International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 79-85, May-2026

https://doi.org/10.47001/IRJIET/2026.105011

Glaucoma Detection from Fundus and OCT Images Using Attention-Based CNN and Transfer Learning Architectures

[1]Kande Srija, [2]Bale Bhavani, [3]Tummati Swapna, [4]Mada Sharanya, [5]Sheelam Sushma Reddy

[1,2,3,4,5]Department of Computer Science Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

E-mail: [1]srija.k0412@gmail.com, [2]bhavani.bale222@gmail.com, [3]t.swapna@gnits.ac.in, [4]madashraranya12@gmail.com, [5]sheelamsushmareddy@gmail.com

***Abstract* - Glaucoma is a chronic eye disease that is leading cause of irreversible vision loss worldwide. Early and accurate classification of glaucoma is crucial for timely intervention and effective management. This study is undertaken to automate the classification of glaucoma using the dataset that contain fundus images and OCT images. The proposed work targets on deep learning-based architectures for glaucoma classification: CNN-based transfer learning models and CNN with an Attention mechanism. The transfer learning models leverage pretrained networks for efficient feature extraction, while the attention-based CNN enhances focus on glaucoma-specific regions. A 2D CNN combined with attention layers to enhances the capacity of the model to attend to the key areas of medical images. ResNet-18 is a type of deep learning model where every block in the network has a skip connection that adds the input back to the output. DenseNet121 is a deep convolutional neural network that improves feature reuse and gradient flow by introducing dense connections between layers. The models DenseNet121, ResNet18 and Attention based 2D CNN are trained separately on fundus and OCT data to classify whether an eye is affected by glaucoma. By comparing the accuracies and performance metrics of all the trained models, we determine which model and image type is more suitable for glaucoma classification. This helps in identifying the most effective imaging method for future AI-based diagnostic tools.**

Keywords: Glaucoma detection, Deep learning, Fundus images, Optical Coherence Tomography (OCT), Convolutional Neural Network (CNN), Transfer learning, Attention mechanism.

I. INTRODUCTION

Glaucoma is a chronic and progressive eye condition that causes damage to the optic nerve and often leads to irreversible vision loss. Glaucoma is typically progressive and asymptomatic and early detection is important. Medical imaging is significant for diagnosis of glaucoma, namely fundus photography and Optical Coherence Tomography (OCT) imaging. Fundus images can reveal structural changes, such as cupping of the optic disc, as well as alteration of the cup-to-disc ratio, which are known early indicators of glaucoma. Contrarily, OCT imaging provides high resolution cross-sectional images of the retina, which enables an accurate measurement of regional retinal nerve fiber layer (RNFL) thickness. Thinning of the RNFL is a primary indicator of optic nerve damage associated with glaucoma.

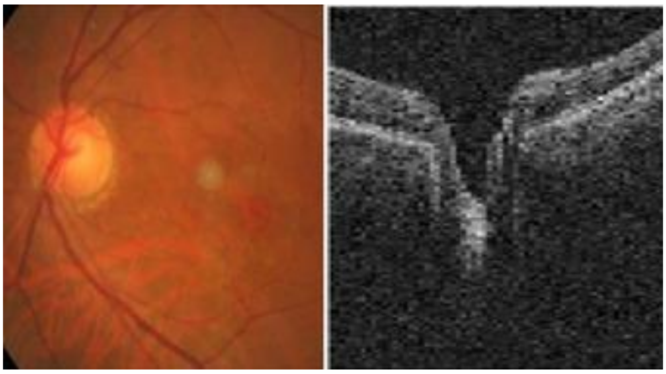

Figure 1: Fundus and OCT image of normal eye

Deep learning is becoming increasingly significant and more importantly, artificial intelligence has evolved to a level where it can automate the detection of glaucoma based on various medical images. Application of Fundus and OCT images together provides a thorough representation of retinal health and can lead to precision diagnostics. This project will use transfer learning and attention-based CNN models to classify glaucoma using fundus and OCT images in combination. Recent glaucoma detection protocols typically employ a single modality input, either fundus photographs or OCT scans. This one-sided approach fails to recognize complementary information from either modality; fundus photographs can identify structural changes such as optic disc cupping, while OCT can view more fine-grained structural detail, like RNFL thinning. Finally, there has yet to be a study comparing which modality, or potential combination of modalities, is more reliant as there has been a lack of comparative studies between these types of systems.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 79-85, May-2026
https://doi.org/10.47001/IRJIET/2026.105011

II. PROBLEM STATEMENT

Glaucoma is a leading cause of irreversible blindness, where early and accurate detection is critical to prevent vision loss. There is a need for a comprehensive and reliable method that can effectively utilize both imaging modalities. Using attention-enabled CNNs combined with transfer learning techniques has the potential to improve a model's ability to pay attention to clinically relevant areas of images, as well as its effectiveness at classifying them accurately. The analysis of fundus and OCT images order to create a strong and viable glaucoma detection framework. Conventional methods for diagnosing glaucoma have historically used a subjective and independent assessment of fundus or OCT images.

Unfortunately, the lack of automation leads to inconsistent results and limits the amount of diagnostic information available to clinicians for making treatment decisions. Consequently, a multi-modal and automated glaucoma diagnostic system is needed to ensure timely intervention and prevent irreversible vision loss. The present study will utilize a deep learning methodology using convolutional neural networks with attention mechanisms and transfer learning techniques for classifying glaucoma. Furthermore, attention mechanisms will allow the network to focus only on clinically significant areas and improve both feature extraction and classification features. This research provides contributions to:

- Integrating fundus imaging and OCT imaging modalities for complementary structural information.
- Using attention-based CNN architects to focus on diagnostically relevant retinal regions, increasing model interpretability.
- Applying transfer learning to pre-trained deep learning architectures, thus reducing time spent training models while improving accuracy when working with smaller medical image datasets.
- Developing an automated glaucoma detection system to give ophthalmologists an efficient and reliable tool for early detection of disease and better supporting focused clinical decision making.

III. LITERATURE REVIEW

Convolutional neural networks (CNNs) have emerged as a powerful technique for glaucoma detection using retinal images. Some studies show that CNN models are able to learn structural patterns from fundus images to detect glaucoma at an early stage [1], while others integrate feature extraction and classification methods to enhance detection accuracy [2]. Transfer learning has also been crucial, where models are fine-tuned for glaucoma detection, improving accuracy with smaller datasets [4], [5].

Moreover, tailored models like Deep-GlaucomaNet also improve detection accuracy by specifically designing models for retinal images [8]. Beyond fundus images, other imaging techniques like Optical Coherence Tomography (OCT) have been used to provide depth-related information for detecting certain forms of glaucoma [3]. Benchmarking studies show that some models perform better than others on different datasets, underscoring the need for careful model design and evaluation [9].

Additionally, assessment of several transfer learning models demonstrate their feasibility for clinical use, but results may be influenced by dataset features and preprocessing methods [6]. In summary, current literature suggests that deep learning has made substantial advances in the detection of glaucoma, with both fundus and OCT-based methods [7]. But there remain issues with data variability, model transferability and interpretability. Overall, while current approaches are promising, there is a need for more reliable models that can be applied in the clinical setting.

IV. METHODOLOGY

The method used for completing the project includes an organized systematic pipeline that consists of a series of steps that begin with pre-processing image data to allow for model training. When looking at medical imaging data sets of fundus and/or OCT types, there are typically multiple variations of illumination, noise, and low contrast that can be seen with any of the images in the entire data set. Therefore, applying various enhancement image techniques to improve image quality and make important structures within the retina more visible The Contrast Limited Adaptive Histogram Equalization (CLAHE), median filtering was also applied to suppress the amount of noise in the image. Lastly, using a bilateral filter to suppress noise within an image will also allow for the preservation of the majority of edges and structural information.

In the next step, deep learning architectures are created using transfer learning and an attention-based CNN approach. The CNN for fundus and OCT image datasets will be fine-tuned independently against pre-trained architectures such as ResNet, DenseNet, and an attention-based 2D CNN. As both long-established models were trained originally on very large datasets such as ImageNet, they both have strong image feature extraction characteristics that can also be leveraged for medical image analysis.

The overall image data flow for the proposed system is shown in Figure 2.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 79-85, May-2026
https://doi.org/10.47001/IRJIET/2026.105011

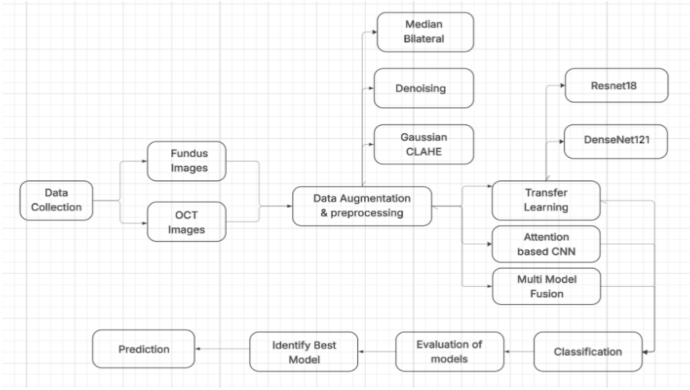

Figure 2: Visualization of the proposed model

V. IMPLEMENTATION

A fully automated glaucoma detection system will utilize fundus (retinal) images and OCT images to detect glaucoma. The proposed system will begin by creating the image database fundus and OCT images will be divided into two groups for training data and test data and the database will contain both normal fundus with normal OCT images and fundus with abnormal OCT images containing evidence of glaucoma. After all images have been preprocessed, attention-based CNNs, DenseNet121 and ResNet18 are trained to learn and identify attributes of images where glaucoma is present. By combining all of these processes into one system architecture, the automated glaucoma detection system will provide a reliable tool for ophthalmologists and assist in providing bar evidence of the presence of glaucoma.

5.1 Image preprocessing

Data preprocessing is necessary to analyze medical images because it processes raw medical images before analysis by removing noise and making critical structures visible. For this study, all images were first normalized to the same intensity range in order to minimize the differences resulting from variations in imaging conditions.

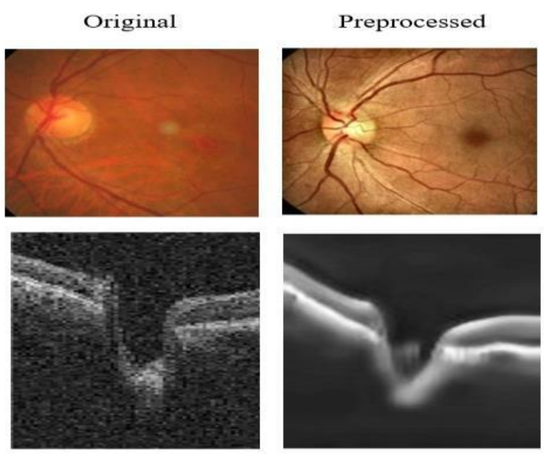

Figure 3: Original and Preprocessed images

Images were subsequently processed with a Bilateral Filter to smooth the uniform portions of the image and preserve sharp edges around structural features such as the optic disc and retinal blood vessels. Using CLAHE, local contrast was elevated without adding any unwanted artifacts to each image. All of these image-preprocessing steps improve the quality of each image and allow improved visibility of retinal structures, thereby making the images better candidates for reliable feature extraction and deep learning-based determination of the presence of glaucoma.

5.2 Attention based 2D-CNN architecture

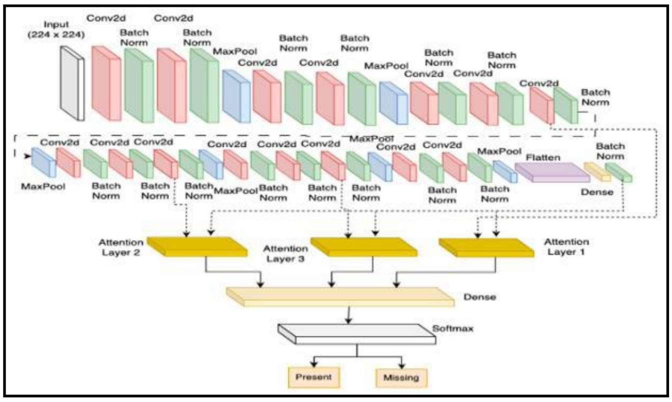

Figure 4: Attention based 2D-CNN Architecture

An attention based 2D CNN is designed to enhance feature extraction through increased attention on the informative areas of an image. The model architecture is a conventional CNN workflow consisting of convolutional and pooling layers to extract hierarchical spatial features from 2D imagery. Also employed are attention mechanisms within the neural network to ensure the model places emphasis on clinically relevant areas of the image while reducing the influence of non-meaningful background information on the model's performance. Focusing on important retinal structures such as the optic disc and retinal nerve fibres, which are markers of glaucoma, the model will be able to identify normal vs glaucomatous images based on their respective features. As such, by focusing attention on the most pertinent areas of interest, an attention based 2D CNN will improve the quality and accuracy of medical imaging classification tasks relative to other non-attention based approaches.

5.3 ResNet-18 architecture

The ResNet-18 model uses deep CNN architecture to solve the degradation problem typically seen in very deep neural networks. The ResNet-18 model has a total of 18 layers consisting of residual learning blocks which use shortcut (also called skip) connections. In this type of network, the input to the layer is sent directly to output of the layer, thereby allowing the network to learn a residual mapping rather than

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 79-85, May-2026

https://doi.org/10.47001/IRJIET/2026.105011

learning a total transformation. ResNet-18 is composed of an architecture that allows for improved gradient flow through backpropagation in order to provide greater ease of training when it comes time to train larger networks.

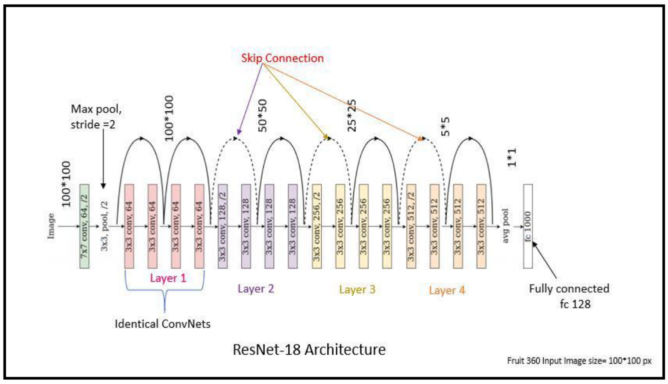

Figure 5: ResNet-18 Architecture

Additionally, with the assistance of ResNet-18's ability to learn hierarchical representation of the features of an image without sacrificing too much in terms of the amount of computational power needed. First, ResNet-18 has proven to be an excellent source of reliable and accurate representations of images on which to perform classification based on pre-trained weights well as help expedite the time it takes to learn to classify images for newly created locations that utilize medical imaging. Therefore, ResNet-18 is ideal for use in real-time clinical applications due to its efficiency, adaptability, and significant feature extractor capabilities.

5.4 DenseNet-121 architecture

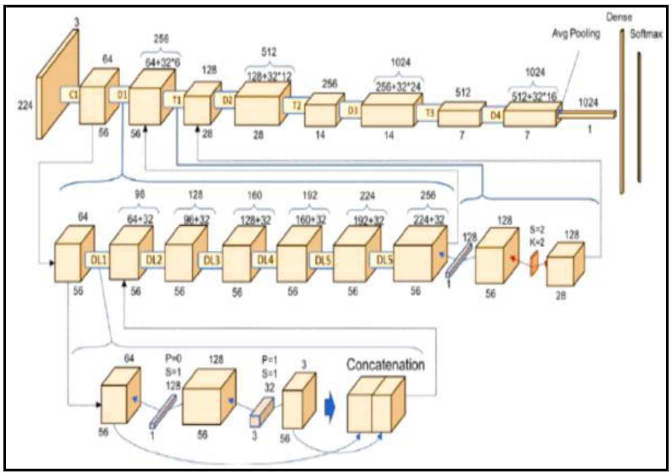

Figure 6: DenseNet-121 Architecture

The goal of DenseNet121 is to improve feature propagation throughout the network, by allowing all layers to connect to all previous layers. By connecting all layers in this way, DenseNet121 is able to help reduce issues that arise when training deep learning models, such as vanishing gradients, and allows for both low and high-level feature representations to be learned from the images. The architecture of DenseNet121 consists of 121 layers, which are divided into dense blocks separated by transition locations that reduce the size of the feature maps and limit the complexity of the model. While still having an ability to represent representations, this architecture provides a way for gradients to pass more easily through the model to earlier layers, which provides a form of implicit deep supervision and increases the stability of the learning process. Using features from different depths of the model, DenseNet121 can obtain different and richer representations for image classification tasks.

5.5 Training and testing

Datasets were created for the training and evaluation of attention-based CNNs in order to compare images of fundus and OCT showing both healthy and glaucomatous retinas. To allow the model's performance to be fairly assessed, the datasets were divided into training, evaluation, and testing datasets. The training datasets were used to train attention-based CNNs and ResNet-18 and DenseNet-121 models by utilizing multiple epochs with training dataset images displaying features related to glaucomatous retinal structures. Additionally, during the training process, the models progressively converged upon optimal parameters designed to identify distinguishing features between healthy retinal structures and those with glaucomatous alterations. The evaluation dataset was used during training to monitor model performance and minimize overfitting. Once trained, each attention-based CNN, ResNet-18, and DenseNet-121 were tested with previously unseen fundus images, OCT images, and combined retina images to evaluate their generalization performance capability. Performance metrics were calculated whereby predicted outputs were compared against the ground truth labels providing an objective means of measuring classification accuracy of each class of glaucoma. Ultimately, the model would provide the best detection performance will be selected as most promising for reliable glaucoma classification and therefore could be utilized in future evidence-based clinical practice.

VI. EVALUATION METRICS

The evaluation metrics will be used to measure how well the model performs on segmenting diabetic retinopathy. This will include the use of Accuracy, Recall, Precision, and F1 score to evaluate the model's performance. Below are the formulas for each of these evaluation measures:

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 79-85, May-2026

https://doi.org/10.47001/IRJIET/2026.105011

- $\text{Accuracy} = \dfrac{TP + TN}{FP + FN + TP + TN}$
- $\text{Precision} = \dfrac{TP}{TP + FP}$
- Recall (Sensitivity or True Positive Rate):

  $\text{Recall} = \dfrac{TP}{TP + FN}$
- $\text{F1\_Score} = \dfrac{2 \text{ X Precision X Recall}}{\text{Precision} + \text{Recall}}$

Use of these metrics in relation to obtained and true labels enables evaluation of 'model performance' regarding Glaucoma classification.

VII. RESULTS & DISCUSSION

7.1 Qualitative analysis

The qualitative analysis involved visualizing the classified images and their predicted classes generated by the trained models for classification. For every new image from the dataset, the classification model produced classification labels. And then, the best model is applied to new dataset taken from Mendeley consisting of respective fundus and oct image of eye.

The predicted and actual classification of test images from datasets and new test images are shown below.

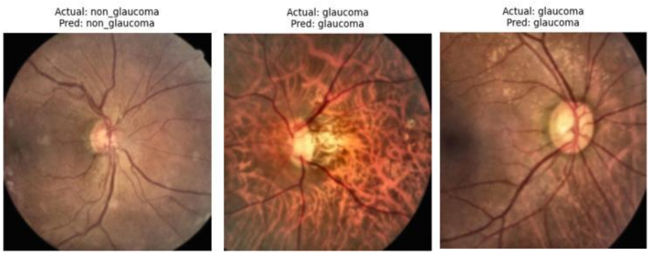

Figure 7: Predicted vs Actual classification using Fundus images

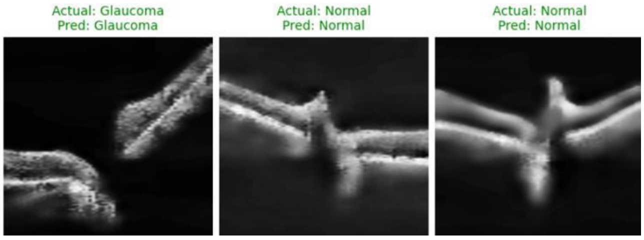

Figure 8: Predicted vs Actual classification using OCT images

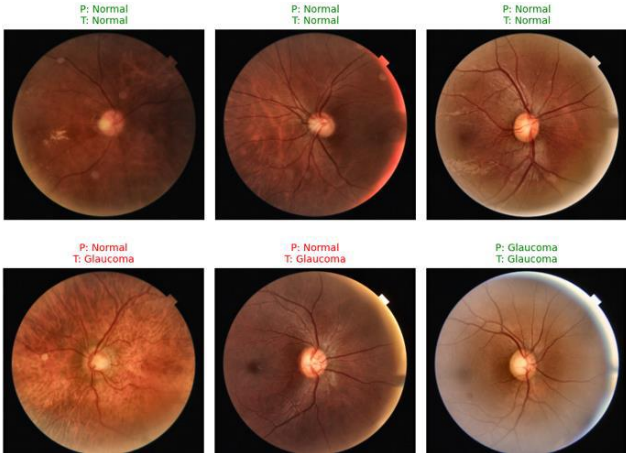

Figure 9: Predicted vs Actual classification using test Fundus dataset

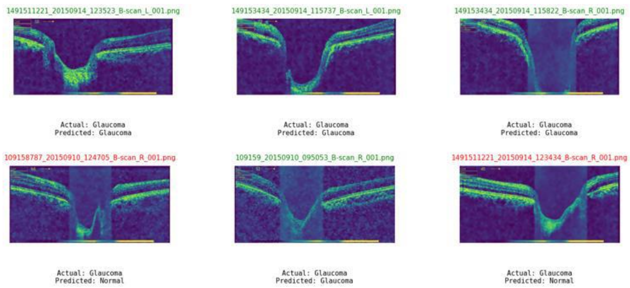

Figure 10: Predicted vs Actual classification using test OCT dataset

7.2 Quantitative analysis

The quantitative analysis involved computing performance metrics such as accuracy, precision, recall, and F1-score to objectively evaluate the model performance. This compares the performance metrics of a proposed classification model with established pre-trained models Attention based 2D CNN, ResNet-18, DenseNet 121 for glaucoma classification as shown in the table 1.

Table 1: Comparison of evaluation metrics for Classification model using different deep learning techniques

	Fundus			OCT		
	Attention based 2D-CNN	ResNet18	DenseNet121	Attention based 2D-CNN	ResNet18	DenseNet121
Accuracy	0.9443	0.9670	0.9753	0.9029	0.9145	0.9261
Precision	0.9201	0.9632	0.9675	0.8838	0.9116	0.9254
Recall	0.9732	0.9711	0.9835	0.98	0.9622	0.9644
F1 Score	0.9459	0.9671	0.9755	0.9294	0.9362	0.9445
Specificity	0.9155	0.9711	0.9670	0.98	0.9622	0.8542
Sensitivity	0.972	0.9629	0.9834	0.7583	0.825	0.9644
ROC Curve Area	0.989	0.996	0.996	0.949	0.963	0.951

The comparison of the three models, Attention-based 2D-CNN, ResNet18, and DenseNet121 was performed on two different datasets for Fundus and OCT images as in Table 1, to identify the best architecture. On the Fundus dataset, the

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 79-85, May-2026
https://doi.org/10.47001/IRJIET/2026.105011

DenseNet121 model performed great, delivering the best performance on all metrics tested. It had a top-performing Accuracy of 0.9753, an F1 Score of 0.9755, and a Recall of 0.9835. Followed by ResNet18 and Attention-based 2D-CNN. On the OCT dataset, DenseNet121 excelled in most of the key parameters with an Accuracy of 0.9261, a Precision of 0.9254, and the highest F1 Score of 0.9445. ResNet18 performed equally well, falling in the middle among the other two models with an accuracy of 0.9145 and an F1 Score of 0.9362. Lastly, on overall performance across the two datasets, DenseNet121 is the overall winner. It had the best performance in the most critical metrics such as accuracy, F1 score, and recall across the board. Its continued dominance, especially its highest F1 scores of 0.9755 on Fundus and 0.9445 on OCT, also to its consistency and robustness it stood as the best and most well-rounded model among the three for the given tasks.

Attention-based 2D CNN is highly accurate for both image categories, correctly classifying most normal and abnormal cases. In general, all three models exhibit high reliability with marginal performance differences between image modalities, reflecting their robust ability in glaucoma classification.

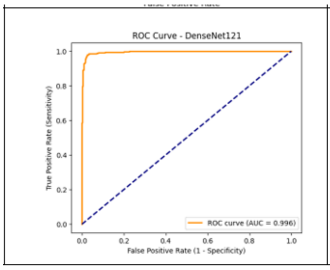

Figure 11: DenseNet121 ROC curve for Fundus images

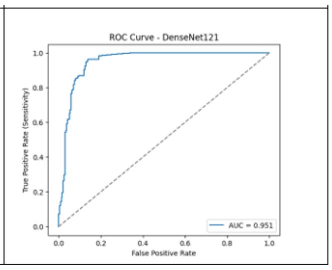

Figure 12: DesnseNet121 ROC curve for OCT images

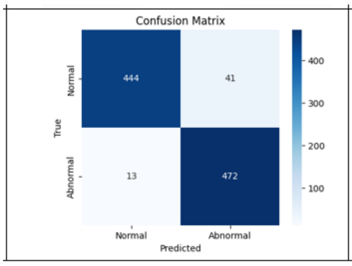

Figure 13 : Confusion Matrix of DenseNet 121 for Fundus images

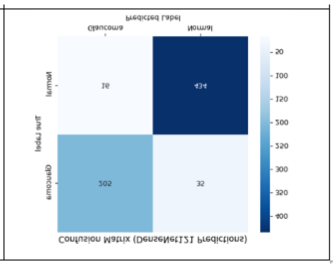

Figure 14 : Confusion Matrix of DenseNet 121 for OCT images

VIII. DISCUSSION

The results show that deep learning models are effective for glaucoma detection using fundus and OCT images. Models such as DenseNet121, ResNet-18, and attention-based CNN were able to learn important features from the images and perform reliable classification. DenseNet121 performed better due to its ability to reuse features and maintain strong gradient flow. The attention-based CNN helped the model focus on important regions, especially in fundus images. Fundus images provided useful information about optic disc changes, while OCT images gave detailed structural information of the retina. However, the performance of these models depends on the quality and size of the dataset. Overall, the study shows that deep learning can support automated glaucoma detection using individual imaging modalities.

IX. CONCLUSION

Deep learning techniques have been highly effective at detecting glaucoma using fundus and OCT images. Transfer learning and attention-based CNN models were tested on separate imaging modalities, thus allowing us to demonstrate the advantages of using each of the imaging modalities. Fundus images provided useful information about the structure of the optic disc while providing detailed structural information about the retina via OCT images to assist with accurate glaucoma detection. AI algorithms provide clinicians with the reliability and accuracy needed to diagnose glaucoma. The models are automated to assess medical image databases and can assist ophthalmologists with automated assessment of medical image databases for early detection and treatment of glaucoma. The results indicate that deep learning has the potential to be an effective technique in the analysis of medical images and is a positive advancement toward developing more efficient and simpler glaucoma screening systems. Therefore, this implementation supports the integration of multimodal data processing, training, and validation of models into a single glaucoma detection system, therefore providing clinicians with the ability to make accurate and reliable diagnoses of early-stage glaucoma as they make decisions on how to treat patients clinically.

REFERENCES

[1] A Kaur, “A Deep Learning Approach to Glaucoma: CNN Role in Early Detection,” *in Proc. 2nd Int. Conf. Computational and Characterization Techniques in Engineering & Sciences (IC3TES),* Lucknow, India, 2024.

[2] A Kukker, G. Pandey, and S. Hudnurkar, “Fundus Image-Based Glaucoma Classification Using Feature Extraction and Classification Techniques,” in *Proc. Int. Conf. Sustainable Communication Networks and Application (ICSCNA),* Theni, India, 2024.

[3] F. Teke, O. Kaynar, and Y. Görmez, “Detection and Classification of Closed Angle Glaucoma Using Optical Coherence Tomography Images,” *in Proc. Innovations in Intelligent Systems and Applications Conf. (ASYU),* Sivas, Turkiye, 2023.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 79-85, May-2026

https://doi.org/10.47001/IRJIET/2026.105011

[4] J. J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. I. Sánchez, and M. J. Ledesma-Carbayo, "Automatic Glaucoma Classification Using Color Fundus Images Based on Convolutional Neural Networks and Transfer Learning," *Biomed. Opt. Express*, vol. 10, pp. 892–913, 2019.

[5] M. F. Karim, M. A. Hossain, A. Y. Srizon, and N. Haque, "Early Detection of Glaucoma from Cropped Fundus Images Using Transfer-Learned Convolutional Neural Network," *in Proc. IEEE Int. Conf. Power, Electrical, Electronics and Industrial Applications (PEEIACON), Rajshahi, Bangladesh*, 2024.

[6] M. Phankokkruad, "Evaluation of Deep Transfer Learning Models in Glaucoma Detection for Clinical Application," in *Proc. 4th Int. Conf. Information and Communications Technology (ICOIACT),* Yogyakarta, Indonesia, 2021.

[7] R. Kiefer, J. Steen, M. Abid, M. R. Ardali, and E. Amjadian, "A Survey of Glaucoma Detection Algorithms Using Fundus and OCT Images," *in Proc. IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver*, BC, Canada, 2022.

[8] S. K. Sharma and D. Muduli, "Deep-GlaucomaNet: A Deep Learning-Based Approach for Glaucoma Detection in Fundus Images," *in Proc. 14th Int. Conf. Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023..

[9] T. Swapna, Y. S. R. Varshitha, K. L. Sudeepthi, B. Manavika, and T. Saishree, "A Comparative Study of Deep Learning Algorithms for Glaucoma Classification Using Retinal Images," *Lecture Notes in Electrical Engineering, Springer*.

[10] https://www.kaggle.com/datasets/jeremypoveda/fondo-del-ojo-normal-glaucoma-retinopatia

[11] https://zenodo.org/records/1481223

[12] https://data.mendeley.com/datasets/2rnnz5nz74/2

Citation of this Article:

Kande Srija, Bale Bhavani, Tummati Swapna, Mada Sharanya, & Sheelam Sushma Reddy. (2026). Glaucoma Detection from Fundus and OCT Images Using Attention-Based CNN and Transfer Learning Architectures. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 79-85. Article DOI https://doi.org/10.47001/IRJIET/2026.105011

****