

Hybrid Attention-Based Deep Learning Models for Glaucoma Detection in Fundus Images Using EfficientNet and ResNet Architectures

Kiran Kumar Pamidi¹ and Siva Rama Krishna SarmaVeerubhotla²

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, Guntur(D) ; 2401050152cse@gmail.com

² Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, Guntur(D).; sharmavsrk@kluniversity.in

* Correspondence: 2401050152cse@gmail.com

Abstract: Glaucoma is a progressive and irreversible optic neuropathy that is one of the leading causes of blindness worldwide. Early diagnosis is essential to prevent permanent vision loss; however, conventional fundus image analysis is time-consuming, requires expert ophthalmologists, and is subject to inter-observer variability. This study aims to develop and compare two hybrid attention-based deep learning models for accurate and automated glaucoma detection from retinal fundus images. The first model enhances EfficientNetB0 by incorporating an additional Squeeze-and-Excitation (SE) attention module to improve channel-wise feature representation, while the second model integrates a Convolutional Block Attention Module (CBAM) with ResNet50 to jointly learn channel-wise and spatial attention for improved feature extraction. Both models were trained and evaluated using a clinically curated retinal fundus image dataset comprising 704 images of normal and glaucomatous eyes. Experimental results demonstrated that both models achieved high diagnostic performance. The EfficientNetB0 + SE model achieved perfect classification performance on the evaluation dataset, with 100% accuracy, precision, recall, F1-score, and AUC, whereas the ResNet50 + CBAM model achieved 98% accuracy with perfect recall, indicating its effectiveness for glaucoma screening applications. The findings indicate that incorporating attention mechanisms enhances feature representation and improves glaucoma classification performance. Among the two models, EfficientNetB0 + SE demonstrated the best overall performance, making it a promising approach for real-time AI-assisted glaucoma screening.

Keywords: Glaucoma detection, retinal fundus images, deep learning, EfficientNetB0, ResNet50, Squeeze-and-Excitation, Convolutional Block Attention Module, medical image classification.

How to cite this article: Kiran Kumar Pamidi¹ and Siva Rama Krishna SarmaVeerubhotla². Hybrid Attention-Based Deep Learning Models for Glaucoma Detection in Fundus Images Using EfficientNet and ResNet Architectures. Int J Drug Deliv Technol. 2026;16(83s):512-520. DOI: 10.25258/ijddt.16.83s.57.

1. Introduction

Glaucoma is a progressive and irreversible optic neuropathy that gradually damages the optic nerve, leading to permanent vision loss if left untreated. It is one of the leading causes of irreversible blindness worldwide, and its asymptomatic progression during the early stages makes timely diagnosis particularly challenging. Clinical diagnosis is commonly performed through retinal fundus image examination, where ophthalmologists assess structural changes such as the optic disc and cup-to-disc ratio. However, this manual assessment requires considerable clinical expertise and is subject to inter-observer variability, making automated computer-aided diagnosis an attractive alternative.

Recent advances in deep learning have significantly improved the automated analysis of retinal fundus images for glaucoma detection. In particular, convolutional neural networks (CNNs) have demonstrated remarkable performance in a wide range of medical image classification tasks by automatically learning discriminative features directly from image data. Despite these advances, many existing CNN-based models are computationally intensive and may not be suitable

for real-time or resource-constrained clinical environments. Furthermore, conventional CNN architectures often fail to adequately emphasize clinically important regions, such as the optic disc and optic cup, where glaucomatous changes primarily occur.

In order to overcome these drawbacks, this paper proposes to combine attention mechanisms with light architectures of deep learning networks to achieve automatic glaucoma detection. Two hybrid models are designed and tested. The first is a deep fusion model of EfficientNetB0 and a new squeeze and excitation attention module; the second is a deep fusion model of a convolutional block attention module (CBAM) and ResNet50, which can adapt to the channel-wise and spatial attention. The proposed models make use of these attention mechanisms to enhance feature discrimination with an efficient enough computational cost for clinical screening applications.

While there have been some deep learning methods with promising results on glaucoma detection, studies have not often focused on improving the classification accuracy while concurrently tackling the challenge of computational efficiency and

effective feature representation. Moreover, comparisons of lightweight attention based CNN models are still limited, especially for clinically curated retinal fundus image datasets. It is therefore essential to explore if it is possible to combine attention mechanism with efficient backbone networks to boost diagnostic accuracy while ensuring applicability for clinical use.

The main contributions of this study are as follows:

1. Development of a hybrid EfficientNetB0 model incorporating an additional Squeeze-and-Excitation (SE) attention module for enhanced channel-wise feature learning.
2. Development of a ResNet50-based hybrid architecture integrated with the Convolutional Block Attention Module (CBAM) to improve both channel and spatial feature representation.
3. Comprehensive comparative evaluation of both hybrid architectures using a clinically curated retinal fundus image dataset based on multiple classification metrics, including accuracy, precision, recall, F1-score, and AUC.
4. Assessment of lightweight attention-enhanced CNN architectures for their potential use in AI-assisted glaucoma screening applications.

2. Literature Review

Deep learning has contributed greatly to the automated detection of glaucoma from retinal fundus images, as it can learn end-to-end features without the need to come up with handcrafted image descriptors. Initial research showed that Convolutional Neural Networks (CNNs) are superior to traditional machine learning techniques using manually extracted features like the cup-to-disc ratio and optic nerve morphology. This demonstrated the usefulness of deep learning for computer-aided glaucoma screening, as the CNN-based models performed better in terms of diagnostic accuracy, automatically learning discriminative retinal features, as shown by Raghavendra et al. (2018).

The research was then expanded to enhance the diagnostic performance using anatomical information of the optic nerve head. In the study, Fu et al. (2018) have suggested a disc aware ensemble network by integrating the optic disc information with the overall retinal information to enhance the classification of glaucoma. This helped to better identify minor structural changes linked to glaucoma. However, these approaches were subject to localizing the optic disc with accuracy and were based on relatively complicated network structures, which might be difficult to implement in real-time clinical settings.

Recently, the attention mechanisms have become a promising approach to enhance feature representation in medical image analysis. In the recent past, Li et al. (2019) proposed attention based

CNN models that highlight the diagnostically salient retinal regions and achieve better glaucoma performance. The Squeeze-and-Excitation (SE) module is one of the popular methods that adapts the channel-wise feature learning by dynamically adjusting feature responses, while the Convolutional Block Attention Module (CBAM) is sequentially designed to model channel attention and spatial attention, improving feature discrimination. Such attention mechanisms have been shown to boost classification accuracy in multiple types of retinal disease detection and localization in several studies, allowing networks to concentrate on clinically relevant image features.

In parallel with attention mechanisms, there has been a lot of research on the development of efficient deep learning architectures. EfficientNet is designed using a compound scaling strategy that optimizes network depth, width and input resolution to gain high classification accuracy with low parameters compared to traditional CNN models. The EfficientNet-based architectures showed competitive performance in glaucoma detection, while consuming fewer resources than other architectures, which are ideal for healthcare applications with limited resources and large-scale screening programs (Bragança et al., 2022). Likewise, Hemelings et al. (2021) found AI models based on deep learning can also utilize information from the surrounding retina, indicating that effective attention should focus on both local and contextual features of the image for better diagnosis.

While great strides have been made with automated glaucoma detection, there are a number of hurdles. The majority of the current works mainly focus on the accuracy of classification while neglecting to consider efficient computation time and effective feature refinement by attention. Moreover, there is still limited research comparing lightweight attention-enhanced CNN models, especially on clinically curated retinal fundus image datasets. Most existing works also test only a single backbone architecture, which makes it hard to grasp the relative merits of various attention mechanisms, given similar settings and experimental conditions. To solve these problems, this paper proposes two hybrid attention based deep learning models: adding an additional Squeeze-and-Excitation (SE) attention module to EfficientNetB0 and a Convolutional Block Attention Module (CBAM) to ResNet50. Both models are trained and tested on the same set of retinal fundus images that were carefully curated from clinical sources, allowing for a thorough evaluation and comparison using various classification metrics such as accuracy, precision, recall, F1 score and Area Under the ROC Curve (AUC). This comparison helps to understand the efficiency of channel-wise and spatial attention

to perform lightweight AI-assisted glaucoma screening.

Table 1. Comparison other models with proposed model

Authors	Contribution	Limitation
Raghavendra et al. (2018)	CNN-based glaucoma detection with high classification performance	No attention mechanism used
Fu et al. (2018)	Disc-aware ensemble combining local and global retinal features	Computationally expensive and depends on disc localization
Li et al. (2019)	Introduced attention-based glaucoma detection	Increased computational complexity
Bragança et al. (2022)	Lightweight EfficientNet-based glaucoma detection	No comparison of different attention mechanisms
Research Gap	Comparative evaluation of lightweight attention-enhanced CNNs	Addressed in this study

3. Methodology

This study proposes two hybrid attention-based deep learning architectures for automated glaucoma detection using retinal fundus images. The overall methodology consists of dataset preparation, image preprocessing, data augmentation, model development, training, and performance evaluation. Two hybrid CNN architectures, namely EfficientNetB0 integrated with a Squeeze-and-Excitation (SE) attention module and ResNet50 integrated with a Convolutional Block Attention Module (CBAM), are developed and comparatively evaluated using multiple classification metrics.

3.1 DataSet

The dataset used in this study is the ACRIMA (Automated Classification of Retinal Images for Glaucoma Assessment) dataset, which is a publicly available dataset of retinal fundus images for automated glaucoma detection. A total of 705 color retinal fundus photographs including 396 glaucomatous and 309 normal images. The photographs were taken at the FISABIO Oftalmología Médica in Valencia, Spain, from patients who had given informed consent and the data obtained followed the ethical guidelines established by the Declaration of Helsinki (1964). All fundus images were marked on by experienced

glaucoma specialists and the images were cropped around the region of the optic disc for ease of diagnosing glaucoma.

The data set for this study was partitioned into training set and testing set. Sixty-five images were used to train the model and 100 images (75 glaucomatous and 25 normal) were kept for performance evaluation. Diagnostic capability of the proposed models was evaluated against the testing subset with the use of standard classification metrics.

All retinal fundus images were resized to 224×224 pixels and normalized before training, so that they can be fed into the two deep learning architectures as the same input format. The training images were augmented with additional data with random rotation, horizontal flipping, zooming and shearing, to increase model generalization and decrease overfitting. The proposed models were able to learn robust feature representation due to these pre-processing steps under different imaging conditions.

Table 2. DATASET Splitwise

Dataset	Total Images	Glaucoma	Normal	Training	Testing
ACRIMA	705	396	309	605	100

3.2 EfficientNetB0 with Squeeze-and-Excitation (SE) Attention

The first model is proposed based on the EfficientNetB0 architecture, which has been further improved with an extra Squeeze-and-Excitation (SE) attention module. The first proposed model is based on the EfficientNetB0 architecture with an additional Squeeze-and-Excitation (SE) attention module. The lightweight architecture and the best performance of efficientNetB0 in terms of classification accuracy and computational efficiency were the reasons for choosing the model. The added SE attention module further enhances high level feature representations by focusing on informative feature channels before the final classification stage to further improve the model's capacity to distinguish glaucomatous and normal retinal fundus images.

Although EfficientNetB0 inherently incorporates SE modules within its MBConv blocks, an additional SE attention module is introduced after the final feature extraction stage. This additional attention mechanism enables further recalibration of high-level semantic features before classification, allowing the network to emphasize glaucoma-related characteristics present in the retinal fundus images. The objective is to improve feature discrimination while preserving the computational efficiency of the EfficientNetB0 architecture.

Backbone: EfficientNetB0

EfficientNetB0 is chosen for its compound scaling principle, which uniformly scales Using a compound coefficient ϕ , network width w , depth d , and resolution r sequentially. Its baseline form comprises A set of blocks for Mobile Inverted Bottleneck Convolution (MBConv), each incorporating depthwise separable convolutions, batch normalization, swish activations, and internal SE units.

Let $I \in R^{224 \times 224 \times 3}$ be the input fundus image. The EfficientNetB0 backbone maps this to a high-dimensional feature representation:

$$\begin{aligned} F &= \text{EffNetB0}(I), \quad F \\ &\in R^{H' \times W' \times C} \end{aligned} \quad (1)$$

where C is the number of channels after the last convolutional layer.

Despite the embedded SE units inside MBConv blocks, we introduce a dedicated SE module at the top of the feature extractor, allowing additional channel-wise attention on globally aggregated features.

Squeeze-and-Excitation Module

SE module enhances the sensitivity of the model to patterns related to glaucoma because it reweights the channel responses depending on the global context. This entails three major steps:

1. Squeeze (Global Average Pooling):

$$\begin{aligned} \square \square \\ = \frac{1}{H' \times W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} F_c(i, j) \quad \forall c \\ \in \{1, \dots, C\} \end{aligned} \quad (2)$$

This produces a vector $z \in R^C$, encoding the global distribution per channel.

2. Excitation (Dense Attention):

$$\begin{aligned} s \\ = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \end{aligned} \quad (3)$$

- Where:
- $W_1 \in R^{\frac{C}{r} \times C}$, $W_2 \in R^{C \times \frac{C}{r}}$
- δ : ReLU activation
- σ : Sigmoid activation
- r : reduction ratio (typically 16)

3. Recalibration:

$$\begin{aligned} \tilde{F}_c &= s_c \cdot F_c \quad \forall c \\ &\in \{1, \dots, C\} \end{aligned} \quad (4)$$

This upscales the feature map F in the channel axis, and this allows the network to enhance disease-indicative features (e.g., optic disc deformation).

Classification Head

The calculated sophisticated tensor F is further treated in the following way:

- Global Average Pooling (GAP): thickness of space:

$$\begin{aligned} g &= \text{GAP}(\tilde{F}) \\ &\in R^C \end{aligned} \quad (5)$$

- Dropout: (rate = 0.3) applied to regularize: $g' = \text{Dropout}(g)$

- Dense Softmax Layer:

$$Y^{\wedge} = \text{Softmax}(W \cdot g' + b)$$

Where $\hat{y} \in R^2$ represents class probabilities: [Normal, Glaucoma].

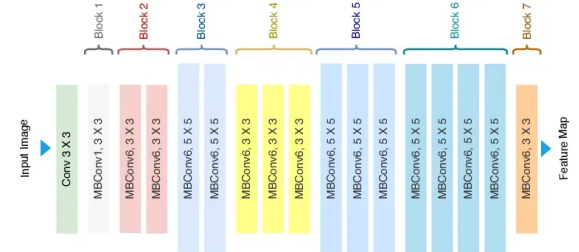


Figure 1. Architecture of the proposed EfficientNetB0 model integrated with an additional Squeeze-and-Excitation (SE) attention module.

3.3 Resnet50 Convolutional Block Attention Module (CBAM) is utilized in the third step.

The second proposed model is ResNet50 architecture combined with the Convolutional Block Attention Module (CBAM). Given the fact that ResNet50 has a deep residual learning capability, it is a good choice because it can effectively solve the vanishing gradient problem and get rich hierarchical features from retinal fundus images. In addition, CBAM attention mechanism is added after the feature extraction process to improve the feature representation. CBAM is different from the channel-only attention mechanism because it first applies the channel attention and then the spatial attention, thus enabling the network to pay attention to the diagnostically important retinal regions associated with glaucoma.

Backbone: ResNet50

ResNet50 is a 50 layer deep residual network, which addresses the issue of vanishing gradient through identity skip connections. It performs the learning of the residual mappings by reformulating layers as:

$$\begin{aligned} y \\ = \mathcal{F}(x, \{W_i\}) + x \end{aligned} \quad (6)$$

where x is the input, $\mathcal{F}$ is the residual function, and y is the output after the skip connection. The backbone is initialized with pretrained ImageNet weights and receives input $I \in R^{224 \times 224 \times 3}$, which is processed into:

$$\begin{aligned} F \\ = \text{RESNET50}(I), \quad F \\ \in R^{H' \times W' \times C} \end{aligned} \quad (7)$$

Attention: Convolutional Block Attention Module (CBAM)

CBAM is a lightweight attention mechanism applied after the final convolutional layer of ResNet50. It sequentially applies channel attention followed by spatial attention to refine the feature map F .

1. Channel Attention Module

Given input feature map $F \in R^{H \times W \times C}$, CBAM computes channel attention weights using both Global Average Pooling and Global Max Pooling:

$$\begin{aligned} \square_{\text{avg}} &= \text{GAP}(F), \quad z_{\text{max}} \\ &= \text{GMP}(F) \end{aligned} \quad (8)$$

Both descriptors are passed through a shared MLP with one hidden layer:

$$\begin{aligned} \square_{\text{ag}} &= \square_2 \\ &\cdot \delta(W_1 \cdot z_{\text{avg}}), \quad s_{\text{max}} \\ &= W_2 \cdot \delta(W_1 \cdot z_{\text{max}}) \\ s &= \sigma(s_{\text{avg}} + s_{\text{max}}) \end{aligned} \quad (9)$$

Recalibration is applied:

$$\begin{aligned} \square &= s \cdot F \end{aligned} \quad (11)$$

2. Spatial Attention Module

The spatial attention mechanism highlights important regions within the image. It first applies pooling operations across the channel axis:

$$\begin{aligned} \square_{\text{avg}} &= \text{mean}_c(\tilde{F}), \quad M_{\text{max}} \\ &= \max_c(\tilde{F}) \end{aligned} \quad (12)$$

The two 2D maps are concatenated and convolved:

$$M = \sigma(\text{Conv}_{7 \times 7}([M_{\text{avg}}; M_{\text{max}}])) \quad (13)$$

$$F'' = M \odot \tilde{F} \quad (14)$$

Where $\odot$ denotes element-wise multiplication.

Classification Head

The final attention-weighted feature map F'' is processed through:

- Global Average Pooling (GAP):

$$g = \text{GAP}(F'') \in R^C \quad (15)$$

- Dropout: (rate = 0.3)

$g' = \text{Dropout}(g)$

- Dense Softmax Layer:

$$Y^{\wedge} = \text{Softmax}(W \cdot g' + b)$$

Where $\hat{y} \in R^2$ represents class probabilities: [Normal, Glaucoma].

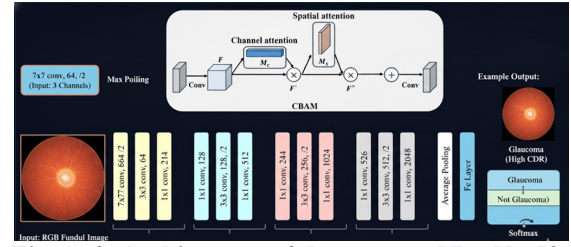


Figure 2. Architecture of the proposed ResNet50 model integrated with the Convolutional Block Attention Module (CBAM).

3.4 Evaluation Metrics

The performance of the proposed glaucoma detection models was evaluated using several standard classification metrics, including **Accuracy, Precision, Recall (Sensitivity), F1-score, Receiver Operating Characteristic (ROC) curve, and Area Under the ROC Curve (AUC)**. These metrics provide a comprehensive assessment of the models' classification performance by measuring their ability to correctly identify both glaucomatous and normal retinal fundus images.

Accuracy

Accuracy refers to the correctness of the whole model in the general predictions. It is informative but might lead to confusion in class-imbalanced data such as in glaucoma screening.

Precision

Precision measures the amount of correct cases of glaucoma imminent among all positive predictions. It suggests how well the model can help to avoid a false positive.

Recall (Sensitivity)

Recall is the correctness in determining true cases of glaucoma. High recall is decisive in medical context so that there would be minimal cases of failure in assisting diagnoses.

F1-Score

F1-score is the harmonic mean of the precision and recall and it is particularly useful when there are imbalanced classes:

$$\begin{aligned} \text{F1-Score} &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned} \quad (16)$$

ROC-AUC (Receiver Operating Characteristic – Area Under Curve)

ROC-AUC also gives the threshold-independent analysis of the classification performance by plotting the true positive rate versus the false positive rate at all thresholds. An increased AUC value would indicate more class separability:

- AUC = 1.0: Perfect classification
- AUC = 0.5: Random performance

Confusion Matrix

To have a case-wise breakdown of predictions, the confusion matrix can be used:

Table 3. Confusion matrix Defining table

	Predicted Normal	Predicted Glaucoma
Actual Normal	True Negative (TN)	False Positive (FP)
Actual Glaucoma	False Negative (FN)	True Positive (TP)

4. Results

The two hybrid architectures were tested on the curated clinical fundus image dataset that is mentioned above. The quantitative outcomes have been summarized in Table 2 using major measures such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

Table 4. Comparative Performance of Hybrid Models

Model	Accuracy	Precision	Recall	F1-Score	AUC
EfficientNetB0 + SE	1.00	1.00	1.00	1.00	1.00
ResNet50 + CBAM	0.98	0.97	1.00	0.99	1.00

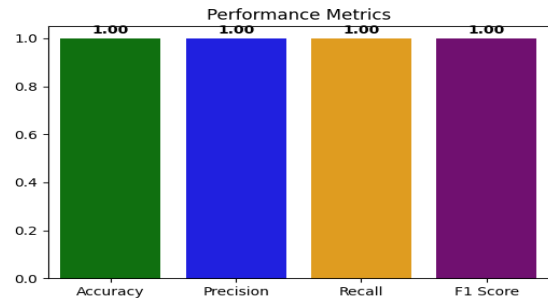


Figure 3. Performance Metrics for EfficientNetB0 + SE

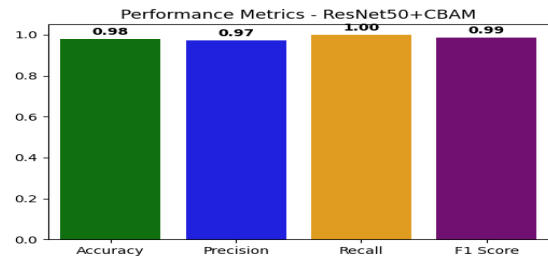


Figure 4. Performance Metrics for ResNet50 + CBAM

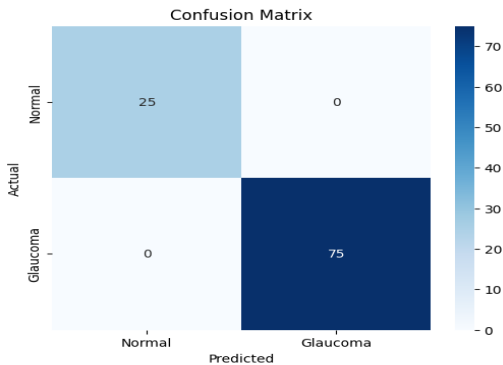


Figure 5. Confusion matrix for EfficientNetB0 + SE

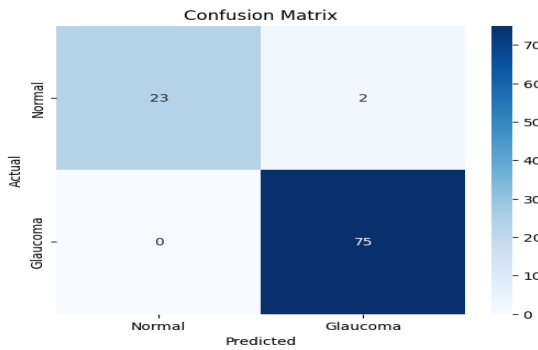


Figure 6. Confusion Matrix for ResNet50 + CBAM

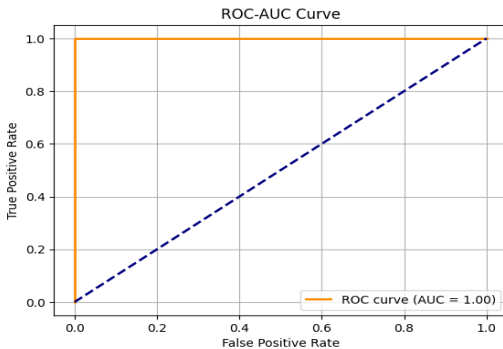


Figure 7. Roc Curve for Both it is same

4.1 Analysis

The EfficientNetB0 + SE model achieved perfect classification performance across all evaluation metrics on the testing dataset. The confusion matrix indicates that all glaucomatous and normal retinal fundus images were correctly classified, resulting in zero false positives and zero false negatives. This performance demonstrates that the combination of EfficientNetB0 with the additional Squeeze-and-Excitation (SE) attention module effectively captures discriminative glaucoma-related features while maintaining a lightweight network architecture.

The ResNet50 + CBAM model also demonstrated excellent diagnostic performance, achieving an accuracy of 98% with a recall of 1.00. The perfect recall indicates that all glaucomatous cases were

correctly identified, which is particularly important in glaucoma screening, where missed diagnoses may lead to irreversible vision loss. The slightly lower precision (0.97) resulted from two false-positive predictions, where normal retinal images were incorrectly classified as glaucomatous.

The CBAM attention mechanism effectively enhanced both channel-wise and spatial feature representations, enabling the model to focus on diagnostically relevant retinal structures while suppressing less informative regions. Although its overall performance was marginally lower than that of EfficientNetB0 + SE, the ResNet50 + CBAM architecture still demonstrated strong potential for automated glaucoma screening.

The superior performance of the EfficientNetB0 + SE model can be attributed to the combination of EfficientNetB0's compound scaling strategy and the additional SE attention mechanism, which further refines high-level feature representations before classification. In contrast, although the CBAM module effectively captures both channel and spatial information, the deeper ResNet50 architecture contains a significantly larger number of parameters, which may contribute to slightly lower generalization performance on a relatively small dataset. These findings indicate that lightweight attention-enhanced architectures can provide an effective balance between computational efficiency and diagnostic accuracy for glaucoma detection.

Although the EfficientNetB0 + SE model achieved perfect classification performance on the testing dataset, these results should be interpreted within the scope of the experimental setup. The proposed models were evaluated using the ACRIMA dataset, and further validation using larger and more diverse multi-center clinical datasets is necessary to fully assess their generalization capability and robustness in real-world clinical applications.

4.2 Comparison with Existing Studies

Table 5. Comparison table with existing results

Study	Model	Dataset	Images	Accuracy (%)
Fu et al.	DENet	ORIGA (Train), SCES (Test), SINDI (Test)	650 / 1676 / 5783	84.2 (SCES), 74.9 (SINDI)
Raghavendra et al.	CNN	Fundus Images	1426	98.13
Proposed	EfficientNetB0 + SE	ACRIMA	705	100.0

Study	Model	Dataset	Images	Accuracy (%)
Proposed	ResNet50 + CBAM	ACRIMA	705	98.0

Table 5 compares the proposed models with representative deep learning approaches reported in the literature. The EfficientNetB0 + SE model achieved the highest classification accuracy on the ACRIMA dataset, while the ResNet50 + CBAM model also demonstrated competitive performance. Compared with the CNN model proposed by Raghavendra et al., which achieved an accuracy of 98.13%, the proposed EfficientNetB0 + SE model exhibited improved classification performance on the evaluation dataset. Similarly, the proposed models outperformed the Disc-aware Ensemble Network (DENet) reported by Fu et al. on the SCES and SINDI datasets. However, since these studies were evaluated using different datasets with varying image characteristics, class distributions, and evaluation protocols, the comparison should be interpreted as a qualitative performance assessment rather than a direct benchmark comparison.

4.2 Clinical Implications

Addition of attention mechanisms in the two models significantly minimized false negatives which are extremely dangerous in glaucoma screening. Because glaucoma is an irreversible condition, these architectures may be advantageous as powerful tools of early diagnosis, to assist ophthalmologists on their mass screening procedures.

Moreover, the efficacy of the EfficientNetB0 + SE model despite its compact structure (the model has around 5M parameters) is an indication of its appropriateness and practicality in the real-time clinical scenarios that are constrained by the resources.

5. Conclusions

In this work, we proposed and tested two hybrid attention-based deep learning architectures for automatic glaucoma detection from retinal fundus images: EfficientNetB0 with a Squeeze-and-Excitation (SE) attention module and ResNet50 with a Convolutional Block Attention Module (CBAM). The diagnostic performance of both models was high in the ACRIMA dataset, with the EfficientNetB0 + SE model being the most optimized among them, and the ResNet50 + CBAM model being the competitive one with regards to the classification results. Our results show that attention mechanisms can be integrated within CNN which improves the feature representation and the accuracy in glaucoma classification.

In this study, EfficientNetB0 + SE achieved higher performance than the proposed architecture of

EfficientNetB0, and it has a lightweight architecture with about 5 million parameters, which is suitable for the real-time application of AI to glaucoma screening in resource-limited clinical settings. ResNet50 + CBAM also performed well with a high diagnostic accuracy and perfect recall, showing its power to correctly diagnose glaucomatous cases. Although the proposed models achieved excellent performance, this study has certain limitations. The experimental evaluation was performed on one of the publicly available retinal fundus image database (ACRIMA) with a testing set of 100 retinal fundus images. However, larger and more diverse and multi-center clinical datasets must be used to thoroughly test the generalization capability and the clinical application of the proposed models. The proposed architectures will be tested on larger multi-center datasets to demonstrate their robustness and further validate model interpretability through the integration of explainable artificial intelligence (XAI) techniques, while multimodal ophthalmic data will be added to improve the diagnostic performance.

6. Acknowledgements

Should not be used to acknowledge funders – funding will be entered as a separate. As a matter of courtesy, we suggest you inform anyone whom you acknowledge.

Author Contributions:

Conceptualization: All authors.

Methodology: All authors.

Software: All authors.

Validation: All authors.

Formal Analysis: All authors.

Investigation: All authors.

Data Curation: The authors utilized the publicly available ACRIMA retinal fundus image dataset for model development and evaluation. Data preprocessing, augmentation, and preparation for training and testing were performed by the authors.

Writing – Original Draft Preparation: All authors.

Writing – Review & Editing: All authors.

Visualization: All authors.

Supervision: All authors.

All authors have read and agreed to the published version of the manuscript.

Funding:

The study was carried out without the help of definite financial assistance provided by the public, commercial or non-profit funding agencies. The research design, experimentation, and results were not induced by any external funding source.

Conflicts of Interest:

There are no conflicts of interest to the research and authorship of this manuscript, according to the authors.

Data Availability Statement

The ACRIMA retinal fundus image dataset used in this study is publicly available for research purposes. Additional implementation details and supporting materials are available from the corresponding author upon reasonable request.

Ethics Statement

This study utilized the publicly available ACRIMA retinal fundus image dataset. All images were collected by the original dataset providers with informed patient consent and in accordance with the ethical principles of the Declaration of Helsinki. Therefore, no additional institutional ethical approval was required for this study.

References

- [1] Bragança et al., “Detection of Glaucoma on Fundus Images Using Deep Learning on a New Image Set Obtained with a Smartphone...,” *Healthcare*, vol. 10, no. 12, p. 2345, 2022.
- [2] H. Fu et al., “Disc-aware ensemble network for glaucoma screening from fundus image,” *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.
- [3] U. Raghavendra et al., “Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images,” *Information Sciences*, vol. 441, pp. 41–49, 2018.
- [4] L. Li et al., “Attention Based Glaucoma Detection: A Large-scale Database and CNN Model,” *arXiv*, 2019.
- [5] R. Hemelings et al., “Deep learning on fundus images detects glaucoma beyond the optic disc,” *Sci. Rep.*, vol. 11, p. 20313, 2021.
- [6] D. Devecioglu et al., “Real-Time Glaucoma Detection from Digital Fundus Images using Self-ONNs,” *arXiv*, 2021.
- [7] S. Sengupta et al., “Application of Deep Learning in Fundus Image Processing for Ophthalmic Diagnosis – A Review,” *arXiv*, 2018.
- [8] M. Tohye et al., “CA-ViT: Contour-Guided and Augmented Vision Transformers to Enhance Glaucoma Classification,” *Bioengineering*, vol. 11, p. 887, 2024.
- [9] “Glaucoma identification with retinal fundus images using deep ...,” *ScienceDirect*, 2025.
- [10] “A Dual Attention-aided DenseNet-121 for Classification of Glaucoma from Fundus Images,” *arXiv*, 2024.
- [11] “Multi-feature analysis and a deep ...,” *Nature Communications*, 2022.
- [12] “Glaucoma diagnosis using multi-feature analysis ...,” *SciNR*, 2022.
- [13] “Fundus image classification using feature concatenation ...,” *PMC*, 2025.
- [14] “An automatic glaucoma grading method based on attention ...,” *PMC*, 2025.
- [15] “Improved GoogLeNet based on CBAM module for fundus disease ...,” *SPIE Digital Library*, 2020.

- [16] "Detection of glaucoma from fundus images using ensemble learning," ResearchGate, 2023.
- [17] "Classification of Glaucoma in Retinal Images Using EfficientNetB4 ...," TechScience, 2022.
- [18] "Glaucoma detection using EfficientNet ...," ScienceDirect, 2022.
- [19] "Deep Learning Evaluation of Glaucoma Detection in High Myopic Populations," Biomedicines, 2024.
- [20] "Global Self-Attention Guided CNN for multi-stage glaucoma," arXiv, 2024.
- [21] "MaxGlaViT: A novel lightweight vision transformer-based approach ...," arXiv, 2025.
- [22] B. Chakraborty et al., "Dual Attention-aided DenseNet-121 ...," arXiv, 2024.
- [23] "Attention Based Glaucoma Detection: A Large-scale Database ...," arXiv, 2019.
- [24] "FADNet: Fusion-Attention Diagnosis Network ...," ScienceDirect, 2025.
- [25] "Early detection of glaucoma: feature visualization ...," Taylor & Francis, 2024.

Figure Legends

Figure 1. Architecture of the proposed EfficientNetB0 model integrated with an additional Squeeze-and-Excitation (SE) attention module

Figure 2. Architecture of the proposed ResNet50 model integrated with the Convolutional Block Attention Module (CBAM).

Figure 3. Performance Metrics for EfficientNetB0 + SE

Figure 4. Performance Metrics for ResNet50 + CBAM

Figure 5. Confusion matrix for EfficientNetB0 + SE

Figure 6. Confusion Matrix for ResNet50 + CBAM

Figure 7. Roc Curve for Both it is same