engineering proceedings

MDPI

Proceeding Paper

Efficient and Explainable Glaucoma Detection Using an Attention-Enhanced Custom CNN on Retinal Fundus Images †

Vijaya Kumar Velpula [1,*], Jonnala Naga Surekha [2], Nekkanti Gowthami [3], S. Anuradha [4], Muvva Venkateswara Rao [5], J. Rajendra Prasad [6] and Jyothisri Vadlamudi [4]

[1] Department of Electronics and Communication Engineering, MLR Institute of Technology, Hyderabad 500043, India
[2] Department of Electronics and Communication Engineering, Dr RVR NRI Institute of Technology (Deemed to be University), Vijayawada 521212, India; jonnalasurekha666@gmail.com
[3] Department of Internet of Things, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru 521356, India; gowthami.nekkanti@gmail.com
[4] Department of Electronics and Communication Engineering, Mallareddy (Deemed to be University), Hyderabad 500100, India; anukr2424@gmail.com (S.A.); vadlamudijyothisri6@gmail.com (J.V.)
[5] Department of Computer Science and Engineering, Dr RVR NRI Institute of Technology (Deemed to be University), Vijayawada 521212, India; muvva.venky@gmail.com
[6] Department of Information Technology, Dr RVR NRI Institute of Technology (Deemed to be University), Vijayawada 521212, India; rp.rajendra@rediffmail.com
* Correspondence: vijaykumarv@mlrit.ac.in
† Presented at the 6th International Electronic Conference on Applied Sciences, 9–11 December 2025; Available online: https://sciforum.net/event/ASEC2025.

Abstract

Glaucoma is an increasing ocular disease and one of the major causes of irreversible blindness globally. Thus, it is of utmost importance to diagnose glaucoma accurately and in a timely manner for appropriate clinical intervention. Retinal fundus image analysis is a non-invasive approach for glaucoma screening. However, it is a tedious and subjective process for accurate glaucoma detection. In this work, a customized convolutional neural network (CNN) model is proposed for glaucoma classification using fundus images, and an attention mechanism is employed for improved discriminative feature learning. Experiments were carried out on two publicly available datasets, namely Drishti-GS1 and ACRIMA, for unbalanced and balanced datasets, respectively. Data augmentation and hyperparameter tuning were employed for improved model generalization. To increase model explainability and trust for accurate glaucoma screening, gradient-weighted class activation mapping (Grad-CAM) is employed for accurate glaucoma screening. For accurate glaucoma screening, the proposed model attained 90.32% and 96.45% accuracy on the Drishti-GS1 and ACRIMA datasets, respectively, with higher sensitivity and AUC scores. Thus, it is evident that the proposed model employing attention mechanisms and explainable AI attained higher accuracy for accurate glaucoma screening.

Keywords: AlexNet; attention module; convolutional neural network; deep learning; explainable AI; glaucoma; Grad-CAM; ResNet-50

Check for updates

Academic Editor: Lucia Billeci

Published: 21 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Glaucoma is a progressive optic neuropathy and has become one of the top causes of irreversible blindness in the world, with millions of people of all ages being affected. The World Health Organization (WHO) estimates that glaucoma is the main cause of almost 12 percent of blindness in the world, and, with the trend of increasing age, it is

likely to increase in the future [1]. One of the main problems associated with glaucoma is that this disease is asymptomatic at the initial stages, so it is often diagnosed and treated late. Therefore, early diagnosis of glaucoma is important in preventing and efficiently managing glaucoma.

Non-invasive modalities for screening glaucoma, such as retinal fundus imaging, have proven to be a common practice. The cup-to-disc ratio may be considered a significant biomarker of glaucoma, and this ratio is a key area of clinical diagnosis using fundus images. Nonetheless, these images are very subjective, time-consuming and liable to inter-observer discrepancy among ophthalmologists when analyzed manually. Automated glaucoma detection systems should be able to offer a solution that is objective, reproducible and effective to help clinicians in mass screening and diagnosis.

Machine learning and computer vision algorithms have become more widely used in recent years to compute automatically detected retinal diseases [2]. The conventional methods were based on hand-made aspects of features like texture descriptions, vessel patterns or morphological measurements of the optic disc and cup. Although these techniques performed well, they were limited by their reliance on manual feature engineering. With the development of deep learning, especially convolutional neural networks (CNNs), the analysis of medical images has been revolutionized because now this feature can learn automatically, without human involvement, through raw image data. Because CNNs can extract hierarchical features, which extract local and global patterns, they are very appropriate for classifying glaucoma. A number of research works have examined the applicability of pretrained deep learning models like AlexNet, VGG and ResNet in glaucoma detection. Despite good results of these transfer learning methods, they usually involve high computing power and are not optimally geared towards retinal fundus images. Additionally, the problems of dataset imbalance, overfitting, and generalization of various clinical datasets are significant challenges. Therefore, a lightweight but efficient CNN model is necessary for glaucoma detection.

In this paper, we present a tailor-made CNN model enhanced with an attention mechanism developed to categorize fundus images as glaucoma and healthy. The model was fully trained and tested on two benchmark datasets: Drishti-GS1 is a small dataset of unbalanced samples, and ACRIMA is a balanced dataset of larger size. The data augmentation approaches were used to solve the issue of data scarcity, and hyperparameters were optimally tuned to achieve the best performance. The findings reveal that the proposed CNN has a better accuracy rate, as noted when compared with popular pretrained models, and the sensitivity and AUC results are very high, thus qualifying it to be used in clinical screening processes.

The main contributions of this paper are summarized as follows:

- The customized CNN is developed and integrated with an attention mechanism for efficient glaucoma classification.
- The model is fully tested on two publicly available datasets, Drishti-GS1 and ACRIMA, considering different train–test combinations.
- The performance of the model is compared with pretrained convolutional neural networks, such as AlexNet and ResNet50, and improved results are observed in terms of accuracy, sensitivity, and AUC.
- The robustness of the model is validated by increasing the number of images and tuning the hyperparameters.
- The interpretability of the model is improved using Grad-CAM, enabling better understanding of the classification results and facilitating glaucoma screening systems.

The rest of the paper will be organized as follows: the literature review of the relevant research in the field of glaucoma detection and deep learning will be discussed in Section 2.

https://doi.org/10.3390/engproc2026124110

The methodology will be presented in Section 3, which includes the dataset, data augmentation, the structure of the CNN with the attention module, and so on. The experimental results and discussion will be shown in Section 4. At last, the summary of the conclusion of the paper and the way to go in the future will be presented in Section 5.

2. Related Work

The techniques of glaucoma detection using retinal images have also been extensively explored in terms of traditional machine learning as well as deep learning techniques. The traditional techniques are generally based on preprocessing techniques, division of the disc and cup, and feature detection using techniques such as texture features and morphological/wavelet evaluations. Although these techniques have been found to be useful in providing information, they are generally restricted by their capacity for generalization in different datasets.

The structural parameters that are generally used for making clinical diagnoses of glaucoma are cup-to-disc ratio (CDR) [3], rim-to-disc area ratio (RDAR), and disc diameter. The most commonly used parameter is CDR as higher values of CDR are associated with a higher risk of glaucoma. To measure CDR correctly, it is therefore important to correctly segment the optic disc and cup. Three-dimensional optical coherence tomography (OCT) is able to quantify correctly but is costly and less available, and fundus imaging is the diagnostic modality of choice in most situations. Boundary detection, region-based segmentation and color/contrast thresholding techniques are some of the techniques that have been used in the localization of optic discs and cups. Moreover, the application of machine learning algorithms such as random forest in glaucoma detection results in an AROC of 93% and is better than stepwise, LASSO, and ridge regression models in identifying glaucoma patients [4]. Angle closure glaucoma detection has also been carried out by using the MultiContext Deep Network (MCDN) on the AS-OCT dataset. The application of data augmentation and integration of clinical factors are also carried out in this approach. This approach results in 89.26% accuracy, 0.9456 AUC, 88.89% sensitivity, and 89.63% specificity by using TensorFlow and VGG-16 [5].

To improve handcrafted feature approaches, such as having fewer features and poor performance on complex datasets, several deep learning approaches are also discussed in the literature. The application of CNN-based approaches, such as AlexNet, VGG, and ResNet, is discussed in detail. The application of AlexNet is carried out for feature transfer but results in poor performance because of overfitting in small medical datasets. The application of ResNet is also discussed for feature extraction, but it results in increased computational complexity. For instance, Ref. [6] proposed a two-stage method using GoogleNet for region-of-interest (ROI) extraction followed by classification. Similarly, Ref. [7] combined QB-VMD decomposition with PHOG and Haralick feature extraction, and Ref. [8] utilized an adapted GoogLeNet for glaucoma detection using ROI localization.

Other works have focused on integrating deep learning [9] with traditional classifiers. For example, CNN-extracted features have been classified using algorithms such as AdaBoost, k-nearest neighbor (kNN), random forest (RF), multilayer perceptron (MLP), support vector machine (SVM), and naive Bayes (NB), demonstrating competitive diagnostic performance [10]. Investigations have also been conducted on hybrid feature extraction strategies that combine histogram of oriented gradients (HOG), local binary patterns (LBP), gray-level co-occurrence matrix (GLCM) and chip histogram features. Comparative tests demonstrated that AdaBoost was more accurate than SVM, especially when texture-based features, as well as shape-based features, are combined. Generalized CNN architectures [11] and CoG-NET [12] have also been proposed, as well as explainable AI models that can use pretrained CNNs with machine learning classifiers, to improve interpretability [13].

https://doi.org/10.3390/engproc2026124110

Segmentation-driven pipelines, where the optic disc is initially segmented, followed by classification with deep learning networks [14], have been studied by other researchers, but channel weighting on RGB fundus images has been seen to bring only small gains to such pipelines by a very small margin [15].

Recent studies have explored lightweight and attention-based deep learning architectures to improve the efficiency and reliability of automated glaucoma detection. Cem Baydogan et al. [16] proposed a hybrid framework that combines lightweight MobileNet architectures with traditional machine learning classifiers. The study demonstrated that MobileNet-based deep feature extraction, when integrated with classifiers such as SVM, can achieve high diagnostic performance while maintaining low computational cost. Mustafa Yurdakul et al. [17] introduced MaxGlaViT, a vision transformer-based model derived from MaxViT, which integrates attention mechanisms at multiple levels to enhance feature representation. In another recent work, Arjun Kumar Bose Arnob et al. [18] proposed a lightweight attention-augmented CNN incorporating squeeze-and-excitation and global-context attention mechanisms for retinal disease screening, with Grad-CAM-based visual explanations used to improve interpretability. This approach addressed challenges related to model transparency and computational efficiency, highlighting the growing importance of explainable AI in ophthalmic image analysis.

Although these approaches have been proposed, high sensitivity in early-stage detection, dataset imbalance, and computation requirements remain issues. Ready-made networks like AlexNet and ResNet provide a good starting point. Our research is motivated based on these constraints and hence proposes a modified CNN framework that can be used in glaucoma.

Artificial intelligence techniques have increasingly been used in the analysis of eye-related diseases for early and accurate diagnosis. Multiple instance learning (MIL) [19] techniques have been used in handling visually similar lesions and weakly labeled medical images. It was found that these techniques are useful in automated discrimination of lesions. At the same time, diabetic retinopathy has been employed using AI-based techniques in analyzing images, demonstrating the potential of deep learning and machine learning models [20,21] in cost-effective screening.

3. Materials and Methods

Figure 1 illustrates the workflow diagram of the proposed method. The methodology proposed in this paper focuses on the development of a stable and efficient model for the structure of glaucoma detection with the assistance of retinal fundus images. The proposed methodology has been tested with two publicly accessible datasets [22], namely Drishti-GS1 and ACRIMA, to enable the evaluation of imbalanced and balanced data. For the purpose of diversification of the training set and the achievement of better generalization of the models with minimal possibilities of overfitting, various preprocessing and augmentation techniques have been employed. The proposed custom CNN model consisted of a number of convolutional and pooling layers for the extraction of features. The important hyperparameters were optimized in order to ensure the best performance, including learning rate, batch size, and the number of epochs. The system was tested on the basis of traditional performance measures that include accuracy, sensitivity, specificity, precision, F-score, and AUC, giving a clear overall picture of the effectiveness of the given system in glaucoma classification in an automated environment.

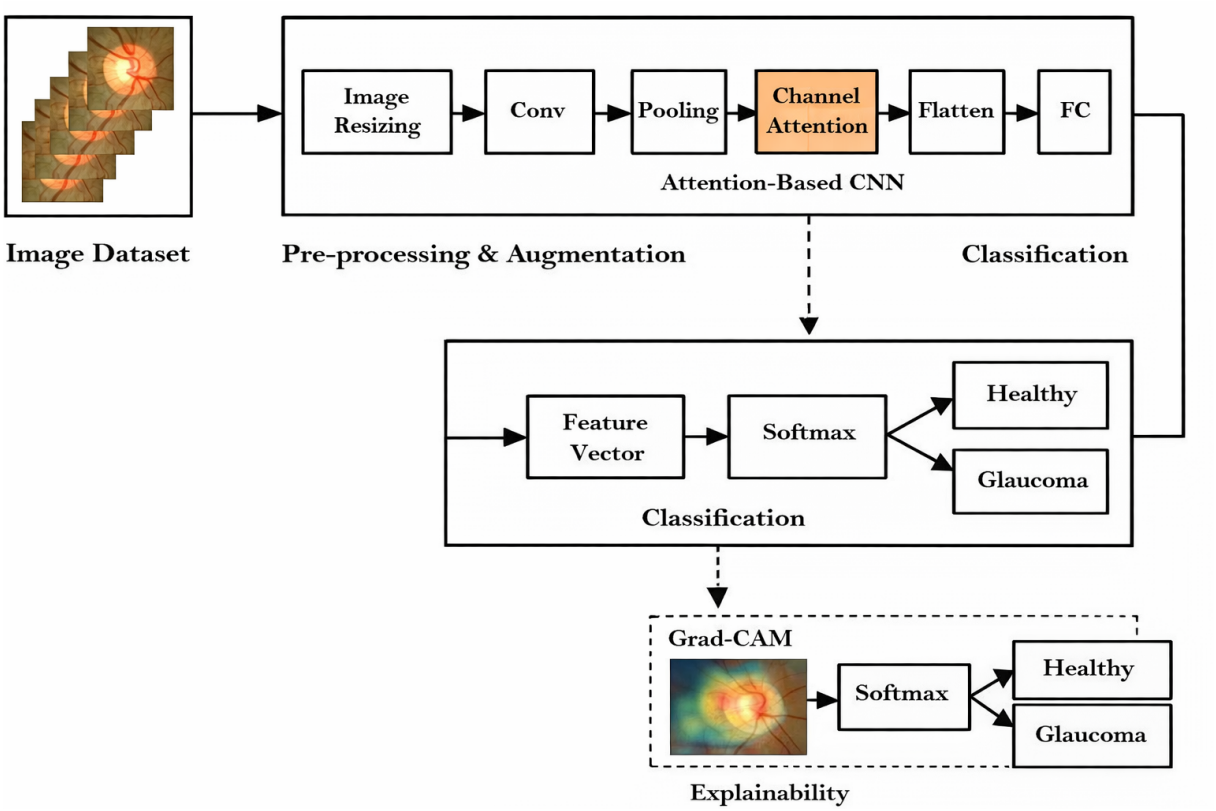


Figure 1. Workflow block diagram.

3.1. Dataset Collection

To validate the proposed approach, experiments were conducted on two publicly available retinal fundus datasets.

Drishti-GS1 dataset is one popularly used publicly available unbalanced dataset that is collected and annotated by Aravind Eye Hospital, Madurai, from the visitors with their consent. This dataset consists of 101 images (70 glaucoma, 31 healthy) divided into 50 training and 51 testing images. All the images are centered on the optic disc (OD) with 300 field of view (FOV) and dimensions of 2896 × 1944 pixels, and all images are available in .png format. This dataset also includes ground truth images collected from data experts with different clinical experiences. This can be downloaded through the https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php (accessed on 10 August 2025) or https://www.kaggle.com/datasets/lokeshsaipureddi/drishtigs-retina-dataset-for-onh-segmentation (accessed on 10 August 2025).

ACRIMA is a balanced public dataset introduced to aim for the development of automatic retinal disease assessment algorithms. It consists of 705 images (396 glaucoma, 309 healthy), and all images are captured with Topcon TRC retinal camera and ImageNet capturing system with 350 FOV and are available in .jpg image format. The images in this dataset are annotated by two clinical experts with 8 years of experience. This can be downloaded through the https://figshare.com/articles/dataset/CNNs_for_Automatic_Glaucoma_Assessment_using_Fundus_Images_An_Extensive_Validation/7613135 (accessed on 10 August 2025) or https://www.kaggle.com/datasets/toaharahmanratul/acrima-dataset (accessed on 10 August 2025).

3.2. Data Preprocessing and Augmentation

Since medical datasets are often limited in size, various augmentation techniques were applied to artificially expand the dataset and enhance the model's generalization capability.

To address the limited size of medical image datasets and improve the robustness of the proposed CNN model, several augmentation techniques were applied to the training images. These included random rotations of up to 40°, shear transformations with a factor of 0.2, zoom variations of 0.2, and brightness adjustments in the range of 0.5 to 1.5. These transformations create the ability of the model that is exposed to diverse variations in the optic disc and surrounding regions, thereby enhancing its ability to generalize across different imaging conditions. The details of the augmentation strategies are represented in this study in Table 1.

Table 1. Data augmentation techniques applied.

Augmentation Technique	Parameter Range/Value
Rotation	Up to 40°
Shear	0.2
Zoom	0.2
Brightness Adjustment	(0.5, 1.5)

3.3. CNN Architecture with Attention Mechanism

The convolutional neural network (CNN) proposed in this study is specifically designed to discriminate between glaucoma-affected and healthy retinal fundus images. The architecture consists of six convolutional layers, each followed by a max-pooling operation, to progressively reduce spatial dimensions while preserving discriminative visual features. All convolutional layers employ 3 × 3 kernels with a stride of 1, and the ReLU activation function is used to introduce nonlinearity and enhance feature learning. Input fundus images are resized to 256 × 256 × 3 before being fed into the network. The initial convolutional layer extracts 32 feature maps, which are downsampled using max-pooling. As the network depth increases, the number of feature maps is expanded to 64, enabling the extraction of more complex and representative patterns related to optic disc structure and texture variations associated with glaucoma.

To further improve feature discrimination, an attention module is integrated into the CNN architecture after selected convolutional blocks. The channel attention mechanism adaptively recalibrates feature responses by explicitly modeling inter-channel dependencies. Given an input feature map, global average pooling is first applied to generate a compact channel-wise descriptor. This descriptor is then passed through two fully connected layers with a reduction ratio to capture nonlinear channel relationships, followed by a sigmoid activation to obtain normalized attention weights. These weights are subsequently used to rescale the original feature maps, enhancing informative channels while suppressing less relevant ones. This attention-driven feature refinement enables the network to focus on clinically significant structures in retinal fundus images, thereby improving glaucoma classification performance. The mathematical formulation of the channel attention module is given as follows:

Let the input feature map be represented as

$$\mathbf{F} \in \mathbb{R}^{H \times W \times C}, \tag{1}$$

where H, W, and C denote the height, width, and number of channels, respectively.

Global average pooling is applied to embed global spatial information into a channel-wise descriptor:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j), \quad c = 1, 2, \ldots, C. \tag{2}$$

This results in a channel descriptor vector

$$\mathbf{z} \in \mathbb{R}^C. \tag{3}$$

The channel descriptor is passed through two fully connected layers with a reduction ratio r to model inter-channel dependencies:

$$\mathbf{s} = \sigma(\mathbf{W}_2\,\delta(\mathbf{W}_1\mathbf{z})), \tag{4}$$

where $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C\times\frac{C}{r}}$ denote the learnable weight matrices, $\delta(\cdot)$ represents the ReLU activation function, and $\sigma(\cdot)$ denotes the sigmoid activation function.

The final attention-refined feature map is obtained by channel-wise scaling:

$$\tilde{\mathbf{F}}_c = s_c \cdot \mathbf{F}_c, \quad c = 1, 2, \ldots, C. \tag{5}$$

The overall channel attention operation can be expressed as

$$\tilde{\mathbf{F}} = \mathbf{F} \odot \sigma(\mathbf{W}_2\,\delta(\mathbf{W}_1\,\mathrm{GAP}(\mathbf{F}))), \tag{6}$$

where $\odot$ denotes channel-wise multiplication.

Following the attention-enhanced convolutional and pooling stages, the refined feature maps are flattened into a one-dimensional feature vector to facilitate classification. To improve generalization and mitigate overfitting, a dropout layer with a rate of 0.5 is applied before the fully connected dense layer. The final classification layer employs a softmax activation function to produce probability scores corresponding to the two target classes, namely glaucoma and healthy. Overall, the proposed attention-integrated CNN architecture effectively balances representational depth and computational efficiency, enabling robust feature learning while maintaining suitability for practical glaucoma screening applications. The detailed configuration of the CNN with the attention module is summarized in Table 2.

Table 2. Proposed CNN architecture for glaucoma classification.

Layer	Feature Maps	Output Shape	Parameters
Input Layer	3	256 × 256 × 3	0
Conv-1	32	254 × 254 × 32	896
Max Pool-1	32	127 × 127 × 32	0
Conv-2	64	125 × 125 × 64	18,496
Max Pool-2	64	62 × 62 × 64	0
Conv-3	64	60 × 60 × 64	36,928
Max Pool-3	64	30 × 30 × 64	0
Conv-4	64	28 × 28 × 64	36,928
Max Pool-4	64	14 × 14 × 64	0
Conv-5	64	12 × 12 × 64	36,928
Max Pool-5	64	6 × 6 × 64	0
Conv-6	64	4 × 4 × 64	36,928
Max Pool-6	64	2 × 2 × 64	0
Flatten	2	256	0
Dropout	2	128	32,896
Output Layer (Softmax)	2	2	258

The selection of hyperparameters is very critical in the performance of a CNN model as they manage the learning processes during the training process. In the given work, the size of the input image was held at 256 × 256 to ensure that both sets of data had

the same size whilst not losing any important retinal details. The convolutional layers were based on a kernel size of 3 × 3 with a 1 stride rate since this type is commonly used to extract fine-grained spatial information in medical images. The amount of filters was adjusted to 32 and 64 cross layers so that low- and high-level features can be extracted. The model was trained for 20 epochs, and this choice was made to have the right balance between computational efficiency and model convergence. Various batch sizes (16, 32, and 64) were adopted to examine their performance. On the same note, initial learning rate was fluctuated (0.01, 0.001, and 0.0001) to see which one was the most stable to optimize. Such controlled experiments were used to make sure that the model performed optimally without overfitting or underfitting. The information about the selected hyperparameters is outlined in Table 3.

Table 3. Hyperparameters that are used in CNN training.

Hyperparameter	Value/Range
Input Image Size	256 × 256
Kernel Size	3 × 3, stride = 1
Number of Kernels (Filters)	32, 64
Number of Epochs	20
Batch Size	16, 32, 64
Initial Learning Rate	0.01, 0.001, 0.0001

3.4. Performance Evaluation

The results of the recommended modified proposed model were measured on the basis of a robust array of statistical measures based on the confusion matrix and ROC analysis. It was tested on hidden samples, and the following measures were calculated:

- Accuracy (ACC): The proportion of correctly classified samples among all test samples.
- Sensitivity (Recall or True Positive Rate, TPR): The model's ability to correctly identify diseased leaves.
- Specificity (True Negative Rate, TNR): This is the capacity of the model to detect healthy leaves.
- Precision (Positive Predictive Value): The percentage of the number of correct positive predictions to the number of positive predictions.
- G-Mean: Geometric mean of sensitivity and specificity, emphasizes the trade-off between identifying diseased and healthy samples.
- F1-Score: the balanced precision and recall value, calculated as a harmonic average of the two.
- ROC: D shows the discriminatory power of the model among classes.

4. Results

The results of the suggested model were tested on the test dataset of the dataset. The model trained on augmented data was found to be more effective and show no overfitting.

Table 4 shows the relative performance of the suggested CNN model on the Drishti-GS1 and ACRIMA datasets with two train–test splits (70:30 and 80:20). The model had accuracies of 87.10% (70:30) and 90.32% (80:20) on the Drishti-GS1 dataset with a perfect sensitivity (100%). This shows that the model has good capability of accurately diagnosing cases of glaucoma, but the specificity values (63.64 percent and 72.73%) are much lower, which can be explained by the fact that the dataset size and distribution are smaller. The model exhibited better results on the bigger and balanced ACRIMA, which scored 94.81 (70:30) and 96.45 (80:20) in accuracy. Sensitivity and specificity were also greater than 95 percent, and this demonstrated how strong the model would be in identifying

glaucoma and reducing false positives. The values of F-score and AUC also support the discriminative force of the model, and the AUC values are more than 0.99 in both splits, which indicates that this model is highly applicable to clinical screening.

Table 4. Performance metrics of the proposed CNN model.

Dataset	Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F-Score (%)	AUC
Drishti-GS1	70:30	87.10	100	63.64	83.33	90.91	0.8181
	80:20	90.32	100	72.73	86.96	93.02	0.8636
ACRIMA	70:30	94.81	95.76	93.62	94.96	95.36	0.9947
	80:20	96.45	96.20	96.77	97.44	96.82	0.9937

Table 5 provides the performance of two popular pretrained CNNs, AlexNet and ResNet-50, on the glaucoma classification problem. With a train–test split of 70:30, AlexNet had an accuracy of 71.43% and sensitivity of 80 percent, but specificity was low (50%), which means that it produces many false positives and falsely classifies healthy cases. At 80:20 division, AlexNet has significantly improved its performance, with 90.07% accuracy, sensitivity of 94.94, and AUC of 0.8940. This indicates that AlexNet can make better use of larger training data, but its accuracy continues to vary between datasets.

Table 5. Performance metrics of pretrained CNN models.

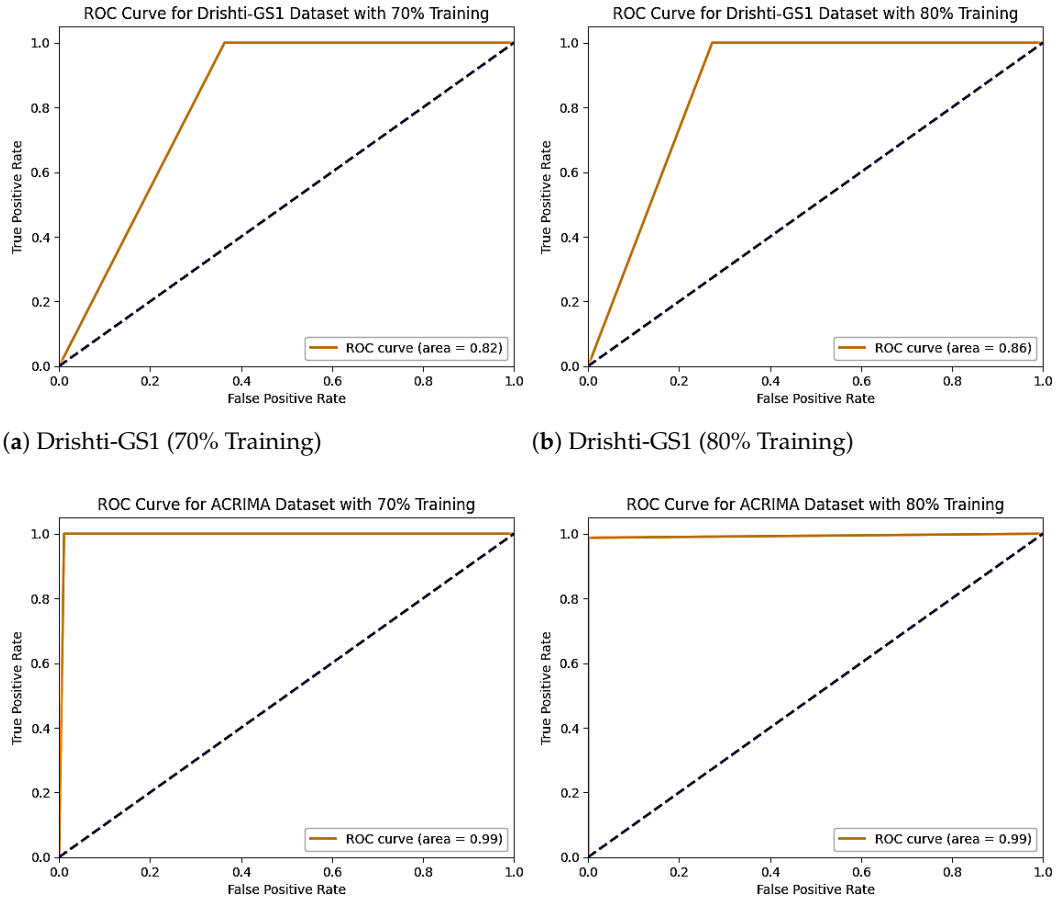
CNN Model	Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F-Score (%)	AUC
AlexNet	70:30	71.43	80.00	50.00	80.00	80.00	0.6500
	80:20	90.07	94.94	83.87	88.24	91.46	0.8940
ResNet-50	70:30	85.71	93.33	66.67	87.50	90.32	0.8000
	80:20	91.49	96.20	85.48	89.41	92.68	0.9084

In most instances, ResNet-50 performed better as compared to AlexNet because it had a more intricate structure and residual connections. The 70:30 percentage produced 85.71% accuracy, 93.33% sensitivity and F-score of 90.32%. ResNet-50 was better, with 80:20, including a precision of 91.49, a sensitivity of 96.20, and an AUC of 0.9084. These findings indicate that ResNet-50 has a superior sensitivity and specificity ratio over AlexNet, so it is more reliable when it comes to glaucoma classification operations. Both AlexNet and ResNet-50 performed better as compared to the proposed customized CNN model. In the Drishti-GS1 dataset, the updated CNN obtained accuracies of 87.10% (70:30) and 90.32% (80:20), which are higher than AlexNet and ResNet-50. In both splits the sensitivity of the suggested approach was 100%, meaning that it did not miss any cases of glaucoma, which is essential in clinical screening. The modified CNN performed better on the balanced ACRIMA dataset compared to AlexNet and ResNet-50, with precision rates of 94.81% (70:30) and 96.45% (80:20), respectively. Moreover, the tailored CNN achieved higher AUC values than 0.99, including 0.8940 in the case of AlexNet and 0.9084 in the case of ResNet-50, indicating that it has better discriminatory ability.

In general, the use of pretrained CNNs is rather reasonable, but the customized CNN shows a distinct advantage, especially regarding accuracy, sensitivity, and AUC. These findings confirm that a CNN model with dataset-centric training policies is more useful when it comes to glaucoma detection.

Figure 2 shows the ROC Curve Analysis of the Proposed Model on Drishti-GS1 and ACRIMA Datasets. The high sensitivity and AUC values achieved by the proposed

model indicate strong discriminative capability for glaucoma detection. However, such performance metrics may be influenced by dataset-specific characteristics, including class distribution and image acquisition protocols. The consistent performance observed across both the imbalanced Drishti-GS1 dataset and the balanced ACRIMA dataset suggests that the model generalizes well under different data conditions. Nonetheless, slight variations in performance between datasets highlight the impact of dataset composition and reinforce the importance of evaluating glaucoma detection models on diverse datasets.



(**a**) Drishti-GS1 (70% Training) (**b**) Drishti-GS1 (80% Training)

(**c**) ACRIMA (70% Training) (**d**) ACRIMA (80% Training)

Figure 2. ROC Curve Analysis of the Proposed Model on Drishti-GS1 and ACRIMA Datasets under Different Training Splits.

A comparative analytical study of the various existing approaches and the proposed customized CNN model in detecting glaucoma is shown in Table 6. Previous researchers have utilized various approaches, including classical CNN models all the way to explainable and hybrid AI models. As an example, Ref. [6] used GoogleNet, where a region of interest (ROI)-based classification strategy was employed, and the accuracy was 95.31%. Ref. [8] and Ref. [10] also suggested the modified version of the GoogLeNet framework and achieved 86.40 and 92.96% accuracies, respectively. More current developments, such as CoG-NET (Ref. [12]) and Self-ONNs (Ref. [5]), have achieved high scores of 95.30 and 94.50, respectively. Ref. [4] tested ImageNet-pretrained models (VGG16 and VGG19), but the performance was not so high (70.21%).

Recent lightweight and attention-based approaches reported in the literature demonstrate competitive performance in glaucoma detection. The hybrid MobileNet-based framework proposed by Ref. [16] achieved a maximum classification accuracy of 94.09%, with the MobileNetv2 combined with an SVM classifier showing the best results. Ref. [17] intro-

duced the MaxGlaViT vision transformer model, which demonstrated high classification accuracy with improved parameter efficiency, particularly excelling in early-stage glaucoma detection. In another study, Ref. [18] employed a lightweight attention-augmented CNN integrated with squeeze-and-excitation and Grad-CAM explainability and achieved an accuracy of 87.9% on a multi-class retinal fundus dataset. Compared to these methods, the proposed attention-enhanced CNN achieved higher accuracy on benchmark glaucoma datasets, with 90.32% on Drishti-GS1 and 96.45% on ACRIMA, while maintaining architectural simplicity and providing interpretable predictions through Grad-CAM. These results support the assumption that the personalized CNN offers an excellent ratio of computational capacity and diagnostic precision and can be used in clinical screening better than generic pretrained CNNs.

Table 6. Performance comparison of existing and proposed methods for glaucoma detection.

Author(s)	Year	Method/Approach	Accuracy (%)
Diaz-Pinto et al. [4]	2019	ImageNet (VGG16, VGG19, etc.)	70.21
Devecioglu et al. [5]	2021	Self-ONNs	94.50
Claro et al. [6]	2019	GoogleNet (ROI + classification)	95.31
Cerentini et al. [8]	2017	Adapted GoogLeNet	86.40
Oguz et al. [10]	2024	CNN + ML classifiers (AdaBoost, SVM, etc.)	92.96
Juneja et al. [12]	2022	CoG-NET	95.30
Cem Baydogan et al. [16]	2025	MobileNetv2 + SVM	94.09
Bose Arnob et al. [18]	2025	MobileNetv2 + SVM	87.90
Proposed Method	-	Customized CNN (fundus images)	96.45 (ACRIMA), 90.32 (Drishti-GS1)

Explainable AI Using Grad-CAM

To improve model transparency and interpretability, Grad-CAM was applied to generate class-discriminative heatmaps for glaucoma predictions. These visual explanations highlight the regions within fundus images that contribute most to the classification decision, enabling validation of whether the model focuses on clinically relevant areas, such as the optic disc and surrounding regions. Figure 3 illustrates the resulting Grad-CAM images taken from the ACRIMA dataset.

Let y^c denote the score for class c before the softmax layer, and let $\mathbf{A}^k \in \mathbb{R}^{H \times W}$ represent the k-th feature map of the final convolutional layer.

The importance weight α_k^c for each feature map is computed by globally averaging the gradients of y^c with respect to $\mathbf{A}^k$:

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \frac{\partial y^c}{\partial A_{ij}^k}. \tag{7}$$

The Grad-CAM localization map is then obtained by computing a weighted combination of the feature maps followed by a ReLU activation:

$$\mathbf{L}_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c \mathbf{A}^k \right). \tag{8}$$

https://doi.org/10.3390/engproc2026124110

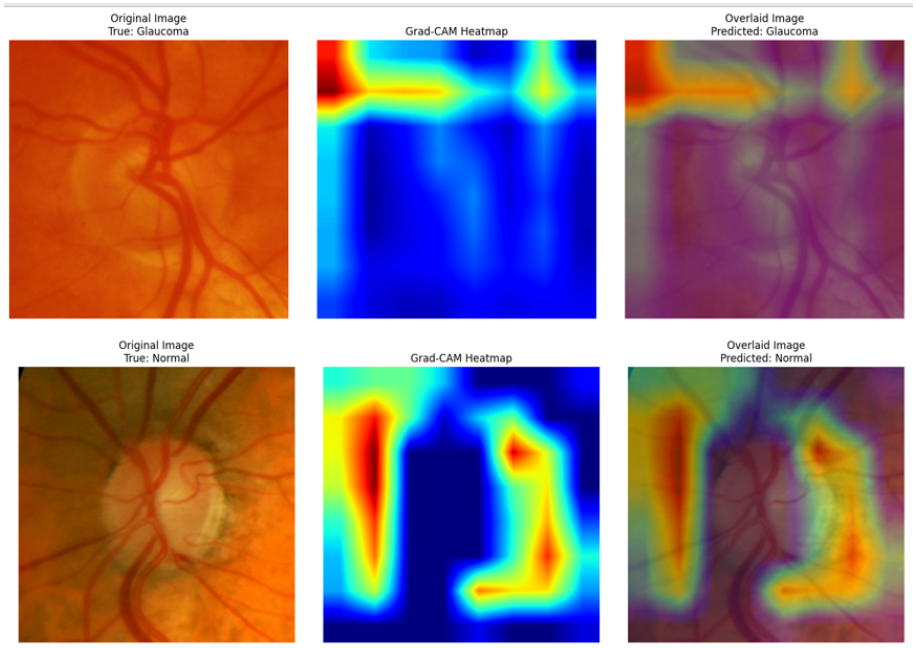


Figure 3. Explainability of the Proposed Model Using Grad-CAM: Warm Colors (Red/Yellow) Highlight Discriminative Regions, Cool Colors (Blue) Indicate Less Contribution.

This explainability component enhances trust in the proposed system and supports its potential use in clinical screening environments.

5. Conclusions

This study presented an attention-enhanced customized convolutional neural network (CNN) for automated glaucoma detection using retinal fundus images. The proposed model was assessed on two datasets, namely Drishti-GS1 and ACRIMA, using various experimental scenarios. Classification results of 90.32% on Drishti-GS1 and 96.45% on ACRIMA, along with high values of sensitivity and AUC, indicate the robustness of the model in correctly classifying glaucoma cases. With the incorporation of the channel attention mechanism, the model was able to selectively focus on clinically significant regions, especially the region surrounding the optic disc, thus improving the learning of discriminative features. Additionally, the Grad-CAM technique was used, which provided explanations regarding the predictions made by the model. It was observed during the comparative evaluation of the proposed method with pretrained CNN architectures like AlexNet and ResNet-50, as well as other state-of-the-art techniques, that the proposed method performs better. It is worth mentioning that the high values of sensitivity on both datasets indicate that glaucoma cases are correctly identified and not missed during the screening process. Thus, the results indicate the effectiveness of the task-specific CNN architecture, along with attention and explainable AI techniques, in surpassing the performance of pretrained models, thus providing a reliable solution for glaucoma screening.

Author Contributions: Conceptualization, V.K.V. and J.V.; methodology, V.K.V.; software, J.V.; validation, J.V., J.N.S. and N.G.; formal analysis, V.K.V.; investigation, M.V.R. and J.R.P.; resources, J.V.; data curation, S.A.; writing—original draft preparation, J.V.; writing—review and editing, V.K.V.; project administration, V.K.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: In this paper two publicly available datasets have been utilized, Drishti-GS1 and ACRIMA. These are downloadable by the links in the subsection referred to as Section 3.1.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CNN	Convolutional Neural Network
DL	Deep Learning
ML	Machine Learning
OD	Optic Disc
OC	Optic Cup
CDR	Cup-to-Disc Ratio
AUC	Area Under the Curve
VGG	Visual Geometry Group
WHO	World Health Organization

References

1. Bourne, R.R.; Stevens, G.A.; White, R.A.; Smith, J.L.; Flaxman, S.R.; Price, H.; Jonas, J.B.; Keeffe, J.; Leasher, J.; Naidoo, K.; et al. Causes of vision loss worldwide, 1990–2010: A systematic analysis. *Lancet Glob. Health* **2013**, *1*, e339–e349. [CrossRef] [PubMed]
2. Velpula, V.K.; Sharma, L.D. Automatic glaucoma detection from fundus images using deep convolutional neural networks and exploring networks behaviour using visualization techniques. *SN Comput. Sci.* **2023**, *4*, 487. [CrossRef]
3. Fernandez, D.C. Delineating fluid-filled region boundaries in optical coherence tomography images of the retina. *IEEE Trans. Med. Imaging* **2005**, *24*, 929–945. [CrossRef] [PubMed]
4. Diaz-Pinto, A.; Morales, S.; Naranjo, V.; Köhler, T.; Mossi, J.M.; Navea, A. CNNs for automatic glaucoma assessment using fundus images: An extensive validation. *Biomed. Eng. Online* **2019**, *18*, 29. [CrossRef] [PubMed]
5. Devecioglu, O.C.; Malik, J.; Ince, T.; Kiranyaz, S.; Atalay, E.; Gabbouj, M. Real-time glaucoma detection from digital fundus images using self-onns. *IEEE Access* **2021**, *9*, 140031–140041. [CrossRef]
6. Claro, M.; Veras, R.; Santana, A.; Araujo, F.; Silva, R.; Almeida, J.; Leite, D. An hybrid feature space from texture information and transfer learning for glaucoma classification. *J. Vis. Commun. Image Represent.* **2019**, *64*, 102597. [CrossRef]
7. Sonti, K.; Dhuli, R. Shape and texture based identification of glaucoma from retinal fundus images. *Biomed. Signal Process. Control* **2022**, *73*, 103473. [CrossRef]
8. Cerentini, A.; Welfer, D.; Cordeiro d'Ornellas, M.; Pereira Haygert, C.J.; Dotto, G.N. Automatic identification of glaucoma using deep learning methods. In *MEDINFO 2017: Precision Healthcare through Informatics*; IOS Press: Amsterdam, The Netherlands, 2017; pp. 318–321.
9. Velpula, V.K.; Vadlamudi, J.; Kasaraneni, P.P.; Kumar, Y.V.P. Automated glaucoma detection in fundus images using comprehensive feature extraction and advanced classification techniques. *Eng. Proc.* **2024**, *82*, 33.
10. Oguz, C.; Aydin, T.; Yaganoglu, M. A CNN-based hybrid model to detect glaucoma disease. *Multimed. Tools Appl.* **2024**, *83*, 17921–17939. [CrossRef]
11. Serte, S.; Serener, A. A generalized deep learning model for glaucoma detection. In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
12. Juneja, M.; Thakur, S.; Uniyal, A.; Wani, A.; Thakur, N.; Jindal, P. Deep learning-based classification network for glaucoma in retinal images. *Comput. Electr. Eng.* **2022**, *101*, 108009. [CrossRef]
13. Velpula, V.K.; Sharma, D.; Sharma, L.D.; Roy, A.; Bhuyan, M.K.; Alfarhood, S.; Safran, M. Glaucoma detection with explainable AI using convolutional neural networks based feature extraction and machine learning classifiers. *IET Image Process.* **2024**, *18*, 3827–3853. [CrossRef]
14. Singh, L.K.; Khanna, M. A novel multimodality based dual fusion integrated approach for efficient and early prediction of glaucoma. *Biomed. Signal Process. Control* **2022**, *73*, 103468. [CrossRef]
15. de Moura Lima, A.C.; Maia, L.B.; Pereira, R.M.P.; Junior, G.B.; de Almeida, J.D.S.; de Paiva, A.C. Glaucoma diagnosis over eye fundus image through deep features. In *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4.
16. Baydogan, C. Lightweight Deep Learning Architectures for Ophthalmic Disease Detection: MobileNet Variants Applied to Glaucoma Classification. *Electron. Lett. Sci. Eng.* **2025**, *21*, 123–135.
17. Yurdakul, M.; Uyar, K.; Taşdemir, Ş. MaxGlaViT: A Novel Lightweight Vision Transformer-Based Approach for Early Diagnosis of Glaucoma Stages From Fundus Images. *Int. J. Imaging Syst. Technol.* **2025**, *35*, E70159. [CrossRef]
18. Arnob, A.K.B.; Chayon, M.H.R.; Al Farid, F.; Husen, M.N.; Ahmed, F. A Lightweight CNN for Multiclass Retinal Disease Screening with Explainable AI. *J. Imaging* **2025**, *11*, 275. [CrossRef] [PubMed]

19. Vocaturo, E.; Zumpano, E.; Giallombardo, G.; Miglionico, G. A multiple instance learning approach for the automatic classification of skin lesions. In *29th Italian Symposium on Advanced Database Systems, SEBD 2021*; CEUR-WS: Bonn, Germany, 2021; Volume 2994.
20. Vocaturo, E.; Zumpano, E. Diabetic retinopathy images classification via multiple instance learning. In *2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*; IEEE: Piscataway, NJ, USA, 2021.
21. Vocaturo, E.; Zumpano, E. AI for the detection of the diabetic retinopathy. In *Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector*; Springer International Publishing: Cham, Switzerland, 2022; pp. 129–140.
22. Velpula, V.K.; Roy, A.; Sharma, L.D.; Sharma, D. Fundus image datasets: A valuable resource for deep neural network-based glaucoma research. In *Intelligent Computation and Analytics on Sustainable Energy and Environment*; CRC Press: Boca Raton, FL, USA, 2024; pp. 84–89.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/engproc2026124110