoma research. Recent studies no longer limit themselves to image-only classification, but attempt to combine fundus photographs with OCT, visual fields, clinical text, or demographic variables. Huang *et al.* proposed in 2025 an interpretable longitudinal multimodal deep learning system that estimates present and future visual fields from color fundus photographs and clinical narratives, demonstrating how multimodal information can support more clinically relevant glaucoma assessment beyond binary screening alone [20]. Likewise, Sivakumar and Penkova reported a multimodal framework integrating retinal images and clinical biomarkers, further confirming that combined representations can improve diagnostic performance and better reflect real clinical decision-making [21].

Beyond multimodality, recent literature also emphasizes fairness, robustness, and trustworthiness. Shi et al. proposed fair identity normalization to reduce subgroup disparities in glaucoma screening, explicitly addressing demographic bias during model training [22].

In light of recent research, lightweight and interpretable multimodal approaches remain highly relevant. Although deep learning and foundation models dominate the field, they are not always the best option for small clinical datasets or resource-limited settings. There is still a clear need for transparent frameworks that combine explicit structural biomarkers with clinical data. Our work follows this direction by relying on clinically interpretable optic-disc measurements and complementary demographic variables for glaucoma screening on a relatively small real-world dataset.

III. Materials and Methods

A. Dataset

This study was conducted on a clinical dataset collected at the Ophthalmology Hospital Hedi Raies in Tunis. The dataset consists of 204 retinal fundus images, of which 113 (55.4%) were labeled as glaucomatous and 91 (44.6%) as normal. All images were acquired in JPEG format with an original resolution of 3008 × 2000 pixels and were manually labeled by ophthalmologists based on clinical examination of the optic nerve head and cup-to-disc ratio assessment. Representative examples of normal and glaucomatous fundus images are shown in Fig. 1. The dataset was selected to reflect real screening conditions and to support the evaluation of a lightweight multimodal glaucoma detection framework on a relatively small clinical sample.

B. Image Preprocessing

To ensure consistent feature extraction, all retinal images underwent a series of preprocessing steps. First, each image was resized to 752 × 500 pixels in order to reduce computational cost and standardize the input dimensions. The images were then converted from RGB to grayscale to simplify subsequent structural analysis.

Noise reduction was performed using a sequence of bilateral, Gaussian, and median filters. This combination was adopted to smooth irrelevant variations while preserving the main anatomical contours of the optic nerve head. After denoising, edge enhancement was applied to facilitate the detection of the optic disc and optic cup boundaries. Since the goal of the study was not to propose a novel segmentation algorithm but to extract clinically meaningful structural biomarkers, preprocessing was designed as a practical and computationally efficient stage supporting reliable feature computation.

C. Structural Biomarker Extraction

The proposed framework relies on explicit structural biomarkers derived from the optic nerve head. After preprocessing, contour-based analysis was used to localize the optic disc and optic cup boundaries. These structures were then approximated to obtain geometric measurements suitable for glaucoma screening.

Two biomarkers were retained for classification:

- Optic Disc Diameter (ODD): a descriptor of the overall size of the optic disc.
- Cup-to-Disc Ratio (CDR): a clinically established biomarker reflecting the relative size of the optic cup with respect to the optic disc.

The CDR was computed as:

$$CDR = \frac{D_{Cup}}{D_{Disc}} \qquad (1)$$

where D_{Cup} and D_{Disc} denote the estimated diameters of the optic cup and optic disc, respectively.

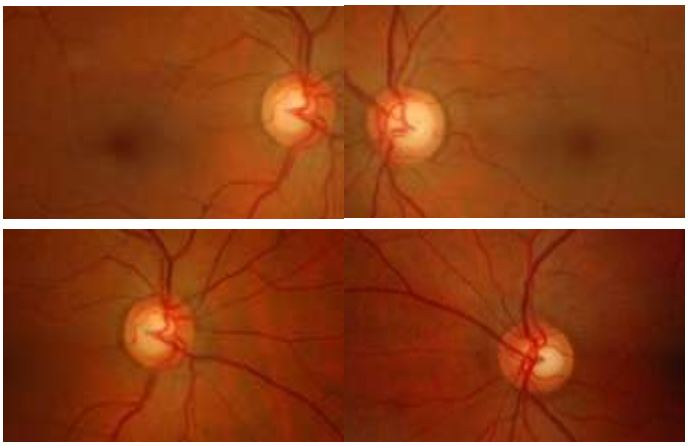

Fig. 1. Representative retinal fundus images.

Among the extracted variables, CDR is particularly important because it reflects structural deformation of the optic nerve head and is widely used in clinical glaucoma assessment. Optic disc diameter was included as an additional morphological descriptor in order to capture complementary structural information. By relying on explicit anatomical measurements rather than latent image embeddings, the proposed method preserves interpretability and maintains a direct connection with ophthalmic reasoning.

The successive stages of the proposed image-processing pipeline are shown in Fig. 2, from the original retinal image to the final localization of the optic disc and optic cup.

D. Clinical Feature Integration

In addition to image-derived biomarkers, demographic and clinical metadata were incorporated into the classification framework. The selected variables were age, sex, and origin. These features were combined with the extracted structural biomarkers to form a unified multimodal representation of each patient.

The rationale behind this integration is that glaucoma diagnosis is not based exclusively on image appearance. Clinical and demographic characteristics may provide complementary information and help improve screening reliability, particularly in borderline cases. The final input vector for each sample therefore consisted of both anatomical and patient-related descriptors, allowing the model to exploit multimodal information while remaining computationally lightweight and clinically interpretable.

E. Classification Models

The multimodal feature vectors were used to train and evaluate three classification models: AdaBoost, Support Vector Machine (SVM), and Logistic Regression. These models were selected to compare different levels of model complexity while preserving interpretability and suitability for small datasets.

AdaBoost was used as the main ensemble classifier because of its ability to combine weak learners into a stronger predictive model through iterative reweighting of misclassified samples. In this study, shallow decision trees were used as base estimators in order to limit overfitting and preserve model simplicity. SVM and Logistic Regression were included as baseline models for comparison, since both are commonly used in medical classification tasks and offer well-established benchmarks for tabular multimodal data.

F. Training Strategy and Hyperparameter Optimization

The dataset was divided into 70% for training and 30% for testing using a stratified random split in order to preserve the class distribution across the two subsets. In practice, the split was performed with shuffle=True, stratify=y, and random_state=42 to ensure reproducibility and balanced partitioning.

Hyperparameter optimization was carried out on the training set using GridSearchCV with five-fold cross-validation. For Logistic Regression, the regularization parameter C was tuned over the set $\{0.01, 0.1, 1, 10\}$. For SVM, different kernels, including linear, radial basis function, and polynomial kernels, were evaluated together with the regularization parameter C. For AdaBoost, the optimization considered decision tree base learners with maximum depths of 1, 2, and 3, along with different numbers of estimators $\{50, 100, 200\}$ and learning rates $\{0.01, 0.1, 1, 10\}$. The best configuration for each model was selected according to cross-validation accuracy.

G. Performance Evaluation

The trained models were evaluated on the independent test set using four standard classification metrics: accuracy, precision, F1-score, and area under the receiver operating characteristic curve (AUC). These metrics were selected to provide a balanced view of classification performance and discriminative ability.

In addition to the quantitative comparison between classifiers, feature importance analysis was used to assess the relative contribution of each variable in the final AdaBoost model. This analysis was intended to determine whether glaucoma prediction was primarily driven by structural optic disc biomarkers or by the added clinical variables. Such interpretation is particularly important in the context of clinical decision support, where understanding the reasoning behind predictions is often as important as achieving high classification performance.

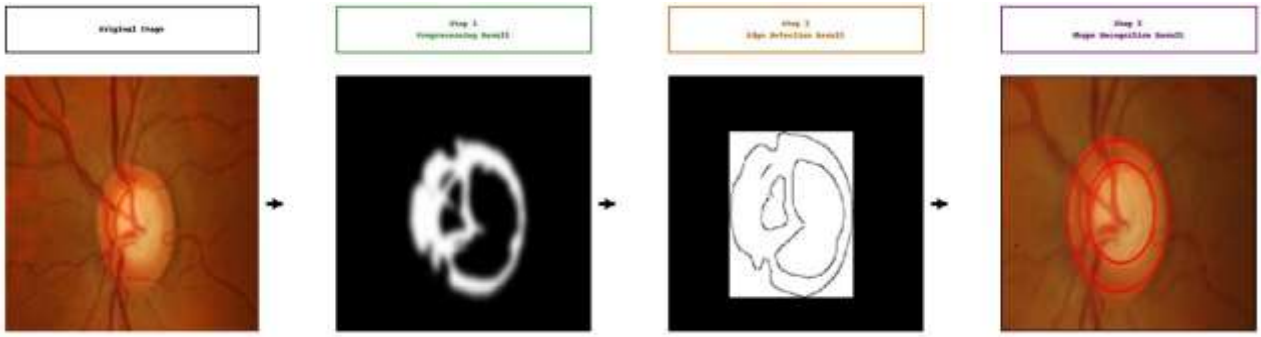

Fig. 2. Main stages of the proposed image-processing pipeline for structural biomarker extraction.

IV. Results and Discussion

A. Comparative Classification Performance

To evaluate the effectiveness of the proposed multimodal framework, three classification models were trained and tested on the extracted structural and clinical features: AdaBoost, Support Vector Machine (SVM), and Logistic Regression. Performance was assessed using accuracy, precision, F1-score, and area under the ROC curve (AUC).

As shown in Table I, AdaBoost achieved the best overall performance, with an accuracy of 0.98, a precision of 0.97, an F1-score of 0.99, and an AUC of 0.99. SVM also produced strong results, reaching 0.95 accuracy and 0.99 AUC, whereas Logistic Regression achieved slightly lower but still competitive performance, with 0.93 accuracy and 0.94 AUC.

These findings indicate that the proposed multimodal feature space is highly informative for glaucoma screening and that ensemble learning is particularly effective in exploiting the complementary contribution of structural and clinical variables.

B. ROC Analysis

The ROC curves shown in fig.3 further confirm the effectiveness of the evaluated models. AdaBoost and SVM both achieved near-perfect discrimination, with AUC values close to 1.0, indicating a strong ability to separate glaucomatous from normal cases. Logistic Regression also showed acceptable discriminative performance, although its ROC curve remained slightly below those of the other two models. Overall, these results suggest that the selected multimodal features provide a stable representation of glaucoma-related patterns and support reliable classification.

While the high AUC values obtained by both AdaBoost and SVM are encouraging, AdaBoost remains more attractive in the context of this study because it also offers a clearer feature-based interpretation of the final decision process. This is an important advantage for clinical decision support, where a transparent relationship between the input variables and the prediction is often preferable to purely black-box performance.

C. Contribution of Structural and Clinical Features

To better understand the behavior of the final model, feature importance analysis was performed on the AdaBoost classifier. The results show that the most influential predictors were the two structural biomarkers extracted from fundus images: cup-to-disc ratio and optic disc diameter. Age contributed moderately to the prediction, whereas sex and origin had a comparatively lower impact. These findings are consistent with clinical knowledge, since glaucoma diagnosis is primarily associated with structural deformation of the optic nerve head, while demographic variables usually play a secondary but complementary role.

This result supports one of the main assumptions of the proposed framework: image-derived optic disc morphology should remain at the center of automated glaucoma screening, but its predictive value can be reinforced by integrating patient-level clinical information. In this sense, the multimodal design does not replace anatomical reasoning but rather extends it with contextual information that may improve robustness in borderline or heterogeneous cases.

TABLE I. Comparative table between different classification algorithms

Metrics	***AdaBoost***	***SVM***	***Logistic Regression***
Accuracy	0.98	0.95	0.93
Precision	0.97	0.95	0.89
F1-Score	0.99	0.96	0.94
AUC	0.99	0.99	0.94

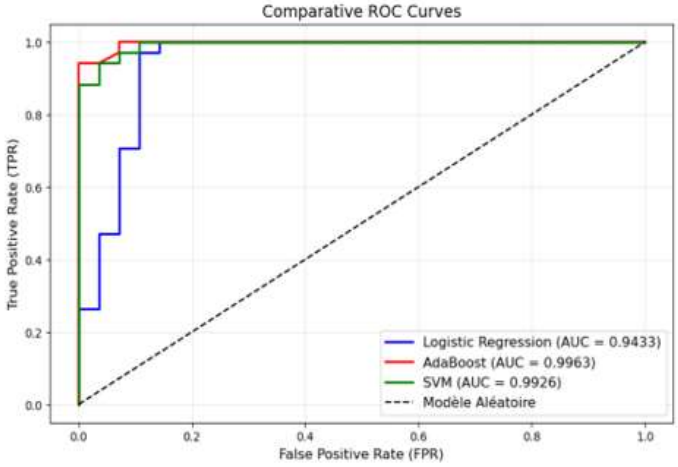

Fig. 3. Comparative ROC Curves

D. Clinical Relevance and Interpretability

One of the main strengths of the proposed framework is its interpretability. Unlike end-to-end black-box models, the present approach is built around explicit variables that are already meaningful in ophthalmic practice, especially cup-to-disc ratio and optic disc diameter. This makes the final decision process easier to interpret and potentially easier to trust in clinical screening settings. In addition, the integration of age, sex, and origin preserves a simple and understandable multimodal representation without The framework is also computationally lightweight. Since it relies on handcrafted structural biomarkers and conventional machine learning models, it does not require large-scale annotated datasets or expensive training procedures. This makes it particularly suitable for small clinical datasets and for healthcare environments where computational resources may be limited. In such contexts, an interpretable and efficient screening tool may be more practical than a heavier deep learning pipeline, even if the latter is potentially more flexible on very large datasets.

substantially increasing model complexity.

As shown in Fig. 4, optic disc morphology, particularly optic disc diameter and cup-to-disc ratio, had the strongest influence on glaucoma prediction, whereas clinical variables provided complementary information.

E. Limitations

Despite the promising results, several limitations should be acknowledged. First, the study was conducted on a relatively small dataset collected from a single clinical center, which may limit the generalizability of the findings. Second, the comparison focused on conventional machine learning models and did not include recent deep learning or foundation-model-based baselines. Although this choice is consistent with the low-compute and interpretable objective of the study, broader comparative experiments would strengthen the external relevance of the results. Third, the current framework depends on the quality of structural biomarker extraction, meaning that segmentation errors may affect downstream classification performance.

Given the high classification scores obtained on a relatively small dataset, special attention was paid to the risk of overfitting. To mitigate this issue, hyperparameter tuning was performed using five-fold cross-validation on the training set, and shallow decision trees were used as base learners in the AdaBoost model to limit complexity. In addition, the final evaluation was conducted on an independent test set obtained through a stratified split. Although the consistency of the results suggests encouraging generalization, the dataset size remains limited, and further validation on larger and multi-center datasets is required to confirm the robustness of the proposed framework.

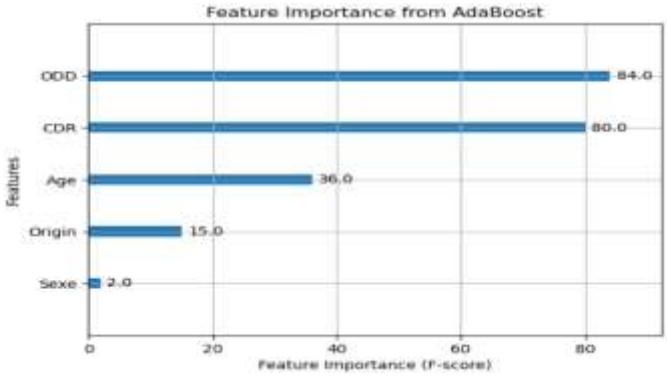

Fig. 4. Feature importance analysis

V. CONCLUSION

This paper presented an interpretable and low-compute multimodal framework for glaucoma screening based on retinal fundus images and clinical data. Structural biomarkers extracted from the optic nerve head, namely cup-to-disc ratio and optic disc diameter, were combined with demographic variables and used for classification.

Among the evaluated models, AdaBoost achieved the best performance, reaching 0.98 accuracy, 0.97 precision, 0.99 F1-score, and 0.99 AUC. Feature importance analysis showed that optic disc morphology was the main predictor of glaucoma, while clinical variables provided complementary information.

Overall, the results indicate that combining explicit anatomical biomarkers with clinical data can provide accurate and interpretable glaucoma screening on small clinical datasets. Future work will focus on validation on larger multi-center datasets and comparison with recent deep learning approaches.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. Leila Jabri and Dr. Ines Malek from Hedi Raies Hospital for providing the clinical dataset and for their valuable support through clinical information and validation of the results.

REFERENCES

[1] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British Journal of Ophthalmology*, vol. 90, no. 3, pp. 262–267, 2006.

[2] F. Topouzis, A. L. Coleman, and A. I. Harris, "Factors associated with undiagnosed open-angle glaucoma: the Thessaloniki Eye Study," *American Journal of Ophthalmology*, vol. 145, pp. 327–335, 2008.

[3] C. W. McMonnies, "Glaucoma history and risk factors," *Journal of Optometry*, vol. 10, pp. 71–78, 2016.

[4] L. Mowatt, "Risk Factors for Rapid Glaucoma Disease Progression," *American Journal of Ophthalmology*, vol. 86, pp. 170–171, 2017.

[5] L. J. Coan, B. M. Williams, V. K. Adithya, S. Upadhyaya, A. Alkafri, S. Czanner, R. Venkatesh, C. E. Willoughby, S. Kavitha, and G. Czanner, "Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review," *Surv. Ophthalmol.*, vol. 68, no. 1, pp. 17–41, 2023, doi: 10.1016/j.survophthal.2022.08.005.

[6] P. C. T. Kuan, "Optic Disk Feature Extraction Via Modified Deformable Model Technique for Glaucoma Analysis," *Pattern Recognition*, vol. 40, pp. 2063–2076, 2007.

[7] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, and H. Fujita, "Determination of Cup and Disc Ratio of Optical Nerve Head for Diagnosis of Glaucoma on Stereo Retinal Fundus Image Pairs," *Proc. SPIE*, vol. 7260, pp. 1–8, 2009.

[8] Y. Hagiwara, J. E. W. Koh, J. Tan, S. V. Bhandary, A. Laude, E. J. Ciaccio, T. Louis, and U. R. Acharya, "Computer-aided diagnosis of glaucoma using fundus images: A review," *Comput. Methods Programs Biomed.*, vol. 165, pp. 1–12, 2018, doi: 10.1016/j.cmpb.2018.07.012.

[9] P. C. T. Kuan, G. H. Ong, M. K. Fadzil, M. A. M. Noor, H. A. Nugroho, and T. W. Ng, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognit.*, vol. 40, no. 7, pp. 2063–2076, 2007.

[10] I. Ashish, M. Partha Sarathi, and M. Kishore Dutta, "An adaptive threshold based image processing technique for improved glaucoma detection and classification," *Computer Methods and Programs in Biomedicine*, vol. 122, no. 2, pp. 229–244, 2015.

[11] F. Târcoveanu, F. Leon, C. Lisa, S. Curteanu, A. Feraru, K. Ali, and N. Anton, "The Use of Artificial Neural Networks in Studying the Progression of Glaucoma," *Scientific Reports*, published Aug. 23, 2024.

[12] H. Xiong, F. Long, M. S. Alam, and J. Sang, "Multi-GlaucNet: A Multi-Task Model for Optic Disc Segmentation, Blood Vessel Segmentation and Glaucoma Detection," *Biomedical Signal Processing and Control*, vol. 99, Art. no. 106850, Jan. 2025.

[13] V. N. Nirgude *et al*., "Glaucoma Detection Using Machine Learning and Deep Learning Approach," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 8, no. 6, Jun. 2024.

[14] O. Kovalyk-Borodyak, J. Morales-Sánchez, R. Verdú-Monedero, and J.-L. Sancho-Gómez, "Glaucoma detection: Binocular approach and clinical data in machine learning," *Artif. Intell. Med.*, vol. 159, Art. no. 103050, 2025, doi: 10.1016/j.artmed.2024.103050.

[15] X. C. Ling, H. S.-L. Chen, P.-H. Yeh, Y.-C. Cheng, C.-Y. Huang, S.-C. Shen, and Y.-S. Lee, "Deep learning in glaucoma detection and progression prediction: A systematic review and meta-analysis," *Biomedicines*, vol. 13, no. 2, Art. no. 420, 2025, doi: 10.3390/biomedicines13020420.

[16] S. Beegam M. K., M. Kalra, and A. Zahoor, "Advancements in deep learning for glaucoma detection from fundus images: A comprehensive analysis," *J. Transform. Technol. Sustain. Dev.*, 2025, doi: 10.1007/s41314-025-00073-6.

[17] M. Ashtari-Majlan, M. M. Dehshibi, and D. Masip, "Glaucoma diagnosis in the era of deep learning: A survey," *Expert Syst. Appl.*, vol. 256, Art. no. 124888, 2024, doi: 10.1016/j.eswa.2024.124888.

[18] P. Sharma, N. Takahashi, T. Ninomiya, M. Sato, T. Miya, S. Tsuda, and T. Nakazawa, "A hybrid multi model artificial intelligence approach for glaucoma screening using fundus images," *npj Digit. Med.*, vol. 8, no. 1, 2025, doi: 10.1038/s41746-025-01473-w.

[19] B. Chuter, J. Huynh, S. Hallaj, E. Walker, J. M. Liebmann, M. A. Fazio, C. A. Girkin, R. N. Weinreb, M. Christopher, and L. M. Zangwill, "Evaluating a foundation artificial intelligence model for glaucoma detection using color fundus photographs," *Ophthalmol. Sci.*, vol. 5, no. 1, Art. no. 100623, 2025, doi: 10.1016/j.xops.2024.100623.

[20] X. Huang, X. Kong, Y. Yan, Z. Gao, Z. Li, C. Zhang, K. Jin, and J. Ye, "Interpretable longitudinal glaucoma visual field estimation deep learning system from fundus images and clinical narratives," *npj Digit. Med.*, vol. 8, no. 1, 2025, doi: 10.1038/s41746-025-01750-8.

[21] R. Sivakumar and A. Penkova, "Enhancing glaucoma detection through multi-modal integration of retinal images and clinical biomarkers," *Eng. Appl. Artif. Intell.*, vol. 141, Art. no. 110010, 2025, doi: 10.1016/j.engappai.2025.110010.

[22] M. Shi, Y. Luo, Y. Tian, L. Q. Shen, N. Zebardast, M. Eslami, S. Kazeminasab, M. V. Boland, D. S. Friedman, L. R. Pasquale, and M. Wang, "Equitable artificial intelligence for glaucoma screening with fair identity normalization," *npj Digit. Med.*, vol. 8, Art. no. 46, 2025, doi: 10.1038/s41746-025-01432-5.