Journal of Distributed Computing and Systems(JDCS)
Vol.8, Issue 1, 2025, Pages 85-91

A Multi-Branch Attention-Based Neural Network for Automated Glaucoma Detection Using Multimodal Clinical and OCT-Derived Features

Mehran Rajaeifar[1], Roshanak Rafiei Nazari[2], Zoleikha Jahanbakhsh Naghadeh[3*]

[1]*Department of Medical Engineering, ST.C., Islamic Azad University, Tehran, Iran.*
[2]*Department of Physics, ST.C., Islamic Azad University, Tehran, Iran.*
[3]*Department of Computer Engineering, Nag.C., Islamic Azad University, Naghadeh, Iran.*

Article History:
Received: 23 May 2025
Received in revised form: 30 July 2025
Accepted: 19 August 2025
Available online: 18 September 2025

Abstract

Glaucoma is a leading cause of irreversible blindness worldwide, often progressing asymptomatically in its early stages. Early and accurate detection is critical to prevent permanent vision loss. While deep learning has shown promise in ophthalmic diagnosis, most existing models process heterogeneous clinical and imaging features monolithically, potentially overlooking group-specific patterns. This study proposes a novel multi-branch feedforward neural network (MBFNN) equipped with an attention mechanism for automated glaucoma detection. The model processes six distinct groups of multimodal features, including clinical parameters, visual field indices, and retinal layer thicknesses from swept-source OCT, through dedicated parallel branches. An attention layer dynamically learns to weight the contributions of each branch. The model was trained and evaluated on a dataset of 132 eyes (62 glaucomatous, 70 healthy). The proposed MBFNN achieved 90.0% accuracy, 100% precision, 77.8% recall, 100% specificity, 87.5% F1-score, and 83.8% AUC-ROC. It outperformed baseline models, particularly in eliminating false positives. Attention weight analysis revealed that the total retinal thickness (TRT) branch contributed most significantly (weight ≈0.37) to the model's decision. The MBFNN provides a robust, framework for glaucoma screening, potentially reducing unnecessary referrals. Future work will integrate fundus images and involve larger, multi-centric validation.

Keywords: Glaucoma, Deep Learning, Multimodal Learning, Multi-Branch Neural Network, Attention Mechanism, Optical Coherence Tomography (OCT), Automated Diagnosis.

I. INTRODUCTION

Glaucoma, a group of progressive optic neuropathies, is the second leading cause of global irreversible blindness [1]. Its primary hallmark is the degeneration of retinal ganglion cells and their axons, typically associated with elevated intraocular pressure (IOP). The insidious and asymptomatic nature of early-stage glaucoma makes its timely detection a significant clinical challenge [2]. Conventional diagnostic methods, including IOP measurement, gonioscopy, visual field testing, and optic disc assessment, often rely on subjective interpretation and may lack the sensitivity to identify early structural damage [3].

The advent of artificial intelligence (AI), particularly deep learning (DL), has revolutionized medical image analysis and disease prediction [4]. In ophthalmology, convolutional neural networks (CNNs) have been extensively applied to detect glaucoma from fundus photographs and optical coherence tomography (OCT) scans [5, 6]. However, a prevalent limitation in many existing DL approaches is the monolithic processing of multimodal data. Clinical parameters (e.g., age, IOP), functional metrics (visual field indices), and structural features (retinal layer thicknesses) are often concatenated into a single input vector, neglecting their inherent heterogeneity. This approach may fail to learn optimal representations for each feature group, can allow irrelevant features to degrade model performance [7], and, more importantly, obscures the distinct contribution of each modality, limiting interpretability; a key factor for clinical trust.

A more biologically plausible strategy involves processing different data modalities through separate, dedicated network pathways. Furthermore, an attention mechanism can be integrated to allow the model to dynamically focus on the most salient features for each specific case [8]. This study addresses the identified research gap by proposing a novel multi-branch feedforward neural network (MBFNN) with an integrated attention layer for automated glaucoma detection. Our model separately processes six clinically relevant feature groups derived from multimodal patient data (including tabular clinical parameters, visual field indices, and quantitative OCT-

* Corresponding Author: Zoleikha Jahanbakhsh Naghadeh (zoleikha.jahanbakhsh@iau.ac.ir)

derived structural measurements) and uses attention to fuse these representations intelligently. The primary objectives are: (1) to develop a high-performance, interpretable model for glaucoma detection, (2) to evaluate its superiority over standard monolithic models, and (3) to analyze the relative importance of different feature groups via the learned attention weights.

II. Related Works in Glaucoma Detection

The application of Artificial Intelligence (AI), particularly Deep Learning (DL), for glaucoma detection has been extensively explored, primarily focusing on two data modalities: fundus photographs and Optical Coherence Tomography (OCT) scans.

A. Fundus Image-Based Approaches

A significant body of research employs Convolutional Neural Networks (CNNs) to classify fundus images. Pioneering work by Al-Bander et al. [5] used a CNN to automatically extract features from fundus images, achieving an accuracy of 88.2%. Manassakorn et al. [12] proposed GlauNet, a custom CNN for OCT angiography, reporting a sensitivity of 88.9% and specificity of 89.6%. More recently, Wassel et al. [13] leveraged Vision Transformers (ViTs) on a large, aggregated fundus dataset, achieving a high AUC of 0.979. While powerful, these methods rely solely on 2D fundus images, missing the detailed, quantitative layer-by-layer structural data provided by OCT.

B. OCT-Based Approaches

To capture structural damage, studies have utilized OCT data. Garcia et al. [6] combined CNNs and LSTMs to model spatial dependencies in 3D SD-OCT volumes, achieving an AUC of 0.885. Gaddipati et al. [14] employed Capsule Networks on 3D OCT scans, reporting an AUC of 0.97. These methods excel at extracting complex spatial features from raw volumetric data but often require substantial computational resources and large datasets.

C. Multimodal and Advanced Architectures

Recognizing the value of diverse data and advanced architectures, some studies have ventured into hybrid pipelines. For instance, Shinde [15] used a U-Net for optic disc segmentation from fundus images, followed by multiple classifiers (SVM, NN, AdaBoost) on the extracted features, reporting high accuracy. Others have used Generative Adversarial Networks (GANs) for data augmentation or image quality enhancement [16]. However, a common limitation in studies that do combine multiple data types persists: heterogeneous features (clinical, functional, structural) are frequently concatenated into a single input vector for a monolithic model. This approach fails to learn optimal, disentangled representations for each feature group and cannot dynamically weight their importance per case.

Our proposed model directly addresses this gap. Unlike monolithic CNNs or ViTs, we introduce a multi-branch feedforward architecture that processes distinct feature groups in parallel, respecting their inherent differences. Furthermore, we integrate an attention-based fusion mechanism, enabling the model to dynamically learn and emphasize the most salient features for each individual diagnosis, thereby enhancing both performance and interpretability, a combination not fully explored in prior works focused on tabular clinical and imaging-derived data.

III. Materials and Methods

A. Dataset and Preprocessing

A retrospective, cross-sectional dataset was used, originating from a prospective SS-OCT study conducted at Seoul National University Hospital [9]. The study adhered to the tenets of the Declaration of Helsinki and received institutional review board approval. Informed consent was obtained from all participants.

The dataset comprised 132 eyes from 132 subjects. This included 62 eyes diagnosed with primary open-angle glaucoma (POAG) and 70 healthy control eyes. To ensure statistical independence, only one randomly selected eye per subject was included. POAG diagnosis was based on the presence of glaucomatous optic neuropathy (neuroretinal rim thinning or notching) with a corresponding reliable visual field defect. Control eyes had no evidence of optic neuropathy, normal visual fields, and IOP < 21 mmHg. Key exclusion criteria were a history of intraocular surgery (except uncomplicated cataract surgery >1 year prior), coexisting retinal diseases, nonglaucomatous optic neuropathy, and unreliable visual field tests.

SS-OCT imaging was performed using the DRI OCT-1 Atlantis device (Topcon, Japan). Three-dimensional macular scans (6.2 mm x 6.2 mm) were acquired. The built-in software automatically segmented four retinal layers: retinal nerve fiber layer (RNFL), ganglion cell-inner plexiform layer (GCIPL), ganglion cell complex (GCC), and total retinal thickness (TRT). The macular area was divided into a 31x31 grid of superpixels (200x200 µm each). Asymmetry features (inter-hemispheric and inter-ocular differences) and "black superpixel" counts (areas of significant thinning) were calculated for relevant layers.

Twenty-six final features were categorized into six groups for model input:

1) *Clinical:* Age, Gender, IOP, Central Corneal Thickness.
2) *Visual* Field: Mean Deviation, Pattern Standard Deviation, Visual Field Index.
3) *RNFL:* Mean thickness and asymmetry features.
4) *GCIPL:* Mean, minimum thickness and asymmetry features.
5) *GCC:* Asymmetry features.
6) *TRT:* Asymmetry features.

Missing values were imputed with the median of the corresponding feature. The dataset was stratified and split into training (70%), validation (15%), and test (15%) sets. Features within each group were standardized (zero mean, unit variance) using the `StandardScaler` from scikit-learn, fitted on the training set.

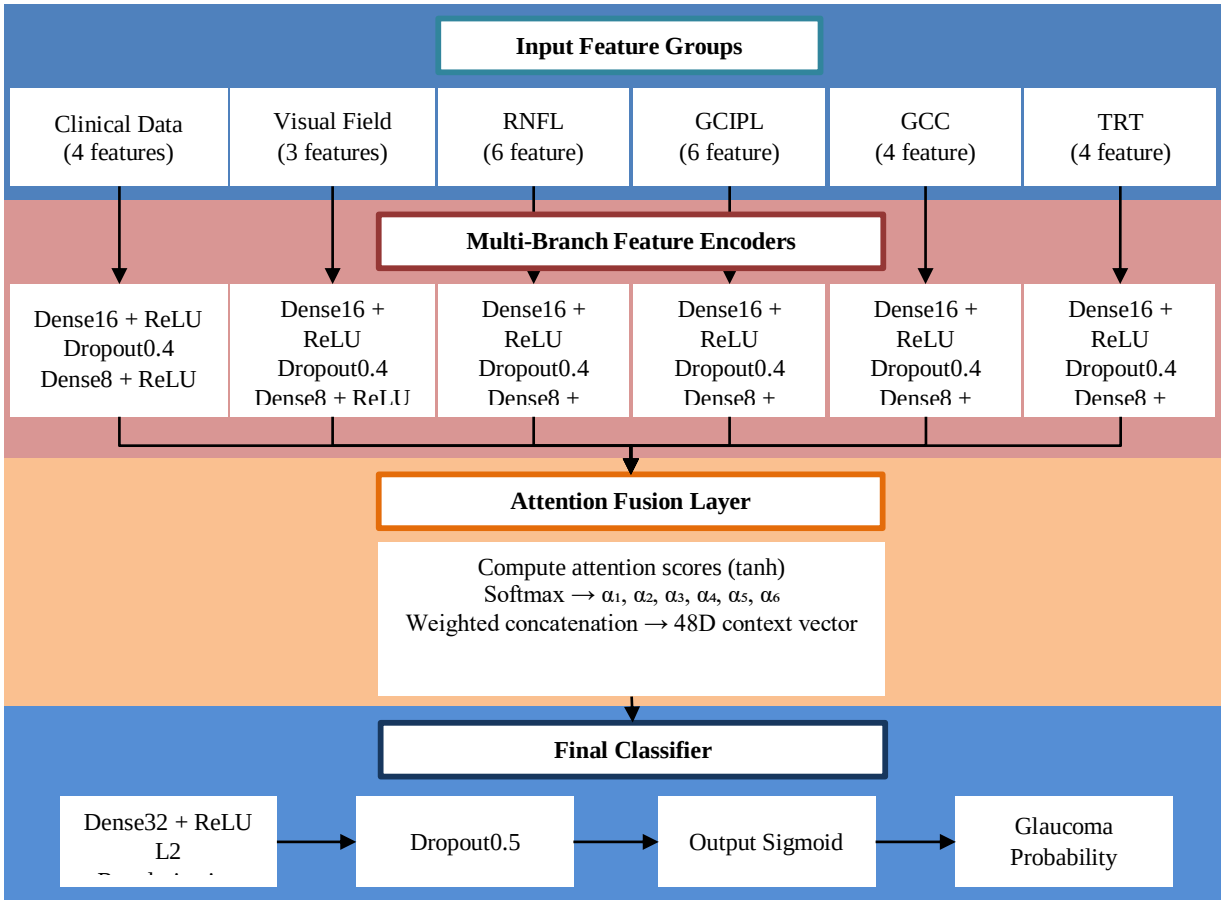

Figure 1. Schematic diagram of the proposed Multi-Branch Feedforward Neural Network with Attention (MBFNN-Attn).

B. Proposed Model Architecture

The proposed Multi-Branch Feedforward Neural Network with Attention (MBFNN-Attn) architecture is illustrated in Figure 1. It consists of three main components designed to process multimodal data effectively.

1) Multi-Branch Feature Encoders

Six independent branch networks correspond to the six input feature groups. Each branch is a small feedforward network comprising two dense layers (16 and 8 neurons, respectively) with ReLU activation and L2 regularization (λ=0.005). A dropout layer (rate=0.4) follows the first dense layer in each branch. Each branch transforms its input features into an 8-dimensional representation vector, capturing group-specific patterns.

2) Attention-Based Fusion Layer

The six 8D representation vectors are fed into an attention mechanism. A single dense layer with tanh activation computes a scalar attention score for each branch. These scores are normalized across all branches using a softmax function to produce attention weights (α1, ..., α6) that sum to 1. Each branch's representation is then multiplied by its corresponding weight, and the weighted representations are concatenated to form a final 48-dimensional context vector.

3) Final Classifier

The context vector is passed through a final classification module: a dense layer (32 neurons, ReLU, L2 regularization), a dropout layer (rate=0.5), and a final output layer with a single neuron and sigmoid activation, producing a probability score for glaucoma.

4) Implementation and Training Details

The model was implemented using TensorFlow/Keras (v2.12) and Python 3.9. Training was performed on Google Colab using a GPU runtime. The model was compiled with the Adam optimizer (initial learning rate = 0.0005) and binary cross-entropy loss. Training proceeded for a maximum of 150 epochs with a batch size of 64, using early stopping (patience=20, monitoring validation loss) and `ReduceLROnPlateau` (patience=7, factor=0.5) callbacks to prevent overfitting and stabilize training.

5) Evaluation and Comparison

Model performance was evaluated on the held-out test set using standard metrics: Accuracy, Precision (Positive Predictive Value), Recall (Sensitivity), Specificity, F1-Score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and Area Under the Precision-Recall Curve (AUC-PR). A confusion matrix was analyzed to understand error types.

To benchmark performance, the proposed MBFNN-Attn was compared against three baseline models trained and tested on the same data split:

- **Single-Branch Feedforward NN (SFNN):** A monolithic network with similar total parameter count, receiving all 26 features as a single input vector.
- **Random Forest (RF):** An ensemble of 100 decision trees.
- **XGBoost (XGB):** A gradient boosting implementation with default parameters.

IV. RESULTS

A. Performance Comparison of Proposed and Baseline Models

The proposed MBFNN-Attn model demonstrated strong and clinically favorable performance on the independent test set, outperforming all baseline models. The comparative results are summarized in Table 1.

Table 1. Comparative performance analysis of the proposed MBFNN-Attn model against baseline models.

Model	Acc.	Pre.	Rec.	Spec.	F1	AUC-ROC
MBFNN-Attn (Proposed Method)	90.0%	100%	77.8%	100%	87.5%	0.838
Single-Branch FFNN	85.0%	87.5%	77.8%	90.9%	82.4%	0.858
Random Forest	85.0%	87.5%	77.8%	90.9%	82.4%	0.858
XGBoost	80.0%	77.8%	77.8%	81.8%	77.8%	0.828

The key strengths of the proposed model are its perfect precision (100%) and specificity (100%), indicating the complete absence of false positives in the test set. This signifies that every eye predicted as glaucomatous was indeed diseased, and all healthy eyes were correctly identified—a critical attribute for a reliable screening tool. While its recall (sensitivity) of 77.8% (identifying 7 out of 9 glaucomatous eyes) indicates room for improvement in catching all cases, the model achieved the highest overall accuracy (90.0%) and F1-score (87.5%), reflecting a superior balance between precision and recall. Although the AUC-ROC of the proposed model (0.838) was slightly lower than that of the Single-Branch FFNN and Random Forest (0.858), its perfect specificity and superior F1-score represent a more desirable clinical trade-off, prioritizing the avoidance of mislabeling healthy individuals.

B. Attention Weight Analysis

The integrated attention mechanism provides inherent interpretability by revealing the relative importance the model assigns to each feature group. Table 2 presents the learned attention weights, averaged over the test set.

Table 2. Learned attention weights for the six feature branches of the proposed model.

Feature Branch	Attention Weight (α)	Clinical Relevance
Total Retinal Thickness (TRT)	0.374	Indicator of pan-retinal neurodegeneration.
GCIPL	0.139	Contains ganglion cell bodies, a primary site of glaucomatous damage.
RNFL	0.128	Contains axons of ganglion cells, a key structural biomarker.
Clinical Parameters	0.122	Demographic and pressure-related risk factors.
Visual Field Indices	0.122	Functional measures of vision loss.
GCC	0.115	A composite (RNFL+GCIPL) measure.

Analysis of Table 2 reveals that the Total Retinal Thickness (TRT) branch was overwhelmingly the most influential, receiving a weight (~0.37) nearly three times greater than any other branch. This suggests the model identifies generalized retinal thinning as a highly discriminative feature for glaucoma in this dataset, which aligns with the understanding of glaucoma as a diffuse neurodegenerative process. The substantial weights assigned to the GCIPL and RNFL branches are consistent with their established pathophysiological importance as locations of primary neuronal injury. The comparable, moderate weights for clinical and visual field data affirm their supportive role. Interestingly, the GCC (a composite of RNFL and GCIPL) received the lowest weight, suggesting the model derives more value from processing these layers separately rather than as a pre-defined combination.

C. Confusion Matrix Analysis

A detailed breakdown of the classification outcomes for the proposed MBFNN-Attn model on the independent test set is provided by the confusion matrix in Table 3.

Table 3. Comparative performance analysis of the proposed MBFNN-Attn model against baseline models.

	Predicted: Healthy	Predicted: Glaucomatous
Actual: Healthy	11 (TN)	0 (FP)
Actual: Glaucomatous	2 (FN)	7 (TP)

As shown in Table 3, the model produced no false positives (FP = 0), directly contributing to its perfect specificity and precision (100%). All misclassifications were false negatives (FN = 2), accounting for the observed recall (sensitivity) of 77.8% (7 out of 9 glaucomatous eyes correctly identified). This analysis confirms the model's primary strength in reliably ruling out disease in healthy subjects (True Negatives, TN = 11) and correctly identifying glaucomatous cases (True Positives, TP = 7), while highlighting the specific area for future improvement: increasing sensitivity to capture the minority of glaucoma cases that were missed.

V. DISCUSSION

This study presented a novel multi-branch neural network with an attention mechanism for the automated detection of glaucoma using multimodal clinical and imaging data. The model's key innovation lies in its structured approach to processing heterogeneous data, which led to superior performance, particularly in eliminating false positive diagnoses.

The perfect specificity (100%) and precision (100%) of our model are its most clinically significant achievements. In a screening or triage context, a tool that never labels a healthy individual as diseased minimizes unnecessary anxiety, referrals, and costly follow-up tests, thereby increasing trust among clinicians [10]. While the sensitivity (77.8%) leaves room for improvement and suggests some early or subtle cases were missed, the overall F1-score (87.5%) reflects a robust balance. The slightly lower AUC-ROC compared to simpler models may stem from the attention mechanism's selective focus, potentially underweighting some complementary patterns that monolithic models use, but this trade-off favors clinical safety.

The attention weight analysis offers valuable interpretability, a crucial aspect for clinical adoption. The model's strong emphasis on Total Retinal Thickness (TRT) is insightful. While RNFL and GCIPL are established biomarkers, our model suggests that generalized retinal thinning (TRT) is a highly discriminative feature in this dataset, possibly capturing diffuse early damage. The lower weight for the GCC branch (a composite of RNFL and GCIPL) implies that the model benefits more from processing these layers separately. The comparable, moderate weights for clinical and visual field data affirm their supportive, but not dominant, role in this structural imaging-focused model.

Our work aligns with the trend towards multimodal and interpretable AI in medicine [11]. Compared to previous studies using single-mode CNNs on fundus images [5, 12] or OCT volumes [6], our model leverages a richer, albeit tabular, representation of the disease. The performance is competitive with recent literature (e.g., [12, 13]), while offering greater transparency through its architecture and attention weights.

Table 4. Comparative summary of the proposed method against selected previous studies for glaucoma detection.

Ref. (Year)	Method	Data Modality	Metric(s)	Strengths
Al-Bander et al. (2017) [5]	CNN	Fundus Images	Acc: 88.2%, Sens: 85%	Early demonstration of automated feature learning from fundus images.
Manassakorn et al. (2022) [12]	Custom CNN (GlauNet)	OCT Angiography	Sens: 88.9%, Spec: 89.6%, AUC: 0.89	Focus on OCTA, robust to noise.
Gaddipati et al. (2019) [14]	Capsule Network	3D OCT Volumes	AUC: 0.97	Effective at learning spatial hierarchies from raw volumes.
Garcia et al. (2021) [6]	CNN + LSTM	3D SD-OCT Volumes	AUC: 0.885	Models spatio-temporal dependencies in OCT scans.
Wassel et al. (2022) [13]	Vision Transformer (ViT)	Fundus Images	Sens: 92.6%, Spec: 96.9%, AUC: 0.979	Leverages global attention on large-scale fundus datasets.
Proposed Method	Multi-Branch FFNN + Attention	Multimodal Tabular (Clinical, VF, OCT layers)	Acc: 90%, Prec: 100%, Spec: 100%, F1: 87.5%, AUC: 0.838	Interpretable, structure-aware processing of multimodal features. Perfect specificity.

A. Comparison with State-of-the-Art Methods

To contextualize our results within the broader field, we compare the performance of our MBFNN-Attn model with a selection of recent and notable studies in Table 4. It is important to note that direct numerical comparison is constrained by differences in datasets, sizes, and evaluation protocols.

As illustrated in Table 4, our model achieves a unique profile of performance. While studies like Wassel et al. [13] report higher sensitivity and AUC on large fundus image sets, and Gaddipati et al. [14] achieve a higher AUC on 3D OCT data, our model's standout achievement is its perfect precision and specificity (100%). This performance is attained not on raw images, but on an efficiently processed set of multimodal tabular data, making it computationally leaner. Furthermore, unlike "black-box" image-based models, our architecture provides explicit interpretability through attention weights, clarifying the contribution of different data types (e.g., highlighting total retinal thickness). Therefore, our work complements the existing literature by offering a highly reliable, interpretable, and efficient model suitable for scenarios where avoiding false alarms is paramount and where multimodal patient data beyond a single image is available.

The proposed model with its multi-branch architecture and attention mechanism, in addition to accurate performance, has the necessary features for clinical application: (1) use of structured clinical routine data, (2) high computational efficiency due to the processing of extracted features, and (3) inherent interpretability. Its 100% accuracy and specificity are particularly suitable for the screening environment where avoiding unnecessary referrals is a priority.

B. Limitations and Future Work

This study has limitations. The dataset, though high-quality, is relatively small and from a single center, which may affect generalizability. The model currently does not process raw fundus or OCT images, limiting its "multimodal" scope to extracted features. Future work will focus on: 1) Validating the model on larger, multi-ethnic datasets, 2) Developing a truly end-to-end multimodal network that jointly processes raw SS-OCT volumes and fundus images with the clinical data, and 3) Employing advanced techniques like self-supervised learning to improve feature representation and boost sensitivity, especially for early-stage glaucoma detection.

Also, the present study was conducted on a single-center dataset and the comparisons presented are with other research studies. To more accurately assess clinical performance, it is essential to evaluate the model in direct comparison with expert judgment on multicenter data. In future work, a multicenter validation study on independent data and comparing the model performance with ophthalmologist judgments as a real-world benchmark are planned. These evaluations, together with a cost-effectiveness analysis, will determine the path for operational deployment of the model.

VI. CONCLUSION

In this study, we introduced and evaluated a novel multi-branch feedforward neural network integrated with an attention mechanism (MBFNN-Attn) for automated glaucoma detection. The core innovation of our approach lies in its structured, biologically-inspired architecture, which processes six distinct groups of multimodal clinical and imaging-derived features through dedicated parallel branches. This design, coupled with an attention-based fusion layer, allows the model to dynamically weigh the contribution of each feature group, moving beyond the monolithic processing common in prior works.

The proposed model demonstrated clinically significant performance, achieving perfect precision and specificity (100%) on the independent test set, thereby eliminating false positives, a critical attribute for a trustworthy screening tool. It also attained the highest accuracy (90.0%) and F1-score (87.5%) among the benchmarked models. The integrated attention mechanism provided a layer of interpretability, revealing that generalized retinal thinning (Total Retinal Thickness) was the most influential feature, followed by established biomarkers like the GCIPL and RNFL. This finding offers valuable, data-driven insight into feature importance for glaucoma diagnosis.

Our work presents a robust, interpretable, and efficient framework for glaucoma assessment using readily available tabular data. It underscores the advantage of modality-aware neural architectures over conventional monolithic models. By prioritizing reliability (zero false alarms) and transparency, this model represents a tangible step towards deployable AI-assisted decision support systems in ophthalmology, with the potential to enhance screening efficiency, reduce unnecessary referrals, and improve patient management. Future efforts will focus on integrating raw image data and validating the framework on larger, multi-centric cohorts to further bolster its generalizability and sensitivity.

REFERENCES:

[1] *Tham, Y. C., et al. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040. Ophthalmology.*
[2] *Weinreb, R. N., et al. (2014). The pathophysiology and treatment of glaucoma. JAMA.*
[3] *Jonas, J. B., et al. (2017). Glaucoma. The Lancet.*
[4] *Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis.*
[5] *Al-Bander, B., et al. (2017). Automated glaucoma diagnosis using deep learning approach. 14th International Multi-Conference on Systems, Signals & Devices.*
[6] *Garcia, G., et al. (2021). Glaucoma detection from raw SD-OCT volumes: A novel approach focused on spatial dependencies. Computer Methods and Programs in Biomedicine.*
[7] *Wang, X., et al. (2022). Multimodal deep learning for biomedical data fusion. Nature Biotechnology.*
[8] *Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems.*
[9] *Lee, S. Y., et al. (2016). Data from: Asymmetry analysis of macular inner retinal layers for glaucoma diagnosis: Swept-source optical coherence tomography study. Figshare. https://doi.org/10.1371/journal.pone.0164866.s001*
[10] *Vollmer, S., et al. (2020). Machine learning and AI research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ.*

[11] *Reyes, M., et al. (2020). On the interpretability of artificial intelligence in radiology. Radiology: Artificial Intelligence.*

[12] *Manassakorn, A., et al. (2022). GlauNet: Glaucoma diagnosis for OCTA imaging using a new CNN architecture. IEEE Access.*

[13] *Wassel, M., et al. (2022). Vision transformers based classification for glaucomatous eye condition. 26th International Conference on Pattern Recognition (ICPR).*

[14] *Gaddipati, D. J., Desai, A., Sivaswamy, J., & Vermeer, K. A. (2019). Glaucoma assessment from oct images using capsule network.* 41st Annual International Conference of the IEEE EMBC.

[15] *Shinde, R. (2021). Glaucoma detection in retinal fundus images using U-Net and supervised machine learning algorithms.* Intelligence-Based Medicine.

[16] *Bisneto, T. R. V., de Carvalho Filho, A. O., & Magalhaes, D. M. V. (2020). Generative Adversarial network and texture features applied to automatic glaucoma detection.* Applied Soft Computing.