Alfarama Journal of Basic & Applied Sciences

Faculty of Science Port Said University

https://ajbas.journals.ekb.eg
ajbas@sci.psu.edu.eg

http://sci.psu.edu.eg/en/

January 2025, Volume 6, Issue I

DOI:https://doi.org/10.21608/AJBAS.2024.324901.1232

ISSN 2682-275X

Submitted: 07/10/2024
Accepted: 26/11/2024

Pages: 75 - 95

Enhancing Glaucoma Detection Using Convolutional Neural Networks: A Comparative Study of Multi-Class and Binary Classification Approaches

Walaa Hassan Hagar[1,*], Nabila Eladawi[2], Dalia sabry[3] and Hosam Salaheldin[1]

[1] Biophysics Research group Physics Department, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
[2] Information Systems Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt.
[3] Ophthalmology Department, Faculty of Medicine, Mansoura University, Mansoura 35516, Egypt.

***Corresponding author: walaahassan201610511@gmail.com**

ABSTRACT

Glaucoma is a leading cause of irreversible blindness globally, primarily characterized by progressive damage to the optic nerve, often associated with elevated intraocular pressure. Early detection is critical to preventing vision loss; however, traditional diagnostic methods depend on specialized equipment and skilled personnel. To address these limitations, this study evaluates and compares Conventional Neural Networks (CNNs) for glaucoma detection, utilizing multi-class and binary classification approaches. Specifically, it investigates the effectiveness of ResNet-50 and DenseNet-201 architectures in classifying retinal images. Additionally, the study assesses the interpretability of these models through Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations, providing insights into how each architecture identifies vital features associated with glaucoma. By integrating advanced CNN architectures and interpretability techniques, this research aims to enhance early glaucoma detection and contribute to more accessible diagnostic methods. For binary classification, the combined ResNet-50 and DenseNet-201 models achieved a precision of 1, recall of 0.92, Specificity of 1, F1 score of 0.958, and accuracy of 0.961. For multi-class classification, the models yielded a precision of 0.8889, recall of 0.8421, Specificity of 0.935, F1 score of 0.8649, and accuracy of 0.9074. Grad-CAM visualizations provided insights into the models' focus areas and decision rationale. The binary classification approach demonstrated superior performance to the multi-class approach, indicating its potential for practical application in glaucoma detection. The use of Grad-CAM enhanced model interpretability, supporting the clinical applicability of AI-driven diagnostic tools. These findings highlight the effectiveness of CNN-based methods in improving glaucoma diagnosis and emphasize the need for further research to optimize and validate these techniques in clinical settings.

Keywords: Glaucoma, Convolutional Neural Networks, ResNet-50, DenseNet-201, Grad-CAM.

1. INTRODUCTION

Glaucoma is a complex, progressive eye condition characterized by damage to the optic nerve, typically linked to increased intraocular pressure (IOP) [1]. As one of the leading causes of irreversible blindness worldwide, it affects millions of individuals. The disease often progresses without noticeable symptoms in its early stages, making early diagnosis essential to prevent vision loss and ensure timely treatment [1]. Traditional methods for diagnosing glaucoma include visual field testing, optical coherence tomography (OCT), and tonometry. Although these techniques are well-established, they require specialized equipment and trained professionals and are also prone to human error[2].

Traditional diagnostic methods, although effective, have limitations that can impact their reliability and accessibility. Visual field tests can be time-consuming and may produce variable results based on patient cooperation and the operator's expertise [3]. OCT provides high-resolution images but is expensive and requires a trained specialist to interpret the results accurately. Tonometry measures intraocular pressure but does not directly assess the optic nerve, which is critical for accurate glaucoma diagnosis. Therefore, there is a growing need for more efficient and accessible diagnostic tools [3].

The advent of artificial intelligence (AI) and machine learning (ML) has the potential to address some of these limitations by providing automated, accurate, and scalable solutions. AI techniques, particularly CNNs, have performed exceptionally well in various medical imaging tasks. CNNs can automatically learn and extract features from images, improving the accuracy of disease detection and classification [4, 5]. These advancements offer promising opportunities for enhancing glaucoma. CNNs are a class of deep learning algorithms specifically designed for processing structured grid data, such as images. They excel in image classification tasks by automatically learning hierarchical features from raw image data. This ability makes them particularly well-suited for medical image analysis, including glaucoma detection. Several CNN architectures, such as ResNet-50 and DenseNet-201, perform superiorly in various image classification challenges [6, 7]. ResNet-50 is a deep residual network that incorporates skip connections to mitigate the vanishing gradient problem in profound networks. This architecture has achieved remarkable success in various medical imaging tasks due to its robustness and high accuracy [8]. DenseNet-201 uses dense connections between layers to enhance feature propagation and reduce redundancy. This architecture has also demonstrated high performance in medical image classification tasks, making it a valuable tool for glaucoma detection.

The glaucoma detection and classification approaches can be broadly categorized into multi-class and binary classification [10]. Multi-class classification involves categorizing images into multiple stages of the disease, such as normal, early glaucoma, and advanced glaucoma. This approach provides detailed insights into the disease's progression but may face challenges with class imbalance and complexity [10]. On the other hand, binary classification simplifies the problem by distinguishing between normal and glaucomatous conditions, which can be advantageous for practical applications due to its simplicity and higher accuracy [11]. Understanding and interpreting AI models is essential for their acceptance and effective use in clinical settings. Techniques like Grad-CAM provide visual explanations of model decisions by highlighting regions of an image that contribute to the model's predictions. This interpretability helps clinicians understand the basis of AI-driven diagnoses and build trust in automated systems [12].

2. PREVIOUS WORKS

The application of CNNs to glaucoma detection has been the subject of extensive research. This section investigates foundational studies and recent advances in the field, presenting a comprehensive overview of associated research through 2024. The application of CNNs for glaucoma detection started with studies that concentrated on binary classification. Early studies highlighted the potential of convolutional neural networks (CNNs) in detecting glaucoma from fundus images, demonstrating promising results [13]. Similarly, deep CNN models have been developed that outperform traditional classifiers, laying the groundwork for further research. Multi-class classification approaches have also

been explored to provide detailed insights into glaucoma progression [14]. Early work in this area utilized CNNs to classify images into normal, early, and advanced glaucoma categories [15]. Building on these efforts, more advanced deep learning models have been developed, achieving high accuracy in multi-class classification by integrating patient demographics with image data [16].Hybrid CNN models with spatial attention mechanisms have recently been proposed, significantly enhancing multi-class classification accuracy [17]. Multi-view CNN architectures have been introduced, aggregating information from various retinal views to enhance the robustness and performance of glaucoma classification [18].

Binary classification is widely used due to its simplicity and high diagnostic accuracy. CNN-based binary classifiers have been developed, achieving excellent sensitivity and specificity, further demonstrating the efficacy of deep learning in glaucoma detection [19]. Ensemble methods combining predictions from multiple deep CNN architectures have been introduced to further enhance classification accuracy [20]. An ensemble of lightweight CNNs optimized for mobile-based glaucoma detection has been proposed, providing a practical solution for low-resource settings [21]. Additionally, transfer learning with pre-trained CNNs has been employed to achieve high accuracy in binary classification, even with limited training data. Pre-trained architectures such as ResNet and DenseNet have become foundational in glaucoma detection models [22]. Researchers have demonstrated the effectiveness of ResNet-50 in glaucoma classification, achieving state-of-the-art performance [23]. DenseNet-201 has been recognized for its ability to capture complex retinal features, making it particularly effective in medical imaging tasks [24]. A variant of DenseNet-201 incorporating self-supervised learning has been introduced, enhancing feature extraction on small, imbalanced datasets. Model interpretability is crucial for the clinical adoption of AI systems, with Grad-CAM emerging as a popular method for visualizing CNN predictions [25]. Grad-CAM has been applied to glaucoma detection, offering valuable insights into the model's decision-making process [26]. Grad-CAM has also been integrated into the training process, enhancing both the accuracy and interpretability of glaucoma detection models [27]. In 2024, Grad-CAM was combined with Local Interpretable Model-Agnostic Explanations (LIME) to enhance the transparency of AI-driven diagnostic tools, making them more acceptable to clinicians. The field continues to see the development of innovative techniques aimed at improving glaucoma detection [28]. A deep reinforcement learning approach has been developed to optimize CNN architectures for glaucoma detection, demonstrating significant improvements over conventional methods [29]. A novel data augmentation technique using generative adversarial networks (GANs) has been introduced to address class imbalance in glaucoma datasets [30].

A diffusion model-based data synthesis method has been proposed, outperforming traditional GANs in generating high-quality synthetic retinal images for training CNNs. This approach effectively addresses class imbalance, a common challenge in multi-class classification [31]. The importance of multi-center datasets for developing robust and generalizable models has been emphasized. Their study explored federated learning, which allows CNNs to be trained on diverse datasets from different regions, enhancing the model's generalizability across populations [32].

Recent studies have explored cross-modality learning, where models are trained on multiple types of data. A multi-task learning framework has been proposed, combining fundus images and OCT data to enhance glaucoma detection. Their approach highlights the benefits of leveraging complementary information from multiple imaging modalities [33]. A graph-CNN approach has been introduced to model the structural relationships within retinal images, providing a novel perspective on feature extraction and disease classification. This method outperforms traditional CNNs by effectively capturing the spatial dependencies in retinal data [34].

Transfer learning has significantly enhanced CNN performance in glaucoma detection. Transfer learning has been applied to adapt pre-trained CNNs for glaucoma classification, achieving high accuracy even with limited training data [35]. This method is particularly valuable in low-resource settings where large, annotated datasets are unavailable. Additionally, domain adaptation techniques have been introduced to improve CNN performance on out-of-distribution data, enhancing the robustness of models

for real-world clinical applications. The development of real-time glaucoma detection systems has been a significant focus in recent years [36]. Compact Self-Organized Operational Neural Networks (Self-ONNs) have been introduced for early glaucoma detection in fundus images, demonstrating superior performance compared to traditional CNNs on datasets such as ACRIMA, RIM-ONE, and ESOGU. These networks provide improved detection accuracy with lower computational complexity, making them particularly well-suited for biomedical applications, especially when data availability is limited [37].

A real-time glaucoma screening system called EyeDr. was developed using Darknet YOLO for optic disc and cup detection. This system, based on criteria like the Vertical Cup to Disc Ratio (VCDR ≥ 0.6) and the ISNT rule, was trained on 1,677 images and validated on 500 images. The AI achieved 92% accuracy, with sensitivity, specificity, and precision exceeding 91%, and can screen for glaucoma in 8 to 12 seconds. Despite these advancements, several challenges remain in the field of glaucoma detection using CNNs. These include the need for large, annotated datasets, the development of more interpretable models, and the integration of AI tools into clinical practice. Future research should focus on addressing these issues to fully harness the potential of CNNs in glaucoma diagnosis [38]. The integration of AI-driven diagnostic tools into electronic health record (EHR) systems was examined, highlighting the importance of interoperability between AI systems and existing healthcare infrastructure [39].

This study aims to compare the performance of multi-class and binary classification approaches for glaucoma detection using advanced CNN architectures, specifically ResNet-50 and DenseNet-201. We evaluate these models' accuracy, precision, recall, and F1 score, and use Grad-CAM visualizations to interpret their decision-making processes. By addressing these objectives, the study seeks to enhance understanding of the effectiveness of different classification approaches and provide insights into the practical application of AI in glaucoma diagnosis.

3. METHODS

The framework for this study includes multiple essential elements to ensure efficient glaucoma detection. The data preparation was performed via loading, augmenting retinal image in order to enhance diversity. Pre-processing steps are followed by employing pre-trained CNNs such as ResNet-50 and DenseNet-201 for feature extraction. These models are combined in an ensemble approach to improve classification accuracy. The Adam optimizer is utilized for efficient training. Evaluation metrics are used to assess performance, including accuracy, sensitivity, specificity, and F1-score. Additionally, Grad-CAM is employed to provide visual interpretability of model predictions. The framework shown in (Fig. 1)describes the ensemble approach with individual models to validate its effectiveness in glaucoma detection.

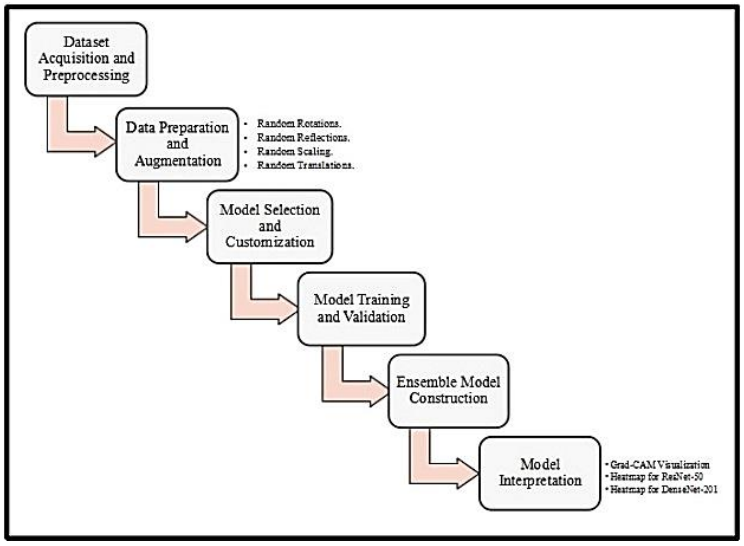

Fig (1): Architecture of the proposed system.

3.1. Data Preparation

The dataset was loaded using MATLAB® image Datastore function, which efficiently manages large image datasets. It was then randomly divided into training and testing sets, with 70% of the images used for training and 30% for testing.

3.2. Data Augmentation

To enhance the model's robustness and mitigate overfitting, data augmentation was applied to the training images, exposing the model to various scenarios. The augmentation techniques used are shown in (Fig. 2) as follows

Random Rotations (-30° to 30°): Address variations in image orientation due to patient positioning or camera alignment differences. The model was trained to recognize glaucomatous features irrespective of their angular disposition by rotating the images within this specified range. Random rotations were applied to increase the robustness of the model to variations in image orientation. This ensures that the model can recognize glaucoma features regardless of how the images are rotated, which is particularly important given that retinal images can be captured at various angles. The model learns to be invariant to such transformations by introducing rotation as a variability during training, thereby improving its generalization capabilities.

Random Horizontal and Vertical Reflections: To simulate natural variations and ensure the model is not biased toward any specific orientation. Horizontal and vertical reflections were used to expose the model to flipped versions of the images. This augmentation simulates real-world scenarios where the eye's orientation can vary due to differences in patient positioning or imaging device. Reflections help the model generalize better by making it less sensitive to the orientation of the input images, ensuring that glaucoma features are correctly identified even when the images are mirrored.

Random Scaling (0.7 to 1.3): To account for variations in the apparent size of retinal structures due to differences in the distance between the eye and the imaging device. This ensured that the model remained scale-invariant, improving its performance across images captured under different conditions. This augmentation simulates different magnification levels that might occur during image acquisition, ensuring that the model can handle images captured at various scales. By learning from scaled versions of the images, the model becomes more robust to variations in image size and can accurately detect glaucoma features across different resolutions.

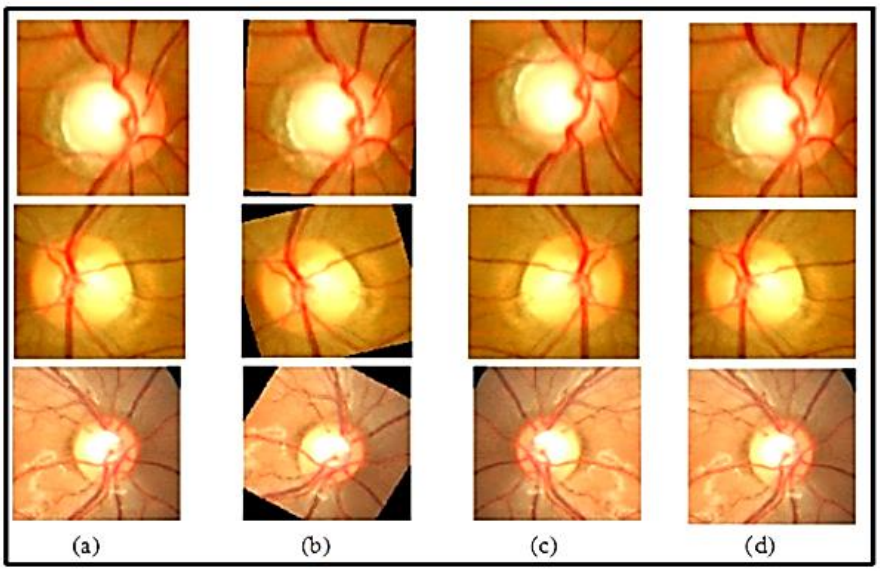

Fig (2): Data argumentation visualization: (a) input image, (b) Random rotation, (c) Random flips, and (d) Random scaling.

These augmentations were implemented to generate a more diverse training dataset, thereby improving the model's ability to generalize across real-world variations.

3.3. Model Selection and Modification

Two pre-trained CNN architectures depicted in (Fig. 3) were selected for this study: **ResNet-50** and **DenseNet-201**. These models were chosen due to their proven effectiveness in image classification tasks. Both models were fine-tuned by replacing the final fully connected layer and the classification layer to match the number of classes in the dataset.

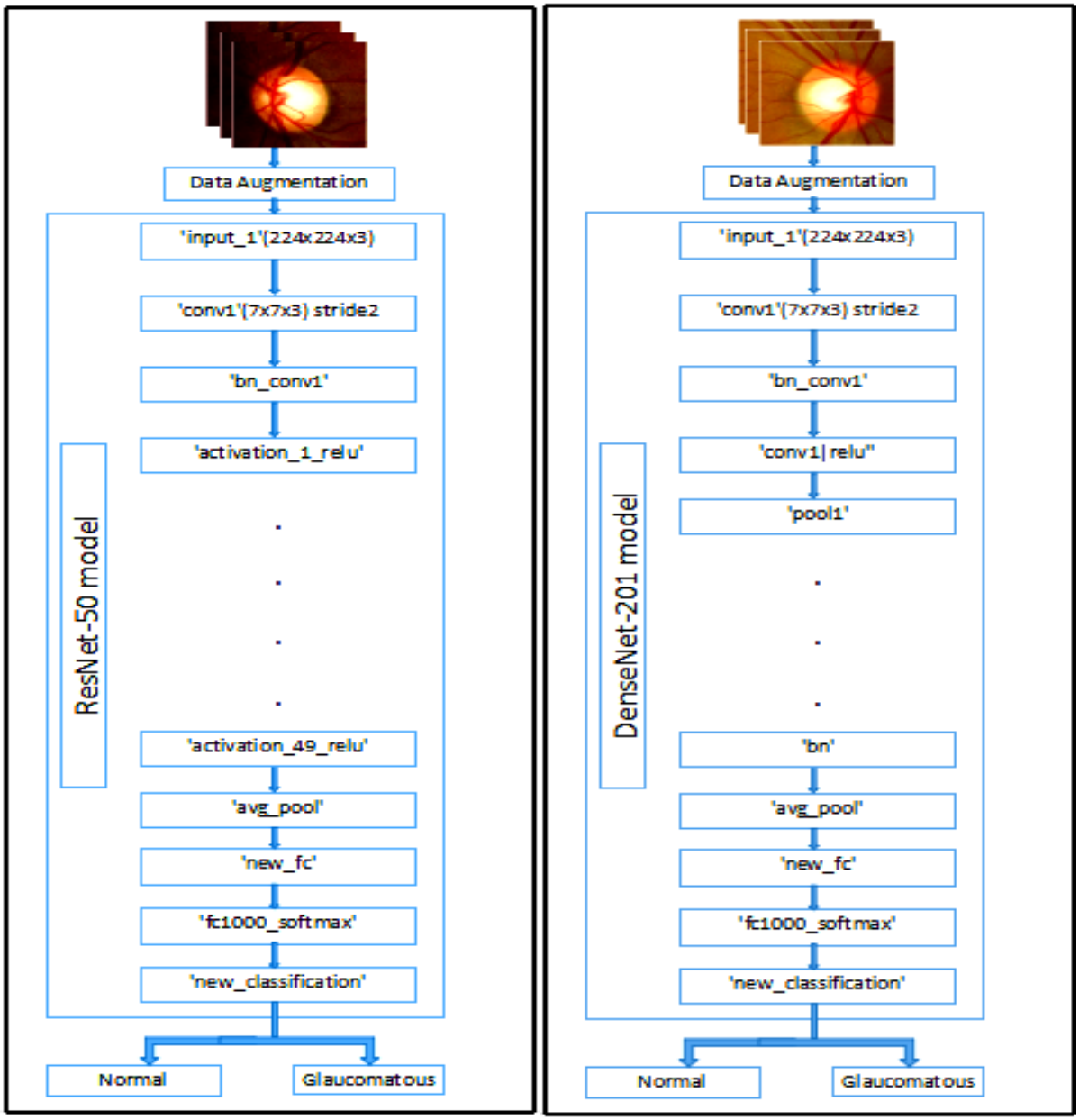

Fig (3): Architecture of the ResNet-50 and DenseNet-201 models.

3.4. Model Training

The modified models were trained using the Adam optimizer, selected for its computational efficiency and capability to handle sparse gradients. Adam merges with the advantages of AdaGrad, which is effective with sparse gradients, and RMSProp, which performs well in online settings. By adjusting the learning rate for each parameter individually, Adam is well-suited for optimizing deep learning models, such as the CNNs used in this study. This adaptiveness helps the model converge faster and more reliably, which is crucial when training on complex and diverse medical images. The training was conducted for 13 epochs with a mini-batch size of 32, and the initial learning rate was set to 0.0001. Model performance was validated using the testing set every 30 iterations.

3.5. Ensemble Model Prediction

To improve classification performance, an ensemble model was developed by averaging the prediction probabilities from both the ResNet-50 and DenseNet-201 models. The final predicted label for each image was assigned based on the class with the highest average probability. The corresponding mathematical equation for averaging the results is the Final Probability which can be defined as the average probability for class *j* for the *i*-th image, which is used to determine the final predicted label by selecting the class with the highest value.

$$\text{Final Probability} = \frac{1}{2}(P_{ResNet}(i,j) + P_{DenseNet}(i,j)) \quad (1)$$

Where $P_{ResNet}(i,j)$ is the predicted probability for class *j* for the *i*-th image using the ResNet-50 model and $P_{DenseNet}(i,j)$is the predicted probability for class *j* for the *i*-th image using the DenseNet-201 model.

3.6. Model Evaluation

The models were evaluated using standard performance metrics to ensure a comprehensive assessment of their capabilities. Metrics such as accuracy, sensitivity, specificity, and F1 score were used to evaluate the models' performance across both binary and multi-class classification tasks. These metrics provide insights into the models' ability to correctly classify images into normal, early glaucoma, and advanced glaucoma categories, as well as their overall effectiveness in detecting glaucoma. The evaluation process was conducted consistently and systematically to ensure the reliability of the results.

3.7. Model Interpretation using Grad-CAM

To gain a deeper understanding of the model's predictions, Grad-CAM was used. Grad-CAM is a visualization technique that helps interpret the decisions of CNNs by highlighting the regions in the input images that most influence the model's predictions. This method creates class-specific heatmaps that indicate where the network is focusing its attention, making it easier to understand which parts of an image are most relevant for classification.

In this study, Grad-CAM was applied to both ResNet-50 and DenseNet-201 models. For each model, heatmaps were generated to visualize the regions in the input images that most significantly contributed to the predicted class. These heatmaps offer insights into the spatial patterns recognized by the models, which are particularly valuable for medical imaging tasks like glaucoma detection.

To enhance interpretability further, a combined heatmap was created. This involves overlaying the heatmaps from the individual models onto the original images, providing a comprehensive view of how different models interact with the same input data. The combined heatmap highlights the areas where the models converge in their attention, offering a clearer picture of which features are consistently deemed necessary across different architectures. Grad-CAM visualizations were created by blending the generated heatmaps with the original images, as illustrated in (Fig. 4), to highlight which parts of the retinal images each model emphasizes more.

By utilizing Grad-CAM in this manner, the study aims to offer a more transparent view of the ensemble model's decision-making process. This approach not only helps verify the model's focus and understand potential biases but also aids in validating the effectiveness of the model's predictions by correlating them with clinically relevant features in the images.

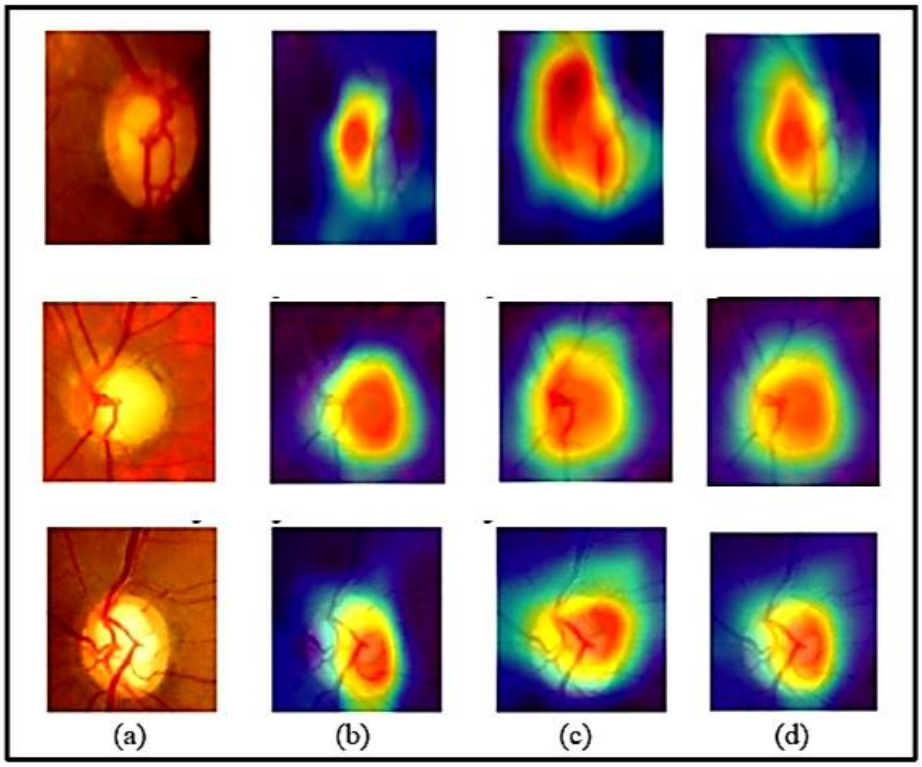

Fig (4): Grad-Cam Visualization: : (a) input image, (b)Blended Image with Grad-CAM (ResNet),

(c) Blended Image with Grad-CAM (DenseNet), and (d) Blended Image with Combined Heatmap.

4. RESULTS AND DISCUSSION

While the dataset provided sufficient variability for initial evaluations, its size and class imbalance pose limitations that may affect model generalizability and performance. The dataset, named 'V1,' ,available online on the Harvard Dataverse (doi:10.7910/DVN/1YRRAC), comprises a total of 1,544 retinal fundus images categorized into three classes: Normal (51%, 788 images), Early Glaucoma (19%, 289 images), and Advanced Glaucoma (30%, 467 images). For binary classification tasks, the Early and Advanced Glaucoma categories were merged into a single 'Glaucomatous' class, creating two categories: Normal and Glaucomatous.

Despite providing valuable insights, the dataset size limits the model's ability to generalize effectively to unseen data from diverse populations. Retinal image variations, stemming from demographic differences, imaging equipment, and clinical conditions, may not be fully captured in this dataset. The limited sample size increases the risk of overfitting, where the model performs well on the training data but struggles to generalize to new images. To address this, data augmentation techniques, such as random rotations, reflections, and scaling, were applied. While these methods improved robustness by simulating variations encountered in real-world scenarios, they cannot fully replicate the benefits of larger, diverse datasets that naturally encompass such variability.

Evaluation Metrics

The performance metrics for the binary and multi-class classification tasks can be derived from specific equations applied to the confusion matrix of the predictions. Here are the equations for each metric:

Binary Classification:

Precision is the ratio of true positives (TP) to the total number of optimistic predictions. True Classifications (TP) are accurate classifications. False positive (FP) measures or counts false positives and classifies negative cases as good. Estimation of model precision was performed using Eq. (2).

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

Recall (Sensitivity or True Positive Rate)

The recall (FN) is the ratio of true positives (TPs) to all true positives and false negatives. The false negative (FN) value refers to the count of true positive cases incorrectly classified as negative.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

Evaluation of performance

To assess the accuracy, specificity, and sensitivity of Glaucoma classification in distinguishing the subjects. Both sensitivity and specificity are quantitative metrics that provide evidence of the importance Of a test in determining the presence or absence of disease.
Calculations for these two parameters are performed using equations (4) and (5) correspondingly.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{4}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

F1 Score

The F1 Score is mathematically defined as the weighted average of Precision and Recall. Thus, this metric considers both true positives and false negatives. While F1 may not be as intuitively comprehensible as accuracy, it is often more valuable than accuracy, mainly when dealing with uneven class-distributed data, and can be estimated from Equation (6). The optimal accuracy is achieved when the costs of false positives and false negatives are comparable.

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

Multi-Class Classification:

For multi-class classification, the same metrics can be calculated for each class, and the overall metrics can be averaged (often weighted by the number of samples in each class).

Given:

TP_i: True positives for class iii (e.g., correctly classified "Normal," "Early Glaucoma," or "Advanced Glaucoma").

FP_i: False positives for class iii (e.g., other classes incorrectly classified as class iii).

FN_i: False negatives for class iii (e.g., class iii incorrectly classified as other classes).

N: Total number of samples.

Precision (for class $\boldsymbol{i}$)**:**

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{7}$$

The overall precision for multi-class classification is often the average of the class-wise precision.

Recall (for class $\boldsymbol{i}$)**:**

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{8}$$

The overall recall for multi-class classification is often the average of the class-wise recall.

Specificity(for class $\boldsymbol{i}$) :

$$\textbf{Specificity}_i = \frac{\boldsymbol{TN_i}}{\boldsymbol{TN_i + FP_i}} \tag{9}$$

The overall **Specificity** for multi-class classification is often the average of the class-wise **Specificity**.

F1 Score (for class $\boldsymbol{i}$)**:**

$$\text{F1 Score}_\text{i} = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{10}$$

The overall F1 score for multi-class classification is often the average of the class-wise F1 scores.

Accuracy**:**

$$\text{Accuracy} = \frac{\sum_i TP_i}{N} \tag{11}$$

The overall accuracy considers the correct predictions across all classes.

Binary Classification Results

In the binary classification task, the dataset was split into two classes: "Normal" and "Glaucomatous," where the "Glaucomatous" class included both "Early Glaucoma" and "Advanced Glaucoma" images. The ensemble model, which combined the predictions from the ResNet-50 and DenseNet-201 models, demonstrated high performance in this classification task.

1) **Precision**: The model achieved a high Precision of **1**, indicating that all the images classified as "Glaucomatous" were correctly identified.
2) **Recall**: The Recall was **0.92**, showing the model's effectiveness in identifying "Glaucomatous" cases, though there were some instances that were not detected.
3) Specificity: A specificity of 1 reflects the perfect performance of a classifier in identifying negative cases.
4) **F1 Score**: The F1 Score of **0.958** reflects a balanced performance, combining both Precision and Recall.
5) **Accuracy**: The overall Accuracy for the binary classification was **0.961**, demonstrating the model's ability to correctly classify many images.

These results highlight the ensemble model's effectiveness in distinguishing between normal and glaucomatous images, making it a robust tool for binary classification in glaucoma detection.

Multi-Class Classification Results

For the multi-class classification task, the dataset was divided into three distinct classes: "Normal," "Early Glaucoma," and "Advanced Glaucoma." The ensemble model was also evaluated on this more complex classification task.

1) **Precision:** The model achieved a Precision of **0.8889** across all classes, indicating a slightly lower yet still strong ability to identify each class correctly compared to the binary task.
2) **Recall**: The Recall was **0.8421**, reflecting the model's capacity to detect images across the three classes, with some misclassification cases.
3) **Specificity:** A specificity of **0.935** indicates that the classifier is highly effective at correctly identifying negative cases, with only a small proportion of negative cases being misclassified as positive. This level of specificity is generally considered excellent and reflects a well-performing classifier in terms of reducing false positives.
4) **F1 Score**: The F1 Score was **0.8649**, suggesting a balanced performance in multi-class classification.
5) **Accuracy**: The overall Accuracy was **0.9074**, demonstrating that the model correctly classified most images across all three classes.

The multi-class classification results indicate that while the model performs well, the complexity of distinguishing between different stages of **glaucoma** presents a more challenging task compared to binary classification.

Confusion Matrix Analysis

Confusion matrices were generated for binary and multi-class classification tasks to understand the model's performance better. The confusion matrix presents a summary of the prediction results, allowing for identifying patterns in misclassification.

Binary Classification Confusion Matrix: The confusion matrix for the binary classification task revealed that the number of misclassified cases was relatively low, further supporting the high Precision and Recall scores.

Table 1: The confusion matrix for Binary classification.

	Normal	Glaucomatous
Normal	236	0
Glaucomatous	18	208

Multi-Class Classification Confusion Matrix: In the multi-class classification task, the confusion matrix showed more variability in misclassifications, particularly between "Early Glaucoma" and "Advanced Glaucoma" images. This suggests that the model had more difficulty differentiating between these two classes, which is consistent with the lower Precision and Recall scores observed in the multi-class task.

Table 2: The confusion matrix for multi-classification.

	Normal	Early Glaucoma	Advanced Glaucoma
Normal	150	5	0
Early Glaucoma	10	70	7
Advanced Glaucoma	7	35	194

The confusion matrix for multi-class classification revealed challenges in distinguishing between Early and Advanced Glaucoma images, indicating a degree of overlap in the features the model identified for these two classes. These misclassifications can be attributed to the dataset's significant class imbalance. Early Glaucoma, which is underrepresented at 19%, provides fewer examples for the model to learn subtle distinguishing features, leading to reduced sensitivity for this class.

The imbalance likely biased the model towards the majority classes, particularly Normal and Advanced Glaucoma, as reflected in the model's lower precision (0.8889) and recall (0.8421) for multi-class classification tasks. Although merging the Early and Advanced Glaucoma classes into a single 'Glaucomatous' category mitigated these challenges in the binary classification task, achieving high precision (1.0) and recall (0.92), this approach simplifies the problem and limits insights into disease progression.

These observations underscore the need for addressing class imbalance in future work, either by incorporating larger, more balanced datasets or using advanced synthetic data generation techniques such as GANs to supplement underrepresented classes. Additionally, implementing strategies like focal loss or class-weighted training could further improve model sensitivity to minority classes, especially Early Glaucoma.

Grad-CAM Visualization Results

To gain insights into the decision-making processes of the models, Grad-CAM was used to visualize the regions of the images most influential in the model's predictions. Heatmaps were generated for both the ResNet-50 and DenseNet-201 models, providing a comprehensive view of the areas contributing most to the classification decisions, as shown in (Fig. 5).

1) ResNet-50 Heatmaps: The heatmaps generated from the ResNet-50 model predominantly highlighted the optic disc and surrounding regions. These areas were identified as critical in distinguishing between normal and glaucomatous images, aligning with clinical knowledge that emphasizes the optic disc as a key region in glaucoma assessment.

2) DenseNet-201 Heatmaps: In contrast, the DenseNet-201 model’s heatmaps not only emphasized the optic disc but also displayed broader activation across the retinal image. This suggests that DenseNet-201 considers more global features in its decision-making process, possibly reflecting additional retinal features that may be relevant to glaucoma detection.

3) Combined Heatmaps: The combined heatmaps, resulting from an ensemble of both ResNet-50 and DenseNet-201, offered a more nuanced understanding of the models' decision-making. By merging the focus areas of both models, the combined heatmaps revealed a comprehensive assessment of the optic disc and surrounding areas, reinforcing the robustness of the ensemble model in capturing relevant features for glaucoma detection.

These Grad-CAM visualizations were crucial for interpreting the decision-making processes of the models, highlighting regions of retinal fundus images that were most influential in their predictions. Both

models consistently focused on the optic disc and surrounding regions, supporting the clinical relevance of these areas in the diagnosis of glaucoma.

Correlation with Clinical Findings

The heatmaps revealed important correlations with clinical findings, particularly focusing on the optic disc, which is a central feature in glaucoma diagnosis.

1) Focus on the Optic Disc: The Grad-CAM heatmaps prominently highlighted the optic disc, a critical region in clinical glaucoma diagnosis. Changes in the optic disc, such as increased cupping or thinning of the neuroretinal rim, are well-documented indicators of glaucomatous damage. The models' emphasis on these regions confirms their reliance on clinically meaningful visual patterns, enhancing their reliability for real-world diagnostic applications.

2) Broader Activation Patterns in DenseNet-201: Compared to ResNet-50, DenseNet-201 demonstrated broader activation patterns, suggesting that it considers additional retinal features beyond the optic disc. This wider focus may reflect sensitivity to peripheral changes associated with glaucoma progression, such as retinal nerve fiber layer thinning. The ensemble model, combining the strengths of both ResNet-50 and DenseNet-201, balanced localized and global features, likely contributing to its superior performance in distinguishing between normal and glaucomatous conditions.

3) Consistency Across Grad-CAM Results: The consistency of the Grad-CAM visualizations with clinical observations especially the increased attention to the optic disc and associated pathological areas in advanced glaucoma cases validates the interpretability of the models. This consistency builds trust in the models' potential for clinical adoption.

Enhancing Practical Implications

The insights gained from Grad-CAM visualizations are crucial for enhancing the practical utility of AI-driven diagnostic tools in clinical settings.

1) Supporting Clinical Decision-Making: Grad-CAM visualizations serve as supplementary tools for clinicians, offering an additional layer of confirmation for diagnoses and helping identify subtle pathological changes. By visually explaining the models' decisions, these heatmaps increase the interpretability and acceptability of AI-driven tools, thereby supporting clinical decision-making.

2) Identifying Potential Biases: The absence of focus on irrelevant regions in the visualizations highlights the robustness of the models. Future iterations of these models can use similar techniques to detect and address biases that may arise from the training data, ensuring further refinement and fairness in predictions.

3) Future Refinements: To enhance the transparency and clinical utility of the models, combining Grad-CAM with other interpretability techniques, such as Local Interpretable Model-Agnostic Explanations (LIME), could provide a more comprehensive understanding of model decisions. This integration would allow for deeper insights into the decision-making processes, ensuring that AI models can be better trusted and adopted in clinical practice.

These findings demonstrate that the predictions made by the models align with clinically relevant features, reinforcing the potential of these AI models as reliable tools for glaucoma detection. By providing clear visualizations of the decision-making processes, Grad-CAM not only enhances model transparency but also supports clinical trust in AI-powered diagnostic solutions.

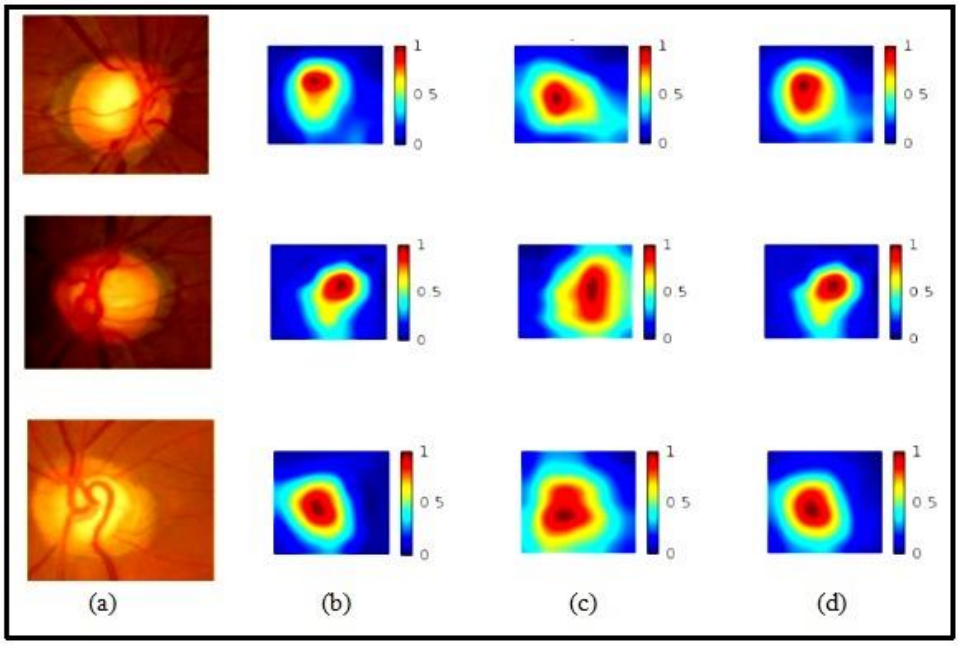

Fig (5): Grad-CAM Visualization Results: (a) input image, (b) ResNet-50 Heatmaps, (c) DenseNet-201 Heatmaps, and (d) Combined Heatmaps.

The Grad-CAM visualizations confirmed that the models were making decisions based on clinically relevant areas of the retinal images, thereby providing confidence in the reliability and interpretability of the ensemble model's predictions.

Comparison with State-of-the-Art Models

This study presents an innovative approach to glaucoma detection by leveraging CNNs, specifically ResNet-50, and DenseNet-201, to enhance diagnostic accuracy. The comparative analysis of multi-class and binary classification approaches in conjunction with Grad-CAM visualization provides a comprehensive understanding of the models' performance and interpretability. The results demonstrate that both ResNet-50 and DenseNet-201 significantly improve glaucoma detection accuracy compared to traditional methods. The application of CNNs for both binary and multi-class classification tasks has shown promising results, aligning with recent advancements in deep learning for medical image analysis. Specifically, the hybrid ResNet-50 and DenseNet-201 models exhibited superior performance in detecting advanced glaucoma, which could be attributed to its dense connectivity and efficient feature reuse. The integration of Grad-CAM for visual interpretability has provided valuable insights into the decision-making process of the models. By overlaying Grad-CAM heatmaps on the input images, we have identified the critical regions influencing the classification outcomes. This aligns with the work of [26], which emphasizes the importance of visual explanations in enhancing the transparency and trustworthiness of deep learning models in medical applications. Recent research has explored various deep-learning techniques for glaucoma detection, each with its strengths and limitations. The comparative analysis of our study's approach with five significant models highlights the advancements achieved in performance demonstrated in Table (3).

Table 3: Comparison table of the proposed model with state-of-the-art models.

Author	Dataset		AUC	Accuracy	Sensitivity	Specificity	F1-Score
Saxena et al.[40]	SCES		88.2%	-	-	-	-
Lin et al. [41]	OHTS LAG		90.4%	93%	-	-	49%
Gomez-Valverde et al. [42]	Local dataset of 2313 images		94%	87.01%	89.01%	89.01%	-
Carlos et al.[43]	RIM-ONE r3		91%	-	-	-	-
Christopher et al. [16]	Local dataset of 14,822 images		97%	88%	95%	95%	-
Proposed Method	Binary classification	V1 Harvard Dataverse	-	96.1%	92.%	100%	95.8%
	Multi classification		-	90.74%	84.21%	93.5%	86.49 %

The proposed ensemble model (ResNet-50 + DenseNet-201) demonstrated exceptional performance in glaucoma detection, particularly in binary classification tasks. The model achieved precision (1), recall (0.92), specificity (1), F1 score (0.958), and accuracy (0.961). In multi-class classification, the metrics included precision (0.8889), recall (0.8421), specificity (0.935), F1 score (0.8649), and accuracy (0.9074). These results compare favorably to or surpass the performance of many state-of-the-art models reported in the literature.

Advantages Over Existing Techniques

1) **Binary Classification Superiority**

Many state-of-the-art models focus on multi-class classification but struggle with class imbalance and overlapping features between stages of glaucoma. Our binary classification strategy addresses these issues by merging Early and Advanced Glaucoma into a single class, resulting in superior performance metrics.

2) **Enhanced Interpretability**

Many existing models lack transparency in decision-making, our integration of Gradient-weighted Class Activation Mapping (Grad-CAM) provides visual explanations for model predictions. This feature enhances clinical trust, making the model more suitable for real-world applications.

Complementing Existing Approaches

1) **Focus on Retinal Fundus Imaging**

Unlike many state-of-the-art techniques that rely on optical coherence tomography (OCT) or advanced imaging modalities, this study highlights the feasibility of achieving high diagnostic accuracy using retinal fundus images alone. This makes the proposed approach more accessible and cost-effective, particularly in resource-limited settings.

2) **Novel Ensemble Model Design**

Ensemble models are underutilized in glaucoma detection. By leveraging ResNet-50's ability to extract subtle local features and DenseNet-201's capability to capture global features, our approach achieves a balance between specificity and sensitivity, improving robustness and overall performance.

Future Directions

1) **Validation on Larger and Diverse Datasets**

Expanding the evaluation to larger datasets with diverse demographics and imaging conditions could provide further evidence of the model's generalizability.

2) **Integration with Multi-Modal Data**

Combining retinal fundus imaging with OCT or other diagnostic modalities may further enhance feature representation and model performance, offering a more comprehensive solution for glaucoma detection.

Clinical implementations

The integration of Convolutional Neural Networks (CNNs) for glaucoma detection into real-world clinical workflows offers significant potential to enhance early diagnosis and prevent irreversible vision loss. Glaucoma, as one of the leading causes of blindness globally, requires timely intervention to mitigate damage to the optic nerve. While the CNN-based models investigated in this study show promising results for both binary and multi-class glaucoma classification, their real-world application in clinical settings, especially in resource-limited environments, comes with several challenges that must be addressed to realize their full potential.

1) **Integration into Clinical Workflows:** A key consideration for implementing AI-driven models like CNNs into clinical practice is how these models can be seamlessly integrated into existing workflows. Current glaucoma diagnostics often rely on specialized equipment such as tonometers and retinal imaging devices, operated by trained clinicians. To be truly effective, AI models must work alongside these traditional methods, providing timely and accurate results that can assist clinicians in making informed decisions. For instance, AI models can be used to analyze retinal images, flagging potential glaucoma cases for further examination by ophthalmologists. The challenge lies in ensuring that AI outputs are presented in a manner that is intuitive for clinicians, allowing them to trust the model's recommendations without replacing their clinical judgment.
2) **Infrastructure and Technology Constraints:** In resource-limited settings, the availability of advanced medical imaging devices and the required computational infrastructure for running CNN models can be a major barrier to implementation. Retinal imaging devices capable of providing high-quality images for analysis are often expensive, and such equipment may not be readily available in low-resource regions. Additionally, the processing power needed to run deep learning models effectively may not be feasible in many healthcare facilities, particularly in rural or underserved areas. For AI models to be useful in these settings, efforts must be made to develop more affordable, portable, and energy-efficient imaging technologies. Cloud-based solutions could be considered to offload computational tasks, enabling local healthcare facilities with limited computational resources to still benefit from AI-powered analysis.
3) **Data Availability and Quality:** Successful deployment of CNN models for glaucoma detection hinges on access to large, high-quality datasets for training and validation. In many resource-limited settings, the availability of annotated retinal images may be scarce, hindering the development of robust AI models. Furthermore, if training data is not representative of the population served in a particular setting (e.g., racial or ethnic differences in retinal characteristics), the model's performance could be compromised. To address this, collaborations between academic institutions, healthcare providers, and AI developers can help generate diverse datasets that reflect the real-world conditions of these regions. In some cases, local healthcare workers could be trained to capture and annotate images, allowing for the creation of tailored datasets that enhance the model's accuracy and generalizability.
4) **Training and Acceptance Among Healthcare Providers:** The introduction of AI models into clinical practice requires adequate training for healthcare providers who will be using these tools. In resource-limited settings, where access to advanced medical training may be limited, this could be a significant barrier. Healthcare providers need to understand not only how to operate the AI tools but

also how to interpret their output effectively. Since AI models often function as decision-support tools, ensuring that clinicians feel confident in using these tools and interpreting their results is crucial. Offering accessible, low-cost training programs through online courses or mobile apps could be an effective way to overcome this challenge.

5) **Cost-Effectiveness and Sustainability:** In resource-limited settings, affordability is a key concern. The cost of implementing AI-based glaucoma detection tools must be considered, including both the initial deployment and ongoing maintenance. If the system is too expensive for widespread adoption, its impact will be limited. However, the potential long-term cost savings associated with early detection and treatment of glaucoma may offset the initial investment. AI models can streamline diagnostic processes, reduce the need for costly human labor, and improve patient outcomes by facilitating early intervention. Public-private partnerships or subsidies from international organizations may be essential to make AI-powered diagnostic tools more affordable and sustainable in low-resource environments.

5. CONCLUSION

This study aimed to enhance glaucoma detection by comparing the effectiveness of binary and multi-class classification approaches using ensemble models based on ResNet-50 and DenseNet-201. Our comprehensive evaluation, employing a range of metrics and visual tools, provides valuable insights into the strengths and limitations of these models in glaucoma detection. For binary classification, the combined ResNet-50 and DenseNet-201 models achieved a precision of 1, recall of 0.92, Specificity of 1, F1 score of 0.958, and accuracy of 0.961. For multi-class classification, the models yielded a precision of 0.8889, recall of 0.8421, Specificity of 0.935, F1 score of 0.8649, and accuracy of 0.9074.

Binary Classification Performance: The ensemble model demonstrated exceptional performance in the binary classification task, achieving a precision of 1, recall of 0.92, Specificity of 1, F1 score of 0.958, and accuracy of 0.961. These results underscore the model’s ability to distinguish between normal and glaucomatous images effectively. The high Precision indicates that most images classified as "Glaucomatous" were accurate, while the high Recall reflects the model's proficiency in identifying glaucomatous cases. The ensemble approach proved to be robust in handling the binary classification task, highlighting its potential for clinical applications where a clear differentiation between normal and glaucomatous conditions is crucial.

Multi-Class Classification Performance: In the multi-class classification task, the ensemble model achieved a a precision of 0.8889, recall of 0.8421, Specificity of 0.935, F1 score of 0.8649, and accuracy of 0.9074. . Although these metrics indicate strong overall performance, the results reveal the increased complexity of distinguishing among "Normal," "Early Glaucoma," and "Advanced Glaucoma" categories. The variability in misclassification patterns, particularly between early and advanced stages of glaucoma, highlights the challenges in differentiating between these closely related conditions. This suggests that while the ensemble model performs well, further refinement and additional features may be necessary to enhance its accuracy in multi-class scenarios.

Confusion Matrix Insights: The confusion matrices provided detailed insights into the misclassification patterns. In binary classification, the error rate was low, reflecting the model’s high effectiveness. In multi-class classification, the confusion matrix highlighted challenges in distinguishing between "Early Glaucoma" and "Advanced Glaucoma," which impacted overall Precision and Recall. This indicates that the model’s ability to differentiate between stages of glaucoma needs improvement.

Grad-CAM Visualization: The Grad-CAM visualizations revealed that both ResNet-50 and DenseNet-201 models focused on the optic disc and surrounding areas, consistent with clinical knowledge of glaucoma assessment. The combined heatmaps from the ensemble model offered a comprehensive view of the regions influential in classification decisions, enhancing our understanding of the model’s focus areas and supporting its robustness in feature extraction.

In conclusion, this study highlights the potential of ensemble CNN models for effective glaucoma detection, particularly in binary classification scenarios. The insights gained from this research contribute to a deeper understanding of model performance and provide a foundation for future advancements in automated glaucoma detection systems.

Conflict of interest:

The authors declare that there is no conflict of interest.

Fund:

No funds are provided.

Data availability:

The dataset used in this study is available online on the Harvard Dataverse (doi:10.7910/DVN/1YRRAC)

Authors contribution:

Conceptualization, W.H., and N.E.; methodology, W.H., N.E., and H.S.; software, W.H.; validation, N.E., and H.S.; formal analysis, W.H., and N.E.; investigation, W.H., N.E., D.S., and H.S.; resources, W.H. and N.E.; data curation, W.H., N.E., and D.S.; writing—original draft preparation, W.H., and N.E.; writing—review and editing, N.E., D.S., and H.S..; visualization, N. E..; supervision, N.E. and H.S.; project administration, N.E., and H.S. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: There is no conflict of interest.

5. REFERENCES

[1] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040," Ophthalmology, vol. 121, pp. 2081–2090, 2014. doi: 10.1016/j.ophtha.2014.05.013.

[2] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma," JAMA, vol. 311, p. 1901, 2014. doi: 10.1001/jama.2014.3192.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015. doi: 10.1038/nature14539.

[4] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,"JAMA, vol. 316, p. 2402, 2016. doi: 10.1001/jama.2016.17216.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in =2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1905.11946.

[9] Z. Li et al., "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," Ophthalmology, vol. 125, pp. 1199–1206, 2018. doi: 10.1016/j.ophtha.2018.01.023.

[10] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017. doi: 10.1109/ICCV.2017.74.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, pp. 84–90, 2012. doi: 10.1145/3065386.

[12] F. A. Medeiros and R. N. Weinreb, "Glaucoma management: What's new in 2012?" Ophthalmology Times, vol. 37, no. 7, pp. 8–9, 2012.

[13] X. Chen, Y. Xu, and D. W. K. Wong et al., "Glaucoma detection based on deep convolutional neural networks," in Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015. doi: 10.1109/EMBC.2015.7318462.

[14] Z. Li et al., "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," Ophthalmology, vol. 125, pp. 1199–1206, 2018. doi: 10.1016/j.ophtha.2018.01.023.

[15] R. Asaoka, H. Murata, A. Iwase, and M. Araie, "Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier," Ophthalmology, vol. 123, pp. 1974–1980, 2016. doi: 10.1016/j.ophtha.2016.05.029.

[16] M. Christopher et al., "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," Scientific Reports, vol. 8, 2018. doi: 10.1038/s41598-018-35044-9.

[17] Q. Zhang and S. Zhu, "Visual interpretability for deep learning: A survey," Frontiers of Information Technology & Electronic Engineering, vol. 19, pp. 27–39, 2018. doi: 10.1631/fitee.1700808.

[18] M. Gende et al., "Robust multi-view approaches for retinal layer segmentation in glaucoma patients via transfer learning," Quantitative Imaging in Medicine and Surgery, vol. 13, pp. 2846–2859, 2023. doi: 10.21037/qims-22-959.

[19] S. H. Kim et al., "Efficacy of a comprehensive binary classification model using a deep convolutional neural network for wireless capsule endoscopy," Scientific Reports, vol. 11, 2021. doi: 10.1038/s41598-021-96748-z.

[20] Y. Ma, T. Xu, S. Han, and K. Kim, "Ensemble learning of multiple deep CNNs using accuracy-based weighted voting for ASL recognition," Applied Sciences, vol. 12, p. 11766, 2022. doi: 10.3390/app122211766.

[21] R. O. Ogundokun et al., "Deep transfer learning models for mobile-based ocular disorder identification on retinal images," Computers, Materials & Continua, 2024. doi: 10.32604/cmc.2024.052153.

[22] R. S. R. Singh and R. K. Sanodiya, "Zero-shot transfer learning framework for plant leaf disease classification," IEEE Access, vol. 11, pp. 143861–143880, 2023. doi: 10.1109/ACCESS.2023.3343759.

[23] M. S. Puchaicela-Lozano et al., "Deep learning for glaucoma detection: R-CNN ResNet-50 and image segmentation," Journal of Advances in Information Technology, vol. 14, no. 6, pp. 1186–1197, 2024. doi: 10.12720/jait.14.6.1186-1197.

[24] H. Fu et al., "Disc-aware ensemble network for glaucoma screening from fundus image," IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2493–2501, 2018. doi: 10.1109/TMI.2018.2837012.

[25] J. Lin, Q. Cai, and M. Lin, "Multi-label classification of fundus images with graph convolutional network and self-supervised learning," IEEE Signal Processing Letters, vol. 28, pp. 454–458, 2021. doi: 10.1109/LSP.2021.3057548.

[26] Y. George et al., "Attention-guided 3D-CNN framework for glaucoma detection and structural-functional association using volumetric images," IEEE Journal of Biomedical and Health Informatics, vol. 24, pp. 3421–3430, 2020. doi: 10.1109/JBHI.2020.3001019.

[27] A. Chattopadhay et al., "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018. doi: 10.1109/WACV.2018.00097.

[28] D. Cian, A. Lengyel, and van Gemert, "Evaluating the performance of the LIME and Grad-CAM explanation methods on a LEGO multi-label image classification task," arXiv, 2020. [Online]. Available: https://arxiv.org/abs/2008.01584.

[29] M. Yan et al., "mixDA: mixup domain adaptation for glaucoma detection on fundus images," Neural Computing and Applications, 2023. doi: 10.1007/s00521-023-08572-3.

[30] Y. Chen et al., "Generative adversarial networks in medical image augmentation: A review" Computers in Biology and Medicine, vol. 144, p. 105382, 2022.

[31] G. Müller-Franzes, J. M. Niehues, F. Khader, S. T. Arasteh, C. Haarburger, C. K. Kuhl, T. Wang, T. Han, T. Nolte, S. Nebelung, J. N. Kather, and D. Truhn, "A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis," Scientific Reports, vol. 13, 2023. doi: 10.1038/s41598-023-39278-0.

[32] A. Rauniyar, D. H. Hagos, D. Jha, J. E. Håkegård, U. Bağcı, D. B. Rawat, and V. Vlassov, "Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions," IEEE Internet of Things Journal, pp. 1–1, 2023. doi: 10.1109/jiot.2023.3329061.

[33] L. Pascal, O. J. Perdomo, X. Bost, B. Huet, S. Otálora, and M. A. Zuluaga, "Multi-task deep learning for glaucoma detection from color fundus images," Scientific Reports, vol. 12, 2022. doi: 10.1038/s41598-022-16262-8.

[34] F. A. Braeu, A. H. Thiéry, T. A. Tun, A. Kadziauskiene, G. Barbastathis, T. Aung, and M. J. A. Girard, "Geometric deep learning to identify the critical 3D structural features of the optic nerve head for glaucoma diagnosis," American Journal of Ophthalmology, vol. 250, pp. 38-48, 2023. doi: 10.1016/j.ajo.2023.01.008.

[35] A. Serener and S. Serte, "Transfer learning for early and advanced glaucoma detection with convolutional neural networks," in 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4. doi: 10.1109/TIPTEKNO.2019.8894965.

[36] Z. Mo and A. Siepel, "Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data," PLoS Genetics, vol. 19, 2023. doi: 10.1371/journal.pgen.1011032.

[37] O. C. Devecioglu, J. Malik, T. Ince, S. Kiranyaz, E. Atalay, and M. Gabbouj, "Real-time glaucoma detection from digital fundus images using Self-ONNs," IEEE Access, vol. 9, pp. 140031-140041, 2021. doi: 10.1109/ACCESS.2021.3118102.

[38] T. H. P. Thanh, T. P. T. Thuy, T. N. Hieu, and M. S. Nguyen, "A real-time classification of glaucoma from retinal fundus images using AI technology," in 2020 International Conference on Advanced Computing and Applications (ACOMP), Quy Nhon, Vietnam, 2020, pp. 114-121. doi: 10.1109/ACOMP50827.2020.00024.

[39] C. Comito, D. Falcone, and A. Forestiero, "AI-driven clinical decision support: Enhancing disease diagnosis exploiting patients' similarity," IEEE Access, vol. 10, pp. 6878-6888, 2022. doi: 10.1109/ACCESS.2022.3142100.

[40] A. Saxena, A. Vyas, L. Parashar, and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," in International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 815-820. doi: 10.1109/ICESC48915.2020.9155930.

[41] M. Lin, B.-J. Hou, L. Liu, M. O. Gordon, M. A. Kass, F. Wang, S. H. Van, and Y. Peng, "Automated diagnosing primary open-angle glaucoma from fundus image by simulating human's grading with deep learning," Scientific Reports, vol. 12, 2022. doi: 10.1038/s41598-022-17753-4.

[42] 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," Biomedical Optics Express, vol. 10, p. 892, 2019. doi: 10.1109/ICESC48915.2020.9155930.

[43] A. Carlos, L. B. Maia, T. P. Braz Junior, and S. Anselmo, "Evolving Convolutional Neural Networks for Glaucoma Diagnosis," Brazilian Journal of Health Review, vol. 3, pp. 9224-9234, 2020. doi: 10.34119/bjhrv3n4-160.