Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GlaucoFusionNet: A Causal Inference Aware Multimodal Explainable AI Framework with Counterfactual Reasoning for Glaucoma Risk Prediction

A. Anushya
University of Ha'il
Harish Kumar Pamnani
Poornima University
Smaranika Mohapatra
smaranika.mohapatra@jaipur.manipal.edu

Manipal University Jaipur

Research Article

Keywords: Casual Inference, Counterfactual Reasoning, Explainable AI, GlaucoFusionNet, Multimodal Deep Learning

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10454812/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GlaucoFusionNet: A Causal Inference Aware Multimodal Explainable AI Framework with Counterfactual Reasoning for Glaucoma Risk Prediction

A.Anushya[1], Harish Kumar Pamnani[2], Smaranika Mohapatra[*3]

1. Department of Artificial Intelligence and Data Science,College of Computer Science and Engineering. University of Ha'il, Ha'il, Saudi Arabia.
   Email- anushya.arol@uoh.edu.sa
   https://orcid.org/0000-0002-5712-4931
2. Faculty of Computer Science and Engineering, Poornima University, Jaipur, Rajasthan, India.
   Email- harish.pamnani@poornima.edu.in
   https://orcid.org/0000-0002-8215-7129
3. Department of IT, Manipal University Jaipur, Jaipur, Rajasthan, India.
   Email- smaranika.mohapatra@jaipur.manipal.edu
   https://orcid.org/0000-0002-5243-4965

Corresponding author: Smaranika Mohapatra
Email: smaranika.mohapatra@jaipur.manipal.edu

ABSTRACT

Glaucoma is one of a top global cause of irreversible vision loss. But timely diagnosis and thorough risk stratification are the keys to interrupting the progression of blindness. But they are correlation-driven, lack causal reasoning, and provide limited interpretability for clinical decision-making. Furthermore, conventional multimodal approaches are combined retinal imaging and clinical data for diagnosis purpose. But they are not addressing the confounding effects and provide actionable explanations for disease progression. To overcome limitations, study proposes GlaucoFusionNet based on a multimodal explainable AI framework with causal inference for automated glaucoma risk prediction and progression analysis. GlaucoFusionNet is designed for images and tabular data. It integrates retinal fundus images and structured clinical data through a feature fusion architecture. ResNet-based CNN extracts deep retinal image representations. An MLP network model is implemented to accumulate the nonlinear relationships among clinical variables. The extracted embeddings are fused to construct multimodal features for glaucoma classification. SHAP provides feature-level explanation and Grad-CAM visualizes prediction of model on images. Directed Acyclic Graphs used to identify meaningful relationships between clinical factors with reduced spurious correlations. Beyond classification, the framework provides a risk stratification score for identifying high risk patients for disease progression.

Keywords: Casual Inference, Counterfactual Reasoning, Explainable AI, GlaucoFusionNet, Multimodal Deep Learning

BACKGROUND AND MOTIVATION

Glaucoma is the silent thief of sight. It is wworsening nerve impairment and lead to permanent vision loss without early intervention. Conventional glaucoma detection requires extensive manual skills in visual field analysis. It is time-consuming and subjective process. AI-driven retinal imaging has significantly optimized detection rates in digital fundus image analysis. To improve the diagnostic performance and robustness, multimodal deep learning approaches integrated retinal images with clinical data. Furthermore, Explainable Artificial Intelligence (XAI) techniques have become increasingly important for improving model transparency and clinician trust by visually. It is explaining model decisions and highlighting important retinal regions. Current research works are reviewed in the following criteria.

Deep Learning Approaches for glaucoma detection

Recent studies have demonstrated the effectiveness of deep learning techniques for utomatic diagnosis of glaucoma from fundus images and the ODIR-5K dataset. Alnuaimi and Albastaki improved glaucoma classification performance by combining Generative Adversarial Networks (GANs) and Region of Interest (ROI) extraction with the InceptionV3 model. It achieved notable testing accuracy [1]. Pathak utilized Convolutional Neural Networks (CNNs) for extracting discriminative retinal features and compared with traditional machine learning methods [2]. Rodríguez et al. hybridized CNN and attention-based to achieve highest accuracies like 95% for capturing the retinal abnormalities [3]. Shoukat et al. demonstrated that ResNet-50 achieved high accuracy, sensitivity, and specificity for detecting early-stage glaucoma [4]. Xu et al. mimicked expert ophthalmic diagnosis for analysing optic disc and retinal nerve fiber layer [5]. In resource-limited healthcare settings, Patil et al worked with EfficientNet-B3 and MobileNet on cropped optic disc images [6]. In [7], Aguilar et al. developed glaucoma classification model using CNN. To distinguish more accurately glaucoma from physiologic large optic cups, Xu and Chai combined the clinical structured features with deep image representations [8].

Explainable AI for Glaucoma Detection and Clinical Trust

To improve the transparency and clinical trust of models for exposing glaucoma, XAI has risen to prominence. Empirical results indicate that deep architectures, namely DenseNet, ResNet-50, EfficientNet, and ensemble models achieve high diagnostic accuracy while incorporating interpretability techniques including Grad-CAM, LIME, and SHAP [9,10, 11]. These Attention mechanisms empower ophthalmologists to map relevant pathological biomarkers. Also, it is **interpreting model logic** behind **algorithmic outputs**. Several research integrated explainability in glaucoma diagnosis systems. to enhance both accuracy and interpretability, [12, 13] extended explainable deep learning frameworks using attention mechanisms. Similarly, Chayan et al. and Shah et al. accentuated that XAI reduces the black-box limitations of models and increases clinician confidence in AI-assisted diagnosis [14,15].

Multimodal Deep Learning Approaches in glaucoma Detection Using Imaging and Clinical Data

To enhance diagnostic accuracy in glaucoma detection, multimodal deep learning approaches adopt diverse data sources including imaging and clinical information. Liu et al. combined retinal fundus images with clinical texts. It achieved an AUC of 87.25% through attention-guided fusion and contrastive learning [16]. Also, Kovalyk et al. explored a binocular approach by integrating data from both eyes alongside clinical information. It improved model performance for yielding an AUC of 0.796 [17]. Huang et al. demonstrated that probabilistic deep learning models for utilizing both fundus photographs and visual fields. It achieved an impressive AUC of 0.94 for highlighting the benefits of combining modalities [18]. FHODLB-RDGC technique showcased effective feature fusion from OCT and fundus images for enhancing glaucoma classification accuracy by Krishna, Nanditha, and K. Nagamani [19]. The Dual-Attention Semantic Parallel-LSTM Network (DA-SPL) combined the fundus imaging with visual inputs. It utilized a dual-attention mechanism to refine cross-modal features and generate precise diagnostic reports in clinical terminology [20]. Ferreira et al., refurbished fundus images and Optical Coherence Tomography (OCT) volumes and achieved a kappa score of 0.886. It indicated the effectiveness in detecting glaucoma at early stages. It identified in severe cases [21]. These studies underscore the potential of multimodal approaches in improving glaucoma detection and management collectively.

Causal and Counterfactual Learning in Multimodal Medical Imaging including Glaucoma

Recent studies are applied the causal inference and counterfactual learning for glaucoma case study. They are enhancing both mechanistic understanding and clinical decision-making. Mendelian randomization (MR) has been widely utilized for establishing causal links between specific traits and glaucoma subtypes [22]. It is identifying 11 causal factors for glaucoma, including intraocular pressure and various metabolic markers. Further MR analyses have demonstrated a bidirectional causal relationship between glaucoma and diabetic retinopathy [23]. Also, it explored the potential in other case studies like albeit complex, influence of the gut microbiota [24] and immune cell phenotypes [25] on disease progression. As well as causal genomics prioritized drug repurposing candidates on metformin for normal-tension glaucoma integrated with generative AI has successfully [26]. In the realm of diagnostic imaging and clinical simulation, LightHCG developed to reduce spurious correlations in detection models [27]. Furthermore, counterfactual simulations and intervention analyses have been proposed to evaluate the validity of surrogate endpoints through information-theoretic metrics in clinical trials [28, 29].

The above studies achieved promising results for solving the challenges such as data imbalance, image quality variations, and accurate diagnosis. But still several critical limitations are unresolved like limited interpretability, lack of causal reasoning with counterfactual analysis for personalized glaucoma risk prediction. These restrictions act as a spur to accelerate progress for durability, enduring complex data and combining different data sources in trustworthy frameworks in for

precise glaucoma screening. This study aims to develop a causality-aware multimodal explainable AI framework for accurate, interpretable, and clinically trustworthy glaucoma risk prediction. GlaucoFusionNet is established in this study as a multimodal explainable AI framework for glaucoma risk prediction. GlaucoFusionNet integrates retinal fundus imaging and clinical data using CNN and MLP feature fusion architecture. GlaucoFusionNet framework incorporates SHAP, Grad-CAM, causal inference, and counterfactual reasoning to provide interpretable, clinically meaningful, and personalized glaucoma diagnosis correspondingly. The framework is intended to improve predictive performance. Also, it enables causal understanding and "what-if" clinical scenario analysis. Therefore, in AI-assisted ophthalmic diagnosis, it applies for transparency, robustness, and clinician trust. The following objectives are established to implement a multimodal explainable AI framework for casual inference and counterfactual reasoning.

1. To exploit a multimodal feature fusion framework with retinal fundus images and structured clinical features using a ResNet-based CNN–MLP for AI-driven glaucoma detection.

2. To encompass Grad-CAM and SHAP for visual and feature-level interpretation of glaucoma predictions.

3. To produce causal inference for identifying clinically meaningful relationships between patient variables and glaucoma outcomes and simulate counterfactual reasoning for "what-if" clinical scenarios in personalized glaucoma risk assessment and progression analysis.

To fulfill the aforementioned goals, GlaucoFusionNet is proposed for automated glaucoma diagnosis and risk prediction through explainability, casual graph and counterfactual reasoning techniques. The detailed methodology is explained in the following section.

METHODOLOGY

This study follows a structured methodology pipeline. It is beginning with data acquisition and ending with counterfactual reasoning, as illustrated in Figure 1. The GlaucoFusionNet framework synthesizes visual and structured data to automatically detect glaucoma. A ResNet-based CNN extracts spatial featuresfrom retinal images, and a Multilayer Perceptron Identifies features from patient metadata independently. The system integrates imaging and demographic data to derive hybrid feature vectors for enhancing glaucoma identification. To augment model transparency, Grad-CAM and SHAP are employed in tandem to offer heatmaps and rank feature significance. Complementing this, the framework adopts graphical causal diagrams and hypothetical "what-if" analysis to map dependencies and model counterfactual outcomes for tailored glaucoma risk forecasting.

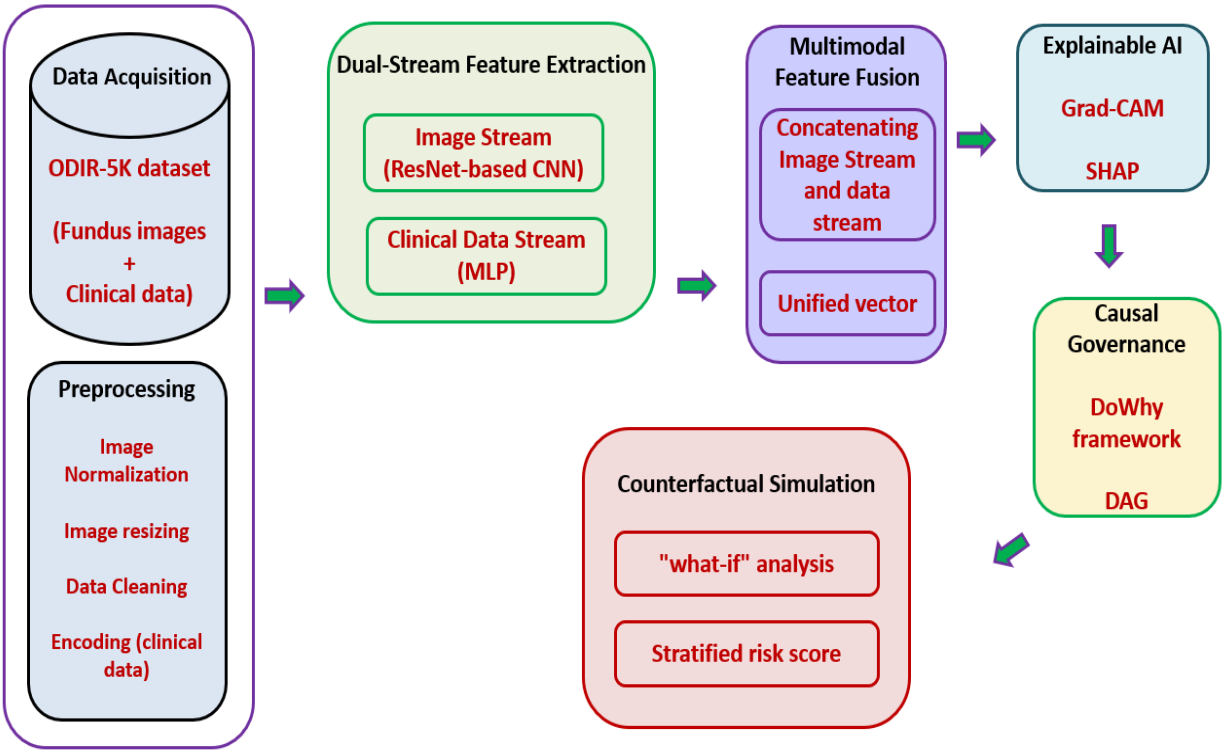


Figure 1: Structural Overview of Methodology

Data Acquisition

For this analysis, the ODIR-5K dataset sourced from Kaggle for this study [30]. It is a publicly available ophthalmic dataset designed for automated ocular disease analysis and classification. The dataset contains clinical data and retinal fundus images. They are collected from approximately 5,000 multi-institutional patients in China. The fundus images are captured using various imaging devices like Canon, Zeiss, and Kowa cameras. Each patient record includes left and right eye fundus photographs, demographic information, and diagnostic annotations provided by trained ophthalmologists under quality-controlled labeling procedures. The dataset consists of eight disease modalities: Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age-related Macular Degeneration (A), Hypertension (H), Pathological Myopia (M), and Other abnormalities (O). As this study focuses casual inference and counterfactual reasoning glaucoma, associated variables are required for further analysis. So, this dataset is adopted for this study.

Preprocessing

To prepare the data for model building, the images are resized to 224×224 and normalized using ImageNet statistics. Most of the clinical data are numerical scales. Also, categorical data are converted into numerical scaling too. As well as one-hot labels converted to class indices.

Multimodal Model

The GlaucoFusionNet is a multimodal deep learning architecture. It is designed for joint feature learning from retinal fundus images and auxiliary clinical information. The image branch is applied a ResNet-style convolutional neural network (CNN). It is consisting of Convolution-Batch Normalization-ReLU block. This is followed by a stacked residual basic block with expanding depth (64 → 128 → 256 → 512 channels). It is enabling hierarchical representation learning while preserving gradient flow through skip connections. Global spatial features are aggregated using adaptive average pooling. Parallelly, a multilayer perceptron (MLP) processes structured clinical input and transforming it into a compact embedding space through two fully connected layers with nonlinear activation. The extracted image features (512-d) and clinical embeddings (16-d) are concatenated to form a unified feature representation (528-d). It is inputted to a densehead. The classifier lies of a dense layer with ReLU activation, followed by dropout regularization ($p = 0.3$) to reduce overfitting. Then they are applied softmax layer for binary glaucoma classification. Overall, the architecture enables effective multimodal fusion of imaging and clinical data for improving discriminative learning to detect accurate glaucoma.

Explainability AI (Grad-CAM and SHAP)

To improve the transparency and interpretability of the proposed glaucoma prediction model, two Explainable Artificial Intelligence (XAI) techniques were employed in this study. Grad-CAM (Gradient-weighted Class Activation Mapping) and SHAP (SHapley Additive exPlanations) are used in this study to unveil the black boxing of predictions by model on fundus images and clinical data respectively. Grad-CAM is a visual explanation technique used for the model. It highlights the important regions in retinal fundus images that influence the prediction of the model. In this study, Grad-CAM generates heatmaps on retinal images to identify glaucoma-related areas such as the optic disc and cup region. This helps clinicians understand where the model focuses while making predictions. SHAP explains the contribution of each clinical variable to the model prediction. SHAP provides both global and patient-level explanations by quantifying how each feature increases or decreases glaucoma risk.

Causal Inference

A DAG is constructed using the DoWhy framework. It represents the causal relationships between clinical variables and glaucoma risk. This graph speak for influenced factors for glaucoma development. This causal model derives the causal inference of specific risk factors.
Counterfactual Simulation

Counterfactual simulation is performed to analyze “what-if” clinical scenarios. It is trailed by modifying individual patient attributes. It is observing the corresponding change in predicted glaucoma risk. If any severity variable is systematically varied while keeping other inputs constant to evaluate its impact on disease prediction. This approach provides interpretable patient-level intervention analysis. It supports clinically meaningful decision-making.

EXPERIMENTAL RESULTS

The proposed multimodal causality-aware with counterfactual reasoning framework was implemented and evaluated using the Python programming language in the Google Colab environment. The experiments are conducted on data preprocessing, multimodal model training, explainability analysis, causal inference modeling, and counterfactual simulation. The obtained outputs demonstrate the effectiveness of proposed framework for glaucoma risk prediction. The following subsections present and discuss the output generated through the experimental analysis.

The training phase of GlaucoFusionNet was operated on mingled retinal fundus images and structured clinical features. The loss and accuracy curves were monitored across 20 epochs to evaluate learning capability as shown figure 2.

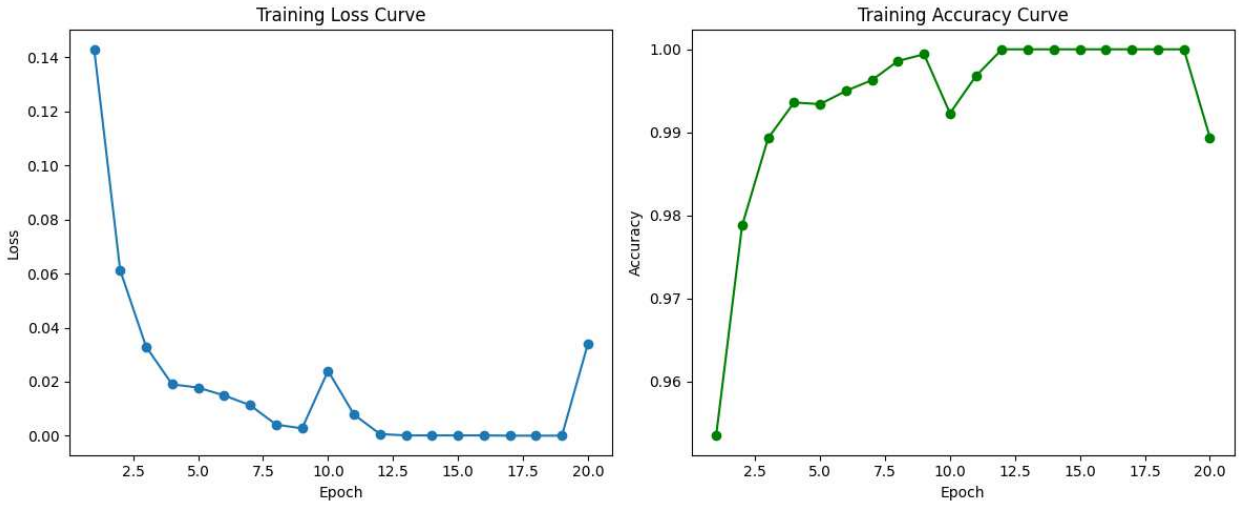


Figure 2: Training Performance of the GlaucoFusionNet for Automated Glaucoma Diagnosis

Figure 2 illustrates the training loss and training accuracy curves of the GlaucoFusionNet model to recognize glaucoma. The training loss decreases rapidly during the initial epochs. It indicates effective learning and optimization of retinal and clinical features by the GlaucoFusionNet. The loss approaches nearly zero after several epochs. The model converged smoothly with minimal performance variance in the final stages. Concurrently, training accuracy climbs steadily from around 95% to nearly 100%. The framework demonstrates superior and data-driven fusion for

yielding highly discriminative representations for improved classification. The results affirm that GlaucoFusionNet captures essential diagnostic patterns accurately. It produces reliable classification results.

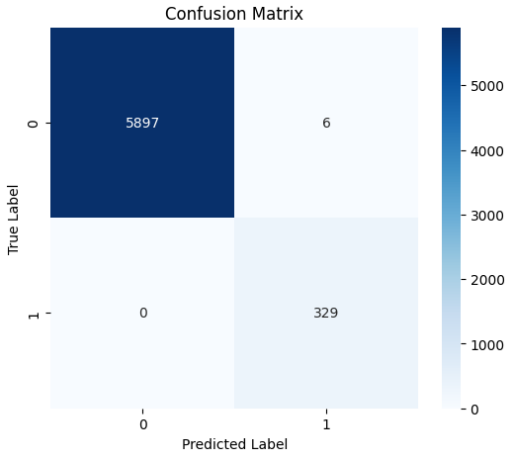


Figure 3: Confusion Matrix of the GlaucoFusionNet for Automated Glaucoma Diagnosis

The confusion matrix in figure 3 highlights the robust classification capability of GlaucoFusionNet. The model correctly classified 5897 normal cases and 329 glaucoma cases. Also, there are only 6 false positive predictions and no false negatives. The test is nearly flawless. Accordingly, GlaucoFusionNet ssegregates healthy and diseased eyes without diagnostic errors.

To evaluate the transparency of GlaucoFusionNet model, Grad-CAM is utilized to generate class activation maps. As shown in Figures 4 and 5 2. GlaucoFusionNet consistently prioritized the optic disc and surrounding retinal nerve fiber layer (RNFL) for identifying glaucoma.

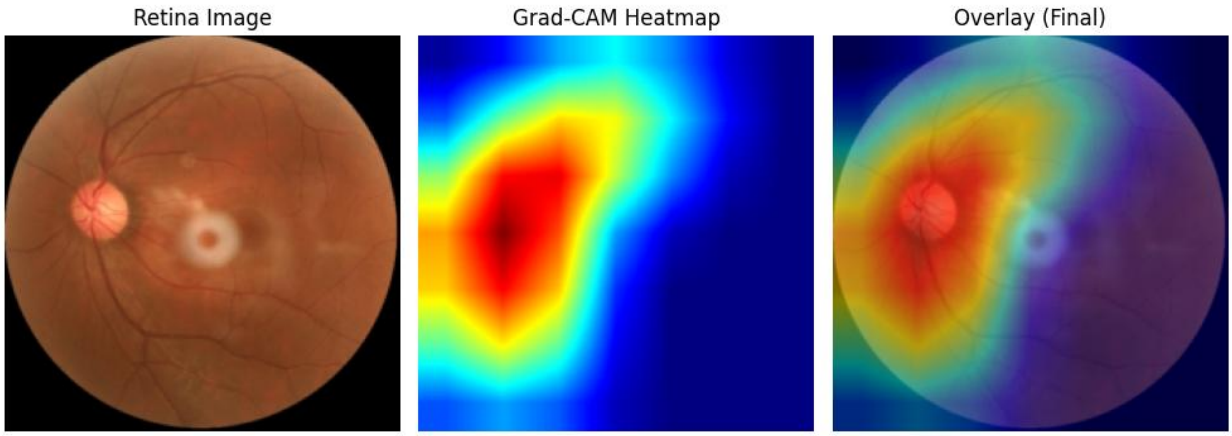


Figure 4: Visualization of the decision-making process by GlaucoFusionNet using Grad-CAM on a representative fundus image.

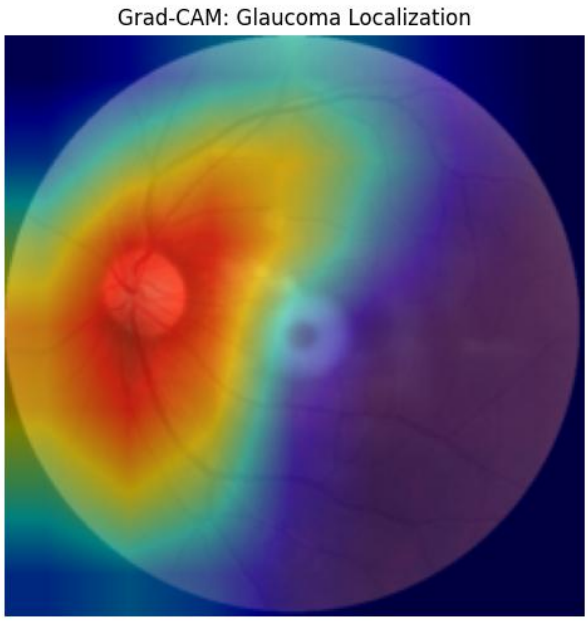


Figure 5: Spatial localization of glaucomatous features highlighting focus of GlaucoFusionNet on the optic disc region.

Figures 4 and 5 demonstrate the interpretability of GlaucoFusionNet for glaucoma detection at clinically relevant anatomical features. Figure 4 illustrates the pipeline of the explainability process including retina image, Grad-CAM Heatmap and overlay. Retina Image shows the raw input data, Grad-CAM Heatmap represents the "attention" of the GlaucoFusionNet. The red regions indicate high-importance areas that the model used to make its decison. Overlay confirms that the model is focusing on the optic disc and neuroretinal rim. They are the primary zones clinicians examine for signs of glaucoma. A successful Glaucoma Localization is presented in figure 5. The intense

red "hotspot" directly overlaps with the optic nerve head. This confirms the ability of model to detect optic disc pathologies. This framework supports the practical adoption by boosting interpretability and trustworthiness.

Figure 6 outlines the SHAP feature importance analysis for glaucoma prediction. This visual depicts the mean impact of clinical variables on classification results. Retinal disease-related attributes such as Normal (N), Diabetes (D), and Other abnormalities (O) show the highest contribution to the prediction process among the features. The results indicate that semantic retinal features substantially drive in accurate glaucoma classification. The glaucoma-related feature (G) also demonstrates a notable contribution for confirming the capability of GlaucoFusionNet to identify retinal characteristics based on disease-specific. Also, clinical variables such as patient age and hypertension (H) are contributing moderately to the prediction outcome. It is highlighting the importance of multimodal clinical information in glaucoma risk assessment. The SHAP analysis improves interpretability by identifying the most influential features affecting model decisions. It provides clinically meaningful insights into how glaucoma prediction is contributed by retinal and patient-related factors. This results demonstrate that the GlaucoFusionNet framework effectively integrates imaging and clinical features and maintaining transparency and explainability in the process of glaucoma diagnosis.

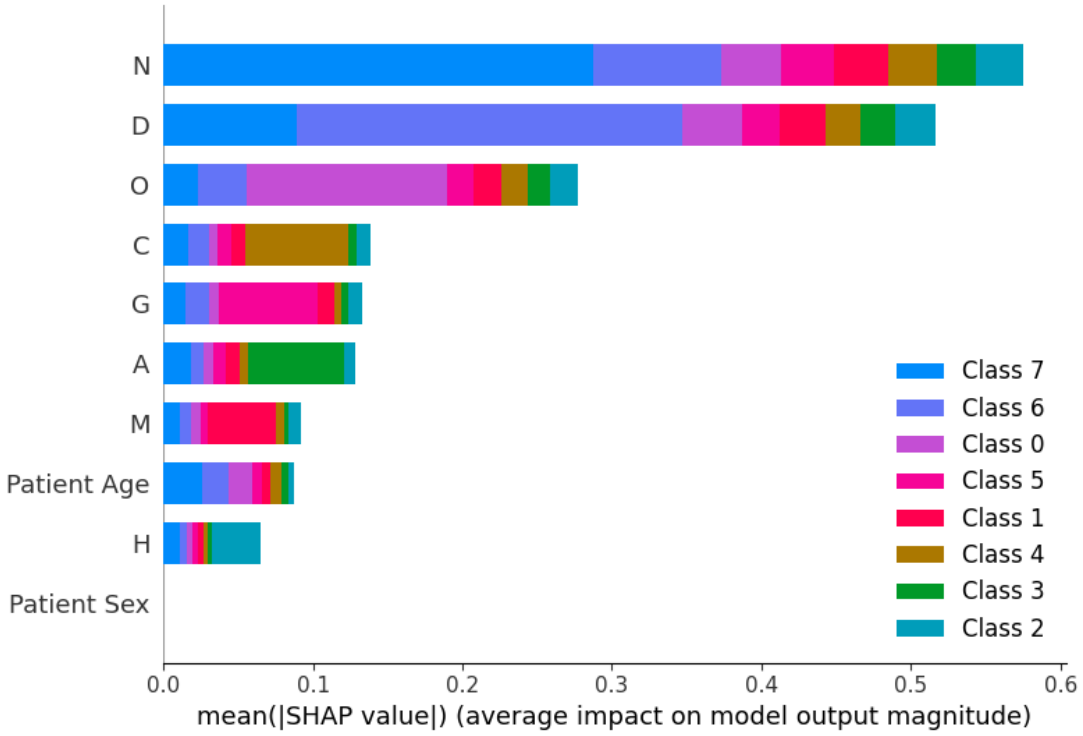


Figure 6: SHAP-Based Feature Importance Analysis for Multimodal Glaucoma Prediction

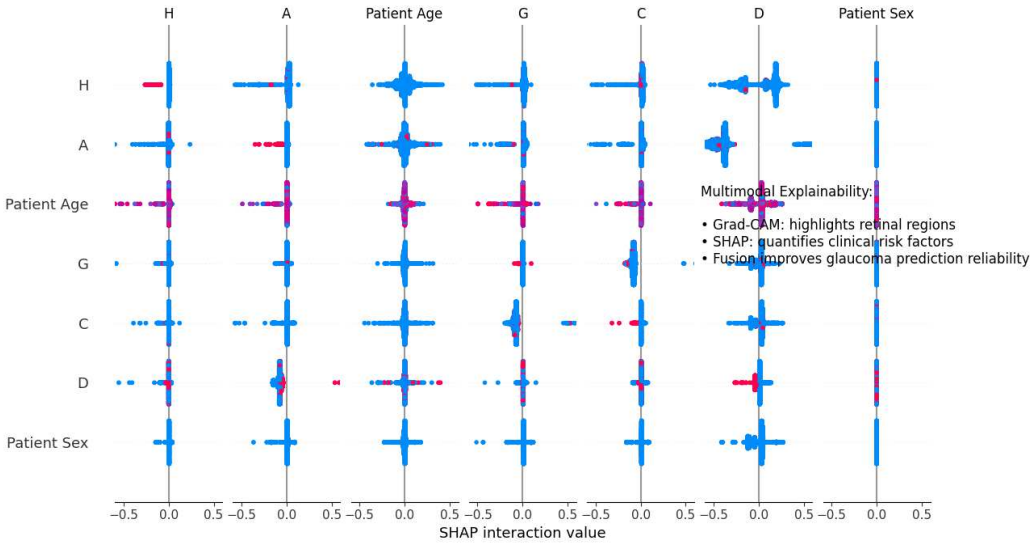


Figure 7: SHAP Interaction Analysis of Clinical and Retinal Features for Glaucoma Prediction

Figure 7 demonstrates the SHAP interaction analysis of retinal and clinical features used in the GlaucoFusionNet framework for glaucoma prediction. This figure visualizes the interaction effects between different variables including glaucoma-related retinal features, diabetes, cataract, hypertension, age, and patient sex, on the model's prediction outcomes. Analysis of SHAP interaction values reveals that the interplay between retinal and clinical factors shapes the estimation of glaucoma risk. Age acts as a significant effect modifier on the impact of retinal characteristics on glaucoma progression. The evidence indicates that both ocular and systemic factors play a role in predicting glaucoma. These findings Spotlighting their contribution of biomarkers on clinical course and and prognostic modeling. Moreover, multimodal explanation method clarifies internal logic by uncovering individual feature importance and relationships between modalities. It delivers actionable, interpretable data to support trustworthy and auditable detection of glaucoma.

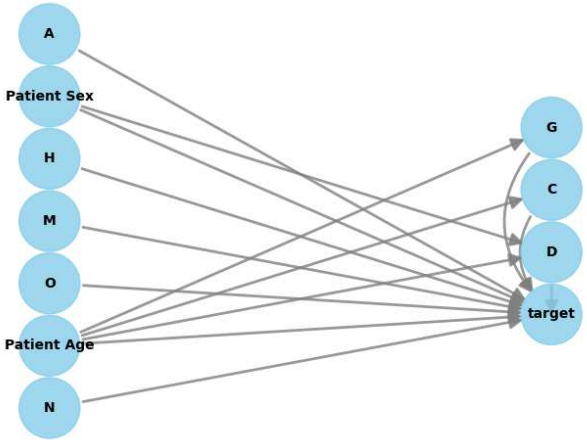


Figure 8: Directed Acyclic Graph for Causal Relationships in Glaucoma Prediction

The Directed Acyclic Graph used in the GlaucoFusionNet framework to model causal relationships between clinical variables, retinal disease features, and glaucoma prediction outcomes in figure 8. This graph illustrates how patient-specific factors glaucoma-related variables and the final prediction. The directed connections represent potential causal pathways between multimodal clinical and retinal features. The DAG-based causal structure helps identify clinically meaningful relationships. It reduces the influence of spurious correlations in the prediction process. In particular, patient age and systemic disease-related variables show direct causal influence on glaucoma-associated outcomes. It is highlighting their importance in glaucoma risk assessment. Causal reasoning is incorporated in multimodal deep learning framework. The GlaucoFusionNet improves interpretability, robustness, and clinical reliability. Furthermore, the DAG serves as the foundation for counterfactual reasoning for enabling "what-if" scenario analysis for personalized glaucoma risk evaluation and decision support.

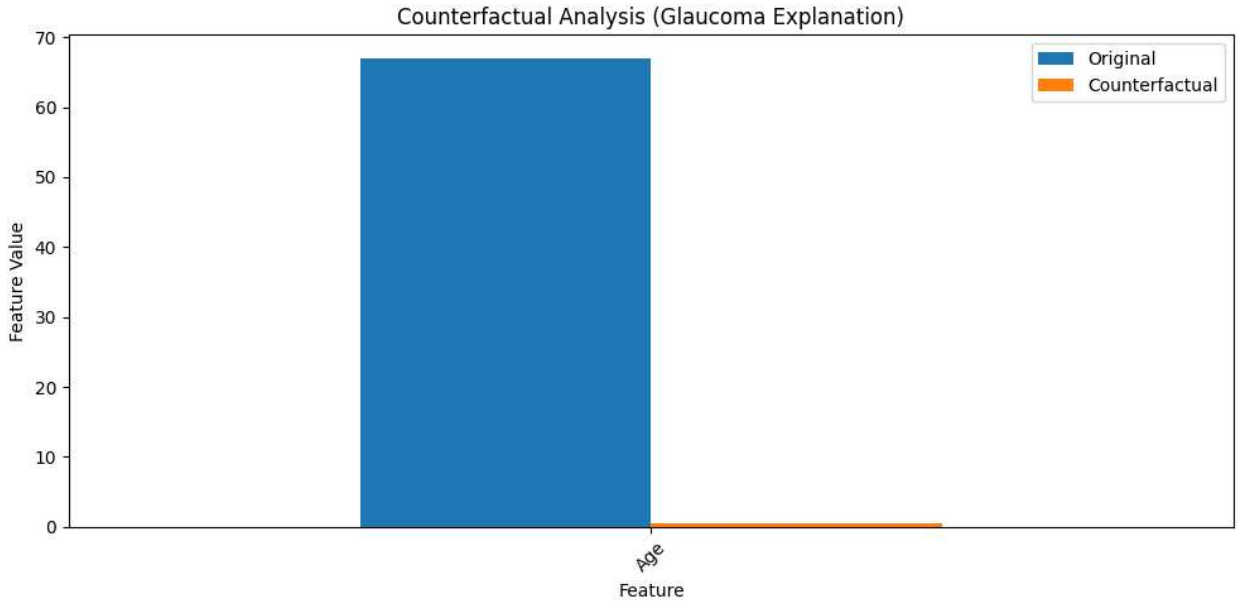


Figure 9: Counterfactual Analysis for Personalized Glaucoma Risk Assessment

With in the GlaucoFusionNet framework, the counterfactual analysis is performed and results is explicated in figure 9. It compares the original clinical feature value with a counterfactual modified value to evaluate how changes in patient-specific variables influence glaucoma risk prediction. In this example, patient age is analyzed as a key contributing factor in the decision-making process

by the model. The comparison demonstrates how altering a clinical variable while keeping other factors unchanged can affect the predicted glaucoma outcome. This "what-if" analysis provides interpretable and clinically meaningful insights into the relationship between patient characteristics and glaucoma risk. The counterfactual reasoning framework enhances transparency. Also it supports personalized clinical decision-making to enable clinicians to understand how modifying a specific risk factors may influence disease prediction and progression assessment.

GlaucoFusionNet framework investigated the effect of changes in clinical variables on glaucoma risk prediction through the intervention simulation in figure 10. Under a simulated intervention setting, the relationship between the clinical feature value and the corresponding predicted glaucoma risk is shown in the graph. It is indicating that a gradual reduction in predicted glaucoma risk while the clinical feature value changes across the intervention range. It demonstrates that the capability of the counterfactual reasoning framework. It explains the modifications in patient-specific clinical conditions may influence glaucoma outcomes. The intervention analysis provides clinically interpretable insights into disease progression. Also, it supports personalized decision-making by enabling "what-if" scenario evaluation. Overall, the figure highlights the usefulness of causality-aware modeling for understanding the impact of clinical interventions on glaucoma risk prediction.

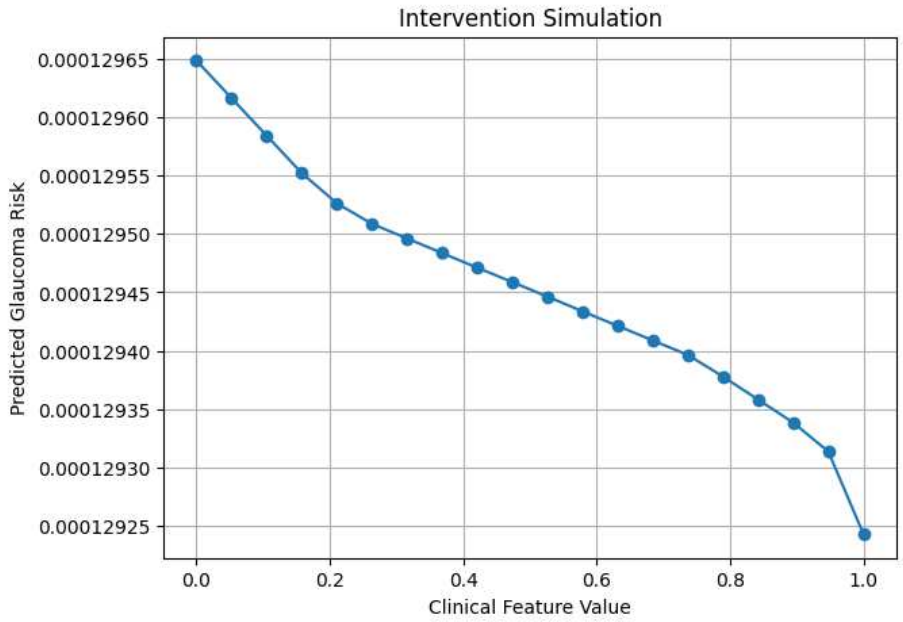


Figure 10: Intervention Simulation for Counterfactual Glaucoma Risk Analysis

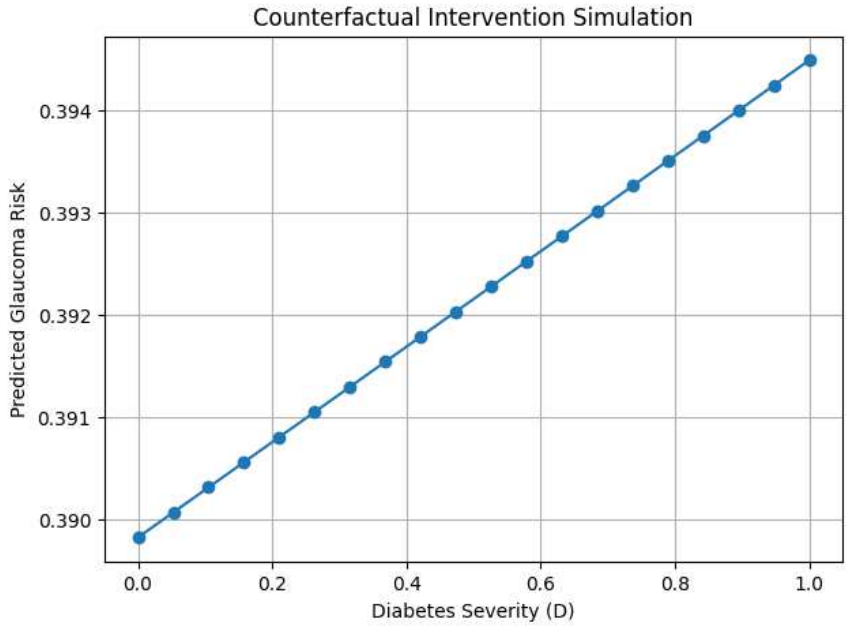


Figure 11: Counterfactual Intervention Analysis of Diabetes Severity on Glaucoma Risk

Figure 11 details the results from the counterfactual intervention simulation. The study investigates the impact of diabetes severity on diabetic progression within the GlaucoFusionNet framework. Findings demonstrate a progressive increase between advancing diabetes severity and elevated glaucoma risk. Diabetes acts as an independent predictor of glaucoma. Using what-if simulation, the system emulates how altering diabetes severity would affect the worsening of glaucoma.

GlaucoFusionNet framework generates the different glaucoma risk groups on dataset and provided in table 1 and figure 12. Based on the cohort analysis, a significant proportion of patients fell within the low-risk category 5,985 individuals. Results show high resistance to glaucoma-inducing ocular markers. Fewer patients were categorized as medium risk (41 cases) and high risk (206 cases). This risk profiling aids enables early identification, tailored surveillance, and clinical reasoning. It flags patients for urgent referral to an ophthalmologist

Table 1: Glaucoma Risk Stratification Based on Predicted Risk Scores

Risk group	**Count**
Low	5985
High	206
Medium	41

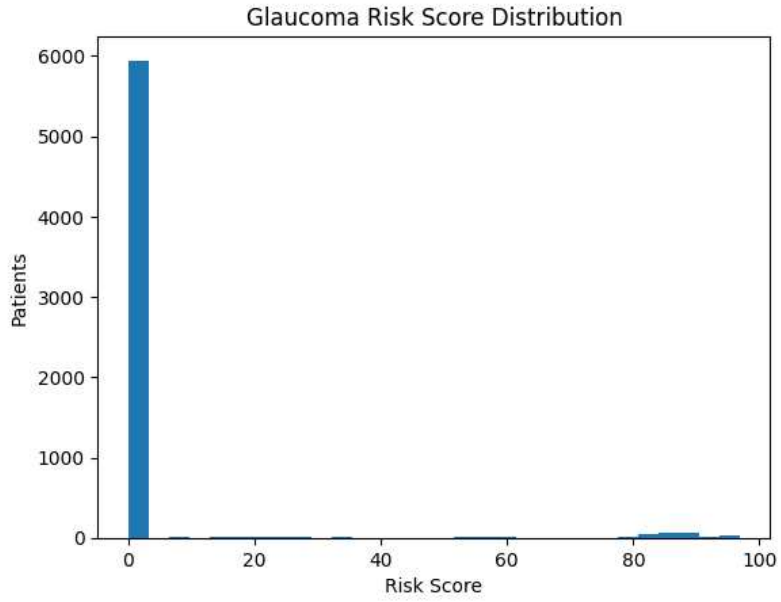


Figure 12: Glaucoma Risk Score Distribution

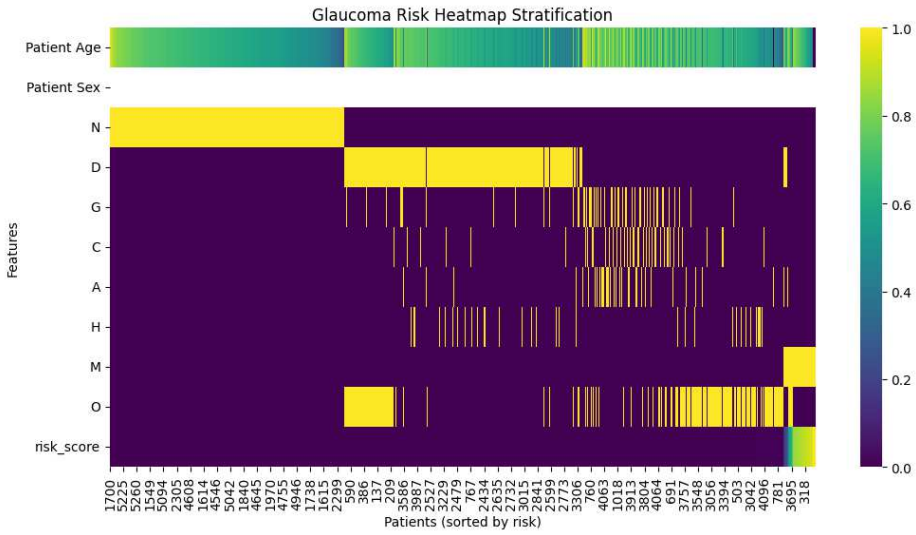


Figure 13: Glaucoma Risk Heatmap Stratification Using Multimodal Clinical and Retinal Features

The Color-coded risk map formulated by fused retinal-clinical inputs shown in Figure 13. It maps the landscape of individual glaucoma risk factors. Patients are **triaged** using predicted risk scores. Rows are categorized by eye diagnosis, including age, sex, and ocular risk factors. Color intensity **indicates** the feature contribution and activation levels. In glaucoma-related and systemic disease

variables, high-level expressions are characteristic of patients requiring close monitoring. The GlaucoFusionNet framework distinctly separates patients by risk profile. It unravels the black box for assisting in targeted screening strategies.

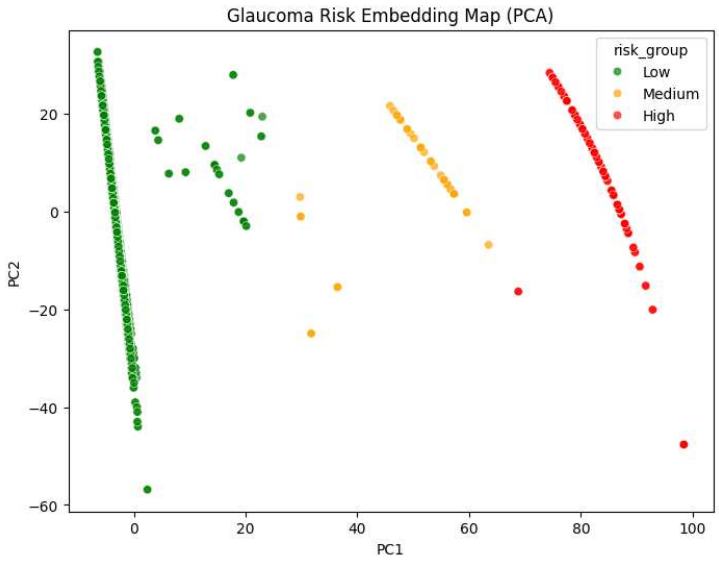


Figure 14: PCA-Based Glaucoma Risk Embedding Map for Patient Stratification

Using PCA to reduce the high-dimensional features extracted by model into a 2D plane, model separates healthy from risk groups as visualized in Figure 14. Patients are risk-stratified based on predictive modeling to differentiate low, moderate, and high susceptibility. A sharp demarcation of risk profiles is shown in this plot. It is pointing to the fact that merged multimodal feature vectors successfully extract distinctive glaucomatous markers.

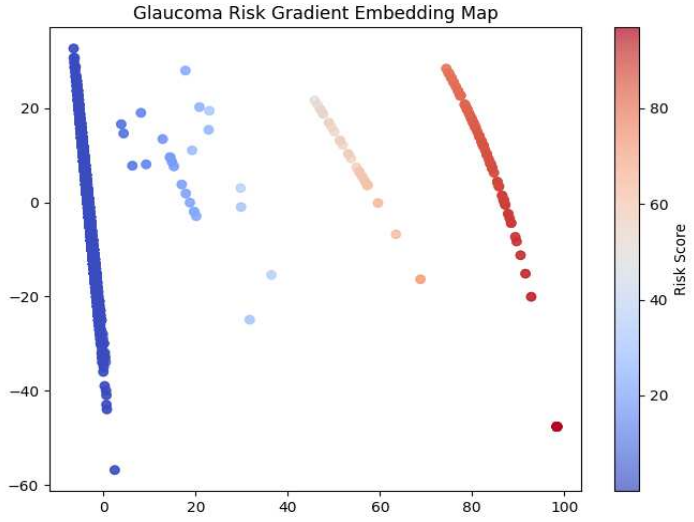


Figure 15: Gradient-Based Glaucoma Risk Embedding Visualization

The multimodal features learned by the GlaucoFusionNet framework yielded the gradient-based glaucoma risk embedding map shown in Figure 15. Color Spectrum indicates a full range of colors from low to high. The embedding shows continuous progression. High-risk glaucoma suspects are partitioned into distinct cohorts whereas lower-risk individuals are clustered separately. This image/map/representation showcases the reliability/stability of the introduced model. This visual evidence showcases the reliability of the GlaucoFusionNet framework.

Figure 16 maps the classification of patients based on the decision boundary model within a 2D feature space. It maps the criteria into a geometric space. to distinguish between risk intensities. It plots the top two principal components to showcase patient similarity. Color saturation signifies the forecasted glaucoma severity. Low-risk individuals concentrate in regions of subtle coloration. High-hazard cohort hotspots in areas of darker color saturation. Decision boundary analysis reveals concatenated retinal and patient metadata high-fidelity representation of pathology-specific markers. It supports reliable patient stratification. Also, it highlights the interpretability and clinical applicability of the proposed framework. It is providing a visual representation of influence glaucoma risk prediction. This supports transparent AI-assisted clinical decision-making and personalized glaucoma assessment.

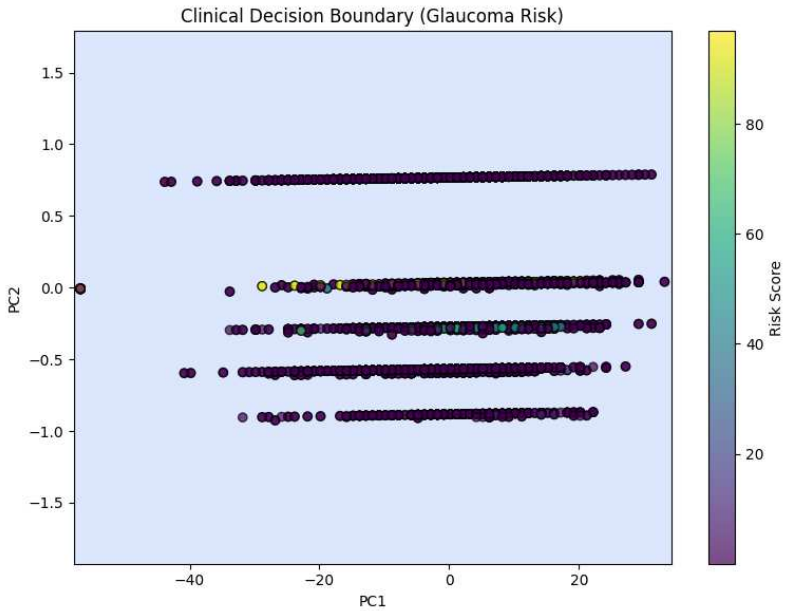


Figure 16: Clinical Decision Boundary Visualization for Glaucoma Risk Prediction

In figure 17, an advanced plot represents a decision boundary adjusted for Causal Relationships. This boundary illustrates the logic of model changes when accounting for the structural

dependencies between variables with the amalgamation of the causal framework. For instance, it shows the causal link between age and disease onset.

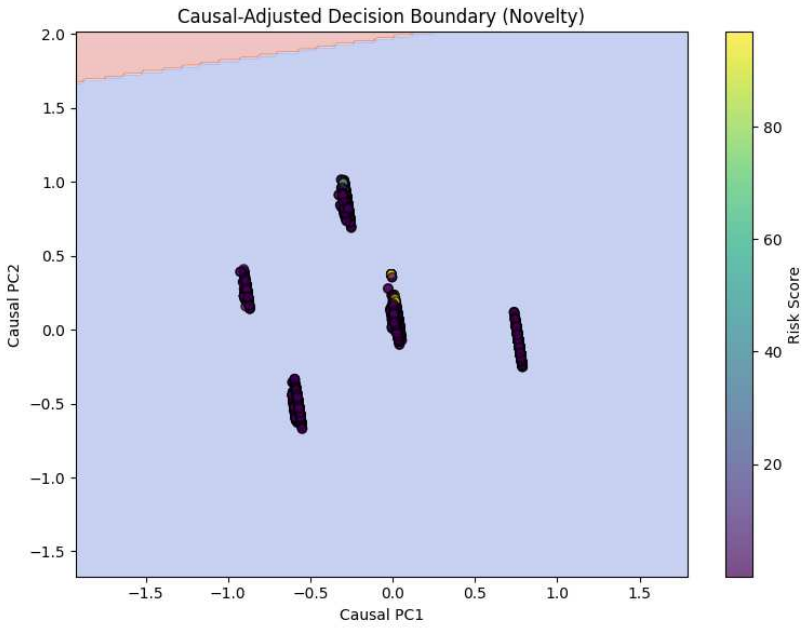


Figure 17: Causal-Adjusted Decision Boundary

DISCUSSION

The experimental findings are ensuring that integrating multimodal deep learning, explainability techniques, causal inference, and counterfactual reasoning in the glaucoma case study. It significantly improves the interpretability and clinical relevance of glaucoma prediction systems. The proposed framework provides accurate disease classification. Also, it provides transparent and causally meaningful explanations. It can support ophthalmologists in clinical decision-making. It significantly enhances the clinical interpretability of glaucoma detection for moving beyond black-box predictions to provide actionable medical insights. Grad-CAM visualization pinpoints pathological changes in the optic disc. SHAP and DoWhy frameworks successfully align its internal logic with established ophthalmological risk factors such as Age and Diabetes Severity. The counterfactual simulation analysis further bridges the gap between AI and clinical practice by enabling "what-if" reasoning. It allows clinicians to quantify how modifications in a patient's systemic health usually influence their predicted glaucoma risk. Besides, the PCA-based embedding reveals well-defined boundaries between glaucoma severity levels. These findings demonstrate that the combined retinal imaging and clinical feature space is exceptionally robust for accurate risk stratification. As well as the counterfactual reasoning framework simulating a different outcome if a patient variable had been different. Manipulating diabetes complication stage and maintaining steady demographic gauge fluctuations in glaucoma risk estimates. This method offers practical clinical intelligence by illuminating advancements in whole-body health.

Sustaining optimal blood sugar levels is lowering glaucoma probability and targeted clinical management.

CONCLUSION

GlaucoFusionNet eemphasizes a major leap forward in Computer-aided ophthalmic screening. It typifies by surpassing correlation-centric approaches toward a causal-inference-based multimodal learning. This platform enables a holistic perspective of glaucomatous changes and susceptibility via the fusion of retinal images and systematic patient info. Results show the model successfully pinpoints relevant biomarkers within the optic disc region via Grad-CAM visualization. Additionally, SHAP analysis determines the impact of systemic and clinical factors on glaucoma prediction. By adding causal frameworks and counterfactual, the system mitigates the influence of confounding variables and boost model clarity. The framework also facilitates clinicians to perform hypothetical scenario testing. It is empowering patient-centric glaucoma risk profiling utilizing dynamic shifts in individual clinical data. Ultimately, it delivers a clear, explainable, and reliable clinical aid to halt permanent blindness.

Capitalizing on the findings of GlaucoFusionNet, subsequent investigations will focus on utilization of temporal data to construct predictive longitudinal models to shift from snapshot-based risk assessment to time-dependent forecasting. Future work should **broaden** the scope to marginalized populations for validating the model across diverse datasets. It is essential to strengthen the robustness of causal relationships and what-if simulations. OCT Data is ccollating En Face Imaging volumetric analysis of optic nerve fiber. Live clinical evaluations can be operationalized the approach in physician-in-the-loop environment to analyze the effect of "what-if" insights ophthalmic clinical determinations.

Author Contribution:

A.A contributed to Introduction, Methdodology and Result analysis; H.K.P. contributed towards the final drafting and Discussion; S.M. has contributed with Literature Survey and Conclusion

Declaration:
- Competing Interests: NA
- Funding Information: NA
- Data Availability Statement: The datasets generated and/or analysed during the current study are available in the : https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k/data
- Research Involving Human and /or Animals: NA
- Ethics declaration: Not Applicable.
- Informed Consent: NA
- Conflicts Of Interest: The authors declared no conflict of interest, financial or otherwise.
- Consent for publication: Not applicable.

References

[1]. Alnuaimi, Sghaira, and Nabil Albastaki. *Automatic Recognition of Glaucoma Using Generative Adversarial Networks and Region of Interest Techniques*. Aug. 2025, pp. 96–101, https://doi.org/10.1109/rtsi64020.2025.11212482.

[2]. Pathak, Divya. "Glaucoma Detection Using Deep Learning Techniques." *Indian Scientific Journal Of Research In Engineering And Management*, vol. 09, no. 05, May 2025, pp. 1–9, https://doi.org/10.55041/ijsrem48881.

[3]. Arrieta Rodríguez, Eugenia, et al. "Deep Learning for Glaucoma Classification and Grading: A Comprehensive Review on Fundus Imaging Approaches." *IEEE Access*, Jan. 2025, p. 1, https://doi.org/10.1109/access.2025.3605819.

[4]. Shoukat, Ayesha, et al. "Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach." *Diagnostics*, vol. 13, no. 10, May 2023, p. 1738, https://doi.org/10.3390/diagnostics13101738.

[5]. Xu, Yongli, et al. *A Hierarchical Deep Learning Approach with Transparency and Interpretability Based on Small Samples for Glaucoma Diagnosis*. no. 1, Mar. 2021, p. 48, https://doi.org/10.1038/S41746-021-00417-4.

[6]. Patil, C. Y., et al. "Deep Learning-Based Glaucoma Detection Using Cropped Optic Disc Regions." *International Journal For Multidisciplinary Research*, vol. 7, no. 2, Apr. 2025, https://doi.org/10.36948/ijfmr.2025.v07i02.43042.

[7]. Aguilar, Kristhian, et al. *Automatic Glaucoma Classification and Feature Detection for the Justraigs Challenge*. May 2024, pp. 1–4, https://doi.org/10.1109/isbi56570.2024.10635775.

[8]. Xu, Jie, and Yidong Chai. *Classifying Glaucoma and Normal Large Optic Cups Using Deep Learning Methods*. Dec. 2023, https://doi.org/10.21203/rs.3.rs-3571001/v1.

[9]. Pathmakumara, Heshan Chandeepa, et al. *Explainable Deep Learning for Glaucoma Detection: A DenseNet121-Based Classification with Grad-CAM Visualization*. Oct. 2025, https://doi.org/10.1101/2025.10.08.25337634.

[10]. Velpula, Vijaya Kumar, et al. "Glaucoma Detection with Explainable AI Using Convolutional Neural Networks Based Feature Extraction and Machine Learning Classifiers." *Iet Image Processing*, Aug. 2024, https://doi.org/10.1049/ipr2.13211.

[11]. Avadhani, Bindu, et al. *Beyond the Black Box: Explainable AI for Glaucoma Detection and Future Improvements*. June 2024, pp. 1–9, https://doi.org/10.1109/icccnt61001.2024.10725963.

[12]. Kumar, Sachin, et al. *Explainable AI for Glaucoma Detection: Integrating Alternet-k and LIME*. June 2024, pp. 1–11, https://doi.org/10.1109/icccnt61001.2024.10724218.

[13]. Deepa, J., et al. "Explainable Deep Learning for Glaucoma Detection: Efficient Dense Net with Quantum-Based Dwarf Mongoose Optimization." *Edelweiss Applied Science and Technology*, vol. 8, no. 6, Oct. 2024, pp. 694–706, https://doi.org/10.55214/25768484.v8i6.2140.

[14]. Chayan, Touhidul Islam, et al. *Explainable AI Based Glaucoma Detection Using Transfer Learning and LIME*. Oct. 2022, pp. 1–6, https://doi.org/10.1109/CSDE56538.2022.10089310.

[15]. Shah, Syed Safi Ullah, et al. *OptiGuard: Generalized, Attention-Driven & Explainable Glaucoma Classification*. July 2025, pp. 1–7, https://doi.org/10.1109/embc58623.2025.11253669.

[16]. Liu, Xiaoyu, et al. *Enhancing Multimodal Glaucoma Screening through Attention-Guided Vision-Language Fusion*. Sept. 2025, https://doi.org/10.21203/rs.3.rs-7481454/v1.

[17]. Kovalyk, Oleksandr, et al. "Glaucoma Detection: Binocular Approach and Clinical Data in Machine Learning." *Artificial Intelligence in Medicine*, vol. 160, Dec. 2024, p. 103050, https://doi.org/10.1016/j.artmed.2024.103050.

[18]. Huang, Xiaoqin, et al. "Detecting Glaucoma from Multi-Modal Data Using Probabilistic Deep Learning." *Frontiers in Medicine*, vol. 9, Sept. 2022, https://doi.org/10.3389/fmed.2022.923096.

[19]. Krishna, Nanditha, and K. Nagamani. "Multi-modal imaging-based feature fusion for accurate glaucoma diagnosis with deep learning." (2023).

[20]. Huang, Cheng, et al. *Automated Glaucoma Report Generation via Dual-Attention Semantic Parallel-LSTM and Multimodal Clinical Data Integration*. Oct. 2025, https://doi.org/10.48550/arxiv.2510.10037.

[21]. Ferreira, Marcos Melo, et al. "Glaucoma Grading Using Multimodal Imaging and Multilevel CNN." *IEEE Latin America Transactions*, vol. 21, Oct. 2023, pp. 1095–102, https://doi.org/10.1109/tla.2023.10255449.

[22]. Julian, T. H., Sanderson, E., Guo, H., et al. (2022). Causal factors in primary open angle glaucoma: A phenome-wide Mendelian randomisation study. *Scientific Reports*, *12*. https://doi.org/10.1038/s41598-023-37144-7

[23]. Chen, S., Lin, M., & Hong, Y. (2025). The causal effect of glaucoma and diabetic retinopathy: A Mendelian randomization study. *Diabetology & Metabolic Syndrome*, *17*(1). https://doi.org/10.1186/s13098-025-01652-5

[24]. Li, J., Ma, X., Zhuo, K., et al. (2024). Investigating the uncertain causal link between gut microbiota and glaucoma: A genetic correlation and Mendelian randomisation study. *Clinical & Experimental Ophthalmology*. https://doi.org/10.1111/ceo.14440

[25]. Han, X., Wang, J., Su, X., et al. (2025). Exploring the causal influence of 731 immune cells on 4 different glaucoma subtypes using a two-sample Mendelian randomization method. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-90545-8

[26]. Jiang, J., Hu, D., Zhang, Q., Lin, Z., & μ-Biomedical Data Investigation Group (Mu-BioDig). (2026). Integrating generative AI and genetic causal inference identifies metformin as a repurposing candidate for normal tension glaucoma. *British Journal of Clinical Pharmacology*.

[27]. Kim, D. (2025). LightHCG: A lightweight yet powerful HSIC disentanglement based causal glaucoma detection model framework. *arXiv preprint arXiv:2512.02437*.

[28]. Alonso, A., Van der Elst, W., & Molenberghs, G. (2016). An information-theoretic approach for the evaluation of surrogate endpoints based on causal inference. *Biometrics*, *72*(2), 669–677. https://doi.org/10.1111/biom.12483

[29]. Alonso, A., Molenberghs, G., & Van der Elst, W. (2018). A reflection on the possibility of finding a good surrogate. Journal of Biopharmaceutical Statistics, 29(1), 101–118. https://doi.org/10.1080/10543406.2018.1559854

[30]. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k/data

[31]. Artamonov, Daniil V., Polina I. Popova, Ekaterina A. Korf, Natalia G. Voitenko, Alisa A. Chernysheva, Pavel V. Avdonin, Richard O. Jenkins, and Nikolay V. Goncharov. "Interpretable Machine Learning with SHAP Identifies Key Biomarkers in a Multi-Factorial Spectrum of Age-Related Neurological and Metabolic Conditions." International Journal of Molecular Sciences 27, no. 4 (2026): 1805.