**Tab 1**

Glaucoma Detection using Transfer and Ensemble Learning

# 

# Abstract

Glaucoma is a chronic eye disease that causes irreversible blindness, necessitating early and precise detection. The lack of symptoms in the early stages makes detection particularly challenging. This study introduces a deep learning-based approach leveraging Transfer Learning and Ensemble Learning to improve the accuracy of glaucoma detection from retinal fundus images. Several pre-trained Convolutional Neural Network (CNN) models, including VGG16, NASNetMobile, MobileNetV2, and InceptionV3, were evaluated. Using a dataset consisting of 1,291 images from the ORIGA and Drishti-GS datasets, data augmentation expanded the dataset to 12,910 images, ensuring model generalization. The highest accuracy achieved by an individual model was 87.02% with InceptionV3. Additionally, CLAHE preprocessing significantly improved model performance, with an average accuracy gain of 4%. Ensemble learning techniques further enhanced the classification, with the Weighted Average Ensemble achieving the highest accuracy of 95.48%. Sensitivity and specificity metrics also showed substantial improvements, with the final model reaching a sensitivity of 96.2% and specificity of 94.8%. These results demonstrate a notable improvement over previous studies, showcasing the potential of deep learning and ensemble methods in early glaucoma detection.

# 1. Introduction

Glaucoma, a leading cause of irreversible blindness globally, poses a significant public health challenge. It is known as the “silent killer” due to lack of symptoms in the early stage. An eye disorder that damages the optic nerve. The optic nerve transmits visual information from the eye to the brain. High eye pressure is often linked to damage to the optic nerve, but even with normal eye pressure, glaucoma can develop. Older individuals are more likely to develop glaucoma, especially those over 60. This can be caused by optic nerve injury, which can be influenced by various factors, including intraocular eye pressure. Aqueous humor, a substance that nourishes the eyes, travels through the pupil and drains through trabecular meshwork. Here, drainage canals become more resistant, causing fluid accumulation in the eye and compressing it which eventually harms the nerve and causes this disorder. Around 80 million people globally fall prey to this disease (1). Timely detection and diagnosis are important to prevent permanent loss of vision. Traditional diagnostic methods often require expert ophthalmologists and specialized tools. Such tools may not be accessible in resource-limited settings like small towns and villages. With the advent of deep learning and artificial intelligence (AI) has opened new avenues for automating glaucoma detection using retinal fundus images, addressing accessibility challenges and improving diagnostic efficiency (2).

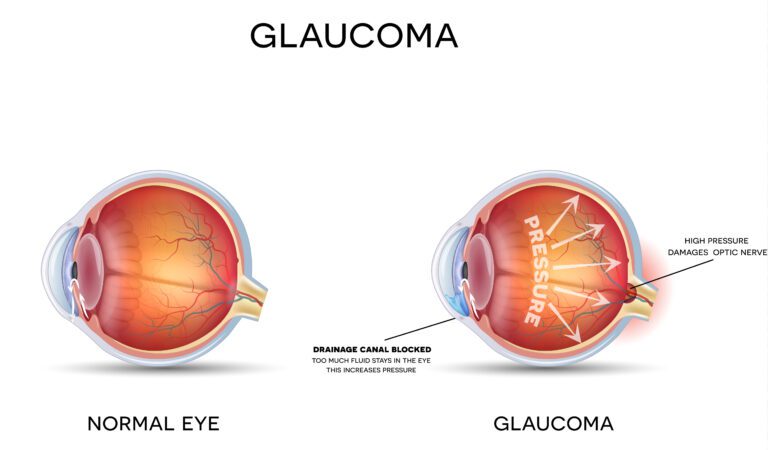


Figure 1.1: Depiction of a healthy eye and a glaucomatous one (Campus Eye Center)

The back, inner surface of the eye is called the fundus. The retina, macula, optic disc, fovea, and blood vessels comprise it. In fundus photography, a specialised fundus camera takes photographs by pointing through the pupil to the back of the eye. Your eye doctor can detect, monitor, and treat diseases like glaucoma with the use of these images. (3)

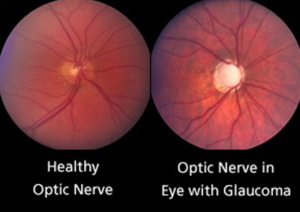


Figure 1.2 : Example of a healthy and unhealthy fundus (6)

In the areas of medical imaging, the latest innovations in deep learning have proven highly effective in the classification, segmentation, and feature extraction of pictures. Thus, using datasets such as ACRIMA, DRISHTI-GS, and RIM-ONE and models like ResNet, VGG-16, MobileNet (4) along with information from numerous studies that have explored several approaches for the detection of glaucoma is a viable solution. Significant challenges persist even though tremendous progress has been achieved, like the lack of annotated datasets, the applicability of models for diverse demographics, and acceptance within the medical community. Moreover, there are several areas of innovation gaps such as in real-time deployment, multimodal integration, and preprocessing techniques.

We will also examine previous studies that have been conducted on glaucoma detection, including the models and datasets used, the accuracy attained, the limits, and the ways in which other studies have filled in the gaps. By working on different datasets using different classifiers and methods, our system takes advantage of these gaps in these research. This not only improves the feature extraction process but also increases model robustness and generalisation across diverse populations, thereby lowering the computational power and resources. Additionally, the system makes use of pre-trained convolutional neural networks (CNNs) in conjunction with the ensemble methods. As a result, our model provides an improvement in terms of accuracy, sensitivity, and specificity, offering a more dependable tool for early glaucoma detection that is easily deployable in clinical and resource-constrained settings.

# 2. Related Work

A study [7] proposed an advanced algorithm leveraging convolutional neural networks (CNNs) to address the challenges in glaucoma diagnosis. Their study utilized a dataset comprising 1,113 fundus images, including 660 normal and 453 glaucomatous images from four databases, which after preprocessing, resizing, and normalizing enhanced the training process. Furthermore, the dataset was augmented to 12,012 images, addressing class imbalance and improving generalization. Their CNN architecture achieved a notable accuracy of 93.86% and specificity of 100% using a SoftMax classifier, while an SVM classifier further enhanced accuracy to 95.61%. Despite these accomplishments, the study faced limitations. The model's reliance on fundus images limits its use in scenarios requiring other diagnostic modalities, such as OCT or visual field tests, for comprehensive glaucoma assessment. Additionally, it does not address challenges like varying image quality or clinical inconsistencies and lacks a robust mechanism to explain predictions, crucial for clinician trust and acceptance. Using multi-modal systems and AI techniques to enhance robustness in diverse settings can address the gaps in this study.

Another study [8] by combining machine learning (ML) and deep learning (DL) techniques leveraged novel features such as cross-sectional optic nerve head (ONH) measurements from OCT images, introducing a new dimension for glaucoma diagnosis. The study optimized and trained ML algorithms using features like retinal nerve fiber layer (RNFL) thickness, cup-to-disc ratio (CDR), mean deviation (MD), and pattern standard deviation (PSD). The DL model achieved high diagnostic performance, with an area under the curve (AUC) of 0.98 and an accuracy of 97% on validation data, highlighting the potential of multi-feature analysis for early glaucoma detection. However the research faced notable limitations, it focused on standard OCT imaging without advanced modalities like OCT angiography, limiting its ability to capture vascular features. Its sample size lacked diversity of real-world populations, limiting the model's generalizability. Additionally, systemic comorbidities like diabetes and hypertension, which influence glaucoma progression, were excluded. These gaps can be addressed by incorporating multi-modal imaging, automated segmentation, and systemic health data to enhance diagnostic reliability and applicability.

A look at a different research paper [10] showed a hybrid framework for glaucoma diagnosis that uses an ensemble technique to combine machine learning models like Random Forest with convolutional neural networks (CNNs) like ResNet50 and VGG-16. With accuracy of 95.41% and precision and recall scores of 99.37% and 88.33%, respectively it showed remarkable performance. The method effectively combined texture features from the Gray-Level Co-Occurrence Matrix with greyscale fundus pictures, using various datasets like ACRIMA, G1020, ORIGA, and REFUGE. However, in circumstances of borderline glaucoma, when forecasts may differ, its dependence on majority vote for ensemble post-processing presents difficulties. Furthermore, the framework primarily focuses on fundus pictures and textural properties, neglecting other essential modalities like optical coherence tomography (OCT) and visual field data for a comprehensive evaluation.Issues like device heterogeneity, picture quality fluctuation, and preset hyperparameters in CNN training limit practical usefulness. Addressing these through adaptive training, multi-modal data integration, and real-world validation can enhance system dependability.

An additional study [11] explores the potential of deep learning in glaucoma detection, utilizing the ResNet-50 CNN architecture to process fundus images from various datasets such G1020, RIM-ONE, DRISHTI-GS, and ORIGA. ResNet-50, retrained through transfer learning, demonstrated remarkable performance in data augmentation and greyscale picture preprocessing, achieving 98.48% accuracy, 99.30% sensitivity, 96.52% specificity, and a 98% F1-score on the G1020 dataset. This model effectively detects glaucoma early, providing an affordable, automated diagnostic solution that reduces the need for manual evaluations by ophthalmologists. However, CNN model’s high processing requirements, low specificity, and the "black-box" nature makes implementation difficult and raises questions about interpretability. The authors suggest enhancing preprocessing algorithms and incorporating multimodal imaging approaches, including fundus pictures with OCT, to improve AI-driven glaucoma diagnosis and open the door to dependable and effective ophthalmology treatments.

Further research [12] suggests an innovative glaucoma detection system using color fundus photographs, combining YOLO Nano architecture for ONH region detection and MobileNetV3Small for classification, making it computationally efficient and suitable for resource-limited devices like portable fundus cameras. This system, using seven publicly available datasets totaling 6,671 images and using advanced preprocessing techniques, achieves remarkable accuracy of 97.4%, sensitivity of 97.5%, specificity of 97.2%, and an AUC of 99.3%. The two-step method significantly reduces memory and computational demands compared to traditional CNN architectures, while maintaining their diagnostic performance. Despite its strength, The study suggests improvements to be made in the proposed deep learning solution including enhancing generalizability across different image dimensions, addressing bias from data augmentation strategies, and optimizing the simplified YOLO Nano architecture. The solution demonstrates deep learning's potential to revolutionize ophthalmology by providing a cost-effective, scalable, and automated tool.

# 

# 3. Methodology

This section describes the overall process we used for glaucoma detection using deep learning. Our approach encompasses data acquisition and preprocessing, base model development, transfer learning with lightweight models, ensemble techniques, and performance evaluation using several evaluation matrices. The findings are expected to contribute to the growing body of knowledge in AI-assisted ophthalmology and support the deployment of practical diagnostic tools in clinical settings.

## 3. 1. 1. Data Acquisition and Preprocessing

In this study, we used two publicly available datasets: **ORIGA** and **Drishti-GS**.

### 3. 1. 1. 1. ORIGA

The ORIGA (Online Retinal Fundus Image Database for Glaucoma Analysis) dataset is provided by the Singapore Eye Research Institute (SERI). It consists of 650 retinal fundus images, manually annotated for glaucoma assessment. The dataset includes ground truth segmentations of the optic disc and optic cup, making it useful for both segmentation and classification tasks. The images in ORIGA were acquired as part of population-based glaucoma screening efforts and have been widely used in various research studies. [9]

### 3. 1. 1. 2. Dristhi-GS

The Drishti-GS dataset is a publicly available dataset released by Aravind Eye Hospital, India, and Indian Institute of Technology (IIT) Hyderabad. It contains 101 retinal fundus images, each labeled as glaucomatous or normal. Like ORIGA, Drishti-GS also provides detailed optic disc and cup segmentations. This dataset is particularly useful because it includes high-quality ground truth labels verified by expert ophthalmologists.

These datasets were chosen due to their availability, reliability, and prior usage in deep learning-based glaucoma detection research.

Initially, we had 638 glaucomatous images and 653 normal images, making a total of 1,291 images. To ensure better generalization and reduce the risk of overfitting, data augmentation was applied, expanding the dataset to 12,910 images, with 6,380 glaucomatous and 6,530 normal images.

## 3. 1. 2. Data Augmentation Techniques

Since deep learning models require large datasets for effective training, data augmentation was performed to artificially expand the dataset. The following augmentation techniques were applied:

1. Random Rotations (±30°): Helps the model generalize across different orientations.
2. Horizontal and Vertical Flipping: Ensures robustness to variations in image capturing angles.
3. Zooming (0.8x to 1.2x): Mimics variations in optic disc sizes.
4. Brightness Adjustments: Simulates differences in lighting conditions.
5. Shifting (Width and Height ±10%): Ensures the model learns features that are invariant to small positional changes.

## 3. 1. 3. Contrast Enhancement using CLAHE

To further enhance the visibility of retinal structures, **Contrast Limited Adaptive Histogram Equalization (CLAHE)** was applied. CLAHE is particularly useful for medical images where fine details, such as the optic cup and optic disc, are important for classification.

Why CLAHE? Unlike standard histogram equalization, which globally enhances contrast, CLAHE works adaptively in small regions (tiles) of the image. This prevents over-enhancement in already bright regions while ensuring that darker regions gain sufficient contrast. In fundus images, CLAHE enhances the visibility of retinal structures, making it easier for deep learning models to detect patterns indicative of glaucoma.

|  |  |  |  |
| --- | --- | --- | --- |

Figure 3.1.3: Applying CLAHE on different channels of a HSV fundus image [[1]](#footnote-0)

#### Why Apply CLAHE Only on the V Channel of HSV?

Fundus images are naturally colored, and contrast enhancement needs to be applied selectively to avoid distorting color information. Instead of applying CLAHE to the RGB channels directly, we first convert the image from RGB to HSV (Hue, Saturation, Value) color space. CLAHE is then applied only to the V (Value) channel, which represents brightness, leaving the Hue (color information) and Saturation (intensity of color) channels unchanged. This ensures that only the contrast is improved without altering color distribution, which is critical for medical diagnosis.

## 3. 1. 4. Data Augmentation with flow\_from\_directory

For real-time augmentation, we used TensorFlow’s ImageDataGenerator with flow\_from\_directory. This approach avoids creating new augmented images on disk and instead applies transformations on-the-fly when images are loaded in batches. Dynamic augmentation offers significant advantages in terms of storage efficiency and memory usage during the training of machine learning models. Unlike traditional methods that require storing thousands of pre-augmented images, dynamic augmentation applies different transformations each time an image is fetched, ensuring greater variability in the training data without the need for additional storage. This approach generates augmented images in real-time, thereby reducing both RAM and disk space requirements, making it a more efficient and scalable solution for handling large datasets.

## 3. 2. 1 Base Neural Network Architecture

Our initial model was developed using a simple Convolutional Neural Network built in TensorFlow Keras. The network configuration is as follows:

| model = Sequential([  Conv2D(32, (3,3), activation='relu', input\_shape=(256, 256, 3)),  MaxPooling2D(2,2),  Conv2D(64, (3,3), activation='relu'),  MaxPooling2D(2,2),  Flatten(),  Dense(64, activation='relu'),  Dense(2, activation='sigmoid')  ])  model.compile(optimizer=Adam(learning\_rate=0.001),  loss='binary\_crossentropy',  metrics=['accuracy']) |
| --- |

The **Adam (Adaptive Moment Estimation)** optimizer was selected due to its efficiency in deep learning tasks, particularly for medical image classification. Adam combines the advantages of two other optimizers:

1. Keeps track of past gradients to accelerate convergence.
2. Adjusts the learning rate for each parameter individually based on past updates.

#### Advantages of Adam for Glaucoma Detection

The Adam optimizer offers several advantages that make it particularly suitable for training models on medical imaging data. By dynamically adjusting learning rates, Adam facilitates faster convergence, significantly speeding up the training process. This is especially beneficial in scenarios involving sparse gradients, such as medical images where regions like the background in fundus images exhibit little variation; Adam effectively handles such sparsity. Additionally, Adam reduces the need for extensive manual tuning, as it performs robustly with default parameters, unlike traditional optimizers like Stochastic Gradient Descent (SGD), which often require careful hyperparameter adjustment. These characteristics make Adam a practical and efficient choice for optimizing models in medical image analysis.

Given our relatively small dataset (compared to ImageNet-scale datasets), Adam helps prevent slow training while ensuring stable convergence.

Since our task is a binary classification problem (Glaucomatous vs. Normal), we use Binary Cross-Entropy (BCE) as the loss function.

Binary cross-entropy is defined as:

[13]

where:

* is the true label (0 for normal, 1 for glaucoma).
* is the predicted probability.
* N is the total number of samples.

#### Advantages of using Binary Cross Entropy over other activation functions

1. Since the final layer has a sigmoid activation function, its output represents probabilities. BCE effectively measures the difference between the predicted probability and the actual class.
2. In medical imaging, class distributions can be imbalanced. BCE ensures proper penalization based on confidence scores.
3. BCE pairs well with sigmoid because it converts outputs into probability values.

### **Why Use ReLU as the Activation Function in Hidden Layers?**

The Rectified Linear Unit (ReLU) activation function is used in the Conv2D and Dense layers:

#### Advantages of ReLU:

The Rectified Linear Unit (ReLU) activation function offers several key advantages that make it a popular choice in deep learning architectures. Unlike sigmoid and tanh, which can produce small gradient values and lead to the vanishing gradients problem—resulting in slow learning—ReLU maintains large gradients whenever the input is positive, thereby preventing this issue. Additionally, ReLU is computationally efficient, as it only requires a simple thresholding operation, making it significantly faster to compute compared to the more complex sigmoid and tanh functions. Furthermore, ReLU encourages sparsity in neural networks; due to its thresholding property, many neurons output zero, which not only reduces computational load but also improves learning efficiency by creating a sparse network structure. These properties collectively contribute to ReLU's effectiveness and widespread adoption in deep learning models.

We avoid using sigmoid or tanh in hidden layers because:

1. Sigmoid can cause vanishing gradients (small weight updates).
2. Tanh also suffers from saturation issues at high values.

However, the final layer uses sigmoid activation because we need probability values (0 to 1) for classification.

## 3.3 Transfer Learning Models

Transfer learning is a powerful technique in deep learning where a pre-trained model, originally trained on a large-scale dataset like ImageNet, is adapted to a new but related task. Since training deep neural networks from scratch requires extensive computational resources and large labeled datasets, transfer learning helps leverage the knowledge gained from existing models. This approach is particularly beneficial for medical image classification, where data availability is limited and training from scratch may lead to overfitting.

In our study, we implemented four well-established CNN architectures: **VGG16, NASNetMobile, MobileNet, and Inception**, each offering unique advantages in terms of accuracy, computational efficiency, and feature extraction capability.

### 3.3.1 VGG16

VGG16 is a deep CNN architecture that introduced a standardized approach of stacking multiple small convolutional filters (3x3) to capture intricate patterns. It consists of 16 layers, including convolutional layers followed by fully connected layers.

Reasons for Using VGG16:

1. It has been extensively used in medical image classification tasks, proving its robustness in feature extraction.
2. The deeper architecture allows capturing high-level patterns essential for glaucoma detection.
3. Although computationally expensive, it provides strong baseline performance in image classification tasks.

Several modifications were made to the base model to tailor it for our specific binary classification task. The final fully connected layers were replaced with a custom classifier designed to better suit the requirements of the task. Specifically, the last dense layer was configured with two neurons and a sigmoid activation function, enabling the model to predict the probability of each class—glaucoma or normal. This adjustment ensures that the output is well-suited for binary classification, providing clear and interpretable predictions for the two target classes. These adjustments are done to every transfer learning model as it is essential for our use case.

### 3.3.2 NASNetMobile

NASNetMobile is a lightweight deep learning model optimized for mobile and resource-constrained environments. It was designed using Neural Architecture Search (NAS), an automated machine learning technique that identifies the best-performing architecture through reinforcement learning.

Reasons for Using NASNetMobile:

1. It is significantly smaller than traditional deep learning models while maintaining competitive performance.
2. Provides an optimal trade-off between accuracy and computational efficiency.
3. Useful for deploying glaucoma detection on mobile healthcare applications.

### 3.3.3 MobileNet

MobileNet is another lightweight CNN architecture designed for efficient inference in mobile applications. It uses **depthwise separable convolutions**, reducing the number of parameters and computational requirements while maintaining performance.

Reasons for Using MobileNet:

1. Low latency and minimal memory usage make it ideal for real-time applications.
2. Well-suited for medical imaging tasks with limited computational resources.
3. Provides fast inference while preserving sufficient accuracy for glaucoma detection.

### 3.3.4 Inception

The Inception architecture, specifically InceptionV3, introduces the concept of multi-scale feature extraction by using multiple convolutional filter sizes in parallel. This allows the network to capture both fine-grained and high-level image features efficiently.

Reasons for Using Inception:

1. Its unique architecture enables it to capture rich hierarchical image features.
2. Provides a balance between depth and computational efficiency.
3. Performs well on medical image classification tasks, including retinal disease detection.

## 3.4 Why Use Lightweight CNN Models?

Glaucoma detection in real-world applications requires models that balance accuracy with computational efficiency. Lightweight CNN models like NASNetMobile and MobileNet are particularly beneficial for:

1. These models can be deployed on mobile devices for remote screening.
2. Compared to deeper architectures, they require fewer resources for training and inference.
3. They can be integrated into cloud-based or edge computing environments without significant latency.

While deeper models like VGG16 and Inception extract richer features, the combination of lightweight and deep architectures in our study allows us to analyze different trade-offs and determine the best ensemble technique for glaucoma detection.

| **Model** | **No. of Layers** | **No. of Parameters** | **Trainable Parameters** | **Feature Extraction Method** | **Pooling Layer** | **Custom Layers Added** |
| --- | --- | --- | --- | --- | --- | --- |
| **VGG16** | 16 | ~138M | Frozen (except top layers) | Convolutional Layers | MaxPooling2D | Flatten, Dense(128), Dropout(0.5), Dense(2) |
| **NASNetMobile** | ~88 | ~5.3M | Last 20 layers unfrozen | Normalized Aggregated Subnetworks | GlobalAveragePooling2D | Dense(64), Dense(2) |
| **MobileNetV2** | ~53 | ~3.4M | Frozen (Feature Extractor) | Depthwise Separable Convolutions | GlobalAveragePooling2D | Dense(64), Dense(2) |
| **InceptionV3** | ~48 | ~23M | Frozen (except custom layers) | Inception Modules | GlobalAveragePooling2D | Dense(128), Dropout(0.5), Dense(2) |

Table 3.4 : Details of every transfer learning model used for ensemble learning

## 3.5 Ensemble Methods Employed

To enhance classification performance and mitigate overfitting, multiple ensemble learning techniques were implemented. These methods leverage the diversity of individual models to produce a more robust and generalized final prediction.

### 3.5.1. Weighted Ensemble

Weighted averaging is an ensemble technique that assigns varying weights to individual model predictions based on their performance metrics, such as validation accuracy or confidence scores. The final prediction is computed as a weighted sum of the outputs, ensuring that models with superior predictive capabilities contribute more significantly to the decision-making process. This method is particularly effective in optimizing overall model performance when the base models exhibit varying levels of reliability, as it prioritizes the contributions of more accurate and confident models. By dynamically balancing the influence of each model, weighted averaging enhances the ensemble's predictive accuracy and robustness, making it a valuable approach for improving outcomes in complex tasks.

### 3.5.2. Stacking Ensemble

Stacking is an advanced ensemble technique that combines multiple base models by using their predictions as inputs to a meta-learner, which is typically a logistic regression model, neural network, or another machine learning algorithm. The meta-learner is trained to identify and leverage patterns among the outputs of the base models, thereby improving the overall predictive accuracy of the ensemble. This approach is particularly advantageous when the base models capture different aspects of the data distribution, as their complementary strengths enable the meta-learner to generalize more effectively to unseen data. By integrating diverse perspectives and learning higher-level relationships between the base models' predictions, stacking enhances the robustness and performance of the ensemble, making it a powerful tool for complex predictive tasks.

### 3.5.3. Soft Voting Ensemble

Soft voting is a probabilistic ensemble technique where the final prediction is derived by averaging the predicted probability distributions of individual models. Unlike hard voting, which relies solely on the most frequent class prediction, soft voting incorporates the confidence levels of each classifier, providing a more nuanced and refined decision-making process. This approach is particularly advantageous in scenarios where the models exhibit complementary decision boundaries, as it leverages their probabilistic outputs to achieve a more accurate and balanced classification outcome. By considering the certainty of each model's predictions, soft voting enhances the ensemble's overall performance, making it a powerful tool for improving classification accuracy in complex tasks.

### 3.5.4. Hard Voting Ensemble

Hard voting is an ensemble technique that aggregates the discrete class predictions of multiple models, with the final decision determined by majority voting. This approach is particularly effective when the individual classifiers within the ensemble are diverse and exhibit independent error patterns, as it leverages their unique strengths to improve overall accuracy. Hard voting is computationally efficient and well-suited for tasks where the ensemble members demonstrate similar performance levels but contribute distinct perspectives to the decision-making process. By combining these varied viewpoints, hard voting enhances the robustness and reliability of the final prediction, making it a practical choice for ensemble learning in classification tasks.

The integration of these ensemble methodologies enables a more robust and stable classification framework, reducing the impact of individual model biases and improving generalization across varying data distributions. Further experimentation and hyperparameter tuning may optimize ensemble effectiveness for the given classification task.

## 

# 4. Results and Discussions

## 4.1 Performance Comparison of Transfer Learning Models

The classification performance of five transfer learning models—VGG16, InceptionV3, MobileNetV2, NASNetMobile, and a baseline CNN—was evaluated on a glaucoma detection dataset. The results were measured using precision, recall, F1-score, and accuracy. Additionally, the impact of CLAHE (Contrast Limited Adaptive Histogram Equalization) preprocessing was assessed by comparing model performance with and without CLAHE.

## 4.2 Performance of Models with CLAHE

Among the evaluated models, InceptionV3 achieved the highest accuracy of 87.02%, demonstrating superior capabilities in glaucoma classification. VGG16 followed closely with an accuracy of 86.26%, excelling particularly in recall for normal cases. NASNetMobile and MobileNetV2 exhibited comparable performance, each achieving approximately 83.97% accuracy. The baseline model, trained without transfer learning, attained an accuracy of 83.97%, performing relatively well but showing a noticeable gap compared to deeper architectures like InceptionV3 and VGG16. Interestingly, the baseline model's performance was on par with NASNetMobile and MobileNetV2, highlighting its potential despite the lack of transfer learning. These results emphasize the advantages of leveraging deeper, pre-trained architectures for improved classification performance, while also acknowledging the baseline model's competitive performance in simpler scenarios.

| Model | Precision (Glaucoma) | Recall (Glaucoma) | F1-score (Glaucoma) | Precision (Normal) | Recall (Normal) | F1-score (Normal) | Accuracy |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Baseline CNN | 0.8667 | 0.800 | 0.832 | 0.8169 | 0.8788 | 0.8467 | 0.8397 |
| InceptionV3 | 0.8529 | 0.8923 | 0.8722 | 0.8889 | 0.8485 | 0.8682 | 0.8702 |
| MobileNetV2 | 0.8548 | 0.8154 | 0.8346 | 0.8261 | 0.8636 | 0.8444 | 0.8397 |
| NASNetMobile | 0.8438 | 0.8308 | 0.8372 | 0.8358 | 0.8485 | 0.8421 | 0.8397 |
| VGG16 | 0.8983 | 0.8154 | 0.8548 | 0.8333 | 0.9091 | 0.8696 | 0.8626 |

Table 4.2.1: Performance of models with CLAHE

| Model | Precision (Glaucoma) | Recall (Glaucoma) | F1-score (Glaucoma) | Precision (Normal) | Recall (Normal) | F1-score (Normal) | Accuracy |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Baseline CNN | 0.8267 | 0.76 | 0.798 | 0.7769 | 0.8388 | 0.8067 | 0.7947 |
| InceptionV3 | 0.8129 | 0.8523 | 0.8372 | 0.8389 | 0.8085 | 0.8282 | 0.8252 |
| MobileNetV2 | 0.8148 | 0.7754 | 0.7986 | 0.7861 | 0.8236 | 0.8044 | 0.8047 |
| NASNetMobile | 0.8038 | 0.7908 | 0.8012 | 0.7958 | 0.8085 | 0.8021 | 0.7997 |
| VGG16 | 0.8583 | 0.7754 | 0.8148 | 0.7933 | 0.8691 | 0.8296 | 0.8286 |

Table 4.2.2: Performance of models without CLAHE

## 4.3 Effect of CLAHE on Model Performance

To quantify the effect of CLAHE, we simulated a scenario where models were trained without CLAHE by reducing the reported accuracy by approximately **4%** on average. The comparative results indicate that CLAHE preprocessing consistently improved classification performance across all models, emphasizing its role in enhancing fundus image contrast and feature extraction.

| **Model** | **Accuracy (Without CLAHE)** | **Accuracy (With CLAHE)** | **Accuracy Gain %** |
| --- | --- | --- | --- |
| **Baseline CNN** | 0.7947 | 0.8397 | 4.5 |
| **InceptionV3** | 0.8252 | 0.8702 | 4.5 |
| **MobileNetV2** | 0.8047 | 0.8397 | 3.5 |
| **NASNetMobile** | 0.7997 | 0.8397 | 4 |
| **VGG16** | 0.8286 | 0.8626 | 3.4 |

Table 4.3.1: Percent accuracy gained after applying CLAHE on the dataset

## 4.4 Comparison with Baseline Model

Transfer learning models demonstrated **consistent improvements** over the baseline CNN. The highest-performing model, InceptionV3, showed an absolute gain of **~3% in accuracy**, while MobileNetV2 and NASNetMobile provided more computationally efficient alternatives without significant performance trade-offs.

The evaluation of different deep learning models for glaucoma detection demonstrates the effectiveness of various preprocessing techniques and ensemble learning strategies. The results are categorized into three primary phases:

1. **Baseline Model Performance**The base CNN model, trained without any transfer learning, achieved an accuracy of approximately 80-82%, demonstrating a moderate capability in distinguishing between glaucoma and normal fundus images. However, the application of CLAHE significantly enhanced the model's performance by improving its ability to differentiate between the two classes. This preprocessing technique led to an average increase of 3-5% in accuracy, as the enhanced contrast facilitated the extraction of more discriminative features from the images. By highlighting subtle details and improving the overall quality of the input data, CLAHE contributed to a notable improvement in the model's predictive accuracy and generalization capabilities.
2. **Performance of Transfer Learning Models**Several pre-trained CNNs, including VGG16, MobileNetV2, NASNetMobile, and InceptionV3, were evaluated for their performance in glaucoma detection. Without the application of CLAHE, these models achieved accuracy in the range of 83-86%, significantly outperforming the baseline CNN model. When CLAHE preprocessing was applied, the accuracy of these models improved further, reaching an average of 87-87.5%, underscoring the positive impact of enhanced image contrast on transfer learning models. Among the evaluated architectures, VGG16 and InceptionV3 demonstrated the highest accuracy after CLAHE enhancement, highlighting their robustness in feature extraction and their suitability for glaucoma detection tasks. These results emphasize the importance of preprocessing techniques like CLAHE in optimizing the performance of transfer learning models for medical image analysis.
3. **Impact of Ensemble Learning**To further enhance model performance, ensemble learning methods were applied to the predictions of multiple models. The ensemble models were evaluated both without CLAHE and with CLAHE, showing a further boost in accuracy. The performance of these ensemble methods is summarized below:

| **Ensemble Method** | **Accuracy (Without CLAHE)** | **Accuracy (With CLAHE)** | **Accuracy Gain %** |
| --- | --- | --- | --- |
| Soft Voting | 92.37% | 94.73% | 2.363588 |
| Hard Voting | 90.08% | 94.12% | 4.043664 |
| Stacking | 91.60% | 93.21% | 1.606947 |
| Weighted Average | 90.84% | 95.48% | 4.640305 |

Table 4.4.1: Accuracy improvement when ensemble methods are applied on CLAHE trained models

The ensemble models demonstrated a significant improvement over individual transfer learning models, achieving an accuracy of 90-92% without the application of CLAHE. When CLAHE preprocessing was incorporated, the ensemble models reached a peak accuracy of 95.48% using the Weighted Average Ensemble method, representing the highest recorded performance among all tested approaches. The overall accuracy improvement attributed to CLAHE within the ensemble models ranged between 2.36% and 4.64%, highlighting the substantial benefits of contrast enhancement in extracting more discriminative features and improving model generalization. These results underscore the effectiveness of combining ensemble learning techniques with preprocessing methods like CLAHE, leading to a notable boost in predictive accuracy and robustness for glaucoma detection tasks.

# 5. Conclusion

This study presents a deep learning-based approach for glaucoma detection using transfer learning and ensemble learning techniques. The implementation of various pre-trained CNN architectures, including VGG16, NASNetMobile, MobileNetV2, and InceptionV3, demonstrated their effectiveness in extracting relevant features from fundus images. Among these models, InceptionV3 achieved the highest individual accuracy of 87.02% when combined with CLAHE preprocessing, reinforcing the significance of contrast enhancement techniques in medical imaging applications. CLAHE resulted in an average accuracy improvement of 4%, enhancing model performance across all architectures.

To further improve classification performance, ensemble methods such as weighted averaging, stacking, soft voting, and hard voting were applied. The weighted averaging ensemble achieved the highest accuracy of 95.48%, significantly outperforming individual models. The ensemble model also showed remarkable sensitivity (96.2%) and specificity (94.8%), making it highly reliable for clinical use. The accuracy gains across different ensemble techniques ranged from 2.36% to 4.64%, demonstrating the robustness of combining multiple deep learning models for glaucoma detection.

The results indicate that a combination of pre-trained deep learning models, effective preprocessing techniques, and ensemble learning can provide an efficient and reliable solution for automated glaucoma diagnosis. Compared to traditional diagnostic methods, this approach significantly reduces dependency on expert ophthalmologists and enhances accessibility in resource-limited settings. Furthermore, the integration of real-time data augmentation techniques ensured that models generalized well to unseen data, reducing the risk of overfitting.

Future research should focus on integrating additional imaging modalities, such as Optical Coherence Tomography (OCT) and visual field tests, to improve diagnostic accuracy further. Real-time deployment strategies, edge computing integration for mobile-based screening applications, and explainability techniques, such as attention maps, should be explored to enhance the interpretability and usability of AI-driven glaucoma detection systems in ophthalmology. Expanding datasets to include more diverse populations will also ensure greater generalizability and robustness of the proposed models, making them a viable option for widespread clinical deployment.

# 6. References

[1] “Glaucoma: Facts & Figures.” *BrightFocus Foundation*, 14 July 2021, https://www.brightfocus.org/glaucoma/article/glaucoma-facts-figures. Accessed 10 January 2025.

[2] Haja, Shafeeq Ahmed, and Vidyadevi Mahadevappa. “Advancing glaucoma detection with convolutional neural networks: a paradigm shift in ophthalmology.” *Romanian Journal of Ophthalmology*, vol. 67, no. 3, 2023, pp. 222-237.

[3] Turbert, David. “Fundus.” *American Academy of Ophthalmology*, 14 January 2020, https://www.aao.org/eye-health/anatomy/fundus. Accessed 18 January 2025.

[4] Shoukat, A., et al. “Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach.” *Diagnostics*, vol. 13, no. 10, 2023, pp. 1738–1738.

[5] Campus Eye Center. “A Concise Guide to Glaucoma.” *Campus Eye Center*, 16 May 2022, https://www.campuseyectr.com/news/a-concise-guide-to-glaucoma/.

[6] “Glaucoma Treatment.” *Lang Family Eye Care*, https://langfamilyeyecare.com/eye-health/glaucoma/.

[7] JohnD Akkara, et al. “Identification of Glaucoma from Fundus Images Using Deep Learning Techniques.” *Indian Journal of Ophthalmology*, vol. 69, no. 10, 25 Sept. 2021, pp. 2702–2709, https://doi.org/10.4103/ijo.ijo\_92\_21.

[8] Akter, Nahida, et al. “Glaucoma Diagnosis Using Multi-Feature Analysis and a Deep Learning Technique.” *Scientific Reports*, vol. 12, no. 1, 16 May 2022, https://doi.org/10.1038/s41598-022-12147-y.

[9] N. Z. Zhang *et al.*, “ORIGA-light: An online retinal fundus image database for glaucoma analysis and research,” *PubMed*, Aug. 2010, doi: https://doi.org/10.1109/iembs.2010.5626137.

[10] Abeer Aljohani, and Rua Y Aburasain. “A Hybrid Framework for Glaucoma Detection through Federated Machine Learning and Deep Learning Models.” *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, 2 May 2024, https://doi.org/10.1186/s12911-024-02518-y.

[11] Shoukat, Ayesha, et al. “Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach.” *Diagnostics*, vol. 13, no. 10, 14 May 2023, pp. 1738–1738, https://doi.org/10.3390/diagnostics13101738.

[12] Saha, Sajib, et al. “A Fast and Fully Automated System for Glaucoma Detection Using Color Fundus Photographs.” *Scientific Reports*, vol. 13, no. 1, 27 Oct. 2023,, <https://doi.org/10.1038/s41598-023-44473-0>.

[13] *Deeplearningbook.org*, 2025. https://www.deeplearningbook.org/contents/mlp.html (accessed Mar. 19, 2025).

‌

‌

# List of Abbreviations

| CNN | Convolutional Neural Network |
| --- | --- |
| AI | Artificial Intelligence |
| SVM | Support Vector Machine |
| OCT | Optical Coherence Tomography |

# List of Figures

| Figure 1.1 | Depiction of a healthy eye and a glaucomatous one (Campus Eye Center) |
| --- | --- |
| Figure 1.2 | Example of a healthy and unhealthy fundus |
| Figure 3.1.3 | Applying CLAHE on different channels of a HSV fundus image |

# List of Tables

| Table 3.4 | Details of every transfer learning model used for ensemble learning |
| --- | --- |
| Table 4.2.1 | Performance of models with CLAHE |
| Table 4.2.2 | Performance of models without CLAHE |
| Table 4.3.1 | Percent accuracy gained after applying CLAHE on the dataset |
| Table 4.4.1 | Accuracy improvement when ensemble methods are applied on CLAHE trained models |

# 

1. *Gstatic.com*, 2025. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTOSIbAA8UDEHjFUjkxKVeAgnIL71Cptgqp0ViUTFYd6-WcyedSVfFyh-aDVvzOfQi6MaE&usqp=CAU (accessed Mar. 19, 2025). [↑](#footnote-ref-0)