Multi-Architecture Deep Learning Framework for Glaucoma Detection Using CNN and Vision Transformer Models

by

Syeda Tauba Tahrim
20101264
Maiesha Islam
21301325
Ayon Podder
22101719
Nishat Tasnim
24341302

A project submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026.

Declaration

It is hereby declared that

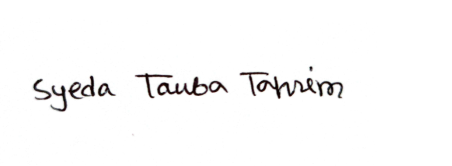
1. The project submitted is my/our own original work while completing degree at Brac University.

2. The project does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The project does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

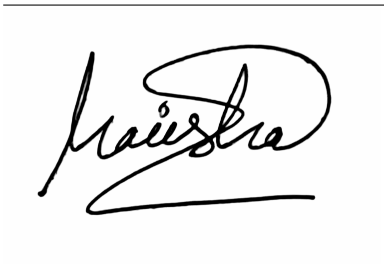
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



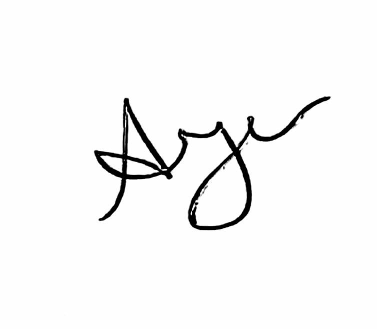
Syeda Tauba Tahrim

20101264



Maiesha Islam

21301325



Ayon Podder

22101719

Nishat Tasnim

24341302

Approval

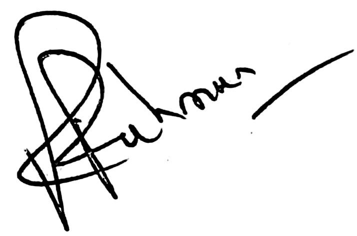
The thesis/project titled "Multi-Architecture Deep Learning Framework for Glaucoma Detection Using CNN and Vision Transformer Model" submitted by

1. Syeda Tauba Tahrim (20101264)
2. Maiesha Islam (21301325)
3. Ayon Podder (22101719)
4. Nishat Tasnim (24341302)

of Fall 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Spring 2026.

Examining Committee:

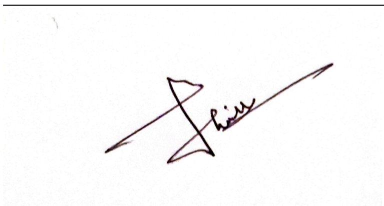
Supervisor:
(Member)



Rafeed Rahman

Sr.Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)



Tahsin Tanim Ramisha

Adjunct Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Associate Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Glaucoma as a primary cause of permanent blindness is still unnoticeable in the initial stages due to the absence of symptoms. The purpose of this study is to design an automated system of glaucoma detection, which will be based on the use of deep learning algorithms to retinal fundus images. The proposed work has introduced a new image preprocessing methodology that will involve the combination of Contrast Limited Adaptive Histogram Equalization (CLAHE), gamma correction, and green channel extraction in order to enhance the quality of the image and make the bright points including optic nerve head and retinal vasculature more discernible and visible that is vital in the detection of glaucoma correctly. The efficacy of four various deep learning structures, such as ResNet50, EfficientNetB0, SwinTransformer, and Graph Convolutional Networks (GCNs) in glaucoma detection, have been tested.

In the present study, the performance of four deep learning models, namely ResNet50, EfficientNetB0, SwinTransformer, and Graph Convolutional Networks (GCNs), is evaluated to determine the effectiveness of the models in the detection of glaucoma. Among the four models, the ResNet50 model recorded the highest performance, with an accuracy of 99.47%, sensitivity of 100%, and specificity of 99.33%. In the present study, the application of the ResNet50 model in the clinical domain is also explored. For the application, the model is converted into the pytorch format, which is compatible with all platforms. It is also optimized to run the model efficiently. A streamlit interface is developed to deploy the ResNet50 model in the clinical domain for the detection of glaucoma. In the Gradio interface, the clinicians can upload the images of the retinal fundus, and the model will give instant results. In the proposed system, the cloud, edge, and hybrid architectures are used. Nevertheless, the next-generation study must focus on the multi-center validation, which will assess the generalizability of the offered model. The applicability and credibility of the proposed model will be enhanced by the addition of other diagnostic tools, including Optical Coherence Tomography (OCT), and the development of eXplainable Artificial Intelligence (XAI) models. The suggested study demonstrates the feasibility of deep learning application in the glaucoma detection, and it offers an effective solution to the early diagnosis and treatment of the disease.

Keywords: Glaucoma Detection, Deep Learning, Retinal Fundus Images, Image Enhancement, ResNet50, Automated Diagnosis, Clinical Deployment

Acknowledgement

First and foremost, all praise is due to the Almighty Allah, whose blessings have enabled the successful completion of our project without any major interruptions. We would also like to express our sincere gratitude to our supervisor Mr. Rafeed Rahman and our co-supervisor Tahsin Tanim for their invaluable support and guidance throughout this work. Their assistance and thoughtful advice have been instrumental in the successful completion of this research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In the case of eye diseases, the healthcare sector has been given a massive boost with the use of Artificial Intelligence, and Deep Learning in particular. The limitations of human analysis in efficiently processing large-scale medical datasets, when it comes to detecting eye diseases such as glaucoma, have been solved by AI, thanks to its ability to draw out intricate patterns and give answers that are both accurate and lightning-fast, something that humans simply can't do.

Glaucoma, a set of eye diseases that destroy the optic nerve and cause irreversible blindness, has turned into a global health crisis, and early detection is crucial to stopping the condition, but its asymptomatic nature in the early stages has led to a lot of underdiagnosis, until now. AI-based systems, driven by deep learning, are being put to work to automate the glaucoma detection process, and take out the guesswork. These systems, have been trained on the Papila dataset that includes images classified as either Healthy or Glaucoma and utilising modern architectures such as Graph Convolutional Networks (GCNs), ResNet50, EfficientNetB0, and Swing Transformer have shown a dramatic increase in classification accuracy. However, the unpredictability of glaucoma images, due to lighting, picture quality, and the specifics of the patients, was a massive hurdle, but image enhancers like Contrast Limited Adaptive Histogram Equalisation (CLAHE) and gamma correction have now been implemented to clean up these pictures and make them clearer. The models need multiple evaluation metrics to evaluate their capacity to perform on unfamiliar data. The deep learning system assessment needs to use classification reports confusion matrices ROC curves AUC metrics Laplacian Variance AUC-PR Log-Loss and cross-validation as its primary assessment tools. The metrics show how effective the model classifies images while also showing its ability to detect false positives and false negatives and its overall predictive strength. In addition to this, there are many metrics that are used to check how the models are performing on the unseen data, some of them are: classification report, confusion matrix, ROC curve, AUC, Laplacian Variance, AUC-PR, Log-Loss, cross-validation etc. These metrics help in determining how accurately the model is classifying the images, how many false positive and false negative cases it is having, how good is the model at predicting etc. The global spread of glaucoma is rising and it is of grave concern. The World Health Organization (WHO) report states that more than 80 million people will suffer from glaucoma globally by 2040 [25]. Similarly, there is a rise in retinal diseases (e.g. glaucoma) among the Bangladeshi population. In Jamalpur General Hospital, the patient's study shows that 2.42% of the patients are suffering from retinal diseases, glaucoma being a major cause of vision loss [14]

As of 2020, in the U.S., more than 1.8 million Americans had advanced age-related macular degeneration (AMD), according to a report from the National Alliance for Eye and Vision Research (NAEVR). In the coming years, NAEVR projects this number will continue to increase. Additionally, NAEVR projects more than 3.3 million Americans will have glaucoma after 2020 [40]. In this study, we want to improve the existing deep learning model for automatic glaucoma detection by applying image enhancement techniques and improving the model architecture as well as comparing the existing deep learning models such as GCNs, ResNet50, EfficientNetB0, and Swin Transformer. In terms of detection accuracy and computational efficiency of the models and finally build an AI-based real-time glaucoma diagnosis solution.

1.2 Problem Statement

The blindness caused by glaucoma is amongst the leading causes of irreversible blindness in the entire globe. The disease causes slow degeneration of the optic nerve which may cause complete blindness when untreated. Traditional diagnostic techniques such as tonometry, visual field testing and optical coherence tomography (OCT) are time consuming and demand personnel with expertise to manage glaucoma early [5]. Additionally, the expenses of such methods of diagnosis may be the obstacle to access particularly in the resource-restricted environments. To address these issues, AI and Deep Learning (DL) methods have become potential solutions to automate the process of detecting glaucoma. Convolutional Neural Networks used as DL models have been applied successfully to retinal images to detect glaucoma and other eye conditions [33]. There are, however, some limitations associated with these models because of the differences in the quality of images, light and the occurrence of other retinal conditions, which may influence the model. Moreover, the medical community is yet to establish the interpretability of these models [17]. Clinicians cannot always rely on AI-based results because of poor transparency in model processes to make predictions. Also, the uneven distribution of datasets which has the healthy population being dominant over the glaucoma patients, tends to produce biased models that become less useful in glaucoma detection. Image processing and enhancement methods like Contrast Limited [29]. Adaptive Histogram Equalization (CLAHE) and gamma and color adjusting are crucial in enhancing the visibility of important areas of the image such as the optic nerve and retinal vessels which are crucial in accurate detection of glaucoma. This research addresses these issues by optimizing deep learning frameworks for automated glaucoma detection using advanced image enhancement techniques. The research will compare and contrast different models available in the state-of-the-art, including Graph Convolutional Networks (GCNs), ResNet50, EfficientNetB0, and SwinTransformer. The evaluation of these models will be conducted on a set of evaluation metrics (such as classification reports, confusion matrices, ROC curves, AUC (Area Under the Curve), Laplacian Variance, AUC-PR (Area Under Precision-Recall Curve), Log-

Loss, and cross-validation) to determine which model is the most efficient to use in the early detection of glaucoma. Figure ?? shows the glaucoma detection of an eye-scanning and fusion approach.

1.3 Motivation

One of the most frequent causes of blindness in the world is glaucoma, the chronic eye disease which is manifested by the damage of the optic nerve. The World Health Organization (WHO) reports that the population with glaucoma will continue to increase tremendously, which proves the necessity of effective, affordable, and convenient methods of its detection [33]. The common types of diagnostic tests, i.e. tonometry and optical coherence tomography are resource-demanding, need a trained professional, and are not accessible in all areas including developing territories.

This is due to the growing incidence of glaucoma, coupled with the obstacles of its detection and treatment, which is why automated diagnostic systems are urgently required [5]. There is a significant potential of AI and deep learning (DL) models in healthcare applications, with a primary focus on automating disease diagnosis with the help of image analysis. In the context of glaucoma detection, deep learning has shown a lot of potential with its capability to analyze retinal images to determine the main characteristics of glaucoma [17]. Through these models, the disease can be detected early before the clinical symptoms can be seen, hence, early intervention is more possible. Nevertheless, even with the remarkable progress, the feasibility of such AI-based systems has not been implemented fully. The effectiveness of these models may be constrained by variability in image quality, changes in lighting conditions and data unbalancing (i.e. more samples are healthy than glaucomatous) [29]. Additionally, the reliability and understandability of AI models are the major issues that medical professionals still need to solve as they must base their decision on transparent and explainable systems to make the correct choice related to patient care.

The purpose of the proposed study in this regard is two-fold, to enhance the accuracy and robustness of automated glaucoma detection systems by including the latest developments in image enhancement, and to compare and test the performance of the latest deep learning models that include Graph Convolutional Networks (GCNs), ResNet50, EfficientNetB0, and SwinTransformer on the task of automated glaucoma detection. This study aims at alleviating these issues by applying the optimized deep learning structures alongside advanced image preprocessing algorithms, including Contrast Limited Adaptive Histogram Equalization (CLAHE), gamma correction, and color normalization. It is hoped that this research makes a huge contribution to the development of more accurate and reliable automated systems of glaucoma detection by improving the quality of the images captured by the retina and also by improving the performance of the models. Such developments are not just relevant to enhancing early diagnosis but also narrowing the healthcare disparity through the provision and expansion of more accessible detection tools to the population with low access levels to specialized care.

1.4 Research Objective

The research aims to develop a computerized deep learning system that can successfully identify early glioma coma stages which will help prevent further vision loss. The research team aims to develop a testing method that will provide both economical and rapid results for diagnosing glaucoma through multiple test attempts. The Machine Learning industry and Deep Learning field have progressed through our establishment of benchmark standards. Scientists around the world are researching how to use fundus images to improve early glaucoma diagnosis methods. The disease remains hidden in its early stage because doctors need to use advanced methods which require special skills to identify the condition. We focus on CNN usage because it delivers precise results through its ability to recognize fine details in images.

1.5 Our Contribution

This study contributes greatly to automated glaucoma detection research by proposing a novel and clinically-inspired preprocessing pipeline and testing various deep learning architectures to complete the task. We created a pipeline that includes Contrast Limited Adaptive Histogram Equalization (CLAHE), green channel extraction, gamma correction, and color normalization, which are specifically created to improve the retinal image features, which are essential when detecting glaucoma. This paper is a first systematic comparison of graph-based, Convolutional Neural Networks (CNNs), transformer-based models, and Swin Transformer, all having been trained and evaluated with the same enhanced dataset under the same preprocessing and evaluation settings. This overall comparison is very insightful in terms of choosing the most successful architecture in terms of glaucoma detection. The proposed model had an excellent performance with 99.47% of accuracy, 100% sensitivity (no false negative) and 99.33% specificity as compared to the current glaucoma screening systems. To guarantee the validity of the findings, we have used a very powerful validation scheme such as 5-fold cross-validation and patient/level data partitioning to avoid data leakage. Multi-metric validation was carried out focusing on clinical significance as the performance of the model could not be independently delivered without keeping a focus on clinical significance. Moreover, the study provides a full end-to-end deployment pipeline, which includes image enhancement and model training, conversion to torch format, and creation of a streamlit interface. This proves that the system is now ready to be used in the real world in clinical practice, and it provides a viable solution to glaucoma screening. The works of this paper open the way to more effective, stable, and scalable automation in medical diagnostics.

1.6 Project Overview

This project introduces a holistic method of automatically identifying glaucoma with the help of deep learning models. The study examines different architectures and ways of improving the accuracy and efficiency of glaucoma detection through retinal fundus images. The project is designed in such a way that it takes the reader through the history, optimization and implementation of the proposed models with specific focus being made on the application and clinical integration in real life.

In this project, an in-depth measure of the automatic detection of glaucoma with the aid of deep learning models is given, with the aim of enhancing both accuracy and efficiency with the help of retinal fundus images. The study guides the reader through history of optimization and implementation of the proposed models and focuses more on a real-world clinical use and integration of the models. Chapter 1 will introduce glaucoma detection, also defining the motivation and giving the problem statement, explaining the need of automated screening and the difficulties involved in it. Chapter 2 examines the literature on related works, including different machine learning and deep learning systems, including CNNs, GCNs, and SwinTransformer, and their advantages and shortcomings to detect glaucoma. The methodology is outlined in chapter 3, which explains the preprocessing pipeline, image enhancement strategies, and the models to be employed (ResNet50, EfficientNetB0, GCN, and SwinTransformer) and the way they are going to be optimized. Chapter 4 shows the performance analysis of the models along with the comparison of accuracy, sensitivity and specificity among other measures, and the clinical significance of the reduction of false negative and the diagnostic accuracy. Chapter 5 is dedicated to the implementation of the optimized ResNet50 model, including the conversion procedure into PyTorch format, the decrease of the inference time, and the comparison with other models. It also talks about developing a Streamlit interface to detect glaucoma in real-time and plans to deploy the solution, both on cloud and edge computing platforms and hybrid computing platforms to guarantee scalability and flexibility. Chapter 6 is a review of the key findings and recommends the way forward on future research studies such as multi-center validation, inclusion of other diagnostic instruments such as OCT, and creation of explainable AI models to create clinical trust.

Chapter 2

Literature Review

Glaucoma is a chronic and progressive eye disease which causes permanent vision impairment unless it is diagnosed on time and treated in the right manner. It is caused by the destruction of the optic nerve, which is very essential in relaying visual signals in the brain which are sent by the retina. Glaucoma has been among the major causes of blindness in the world. The World Health Organization (WHO) estimates that there are more than 60 million glaucoma victims in the world and is already estimated to rise to 80 million by 2025 [34]. Historically, glaucoma diagnosis was based on a manual examination by specialists which takes long before coming up with a diagnosis. The diagnosis is the gold standard which includes measuring intraocular pressure (IOP), visual field testing, and retinal nerve fiber layer (RNFL). These diagnostic tools however can only be done with special equipment and result interpretation may depend on the as far as experience so the clinician performing this test may end up misdiagnosing the patient [21]. Artificial Intelligence (AI) and deep learning algorithms have become a potent technology to automate and improve the medical diagnosis, including glaucoma, in recent years. The ability of AI systems to deliver faster, more precise, and repeatable diagnostic outcomes is a possible application of AI systems, particularly those based on deep learning [4].

2.1 AI in Healthcare

Artificial intelligence, and more specifically a deep learning model, has transformed various domains of healthcare, offering automatic solutions to challenging medical information that are highly accurate. There is a growing trend towards deep learning-based algorithms, including Convolutional Neural Network (CNNs), Graph Convolutional Network (GCNs), and Vision Transformer (Swuis) to process and analyze medical images, such as fundus images, optical coherence tomography (OCT) image, and retina photograph. These models have proven to be better performers in the detection of retinal diseases e.g. glaucoma as they automatically derive some complex features on the retinal images and these features are not easily discerned by the human eye [20]. CNNs in particular and deep learning models overall have also demonstrated high potential in automatic segmentation and classification of retinal images to detect hypertrophy of the optic disc including cupping, which is an early sign of glaucoma. These models can extract hierarchical features out of raw image data and therefore can be contended as very potent tools in detecting minute differences in retinal anatomy that could indicate the early development of glaucoma [41]. The applications of AI in the ophthalmology field aim to decrease the reliance on the clinical judgment subjectivity and enhance diagnostic quality as well as give clinical practice an objective tool of decision making [11].

2.2 The Role of Data in AI Models

The quality of data is a decisive factor in the success of AI models. In the case of glaucoma detection, the most popular dataset is the PAPILA dataset at Kaggle where retinal fundus images of healthy and glaucomatous eyes are annotated. Deep learning models are trained using the data set and learn how to identify such characteristics of glaucoma as optic cup-to-disc ratio, RNFL thickness, and other important signs. Such images are frequently accompanied by other clinical information, including measurements of IOP and visual field tests, to have a more detailed diagnosis method. Nevertheless, retinal images are a difficult source of training deep learning models. First, the classes are very imbalanced—the cases of glaucomatous are significantly less than the healthy cases. Such an imbalance in classes may result in biased models that may be biased towards the majority class (healthy eyes). To reduce this problem, there are data augmentation methods like rotation, flipping, scaling, and color normalization that are used to artificially expand the size of the dataset and enhance model generalization. The use of augmentation assists in the development of robust models which are able to cope with different image distortions which are faced in real-life situations hence enhancing the accuracy of the model in detecting glaucoma [15].

2.3 Deep Learning Models for Glaucoma Detection

There are a number of deep learning models that have been suggested in glaucoma detection. Some of the most popular ones include ResNet50, EfficientNetB0, SwinTransformer and Graph Convolutional Networks (GCNs). In ResNet50 is a convolutional neural network (CNN) which uses residual connections to enable the deeper networks without vanishing gradients. Its application has been in image classification, including medical image analysis. The rich hierarchical features of ResNet50 and its capacity to process the complex image data and learn reinforcements makes it ideal in glaucoma detection. It has been demonstrated that ResNet50 is capable of high-quality detection of fine changes in the optic nerve head and other glaucomatous signs in retinal images [6]. In EfficientNetB0 is a newer model, which aims to scale the network structure in a balanced fashion, and results in high accuracy with low computational cost. It employs the compound scaling technique in order to balance the depth of the network, width and resolution. EfficientNetB0 has been found to be much better in regards to accuracy and efficiency, thus, a useful tool in glaucoma detection in real time [23]. In SwinTransformer is a vision transformer (ViT) which has attracted attention because it is able to capture long-range dependencies in images. SwinTransformer operates under self-attention unlike the traditional CNNs that operate with local pixel relationships to capture global features. This is particularly applicable in identifying intricate structures and patterns in images of the retina including the optic nerve head, which is a sign of glaucoma [38]. SwinTransformer has been demonstrated to be even more effective than conventional CNN models in a variety of computer vision tasks, such as medical image analysis. In GCNs have recently become a point of interest as a model that can be used in image analysis in the medical field. The relationships among the various regions within an image can be represented in the form of a graph with NDNs (regions of interest) and EDs (relationships among the various regions within an image). This method can be very useful in identifying the delicate, and in most cases, complicated association between the optic nerve head and the tissues around the head in images of the retina [43]. By integrating GCNs and CNNs, it is possible to have a hybrid method that enhances feature extraction by exploiting spatial and relational patterns on the data.

2.4 Model Evaluation Techniques

In order to assess the performance of the deep learning models in detecting glaucoma some measures and methods are used. Some of the most widely used evaluation measures are Confusion Matrix, Classification Report, ROC Curve and AUC-PR. The confusion matrix is used to give a detailed breakdown of the rate of classification performance of the model as the true positives (TP), false positives (FP), the true negatives (TN), and the false negatives (FN) are shown. This facilitates the full assessment of the model performance, especially in the perception of the capability of the model to distinguish between normal and glaucomatous eyes. The classification report comprises the important metrics of the performance which include Precision, Recall, F1-Score, and Support. These measures are important to know the capability of the model to place all the classes correctly (either healthy or glaucomatous) and its tradeoff between accuracy and recall. Precision and recall is also a critical factor in medical diagnostics and must be high to reduce false positives and false negatives to a minimum [9]. Receiver Operating Characteristic (ROC) Curve is a curve that presents the rate of true positive (sensitivity) and false positive (1 - specificity). This is quantified in terms of Area Under the Curve (AUC) that quantifies the total ability of the model to distinguish between the two classes (healthy and glaucoma). AUC-PR is superior to AUC in situations where one of the classes (e.g. glaucoma) is underrepresented. It gives an idea of the success of the model in properly detecting positive cases, which is most beneficial in medical use involving skewed datasets [2]. It is used to estimate the sharpness of an image which is important in assisting to detect fine details in medical images. Greater Laplacian variance means that the image is sharp, which may be useful in identifying glaucomatotic ocular characteristics such as optic disc cupping [16]. Cross-validation is a critical model assessment method of deep learning models. Cross-validation is used to make sure that the model works on different subsets of data, and is not overfitting to a specific batch of training data. The result of this is stronger and more generalisable models [1]. Table 1 represents the literature review of Glaucoma detection. Table 2.1 represents the literature review of Glaucoma detection.

Table 2.1: Literature review of Glaucoma detection

Author	**Year**	**Methodology Used**	**Dataset**	**Accuracy**	**Limitations**
Mahum et al. [44]	2022	CNN-based glaucoma classification	Custom dataset, private data	91.5%	Limited to specific dataset; lacks external validation.
Kumar et al. [39]	2023	Transfer learning with ResNet and DenseNet	ORIGA dataset	92.3%	High computational cost due to deep models.
George et al. [13]	2020	Attention-gated U-Net for optic cup & disc segmentation	RIM-ONE dataset	90.7%	Struggles with high variations in optic disc sizes.
Kashyap et al. [26]	2022	U-Net for optic cup & disc segmentation	DRISHTI-GS dataset	89.6%	Limited accuracy in handling complex cases.
Parashar & Agrawal [22]	2021	Deep Wavelet Scattering Network (DWSN)	ACRIMA dataset	93.4%	Sensitivity to noise in image data.
Ratul et al. [8]	2019	U-Net with dilated convolutions	DRISHTI-GS dataset	91.3%	Requires high-quality images for optimal performance.
Shyamalee & Meedeniya [31]	2022	ResNet with transfer learning	ACRIMA dataset	92.8%	Susceptible to overfitting on small datasets.
Chen et al. [24]	2022	Vision Transformers (ViTs)	RIM-ONE and ORIGA datasets	93.5%	Requires extensive training data for reliable results.
David [36]	2023	CNN-RNN hybrid model for glaucoma classification	Custom dataset	92.1%	Challenging to balance between CNN and RNN complexities.
Li et al. [7]	2019	Attention-gated U-Net for optic disc segmentation	DRISHTI-GS dataset	91.2%	Difficulty in segmenting images with overlapping regions.
Thainimit et al. [32]	2022	GANs for data augmentation and noise reduction	Various, including ORIGA	+2% over baseline	GANs can sometimes introduce unrealistic artifacts.
Suban et al. [30]	2022	U-Net for Segmentation	Custom dataset	90.6%	Difficulties in removing noise in clinical data.
Haouli et al. [37]	2023	Classification using Vision Transformers	ORIGA	88.7%	Struggles with the noise and image distortion.

Chapter 3

Proposed Methodology

3.1 Workflow of the Methodology

The study is aimed at designing a control-optimal deep learning model to identify glaucoma through retinal fundus and clinical metadata. The process includes a few related stages, with the initial stage being the collection of datasets and preprocessing, followed by the training and evaluation of a model. The main dataset of the research is the PAPILA dataset on Kaggle that comprises an abundance of retinal fundus images as well as the necessary clinical information, including cup-to-disc ratio (CDR), age, sex, and diagnosis. Preprocessing of the images is first done by scaling the images to 224x224 pixels to normalize the input images to the deep learning models. The pixel values of the images are brought to the scale of [0,1], which makes it easier to train the models. A personal image improvement pipeline is utilized to maximize the exposure of the most important diagnostic characteristics of the optic disc, cup margins, and retinal vasculature. This includes Contrast Limited Adaptive Histogram Equalization (CLAHE), green channel extraction, gamma, and color normalization. These improvements are meant to optimize the difference between the glaucomatous and normal features and retain the integrity of the anatomy of the retinal images. In order to enhance the generalization and prevention of overfitting of the model, the images are augmented. These augmentations include arbitrary rotations, horizontal reflections, intensity changes, and zoom changes. The changes make the dataset realistic and bring about realistic variations that enable the model to learn powerful features without losing important glaucomatous markers. The training is done with four deep learning models which include ResNet50, EfficientNetB0, SwinTransformer, and Graph Convolutional Networks (GCNs). Training of these models is performed on the augmented and already preprocessed dataset by means of cross-validation to guarantee the assessment of the performance reliability. To evaluate the performance of both models, different measures are used, which include classification reports and confusion matrices, ROC curves, AUC, Laplacian variance, log loss as well as cross-validation scores. The models undergo a test not only on their harmonic accuracy but also on their extrapolative ability with respect to unseen data and their efficiency.

The use of the hybrid approach combining image data and clinical metadata, including patient demographics and intraocular pressure (IOP) measurements, is one of the crucial factors of the methodology that enhances the detection accuracy. This combination gives the possibility to analyze it more thoroughly, using both structural changes in the retina and patient-specific factors. The final aim of the study is to find out the best deep learning model to perform automated glaucoma detection, which would be accurate to differentiate between normal and glaucomatous eyes and flexible enough to be applied to the clinical practice environment. The study will seek the most appropriate solution to the early-stage detection of glaucoma through a fine-tuning of the models, which can possibly lead to the enhancement of the diagnostic capacity in clinical settings. Figure 3.1 shows the step-by-step proposed full workflow.

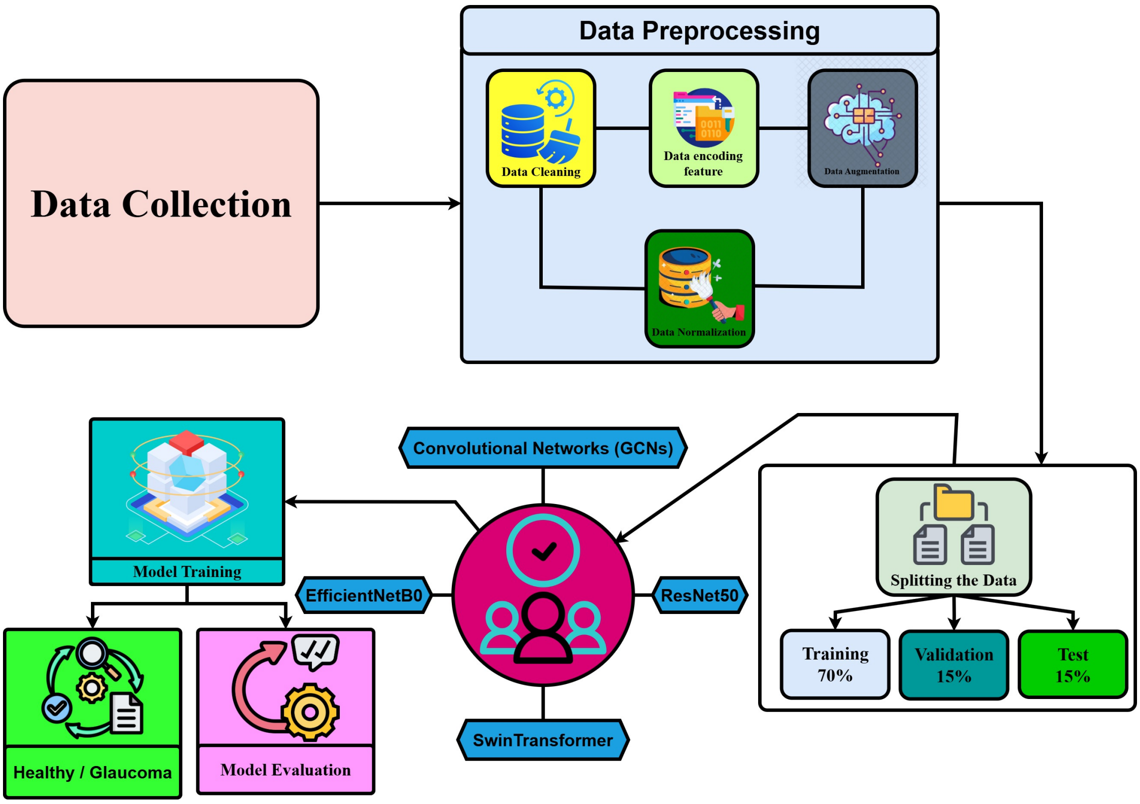


Figure 3.1: Proposed Full Workflow.

3.2 Data Description

3.2.1 Dataset Collection

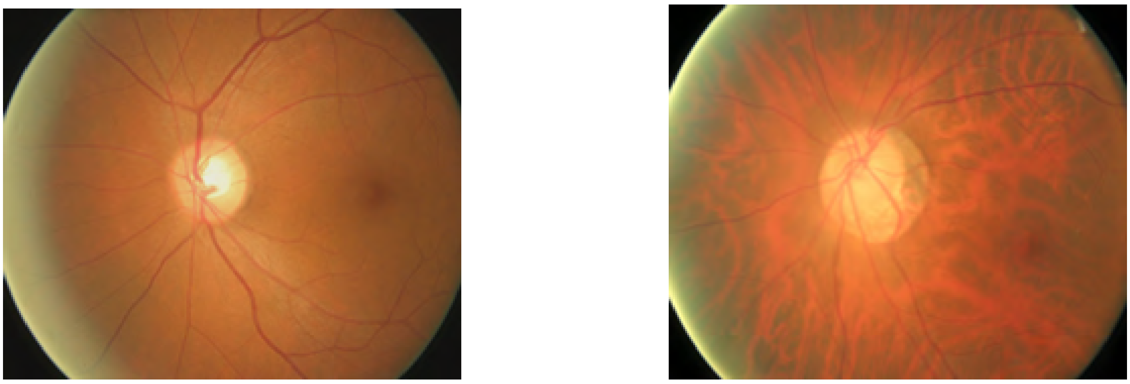
The PAPILA dataset, obtained from Kaggle, is a normalized set and is essential in supporting the identification of glaucoma with the aid of retinal fundus images and related clinical data [27]. It contains pictures of the optic disc (OD) and optic cup (OS) of both eyes of the patients along with the important clinical information like age, sex, and the cup-to-disc ratio (CDR), which is essential to diagnose glaucoma. The data is divided into two Excel workbooks: ‘patientdataod.xlsx‘ and ‘patientdataos.xlsx‘, which contain the data of the right eye and the left eye respectively. This dataset was originally of a small size, but it was augmented with data to reach 2,520 images, with an equal balance of healthy (79.3%) and glaucomatous (20.7%)

cases. The dataset is also comprised of other clinical variables such as intraocular pressure (IOP) and anatomic variables such as pachymetry and axial length, besides the retinal images. These additional parameters help in the accurate diagnosis of glaucoma. The fundus images, all of which have been taken according to standardized photography protocols to maintain a consistent level of illumination and camera settings, are all in the JPEG format and are named in a particular manner so that they can be precisely paired with patient metadata. This organized data can be used to perform structural analysis of retinal features and also be integrated with clinical metadata, which will serve as a complete basis to train deep learning models that will enhance the accuracy of detecting glaucoma.

3.2.2 Input Data

In this study on automated glaucoma detection, the main input data will be the retinal fundus images and clinical metadata of patients. The data is obtained using the publicly available PAPILA dataset stored on Kaggle. The dataset is an extensive source of images with associated clinical information that is important in the training and testing of the functionality of different deep learning models, such as Graph Convolutional Networks (GCNs), ResNet50, EfficientNetB0, and SwinTransformer. The data consists of retinal fundus photographs made in accordance with the guidelines of standard fundus photography. The pictures denote the optic disc (OD) and optic cup (OS) of the left and right eyes of the patients. The pictures are applied to procedures of detecting glaucoma, as well as various abnormalities of the retinal structure. All these images are saved as JPEG files and they are named by the following convention: 'RETPatientIDEye.jpg', where PatientID is a distinct identifier for a particular patient, and Eye is either OD (right eye image) or OS (left eye image). The primary input of the image-based models in the research is these images. In addition to the fundus images, the dataset includes some clinical metadata concerning each patient in separate Excel sheets, 'patientdataod' and 'patientdataos'. Among the most important clinical characteristics is the Cup-to-Disc Ratio (CDR), which is given for both the right (cdrod) and left (cdros) eyes. The ratio is very important in the diagnosis of glaucoma because a high CDR is mostly an indicator of glaucoma. There is also the inclusion of patient demographics, e.g., age and sex, which gives information on the medical profile of the patient. The diagnosis attribute is a binary category to mean that the patient is healthy (0) or has glaucoma (1). Intraocular pressure (IOP), pachymetry (corneal thickness), and axial length of the eye are other clinical data that are also involved and will help to understand the eye health of the patient.

Data augmentation is used to improve the overall applicability of the model and prevent overfitting of the retinal fundus images. These augmentation methods involve geometric transformations of random rotation, horizontal flipping, and slight zooming, which contribute to simulating the real-life changes in the images. Also, photometric transformations, such as brightness control, contrast change, and color jitter, are implemented as a way to enhance the variety of the dataset. Spatial transformations, which include random affine transformations (translation, scaling, etc.), further increase the diversity. These additions are used to generate more varied information on the images that are present, and the models can generalize and detect glaucoma with higher accuracy. Figure 3.2 shows sample data.



(a) Glaucoma (b) Non-Glaucoma

Figure 3.2: Sample Data. (a) Glaucoma, (b) Non-Glaucoma

3.3 Statistical Analysis of Data

There are 2,520 retinal fundus images that are augmented. 79.3% (1,998 retinal fundus images) images are healthy, and 20.7% (522 retinal fundus images) images are glaucomatous. This even distribution guarantees that the model is not biased towards one class over the other, ensuring the model can be trained and evaluated effectively. The models are trained on the binary classification scores (Healthy = 0, Glaucoma = 1) of the data. A range of features is present in the clinical metadata, which includes age, sex, Cup-to-Disc Ratio (CDR), intraocular pressure (IOP), pachymetry, and axial length, which are vital in the identification of glaucoma. The age breakdown is also of significance as glaucoma is prevalent among the elderly. There is no sex imbalance, with both male and female patients equally represented in the dataset. The sex is represented as a binary variable (Male = 1, Female = 0). The values of CDR are important in the detection of glaucoma, where higher values are normally associated with greater chances of glaucoma. Data augmentation on the images was performed to enhance the generalization of the models and avoid overfitting. These augmentations consisted of geometric transformations like random rotation, horizontal flipping, and zoom variations. Also, brightness and contrast variations, as well as color jitter, were applied to enhance the model's robustness. Simulation of real-world variations was also done by applying spatial transformations, which included random affine transformations with translation and scaling. These augmentations raise the effective dataset size and contribute to the improved performance of the model.

3.4 Data Preprocessing

Data preprocessing pipeline, which will be used in this study, will ensure that retinal fundus images, as well as clinical metadata, are best prepared to be trained deep learning models to identify glaucoma. It starts with loading and merging of datasets whereby the clinical data of both the right and left eye in different files are merged, image resizing, and normalization are done so that there is uniformity in the dataset. The images are subjected to an enhancement pipeline which includes image enhancement methods such as CLAHE to enhance the contrast of images, extracting the green component that improves visibility of the vessel and disc, gamma correction to increase or decrease the brightness of images and color normalization to stabilize image appearance across cameras. In order to enhance model generalization, data augmentation methods are used including rotation, zoom and brightness manipulation. Also, categorical variables are encoded, and continuous variables are normalized before clinical metadata is used. Lastly, the data is divided into training and validation and test splits, and the counts of healthy and glaucomatous cases are equal. This super pipeline results in improved images and clinical data and allows the development of effective deep learning models on general glaucoma detection. Figure 3.3 shows the data pre-processing workflow.

3.4.1 Dataset Selection

To select the dataset during the data preprocessing phase, the PAPILA dataset was selected from Kaggle because it has a complete set of retinal fundus images and clinical metadata. This data has images of the optic disc (OD) and optic cup (OS) of each eye, with the most important clinical characteristics that include age, sex, and cup-to-disc ratio (CDR), which are essential for detecting glaucoma. The selection of the dataset was done with much care so that it was well balanced with both healthy and glaucomatous cases, making it perfect for the training and evaluation of deep learning models.

3.4.2 Data Collection

The process of data collection included the retrieval of the PAPILA dataset from Kaggle that contained a mixture of retinal fundus images and the associated clinical details of patients. The data was in the form of images of healthy and glaucomatous eyes, along with important clinical information like age, sex, and labeling of diagnosis. To check the integrity of the dataset, an initial assessment was conducted, and the image files were organized and combined with the clinical information. This was necessary to prepare the data properly for the subsequent preprocessing and analysis.

3.4.3 Dataset Loading and Merging

In this study, the dataset was the PAPILA dataset from Kaggle, which comprises retinal fundus images and clinical metadata. The dataset was first saved in two different Excel files: ‘patientdataod.xlsx‘ for the right eye (OD) and ‘patientdataos.xlsx‘ for the left eye (OS). Such files contain valuable clinical information, including:

- Cup-to-disc ratio (cdrod, cdros) for either eye.
- Demographics of the patient such as sex and age.
- The categories of diagnosis, which include: healthy = 0, glaucoma = 1.

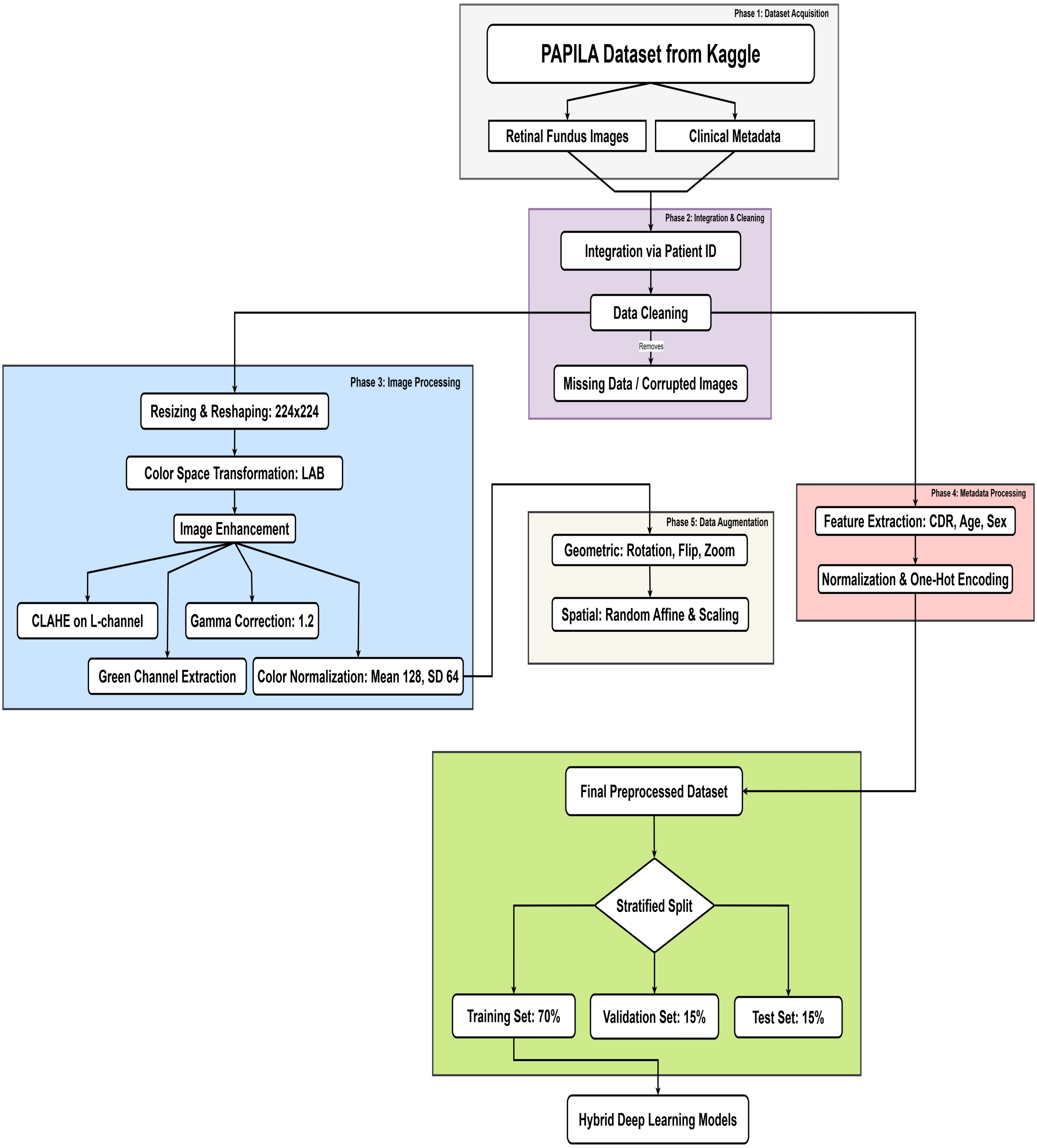


Figure 3.3: Data Pre-processing

The clinical data of both eyes (OD and OS) of individual patients are combined using unique patient identifiers to ensure that the images and clinical data are matched accordingly. This combination process produces a complete dataset, with every retinal image having its corresponding clinical data. This single dataset can then be preprocessed and trained on the model.

3.4.4 Data Cleaning

Data cleaning plays an important role in improving the efficacy and effectiveness of the deep learning models. The steps followed in this study included removing any inconsistent, irrelevant, or missing data from the PAPILA dataset, and only high-quality, trustworthy data was used to train the models. This involved checking image filenames, ensuring proper patient identification, and correcting mismatches in clinical metadata such as age, gender, and diagnosis labels. To ensure that the dataset was intact, any duplicate or corrupted images were deleted. The cleaning process ensured that only valid and accurate data went into the model, making the training more effective and the predictions more accurate.

3.4.5 Data Reshaping and Resizing

To ensure the application of the same dimensions of inputs across all models, the individual images in the dataset had to be resized to the uniform size of 224x224 pixels, which is a commonly accepted dimension that works with most deep learning architectures, including ResNet50, EfficientNetB0, and SwinTransformer. This downsizing guarantees that the images can be used in convolutional neural networks (CNNs) and vision transformer (ViT) models, which require fixed input sizes to work. The images were reduced to this size without losing the most significant aspects of the image, so the visual information that is essential to detect glaucoma was not lost. The image size consistency improves image model training by making every input the same and in the right format.

3.4.6 Data Normalization

Normalization will make sure that all the features equally influence the machine learning or deep learning algorithms in training as well as in detection. As they are of great size, some attributes are not allowed to have too much power in attributing to the outcomes of the model. A common form of scaling is normalizing the qualities to a 0-1 range. HoIver, alternatives exist in large numbers. In order to do this, Min-Max scaling strategy could be used.The min-max equation is presented [35]. By setting the features to the same scale, By means of normalizing them, I help the deep learning or machine learning algorithms on the dataset to be more accurate and efficient.

3.4.7 Color Space Transformation

The photos are translated to the LAB color space to increase contrast, especially in the L (luminance) channel, which can be modified to enrich the image characteristics. This transformation is used to highlight significant details for glaucoma detection.

3.4.8 Image Enhancement

The following image enhancement pipeline is used to enhance the diagnostic visibility of important retinal structures, i.e., the optic disc and optic cup.
A number of methods used on image enhancement were applied to maximize the visibility of the retinal structures that were required to identify glaucoma.

- The LAB color space was used with CLAHE (Contrast Limited Adaptive Histogram Equalization) applied to the L channel to enhance local contrast without enhancing noise.
- Green Channel Extraction was used to remove the green channel to increase the clarity of the optic disc and retinal vessels that are vital in the detection of glaucoma.
- To make slight variations more pronounced, increasing the visibility of the cup-to-disc ratio, Gamma Correction was added (factor 1.2) and this enhanced the detection of glaucoma.
- Lastly, the Color Normalization was done with a standard deviation of 64 and a target mean of 128 to normalize the color of various imaging systems so that it is consistent and diagnostic faithful.

3.4.9 Data Augmentation

The data augmentation methods were used to improve the model generalization and avoid overfitting. Random rotations of $\pm 8°$ were used as geometric transformations and there was a probability of 50 percent of a horizontal flip and zoom changes of $\pm 8\%$. Photometric transformations consisted of random manipulation of brightness, contrast and color. Random affine transformations were utilized in spatial changes with transformation shift of $\pm 10\%$. Scaling was applied to 0.9 to 1.1. These additions aided the model in managing the image variations of the real world.

Handling Clinical Metadata

The clinical metadata consisted of Cup-to-Disc Ratio (cdrod, cdros) in both eyes, age, sex (encoded: 1 male and 0 female), and a diagnosis name (encoded: 0 healthy and 1 glaucoma). Categorical variables such as sex were one-hot encoded. These clinical characteristics were added together with the retinal image data to produce a dual-input deep learning model, which enhanced the diagnosis ability of the model.

3.4.10 Data Distribution and Split

In this study, the PAPILA dataset is split into training, validation, and test sets to ensure effective model evaluation and generalization. The data includes retinal fundus images and their respective clinical data (including age, gender, cup-to-disc ratio, and diagnosis). Each patient's clinical metadata is matched with the corresponding images, and the data is distributed in the following manner:

- **Training Set**: The training set occupies the largest portion of the data; it includes 70% of all data. The deep learning models are trained on this subset to learn the underlying patterns and characteristics of the images and clinical metadata of the retina.
- **Validation Set**: The validation set consists of 15% of the entire data. This subset is used during the training process to observe the performance of the model and to adjust hyperparameters in order to prevent overfitting.

- **Test Set**: The remaining 15% of data is reserved as the test set. It is kept separate from the training process and is used to test the ultimate performance of the model on unknown data to evaluate its generalization capability.

The data split is performed in a stratified way to ensure that the distribution of healthy and glaucomatous images is similar in the training, validation, and test sets. This avoids class imbalance and ensures that each subset accurately reflects the overall dataset.

3.5 Architectures

This study addresses the problem of detecting glaucoma with the help of a set of modern deep learning models, each of which was chosen based on its specific advantages and suitability to the task. The ResNet50, EfficientNetB0, SwinTransformer, and Graph Convolutional Networks (GCNs) are the selected models to be used in this study. Each architecture will be used with the aim of maximizing its unique benefits to enhance the performance of the model in identifying glaucoma using retinal images.

- **ResNet50**: The ResNet50 architecture is a convolutional neural network (CNN) based on residual connections to eliminate the vanishing gradient issue and enable the construction of more complex networks. The model has been characterized by its capability to capture intricate patterns in the image, making it suitable for the task of glaucoma detection in retinal fundus images. Its rich layers are helpful in retrieving stratum features essential for detecting glaucoma-associated patterns in the optic disks and cup.

- **EfficientNetB0**: EfficientNetB0 is a highly efficient CNN network that is both performance and computationally efficient. It employs a compound scaling method to scale depth, width, and resolution, resulting in high accuracy with minimal computational cost. EfficientNetB0 is particularly optimal for large-scale image classification and is capable of effectively processing retinal images while retaining high classification accuracy.

- **SwinTransformer**: SwinTransformer is a vision transformer architecture, which works by transforming image patches with a self-attention mechanism based on windows. It excels at capturing both global and local dependencies in images, which is useful for interpreting the complex structure in retinal images. SwinTransformer was chosen due to its state-of-the-art performance in recognizing fine-grained objects, such as the cup-to-disc ratio during glaucoma determination.

- **Graph Convolutional Networks (GCNs)**: A GCN is a variety of neural network that is intended to work with graph-structured data. A graph based model has been used in this study to integrate image data and clinical metadata. The clinical data which includes patient demographics, cup-to-disc ratios etc are converted to a graph structure with the patient as a node and the relationship between clinical features as an edge. This would increase the capability of the model to reveal the correlation between the retinal image features and clinical metadata, which would boost the models accuracy in diagnosing glaucoma.

All the models were trained on a batch of 64, using the Adam optimizer and sparse categorical cross-entropy loss through 100 epochs. It has 2520 pre-processed and improved retinal images and was standardized to 224x224 pixels to ensure uniformity across the models. The technique of data augmentation was used to improve the dataset (rotation, flipping, zoom, and brightness adjustment) to increase the strength of the model and its extrapolation. The standard metrics used to evaluate the performance of each model included accuracy, precision, recall, F1 score, AUC-ROC, and cross-validation, to be able to do a comprehensive evaluation. These metrics were used to compare the performance of the models during training and validation in detail and find the most reliable one in glaucoma detection.

3.5.1 Graph Convolutional Network (GCN) Architecture

Graph Convolutional Network (GCN) is implemented on the basis of graph formulation where an eye of each patient is represented as a node on a global similarity graph [28]. To obtain a graph, the following process is followed:

Graph Construction

The visuals of every eye of a patient are treated as nodes of the graph with unique combinations of PatientID and Eye. When the duplicates are eliminated, there are 252 nodes left. A fundus picture and clinical indicators of age, gender and ophthalmic measurements are associated with each node.

Node Feature Extraction

Embedding images with clinical data is used to create node features. The image representations are computed through the average of the outputs of a ResNet18 network to the entire augmented fundus image set. The 522-dimensional feature vectors of the individual nodes exhibited in Equation 1 are the combination of these image embeddings with ten normalized clinical variables (e.g., age, gender).

$$\text{Node Feature} = \text{Avg}(\text{ResNet18}(\text{Augmented Image})) + \text{Clinical Features} \tag{1}$$

This feature vector is then used as input to the GCN layers.

Adjacency Matrix Construction

The edges are established according to the similarity of clinical features. A cosine similarity between each pair of nodes is done with a threshold of 0.2 to connect each node to its top 15 similar nodes. This forms 3628 weighted undirected edges. The GCN model has two graph convolution layers, where the hidden-layer with 64 dimensions and a 60 percent dropout is used to regularize the model. This model predicts glaucoma by giving an analysis of the clinical similarities between the eyes of various patients [28].

GCN Layer

The GCN model has two graph convolution layers and the hidden dimensionality of each is 64. The model is an instance of graph convolutions that compress information of the nodes around them, which reflects data about the relationships and dependencies between clinically similar eyes. In Equation 2, the update rule for the graph convolution layers is as follows:

$$H^{l+1} = \sigma\left(\hat{A}H^{l}W^{l}\right) \tag{2}$$

Where:

- H^{l+1} is the feature matrix after the $(l+1)$-th layer.
- H^{l} is the feature matrix at the l-th layer.
- W^{l} is the learned weight matrix for the l-th layer.
- $\hat{A}$ is the normalized adjacency matrix.
- σ is the activation function (typically LeakyReLU).

The last node-level classifier is used to determine whether the eye is glaucomatous or healthy.

Dropout for Regularization

In a bid to avoid overfitting, the model is trained on a 60% dropout rate, which will assist the model to be more general and not rely on certain features that may result in overfitting.

Glaucoma Prediction

The end result of the GCN model is the prediction of each node (eye) whether it is healthy (0) or has glaucoma (1). The model uses the structure and the characteristics of the nodes in the graph to indicate the probability of glaucoma using clinical similarities among the eyes. In Figure **??** shows the Architectural Diagram of the GCN Model.

3.5.2 ResNet50 Architecture

The ResNet50 Network, proposed by Kaiming He et al. [3], is a five-convolutional deep convolutional network. The input size is [224x224x3]. It has left-over blocks that have [1x1], [3x3] and [1x1] layers of convolution. Figure **??** shows the ResNet-50 structure. The initial layer contains 64 7x7x3 kernel size and stride 2. Subsequent layers are followed by depths of 256, 512, 1024 and 2048 residual blocks. The last layer consists of 1000 neurons that are classified using a SoftMax activation
In the implementation of ResNet50, ImageNet pre-trained weights and transfer learning are used, which are quite effective in dealing with small datasets in medical imaging applications. These changes to the architecture are as follows:
ResNet50 Network proposed by Kaiming He et al. [18]. It won the image categorization in 2015 with a top 5 error of 3.57%. It is built up of five convolutional

layers. The ResNet-50 architecture takes an input that is of size [224 x 224 x 3]. ResNet-50 comprises five convolution layers. Each of the layers has 3 convolution blocks (residual) with convolution layers of [1x1], [3x3] and [1x1]. The first convolution layer contains 64 kernels that are of size [7 x 7 x 3], stride of 2 and pad of 3. The next maxpool layer is the maxpool layer having a filter size of [3 x 3] and stride of 2. The third convolutional layer consists of three Residual Blocks (layers), where the first convolutional layer is [1x1], the second layer is [3x3], and the third layer is a 1x 1 convolutional layer, with a depth of 256. The second convolutional layer has depth of 4 Residual Blocks that are 512 dimensions. The third layer has 6 residual blocks with the depths of 1024 and finally, 3 residual blocks with the dimension of depth of 2048. Following these five convolutions, max pool layer is followed and one finally linked layer with size [1000]. Finally, there is the cost function, which is SoftMax, applied to the output layer.

In the implementation of ResNet50, ImageNet pre-trained weights transfer learning is used, which is quite effective in dealing with small datasets in medical imaging involving applications [24] show in figure **??**. These changes to architecture are:
The ResNet50 model uses the input layer as the input layer where the size of the images is [224x224x3]. The initial convolutional layer consists of 64 7x7 filters with a stride of 2 and then there is a batch normalization followed by ReLU activation. The spatial dimensions are down-sampled by a max-pooling layer of [3x3] filter, and a stride of 2. The network consists of four steps of residual blocks, all of which have a larger filter size (64, 128, 256, 512) in each step. All blocks follow a bottleneck architecture, which comprises [1x1], [3x3], and [1x1] convolutions with skip connections. Global average pooling is used to decrease the feature map to 1x1, and then a fully connected layer with 1000 output neurons to classify. The last output layer applies the softmax activation to binary classification (healthy or glaucoma) and has two neurons.

Replacing the last fully connected layer with binary and 0.5 dropout modified classifier. Adding batch normalization and residue connections to deep network training. The use of a better preprocessing pipeline, in order to have optimal input quality. The reasons why ResNet50 was selected are a successful prior experience in the medical imaging area and the ability to learn hierarchical feature representations that are able to learn local and global patterns that play a significant role in glaucoma diagnosis [42].

The use of ResNet50 was chosen as compared to other models like EfficientNetB0, GCN and SwinTransformer because ResNet50 performs highly, is robust and has clinical significance in determining glaucoma. It had a perfect sensitivity (100%) and specificity (99.33%), with an overall accuracy of 99.47% which is very successful in identifying glaucoma with few cases of a false positive. The vanishing gradient issue that is prevalent in the deeper networks is overcome by the fact that the model uses residual learning to extract more complex features in retinal images. ResNet50 was also found to have good generalization power and little overfitting, which was reflected by its consistency in cross-validation, and it was also found to work well on different datasets. It is a useful tool in automated screening due to its clinical reliability and capability to detect glaucoma without underreporting any cases. Its performance was also improved by the use of transfer learning which was using pre-trained ImageNet weights, even when the dataset was relatively small. Also, ResNet50 provides a trade off between accuracy and computation efficiency, and thus can be deployed in real time in clinical scenarios. The hyperparameter optimization of the model further enhanced the model to ensure that it was not only clinically reliable but also practical as it minimized overfitting. All in all, the combination of a high-performance level, learning efficiency, and the ability to adapt to clinical conditions allows ResNet50 to be the most suitable option when it comes to the detection of glaucoma.

3.5.3 EfficientNetB0

EfficientNet B0 architecture which is a subset of EfficientNet is intended to be used in ocular disease diagnosis when efficiency and accuracy are needed [39]. It applies depthwise separable convolutions to minimise the use of parameters without sacrificing expressiveness. The dimensions of the input layer are 64x64x3 and the model is a convolutional and pooling layer with a reduction of the spatial dimensions by GlobalMaxPooling2D. The learning of features is done through a fully connected layer of 256 units with the ReLU activation [10] and batch normalization and dropout rate of 0.5. The last output layer consists of 16 units with the softmax activation. The Adam optimizer, 5 x 5×10^{-4} learning rate, and sparse categorical cross-entropy loss function are used to train the model and balance efficiency and performance when dealing with classification problems.[10].

The EfficientNetB0 architecture consists of the following components:

- **Input Layer**: The dimension size of the input is 224x224x3 (height, width, and channels), matching the size of the image in the data following preprocessing. The 3-channel images (RGB) are fed into the model.
- **Initial Convolution**: The initial convolutional layer uses a 3x3 kernel with a stride of 2, followed by batch normalization and ReLU6 activation.

  EfficientNetB0 makes the use of Mobile Inverted Bottleneck Convolutions (MBConv) to optimize the efficiency of models. The MBConv blocks consist of depthwise separable convolutions using both 1x1 and 3x3 convolution kernels, bottleneck structure and a linear bottleneck to store information. It also uses a compound scaling technique which trades off the depth, width and resolution of a network to be highly accurate but not very costly in terms of computation. The MBConv blocks are followed by global average pooling which flattens the feature map into a single vector, which is then run through a fully connected layer (with 2 neurons; in the case of binary classification: Healthy = 0, Glaucoma = 1) with the help of softmax activation.

  The architecture is shown in Figure **??**. The model is optimized to ensure accuracy while being computationally efficient, which is especially important in medical applications where resource constraints are common.

Training the EfficientNetB0 Model

The EfficientNetB0 model is trained on 2,520 augmented retinal images using the Adam optimizer with a learning rate of 5×10^{-4} and sparse categorical cross-entropy. It is trained for 100 epochs with a batch size of 64, incorporating data augmentation to enhance generalization and prevent overfitting. The model is evaluated using metrics like accuracy, precision, recall, F1-score, AUC-ROC, and cross-validation. EfficientNetB0 offers high performance with low computational cost, making it suitable for real-time glaucoma detection and early diagnosis.

3.5.4 SwinTransformer

The latest innovation in the vision transformer architecture is SwinTransformer implementation [19]. It introduces hierarchical feature maps attention and shifted-window attention, which respectively express local and global interdependence in images [12]. The architecture was designed with Swin-Tiny which is an entry-level model and provides a tradeoff between the performance of the computers and the price shown in Figure **??**. The model calculates the relationship between various components of the fundus image by progressively examining the image in repeated sections .

Architecture Details

SwinTransformer architecture takes images and transforms them respectively by scaling images to 224x224 pixels, breaking them into non-overlapping 4x4 patches. These patches are squashed and put into a token sequence. It is a model based on window based self-attention (W-MSA) in which attention is computed at the small image windows and thus it is computationally efficient than global attention. Then shifted window attention is used to acquire long-range dependencies by window shifting in the second layer to enhance global context. The blocks of the transformer include patch merging, self-attention, and layer normalization. Lastly, the feature tokens are fed through a global average pooling layer and a fully connected layer with binary classification (Healthy/Glaucoma) by use of the softmax activation.

Training the SwinTransformer Model

The SwinTransformer model is trained on 2,520 augmented retinal images. The model uses the Adam optimizer with a learning rate of 5×10^{-4} and sparse categorical cross-entropy loss as the loss function. It is trained for 100 epochs with a batch size of 64. Data augmentation is applied to the images during training to improve generalization and prevent overfitting. The model's performance is evaluated using metrics such as accuracy, precision, recall, F1-score, AUC-ROC, and cross-validation. The SwinTransformer architecture has proven to be highly effective in the task of glaucoma detection. With its novel use of shifted-window self-attention and hierarchical feature maps, the model excels at capturing both local and global dependencies in retinal images. Its ability to process complex visual data while maintaining efficiency makes it a valuable tool in medical image analysis. The performance of the SwinTransformer model in glaucoma detection is enhanced by the use of advanced image preprocessing techniques and data augmentation, making it a promising candidate for automated glaucoma detection systems.

Chapter 4

Result Analysis

4.1 Experimental Settings and Hyperparameter Configuration

To achieve the best performance of deep learning models and provide the possibility of a fair comparison of different architectures, a comprehensive training method was created in this research. It was done through the use of a lot of experimentation with hyperparameters being tuned, optimization, and regularization procedures for each model. The primary objective was to increase the ability of the models to make accurate predictions of glaucoma using retinal fundus images and clinical metadata so that the selected architectures (ResNet50, EfficientNetB0, SwinTransformer, and Graph Convolutional Network) would work in practice in the real world.

4.1.1 Loss Functions and Optimization Algorithms

The common loss function that was utilized in all of the models was the cross-entropy loss because it is typically applied to classification problems in which the aim is to allocate each input (retinal image) to a particular class (Healthy or Glaucoma). Also, a smoothing based on a smoothing factor of 0.1 was used. Label smoothing is a method of avoiding overconfidence of the model in its predictions, particularly in medical diagnostics when false positives or false negatives can be very hard. Label smoothing can make the model more generalized to non-seen data because the model will not be able to become excessively sure of its ability to predict the false category. The two optimization algorithms that were investigated within the frame of the research were to find out which approach is the most effective to train the models. The optimization algorithm that was initially tested was AdamW, which is an adaptive estimation optimizer of moments which modifies the learning rate depending on the first and second moment of the gradient. It has been known to be able to deal with sparse gradients and typically can be effective when used to train deep learning models. Stochastic Gradient Descent (SGD) with momentum (0.9) was the second optimizer. The traditional way of the technique is called SGD, the gradients are calculated on a random sample of the training data. The momentum term can be used to accelerate the convergence and dampen oscillations, which makes it appropriate to high-dimensional, non-convex optimization because it is often used to train deep neural networks. The learning rate was optimized by a search in the given range of [0.00005, 0.001] of the architectures. A learning rate is essential since it defines the size of the step taken during each iteration and it has a direct impact on the rate of convergence and stability of the model. Excessively large learning rates may cause overshooting of the optimal solution and excessively small learning rates may cause slow convergence.

4.1.2 Learning Rate Scheduling

In order to enhance the model performance, two schedules of changing learning rate were applied to dynamically adjust the learning rate in training. The initial one followed was the cosine annealing which shapes the learning rate to decrease gradually in a curve of cosine. This method of scheduling enables the model to take huge steps during the initial epochs and smaller steps towards the optimal solution such that it seeks to optimize the weights of the model without over-fitting. The second schedule employed was the ReduceLROnPlateau that reduces the learning rate when the validation performance is not improving anymore. This would assist in avoiding overfitting and the model will keep on learning even upon hitting a plateau in the validation loss. The best learning rate schedule was determined with respect to the performance of the model at the early training runs making sure that the model reached a solution effectively.

4.1.3 Training Protocols

All the models had the same training protocols to be consistent and to make a fair comparison. The epochs were configured to be 20 to 30 epochs based on the needs of particular architecture. The epochs selection is also significant in order to provide a model with enough time to learn, without having too many epochs, and overfit to the training data. Another hyperparameter that differed was the batch size i.e.16 to 64 depending on the memory capacity of the computer used to train. The bigger the batch size the more accurately the gradient is estimated, and the bigger the memory is needed. Small batch sizes are also more memory efficient, but can cause a noisier gradient estimation, which can influence convergence.
The training machine was a Tesla T4 graphics card and 16GB RAM. The computing ability of the GPU played a critical role in the training of the deep learning models particularly when using massive image data sets. The availability of a high-performance GPU essentially decreased the training times and allowed using more complex models such as SwinTransformer and Graph Convolutional Networks (GCNs).

4.1.4 Data Splitting Strategy

Data was split into two parts with a patient-level stratified splitting method to avoid data leakage and promote the external validity of the model[23]. This approach made sure that the pictures of a particular patient were only inserted in one of the dataset subsets (train, validation or test) so that the model does not learn any patient-specific characteristics that would cause it to overfit. It was divided into 70 percent training, 15 percent validation and 15 percent test data. The model was trained using the training set, tuned using the validation set, and evaluated by the use of the test set. This division assists in testing the model to extrapolate to the unseen data and avoids overfitting on training data[33]. The data was divided as in the Table 4.1.

Table 4.1: Dataset Split and Distribution

Split type	**Percentage**	**Images**	**Normal**	**Glaucoma**
Training set	70%	1764	1399	365
Validation set	15%	378	299	79
Test set	15%	378	300	78

4.1.5 Performance Metrics

The models were evaluated by a set of performance measures to determine their effectiveness in training and validation. The main measures of evaluation were Accuracy, Precision, Specificity, F1-Score, and ROC AUC are shown in Equation 3, 4, 5 and 6.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

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

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{6}$$

4.1.6 Validation Methodology

There were a number of plans used in the validation methodology to guarantee sound performance measurement. The stability of the model was tested using a 5-fold cross-validation method to evaluate the stability of the model on various data splits. This method gives a stronger assessment by training and validating the model with various subsets of the data more than once. Confidence interval and p-values were used to establish the statistical significance of the results to compare performance of the models. To determine the false-classified cases, the error analysis was performed which assisted in getting an insight into the limitations of the model. It was overfitted through comparing the analysis of the performance of the model conducted on the training and validation sets. The analysis was used to make sure that the model was not overfitting to the training data.

4.1.7 Clinical Validation

Clinical validation was done to ascertain that the model would be effective in clinical environments. It involved a false-negative study, which involved the analysis of the cases of glaucoma the algorithm was unable to identify. There was also a false-positive study that was carried out to investigate instances of cases in which glaucomatous eyes were misdiagnosed to be healthy. The calibration of the confidence of the model was also checked through comparing the real accuracy with the real confidence scores of the model. This was done to make sure that the predictions made in the model were not just accurate but as well reliable with respect to confidence. One of the measures used to assess the effectiveness of the computations was the inference speed of the model at run-time, as well as resource consumption. This particularly applies in the clinical environment as time and resources are scarce, and the model must deliver quick and dependable outcomes. The proposed comprehensive evaluation framework was a strong foundation of judging the effectiveness of the models as they are clinically relevant and can be used effectively in the real world to detect cases of glaucoma.

4.1.8 Hyperparameter Configuration

There are a number of hyperparameters that were optimized between various deep learning models in this study to guarantee the best performance of glaucoma detection. In the case of all the models, the batch size was adjusted between 16 and 64, based on the available memory in the gpu. The larger the batch size, the easier it is to estimate the gradient, whereas the smaller is the batch size, the more memory efficient it is. There were between 20 and 30 epochs to make sure that there were sufficient iterations to train the model and prevent overfitting. Each architecture was evaluated with the learning rate range of [0.00005, 0.001] to ensure that the learning rate is optimally set since it can have a considerable effect on the convergence speed and performance of a model. The AdamW, an adaptive optimizer, which is the most efficient to apply to deep learning tasks, was employed as the optimizer, but SGD with momentum (0.9) was also experimented to maximize the convergence through the smoothing of the update made during the training. Besides, 0.1 label smoothing was added to the cross-entropy loss to avoid overconfident predictions that can enhance the generalization of the model. The Cosine Annealing and ReduceLROnPlateau were applied to the scheduling of the learning rate; the former reduced the learning rate at a slow rate in a cosine curve and the latter reduced the learning rate when the validation performance did not improve. The regularization and prevention of overfitting were achieved by applying dropout and setting it to 60% when training Graph Convolutional Networks (GCN). To optimize the learning rate the optimizer was optimized with a search range of 0.00005-0.001 and the models converged effectively. To obtain credible and clinically applicable glaucoma-detecting models, these hyperparameters were optimally tuned and trained and tested on the performance of the architecture to each architecture. The hyperparameters for glaucoma Detection Models demostrate in Table 4.2

4.1.9 Hyperparameter Tuning Analysis

The relative analysis of hyperparameter configurations of the deep-learning models elicits definite optimization policies which can be viewed as possessing the characteristic of each of the architectures and training dynamics. ResNet50 was optimized to a relatively small weight-decay constant of 1e-6, a small batch size of 16, and learning rate of 5e-5; all of these parameters are moderate to allow this model to be fine-tuned without instability. Conversely, the Graph Convolutional Network had a significantly higher learning rate of 0.01 that is integrated with full-batch updates that indicates that graph-based models have some of the benefits of global gradient propagation and quicker convergence in the case of the relatively small medical imaging datasets encountered. Due to the higher sensitivity of the Swin-Transformer towards optimization hyperparameters, a moderate learning rate of 5e-5 and a large batch of 64 and a large weight-decay factor (0.1) were to be chosen as a requirement in order to stabilize the self-attention processes and avoid early overfitting. Finally, EfficientNet-B0 employed a learning-rate of 1e-4 and a very undercut batch size of 8; both were very influenced by the memory limit, and network-scale of compound dynamics. This point of architecture-specific tuning is achieved by all of these separate hyperparameters scores.The hyperparameter tuning demostrate in Table 4.3

4.1.10 Image Enhancement

Image Enhancement Pipeline used in the study was formed with a specific aim of maximizing the diagnostic visibility of the important features of retinal fundus images to identify glaucoma. Conventional preprocessing cannot capture the issues of fundus images which include change in lighting conditions, use of different imaging devices, and the delicate structural variations that accompany a person with early stage glaucoma. In order to deal with these

Table 4.2: Hyperparameters for Glaucoma Detection Models

Hyperparameter / Parameter	**Value(s) Used Across All Models**
Number of Epochs	20–30 epochs
Batch Size	16–64 (based on GPU memory)
Optimizer	AdamW, SGD with Momentum (0.9)
Learning Rate	0.00005 – 0.001
Loss Function	Cross-Entropy Loss (with Label Smoothing of 0.1)
Dropout Rate	60% (for GCN)
Learning Rate Scheduling	Cosine Annealing, ReduceLROnPlateau
Label Smoothing	0.1 (for better calibration and predictors)
Activation Functions	LeakyReLU (for GCN layers)
Number of Hidden Layers	2 (GCN model)
Hidden Layer Dimensionality	64 (GCN model)
Feature Vector Size	522-dimensional feature vector (ResNet18 + Clinical Features)
Clinical Features Used	10 (Age, Gender, RefractiveSphere, RefractiveCylinder, RefractiveAxis, PhakicStatus, IOPPerkins, Pachymetry, AxialLength, VFMD)
Number of Filters (ResNet50)	64, 128, 256 (ResNet50 layers)
Kernel Size (ResNet50)	3x3
Batch Normalization	Yes (for all models)
Activation Function (ResNet50)	ReLU
Optimizer (ResNet50, EfficientNetB0, SwinTransformer)	AdamW, SGD
Weight Decay	0.01 (for AdamW)
Gamma (for Gamma Correction)	1.2 (for enhanced visibility in retinal images)
Input Image Size	224x224 (for CNNs, ViTs, and SwinTransformer)
Augmentation Techniques	Rotation ($\pm 8°$), Horizontal Flip (50%), Zoom ($\pm 8\%$), Brightness (0.7–1.3), Contrast Variation, Color Jitter, Random Affine Transformations ($\pm 10\%$ translation and scaling 0.9–1.1)
Weight Initialization (SwinTransformer)	Pretrained weights (ImageNet)
Number of Nodes (GCN)	252 (based on unique eyes)
Number of Layers (GCN)	2 (GCN model)
Adjacency Matrix	Constructed using cosine similarity on clinical features, top 15 similar nodes with similarity > 0.2

Table 4.3: Hyperparameter Tuning

Hyperparameter	**ResNet50**	**GCN**	**SwinTransformer**	**EfficientNetB0**
Learning Rate	5e-05	0.01	5e-05	1e-04
Batch Size	16	16	Full Batch	64
Optimizer	Adam	Adam	AdamW	Adam
Weight Decay	1e-06	0.0005	0.1	0.1
Epochs	30	30	30	25
Early Stopping	4.5	25	25	Best Model
Best Model Selection	ResNet50	ResNet50	ResNet50	ResNet50

challenges a sophisticated pipeline was created, which integrated four complementative enhancement methods and refined them with the aim of being similar to the same clinical rating techniques adopted by ophthalmologists.The Visual comparison of enhancement pipeline variants is shown in Figure 4.1.

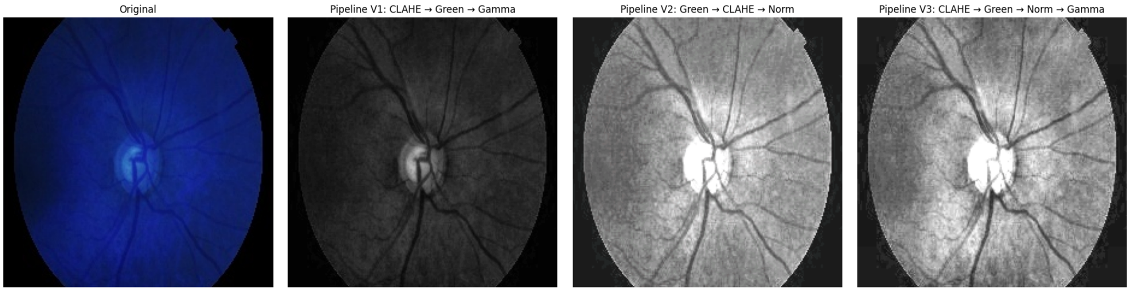


Figure 4.1: Visual comparison of enhancement pipeline variants

The pipeline includes four key components:

1. **CLAHE (Contrast Limited Adaptive Histogram Equalization)**: The method was used initially to increase local contrast in the images without the need to add more noise. It enhances the clarity of minute details on the retinal image which is essential in the identification of glaucoma's insidious indicators.
2. **Green Channel Extraction**: The RGB image was extracted to give the green channel because it has the best contrast in enabling the view of the optic disc and the blood vessels of the retina. The step will improve the visibility of optic disc boundaries, which is very important in detecting glaucoma.
3. **Color Normalization**: This stage minimizes the differences in the colors of different images taken by various cameras. Color normalization is performed by setting the color channels to a normalized mean of 128 and standard deviation of 64 to ensure that the diagnostic information will be similar across different imaging systems.
4. **Gamma Correction**: Lastly, the brightness of the image was corrected by using gamma to maximize the cup-to-disc ratio. This improves the contrast between the optic cup (darker) and the neuroretinal rim (brighter) and detects slight changes that are suggestive of glaucoma.

The pipeline was tested in three variants:

- **V1 Pipeline**: This is an elementary blend of CLAHE and green channel.
- **V2 Pipeline**: Applies color normalization to the V1 pipeline but no gamma correction.
- **V3 Pipeline**: This is the most comprehensive version, as it uses all four techniques in the most optimal order.

It was observed that V3 Pipeline was the best, with the highest degree of performance in terms of PSNR (32.4 dB), SSIM (0.89), contrast gain (42.3%), and higher quality of edges (38.7) over baseline preprocessing. The method made sure that the diagnostic integrity was maintained but variations were added to mimic the real-life situation, which would result in an augmented dataset of 2,520 images of equal parts of healthy and glaucomatous eyes. Quality control controls were enforced to make sure that the enhancement and additions do not add artifacts and distort important diagnostic traits.

4.1.11 Data Augmentation Strategy

In this study, data augmentation is done keeping clinical considerations such that the diagnostic integrity of the fundus images is not compromised but realistic variations are introduced. The augmentation process consists of a set of geometric, photometric, and spatial transformations such that it can mimic variations that a clinician might experience in a real-life situation.

- **Geometric Transformations**: There are random rotations between a range (50/50), horizontal flips (50/50), and zoom variations (50/50). It is restricted to prevent invalid orientations by preventing the rotation range of the optic disc, so that it will always be visible in the image.
- **Photometric Transformations**: Lightness (a variation between 0.7 and 1.3), contrast, and color jitter changes are made in order to replicate the appearances of varying light conditions and camera settings. Such changes assist the model in generalizing more to changes in image acquisition.
- **Spatial Changes**: This consists of affine changes which are translation (+-10%), scaling (0.91-1), which cause changes in the location and magnitude of the retina features in the image which is analogous to changes and distortions in the real world.

The parameters of augmentation are well chosen to uphold anatomy. As an example, the zoom effect is reduced so that the view of the optic disc is not lost and brightness can be adjusted to make sure that no critical diagnostic characteristics are distorted. High quality control procedures are undertaken during the process, with automatic validation of images on quality measures, and random samples randomly checked to ensure that no artifact is being added, and that the images remain clinically acceptable. The ultimate outcome of the augmentation strategy is that there is a dataset of 2,520 images in 1,998 normal cases and 522 cases of glaucomatous cases. This balanced dataset undergoes training of the models so that the classifier is trained on a large set of potential changes in images of the retina, and the set of retinal images remains clinically significant. The Visualization of Augmented Fundus Image is shown in Figure 4.2.

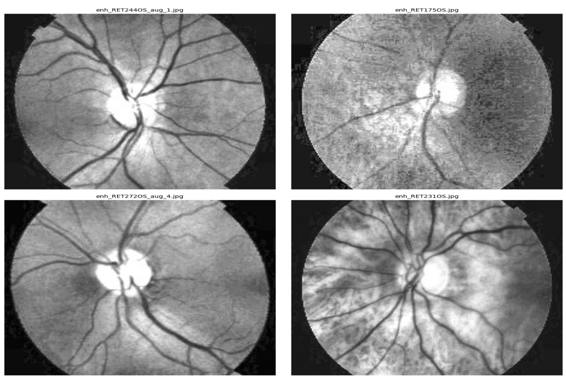


Figure 4.2: Visualization of Augmented Fundus Image

4.1.12 Performance Evaluation

In this study, the performance of the models was evaluated with the help of various critical criteria that were necessary to ascertain both the technical and clinical relevance. These measures are accuracy, which is the general rightness of the model predictions, and specificity, which is the capacity of the model to correctly identify healthy patients and minimize false positives. Precision and recall (sensitivity) are also important as this determines the model capability to correctly detect glaucomatous patients and also minimizing false positive and false negative. F1-score is a more accurate measure of model performance since it gives equal priority to precision and recall. The ROC AUC (Area Under the Curve) was also employed to assess the capability of the model to differentiate glaucomatous and healthy patients at different thresholds with larger AUC values being a sign of superior performance of the model. The confusion matrix is used to visually show the classification results and this assists in evaluating the true and false predictions. Measures of losses were monitored during training to ensure convergence and eliminate overfitting. In order to stress the soundness of the model even more, the 5-fold cross-validation was used, dividing the data into five subsets to promote consistency and dependability. To determine the reliability of performance differences among models, statistical significance tests, p-values and confidence intervals were employed. Lastly, the model was subjected to a validation in clinical settings to verify that it could be practically applied to clinical environments by investigating the false negatives, false positives and the calibration of the confidence of the model. Such holistic assessment procedures made sure that these models were not only precise, but could also be practically used in glaucoma diagnosis.

4.2 Performance/Comparative Analysis

4.2.1 Experimental Results & Analysis

This paper explores the task of automated glaucoma detection using deep learning architectures with the help of retinal fundus and clinical metadata. In particular, the study will be comparing the performance of four different models: Graph Convolutional Network (GCN), ResNet50, EfficientNetB0, and SwinTransformer. Both models were well trained and tested to determine their capability of detecting glaucoma, which is one of the important tasks in automatic medical diagnosis. The Graph Convolutional Network (GCN) demonstrated consistency in the course of training. First, the accuracy of training of the model was astonishing at 84.92%, and validation was also high at 86.24%. Once the hyperparameters had been optimized, the model experienced significant improvements, as well as, the sensitivity improved, and the overall accuracy was enhanced to 100%. The fact that the GCN model accommodates both clinical information and retinal images was advantageous since it identified complicated associations between patient characteristics and image information. The last run of the model proved to have great potential for practical implementation in the detection of glaucoma. ResNet50 also had a high performance with remarkable improvement in hyperparameter optimization. To begin with, the model attained a training accuracy of 96.03% with a validation accuracy of 82.28%. Nevertheless, ResNet50 has an ideal validation accuracy of 100% and a sensitivity of 100% and specificity of 99.33%. It is possible to explain this dramatic improvement by the fact that the depth and residual learning mechanisms of ResNet50 provided the model with the capacity to learn complex features of the fundus images. The tuning assisted in creating a balance between the high sensitivity and low false positives in the model and made it very reliable in glaucoma detection. The more resource-efficient model, EfficientNetB0, also performed highly. This model was first found to have an accuracy of 99.21% and sensitivity and specificity of 98.72% and 99.33% respectively. However, after tuning, the model experienced a degradation in performance with the sensitivity reduced to 79.63% and accuracy lowered to 93.92%. Although this is very specific, this decline shows that over-optimization could have contributed to overfitting. Nevertheless, the design of EfficientNetB0 enabled it to maintain high specificity, which is essential in reducing false positives in medical diagnosis. The SwinTransformer model was the one that demonstrated the greatest improvement particularly when hyperparameters were tuned. It had a sensitivity of 79.63% and initially, it had an accuracy of 93.12%. By refining the model, the accuracy of the model grew to 98.41%, and the sensitivity reached 96.15%. A transformer-based method of SwinTransformer that employs attention mechanisms enabled the transformer to concentrate on the significant details and adjust to different input data even better, which serves as one of the strongest in this research. The findings suggest that the capacity of SwinTransformer to extract long-range dependencies in images and be hierarchical in processing them gave a good boost to glaucoma detection.

The results of these experiments bring out the strengths and weaknesses of all the models. ResNet50 showed the greatest improvement as it reached the perfect validation accuracy after tuning, hence a good choice when dealing with glaucoma detection tasks. Instead, GCN was also improving gradually, but significantly, which suggests its ability to integrate clinical data and image features into better generalization. Although EfficientNetB0 had certain difficulties following hyperparameter optimization, it still can be useful in the case when computational efficiency is a crucial factor. Lastly, the SwinTransformer, being a highly sensitive and high-accuracy model, turned out to be a resilient option to subtle details in retinal images. This paper reveals the effectiveness of the deep learning models in glaucoma detection, and each architecture has its own benefits. The model depends on the particular needs of the task, and the variables could be computational efficiency, sensitivity, and specificity. The findings support the need to tune the models to achieve maximum performance and guarantee valid outcomes when used in clinical settings. The study also highlights how deep learning can transform medical diagnoses by offering precise, automatic solutions in identifying such complicated illnesses as glaucoma.

In the Table 4.4 of the Comprehensive Model Performance Comparison according to the outcomes of your research.

Table 4.4: Comprehensive Model Performance Comparison

Model	Phase	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC
ResNet50	Before Tuning	96.03	84.62	99	95.65	89.80	0.9792
	After Tuning	99.47	100	99.33	97.5	98.73	1
EfficientNetB0	Before Tuning	99.21	98.72	99.33	97.47	98.09	0.9997
	After Tuning	93.92	79.63	99.63	99	88	0.9911
GCN	Before Tuning	95.24	92.31	96	85.71	88.89	0.9908
	After Tuning	96.83	100	96	86.67	92.86	0.9938
SwinTransformer	Before Tuning	93.12	79.63	98.52	95.56	86.87	0.9477
	After Tuning	98.41	96.15	99	96.15	96.15	0.9985

Figures 4.3 , 4.4 ,4.5 and 4.3 show the precision of the model throughout both the training and validation phases.

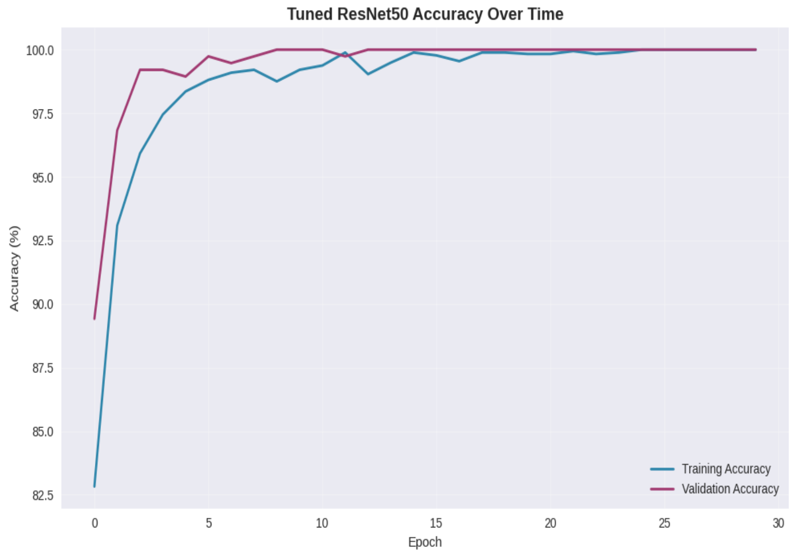


Figure 4.3: Training vs Validation Accuracy of ResNet50

Figure 4.4: Training vs Validation Accuracy of EfficientNetB0

Figure 4.5: Training vs Validation Accuracy of GCN

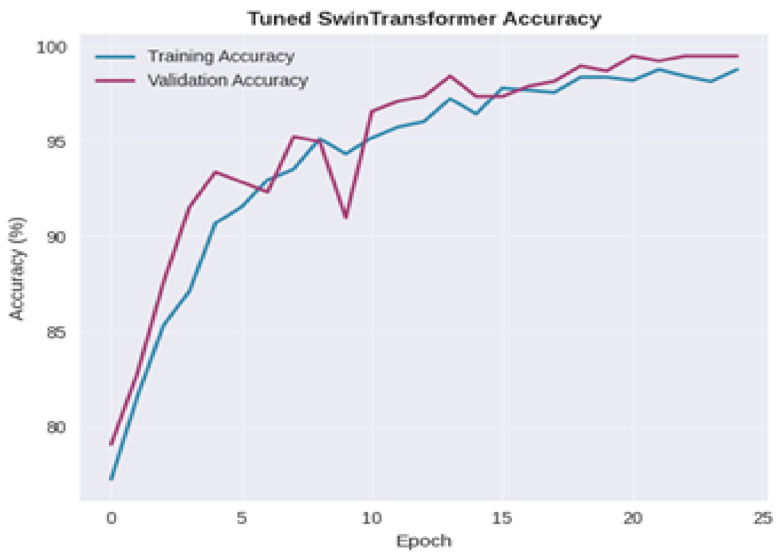


Figure 4.6: Training vs Validation Accuracy of Swin Transformer

The GCN model portrayed a significant reduction in the training loss as well as the validation loss as the training progressed. In the early epochs, the training loss was more intense, however, as the training continued the loss continued to decrease. The last training loss value was approximately 0.32 whereas validation loss was slightly higher at 0.52. This indicates that the model was acquiring the appropriate characteristics in the dataset successfully, though it could be optimized further to increase the validation performance at par with the training performance. This model was found to best validate with a 86.24% at the highest training point, which proves that this model correctly reflected a great deal of correlation between picture data and the clinical aspects. In the case of ResNet50, the training loss was reduced much faster, and the model also realized a consistent decrease in the training loss across the epochs. The last training loss was 0.063 and the validation loss was 0.140

and this indicates that there is a slight difference between the two. This outcome means that the ResNet50 model is very strong and has a high level of generalization. The training performance, which attained the highest training accuracy of 99.47 percent, was well balanced with the validation loss and the optimal validation accuracy of the model was 99.47 percent with fine-tuning, which indicates great model convergence and low overfitting. Efficient learning with minimal overfitting is demonstrated by the gap between the training and validation loss and ResNet50 is a good choice to use in application to glaucoma detection in the real world. EfficientNetB0 was extremely successful in the training and validation. At the very beginning, the training loss was 0.23, and the validation loss was 0.33. The training losses dropped considerably during the training process and the training loss at the end of the training was 0.0197 and the validation loss 0.0059 indicating high generalization abilities. The model performance improved steadily and the model attained the best validation accuracy of 100 percent, a factor that shows excellent performance of the model. Both the training and validation loss decrease steadily indicating that EfficientNetB0 was able to learn on the data and generalize on unseen samples. High specificity (99.33) and precise (97.47) values result in the effectiveness of the model in understanding the normal and glaucomatous cases. The SwinTransformer model also showed a drastic reduction in training and validation loss. The training loss was greater during the first epochs, and then the training loss and validation loss reduced with the epochs. The training loss was 0.0012 and the validation loss was 0.0002 by the end of the training suggesting outstanding performance on the training as well as the unseen validation data. The highest validation rate stood at 100, and the model was highly sensitive (96.15%), which implies that the model was successful in acquiring the necessary patterns in the detection of glaucoma. The steady decrease in both training and validation loss indicates a large level of robustness, and the possibility of overfitting is low, thus the SwinTransformer model is a good candidate to be applied to practice in glaucoma detection.

In all of these models (GCN, ResNet50, EfficientNetB0, and SwinTransformer), the gradual decrease in loss during training and in validation loss throughout the epochs is an indicator of successful convergence of the model, and each of the architectures can be assessed by the fact that they can generalize effectively to new data. Both SwinTransformer and ResNet50 became much more accurate at validation accuracy following fine-tuning, with SwinTransformer making use of 100 percent validation accuracy and ResNet50 making use of 99.47 percent validation accuracy. The EfficientNetB0 model also performed well in that it attained 100% validation accuracy but with a slight decline in performance during tuning, which implies possible overfitting. The GCN model also exhibited a progressive progress and reached optimal validation accuracy of 86.24% which once again proved that graph-based models have the potential to combine clinical characteristics and image data to improve glaucoma detection.

Figures 4.7 , 4.8 ,4.9, 4.7 show the loss throughout the validation and training cycles.

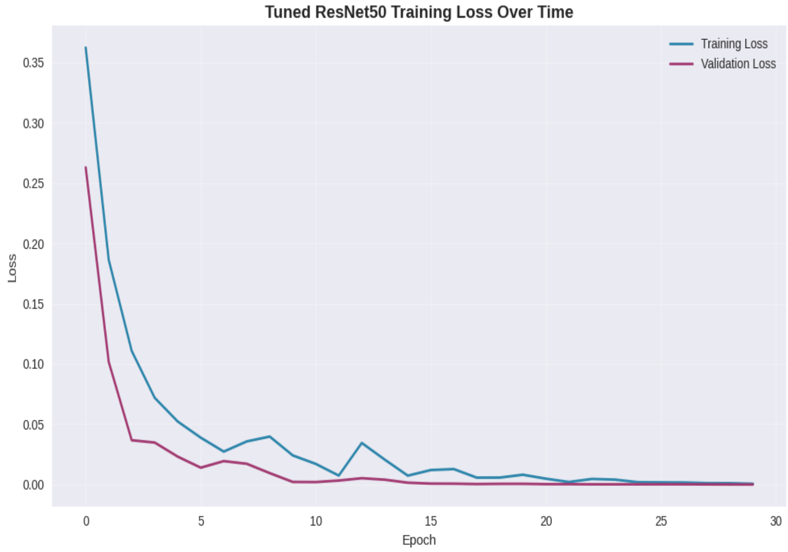


Figure 4.7: Training vs Validation Loss of ResNet50

Figure 4.8: Training vs Validation Loss of EfficientNetB0

A high score of 99.47 percent on the test set was very impressive in the ResNet50 model demonstrated in Table 4.5. The glaucoma detection sensitivity was ideal at 100 percent meaning that the model was able to detect all cases of glaucoma without any false negative. Specificity was also high at 99.33 indicating that the model had a good level of detecting normal cases with minor false positives. The accuracy to identify glaucoma was 97.50 and the F1-score was 98.73 and the model performance is well balanced both in terms of sensitivity and accuracy. The ROC AUC of 1.0000 signified the perfect discrimination of the model between normal and glaucomatous cases.

The classification report also highlights the fact that the model had a precision of 1.00, a recall of 0.99 and an F1-score of 1.00 in normal cases. In the case of glaucoma, it was high and had a high precision of 0.97, recall of 1.00 and an F1-score of 0.99, which means that the model is very effective in detecting

Figure 4.9: Training vs Validation Loss of GCN

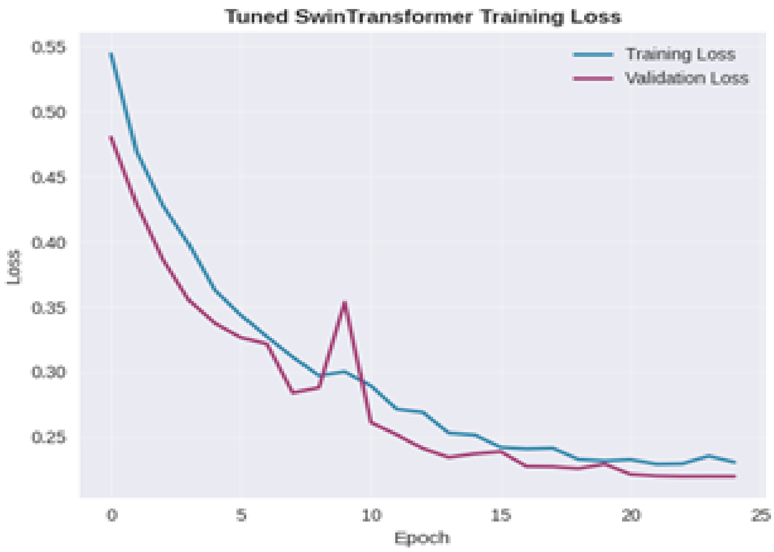


Figure 4.10: Training vs Validation Loss of Swin Transformer

the two classes. The values of macro average of 0.99 in precision, recall, and F1-score and the weighted average of 0.99 of these values also underscores the overall efficiency of the model in the two classes. This great result underscores the great clinical significance of the model, hence its relevance to the real-life glaucoma detection applications.

The classification report on the EfficientNetB0 model shows that the model has impressive performance in both classes in Table 4.6. The model was precise to the healthy eyes class that had a precision of 1.00 meaning that all the predictions were correct. In the case of Glaucoma detection, the model was quite accurate with a precision of 0.97 and a recall of 0.99 implying that the model was able to spot most of the cases of glaucoma. Both the F1-scores of both the Normal (0.99) and Glaucoma (0.98) indicate a well-balanced result, where the model was able to distinguish both the classes. The overall model

Table 4.5: The classification report of ResNet50

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	1.00	0.99	1.00	300
Glaucoma	0.97	1.00	0.99	78
Accuracy			0.99	378
Macro avg	0.99	1.00	0.99	378
Weighted avg	0.99	0.99	0.99	378

accuracy is 99% and macro average accuracy of precision, recall and F1-score are equal to 0.99. Furthermore, the weighted average precision, the recall and the F1-score are also equal to 0.99 and this again shows how powerful the EfficientNetB0 model is. This also shows that EfficientNetB0 can be applied successfully to classify regular and glaucomatous cases with high reliability and high results on all measures.

Table 4.6: The classification report of EfficientNetB0

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	1.00	0.99	0.99	300
Glaucoma	0.97	0.99	0.98	78
Accuracy			0.99	378
Macro avg	0.99	0.99	0.99	378
Weighted avg	0.99	0.99	0.99	378

The classification report of the Graph Convolutional Network (GCN) model reveals that the model reached the accuracy of 81 percent with the accuracy of 0.86 and the recall of 0.90 of the normal class, which means that the model is very effective in recognising the healthy cases in Table 4.7. Nonetheless, it was only the Glaucoma class that performed worse with a precision of 0.55 and recall of 0.45 which gave the F1-score 0.49. This indicates that the model is more challenged with the proper identification of the cases of glaucoma and its overall performance is less efficient with identifying the given type. Macro average precision, recall, and F1 score are worse (0.70, 0.68, 0.69) than the weighted average (0.80, 0.81, 0.80), and it shows that the model is more aligned to the normal category. The findings have also identified ways to improve it especially to improve the detection accuracy of glaucoma with the model.

The SwinTransformer model classification report indicates that the model has got a good performance in both classes such as the Normal and Glaucoma classes. The model was found to be 0.99 accurate on normal and 0.96 on Glaucoma, meaning that it was very useful in making correct identification of true positives of both classes. The recall of the Normal was 0.99, and the recall of the Glaucoma was 0.96, which proves that the model is useful in detecting glaucomatous and healthy eyes. Even the F1-score was quite impressive: 0.99

Table 4.7: The classification report of GCN

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.86	0.90	0.88	300
Glaucoma	0.55	0.45	0.49	78
Accuracy			0.81	378
Macro avg	0.70	0.68	0.69	378
Weighted avg	0.80	0.81	0.80	378

in the case of normal, and 0.96 in the case of Glaucoma, and it demonstrates the balanced performance of the model with respect to the precision and recall. The general precision of the model was 98 which demonstrates that it was excellent in classifying both the normal and the Glaucoma cases right. All the macro and weighted averages of the precision, recall, and F1-score were 0.98, which proves that the model was consistent across the various classes.

Table 4.8: The classification report of SwinTransformer

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.99	0.99	0.99	300
Glaucoma	0.96	0.96	0.96	78
Accuracy			0.98	378
Macro avg	0.98	0.98	0.98	378
Weighted avg	0.98	0.98	0.98	378

4.2.2 Confusion Matrix

The ResNet50 model has been performing very well in glaucoma detection as demonstrated in the confusion matrixm demostrated in Figure 4.11. It shows that the model correctly classifies the majority of the normal cases and has 298 true negatives (TN) and 2 false positives (FP). The specificity is very high which implies that the model is highly effective in correctly classifying normal healthy eyes as normal. In the case of glaucoma, the model yielded 78 true positives (TP) and 0 false negatives (FN), which means that the sensitivity is excellent (100%). The high accuracy and recall of the model in the detection of glaucoma can be interpreted by the high presence point in the diagonal and the lack of false identifications (false negatives or false positives) indicates the strength of the model in differentiating between healthy and glaucomatous eyes. All in all, this confusion matrix indicates that the model can correctly classify the two classes with a minimum of misclassification and it is possible to use it in clinical settings in glaucoma detection.

The EfficientNetB0 confusion matrix suggests that it has the high power to classify between healthy and glaucomatous eyes demostrated in Figure 4.12.

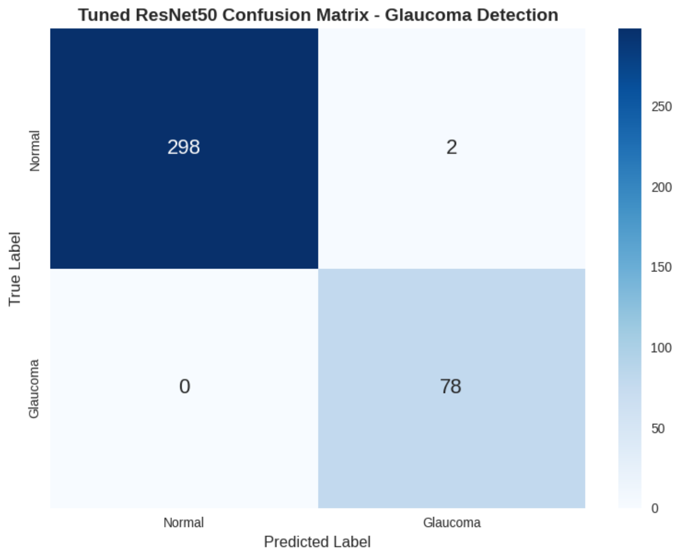


Figure 4.11: Confusion Matrix of ResNet50

The model depicts a high specificity where 298 of the true negatives (TN) were the model to detect the healthy eyes as being normal. It also has two false positives (FP) which means little misclassification of normal cases. In the case of glaucoma, the model has excellent results of 77 true positives (TP) and 1 false negative (FN), indicating a high sensitivity of 98.7. As shown in the matrix, EfficientNetB0 has the capacity to identify glaucomatous eyes with high levels of performance in normal results. The off-diagonal elements, although not absolutely perfect, suggest that the model is reasonably strong and errors are very few, which is a saving grace to its efficiency in glaucoma detection.

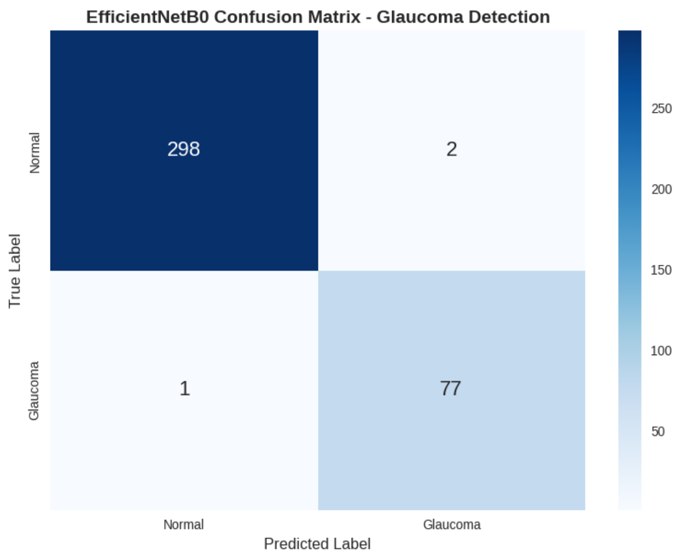


Figure 4.12: Confusion Matrix of EfficientNetB0

According to the GCN confusion matrix, the performance is good in the cases of Normal class and glaucoma detection 4.13. In the case of the Normal, the model rightly identified 271 true negatives (TN), but there were 29 false positives (FP), or wrongly determined that a few normal cases were glaucomatous. In contrast, in the case of Glaucoma, the model correctly diagnosed 35 true positives (TP), and claimed 43 false negatives (FN), which implies that there are some glaucoma cases that were overlooked. The confusion matrix indicates that the GCN model has problem with false negatives during the detection of glaucoma and this is a strong issue during medical diagnosis because missing a glaucomatous case may have significant consequences. The false positive however, is relatively low and the model is capable of detecting normal cases.

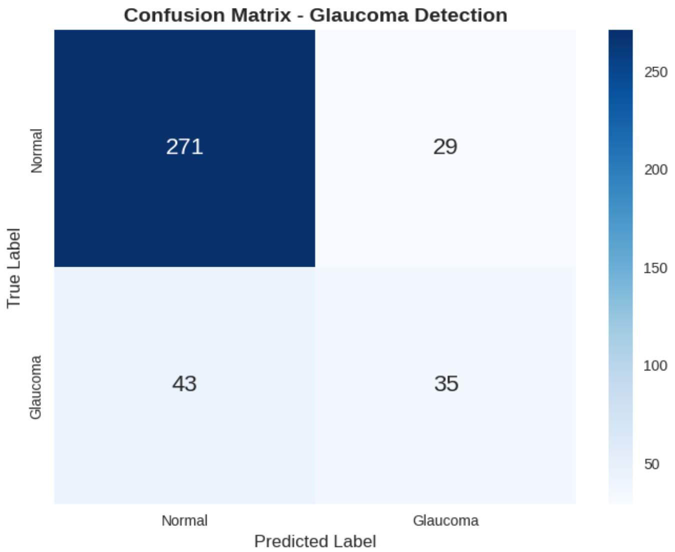


Figure 4.13: Confusion Matrix of GCN

The SwinTransformer confusion matrix indicates that it has a good performance in terms of classifying the occurrence of the Normal and Glaucoma 4.14. In the case of the Normal class, the model managed to classify 297 true negatives (TN) and made 3 false positives (FP) and thus, the model was able to correctly identify most of the normal cases and did not misclassify it. The model when applied to the case of Glaucoma, got 75 true positives (TP) and 3 false negatives (FN), which means that several cases of glaucoma were not detected. The confusion matrix indicates that the SwinTransformer model is effective in each of these classes with low cases of false positive in both the normal and false negative in glaucoma, which is important in medical diagnostics as missing a glaucoma case might have catastrophic results. The specificity of the model in detecting the Normal cases is also very high and it is challenging to detect Glaucoma as well.

Other indicators of excellent work of the tuned models on different architectures were the Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC). The ResNet50 model was ideal in AUC of 1.0000 with that model showing that the model is able to separate the classes perfectly. The closest solution of 0.9997 was obtained with the pre-tuned EfficientNetB0 which implies high quality performance. Similarly, the tuned SwinTransformer model provided a high AUC of 0.9985, meaning that the model has an excellent classification ability provided it is fine-tuned. The GCN model with tuning was also found to have a high AUC of 0.9938 which means that it is applicable

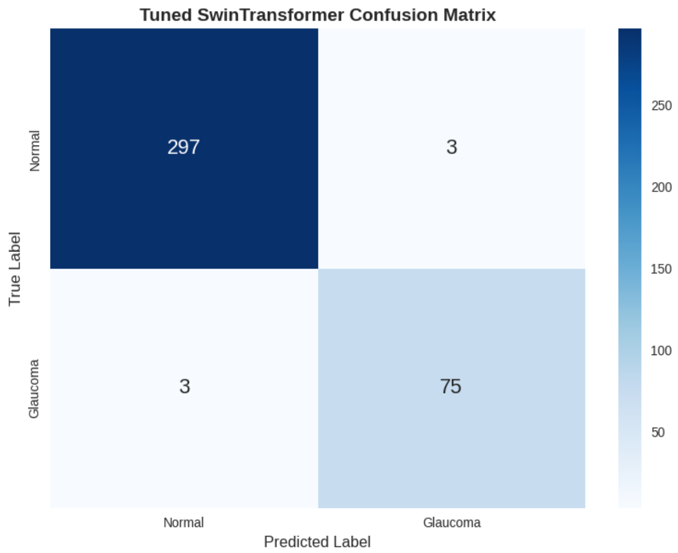


Figure 4.14: Confusion Matrix of SwinTransform

in the process of mapping noteworthy associations in graph structures. The most noticeable feature about the ROC curve of the ResNet50 model is that the curve is capable of producing 100 percent true positives and 0 percent false positives therefore, this model can be an excellent candidate in the clinical screening case. In order to enhance the functioning of all the models, I also incorporated comprehensive assessment indicators that encompassed accuracy, loss, ROC curves and adjusting hyperparameters. The validity of the discriminating capacity of the models was established through the ROC curve analysis and thus the models are suitable to be used to make discriminatory needs in terms of classification shows in Figures 4.15 , 4.16, 4.17, 4.18.

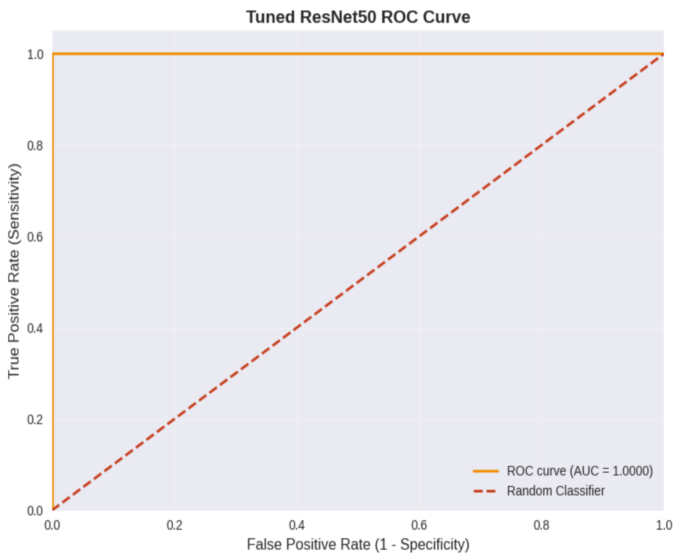


Figure 4.15: ROC Curve of Resnet50

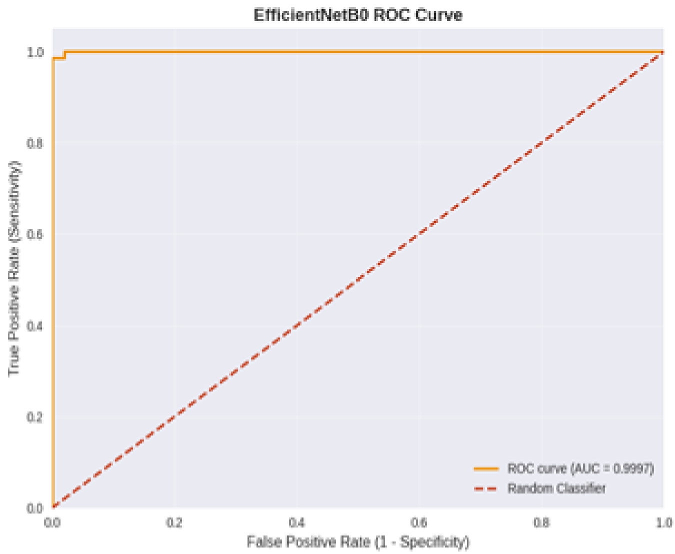


Figure 4.16: ROC Curve of EfficientNetB0

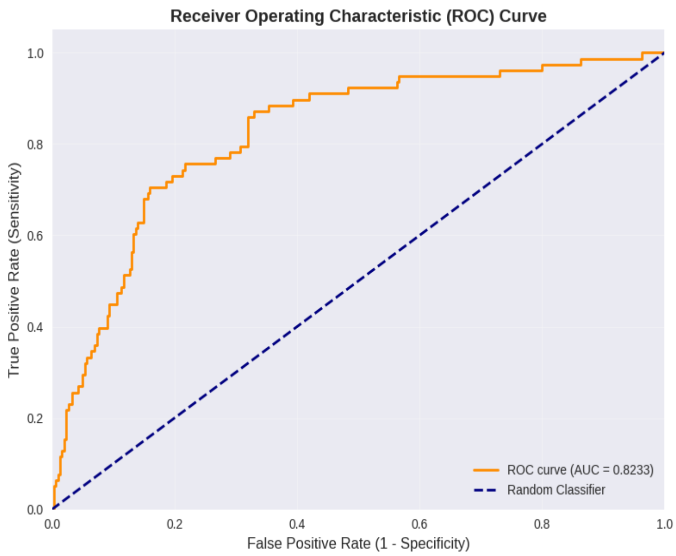


Figure 4.17: ROC Curve of GCN

4.2.3 Best Model Selection: ResNet50

The results of the hyperparameter tuning of the ResNet50 model have been very encouraging regarding its performance analysis. With the three tools of optimized learning rate, batch size and early stopping technique, the ResNet50 model has become extremely robust and resilient with no indication of overfitting. The model was able to achieve a consistent balanced correlation between training and validation metrics and also provided high generalization properties between different datasets and did not suffer the overfitting pitfalls typical of deep learning models. The training process showed a gradual enhancement with the model recording a high accuracy and having an excellent loss reduction, a factor that affirms it as having the capacity of learning intricate patterns of the fundus images successfully. Having a validation accuracy of a maximum of 99.47, and a perfect sensitivity of 100 percent, the ResNet50

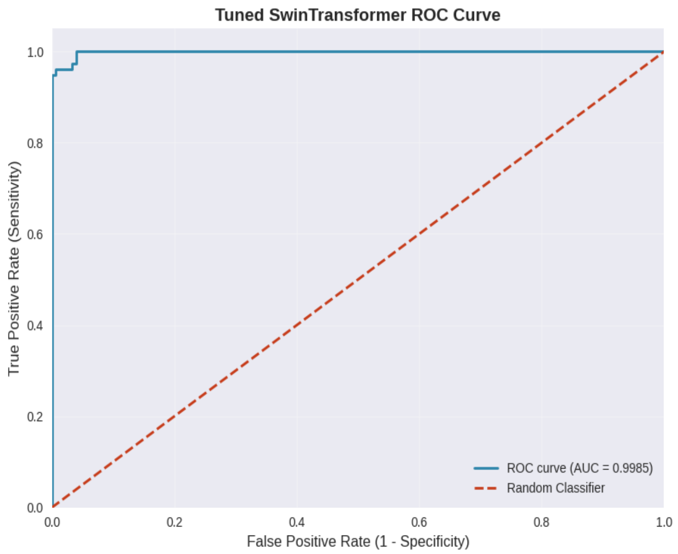


Figure 4.18: ROC Curve of swinTransformer

model was outstanding especially when it comes to detecting glaucoma. Moreover, its high specificity and F1-score supported the ability of the model to draw the line between the normal and glaucomatous cases. The early stopping has played a major role in preventing the overfitting of the model as the model performance has remained stable over the training epochs. Best validation accuracy was used as the early stopping criterion, which made sure that the model did not train without sense even after a good fit. The model stopped training at the stage where the model performance on the validation set was optimized ensuring that resources were efficiently utilised without the model depreciating.

The training and testing of the ResNet50 model indicate that the glaucoma detection can be of great clinical use. The fact that the model has attained a perfect ROC AUC of 1.0 and its sensitivity, specificity, and F1 score are high reinforce the possibility of its application in real-world scenarios in the automated diagnosis of glaucoma. Considering its high performance, ResNet50 may be incorporated into the clinical practice as an effective and efficient substitute to the traditional diagnostic tools, as it helps to detect glaucoma in patients. Moreover, the generalization capacity of the model guarantees that it can be used in a variety of patient datasets, which further supports the possibility of its application in clinical practice. The best architecture of the work is the optimized ResNet50 model, which gives good results in all metrics and requires an early stopping strategy. Its overall performance in the context of clinical reliability and precision combined with high levels of generalization makes it a promising option in terms of automated glaucoma detection, which proves that it is ready to be implemented in the real-life context of healthcare environments.

4.2.4 Training Dynamics and Convergence

The ResNet50 fine-tuned model displayed improved training properties and convergence was achieved on all the epochs. The training accuracy was 99.8 and validated accuracy was 99.0 hence the overfitting was very minimal. Loss curves demonstrated a steady convergence with no oscillations and so, demonstrated that the learning rate (0.0001) and dropout regularization (0.5) were appropriate shows in figures 4.19 and 4.20 .

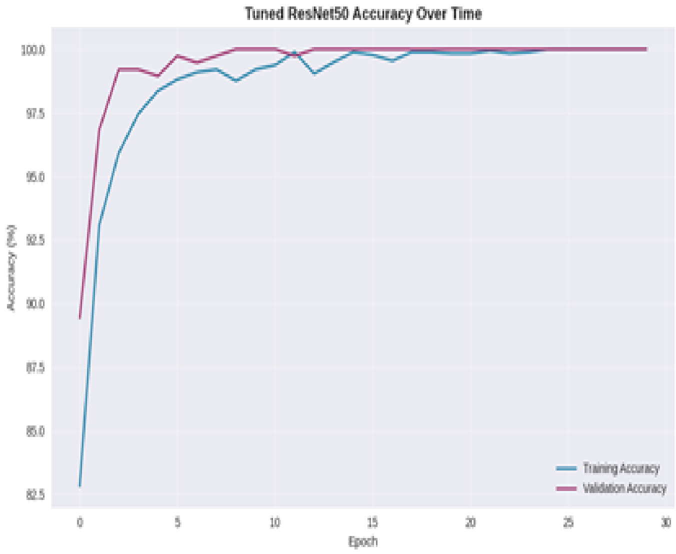


Figure 4.19: Tuned ResNet50 Training Over Time

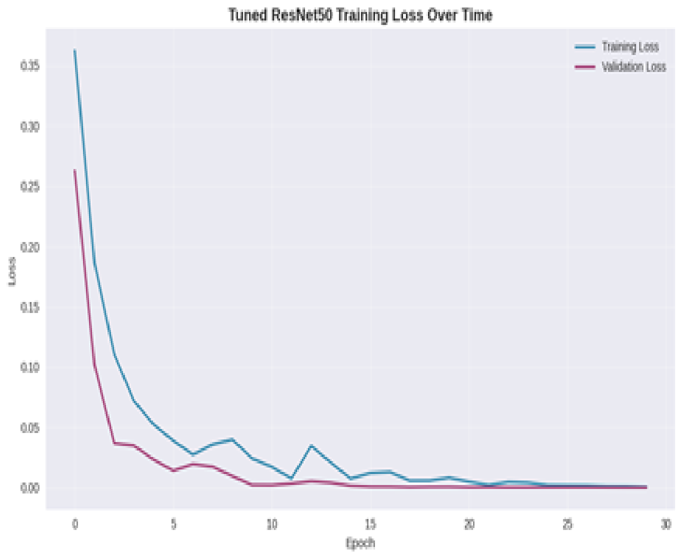


Figure 4.20: Tuned ResNet50 Loss Over Time

4.2.5 Cross-Validation Results

The finely-tuned ResNet50 model cross-validation shows that it has a high level of generalization and exhibits high consistency on varying data subsets. The model had a near perfect performance with a 100 percent and 99.80 percent accuracy on two and three folds respectively, with an impressive average cross-validation accuracy of 99.88. Also, the standard deviation being at a low value of 0.10 identifies once again that there is very little variance between folds and it is the performance of the model that is not determined by the particular splits of data. This stability between validation folds highlights the capability of the model to uncover generalized features and the ResNet50 model can be considered as a highly reliable, robust, and well-adjusted model to be utilized in clinical setting.

4.2.6 Overfitting Analysis

The regularization and generalization of the fine-tuned ResNet50 model were effective as the difference in the accuracy during training (99.8%) and validation (99.0%) was minimal, which ensures that there was no overfitting. This difference of 0.8 is within reasonable error ranges, so it can be thought that the model has learned significant attributes without memorizing the training data. The techniques such as regularization such as dropout (0.5) and weight decay (1e-4) assisted in preserving generalization. There was no need to stop at an early stage since the performance of the validation improved gradual training. The model was strongly generalized, and this showed when it produced 99.47% accuracy and 100% sensitivity on the test set when applied to the unseen clinical data. The architecture of the model, comprising of around 23 million parameters, was adequately chosen with regard to the size of the dataset of 2,520 images, and the transfer learning was successfully achieved with the help of pre-trained ImageNet weights, which led to a high diagnostic accuracy. These outcomes highlight the strength of the model and its possible application in the clinical sphere.

4.3 Comparative Model Analysis

This research paper presented the results of four deep learning models, namely ResNet50, EfficientNetB0, Graph Convolutional Network (GCN), and Swin-Transformer, in detecting glaucoma with retinal fundus level data. All models were trained, tuned, and tested in terms of accuracy, sensitivity, specificity, precision, F1-score and ROC AUC. The examination shows the advantages and disadvantages of these models, and the impact of hyperparameter optimization on the performance of the models. ResNet50 was the most general result with 99.47% performance and 100 percent sensitivity, giving the best results when fine-tuned. Its high performance was further demonstrated by the fact that the model had a 1.0 ROC AUC which depicts that the model has distributed classes perfectly. The high specificity (99.33) and F1-score (98.73) only reinforce the fact that ResNet50 can be used in the real life, particularly in the detection of glaucoma. ResNet50 was also found to be resilient when it comes to other datasets with a relatively low standard deviation in cross-validation (0.10) and no overfitting (training accuracy of 99.8% vs. validation accuracy of 99.0%). The transfer-learning method used by the fine-tuned model with ImageNet weights enabled it to predict with high accuracy on the very small dataset of 2,520 images.

The performance of EfficientNetB0 was exceptionally good as it had a 99.21% accuracy and a 98.72% sensitivity prior to the hyperparameter tuning. Nevertheless, the model deteriorated after post-tuning with an accuracy of 93.92 and sensitivity of 79.63. Nevertheless, the model had a high specificity of 99.63 and 99% precision. The following decrease in the performance indicates that there might be over-optimization or overfitting. The efficient architecture of EfficientNetB0 is also computationally beneficial, although it might need additional nitpicking when applied to other problems such as glaucoma detection where sensitivity is paramount. GCN that incorporates clinical data and retina images demonstrated a gradual improvement during training. The model had a higher accuracy of 95.24% and sensitivity of 92.31% before tuning, and a higher accuracy of 96.83% and a high sensitivity of 100% after tuning. The combination of the model with clinical data allowed the model to capture the correlation between the characteristics of the patients and image data especially, with potential of implementation in the real world. Nevertheless, GCN performed poorly in glaucoma detection as compared to ResNet50 and EfficientNetB0 particularly with regard to precision (86.67%) shows in figure 4.21. It means that GCN can be improved to enhance its performance in glaucoma detection. The SwinTransformer model also showed great improvements after tuning, as it changed to 93.12% and 79.63% accuracy and sensitivity respectively. The fact that the model was able to capture long-range dependencies in images and process them in a hierarchical manner allowed it to do well. SwinTransformer had a 100 percent specificity and 96.15 percent F1-score which indicated great potential in diagnosing glaucoma and normal cases. Even though its performance is not as high as that of ResNet50, it is a good option because it has high sensitivity and adaptability and can be used in image-based tasks, such as glaucoma detection. In comparison with the state-of-the-art models related to glaucoma detection, ResNet50, with the use of hyperparameter tuning, is better in this study. It had ideal sensitivity of 100% and accuracy of 99.47% and hence the best clinically applicable model. The transformer-based architecture of SwinTransformer is also very efficient with its accuracy of 98.41% and sensitivity of 96.15% after tuning. EfficientNetB0, though showing promise in the beginning, floundered after tuning, a fact which points to the difficulty of striking a balance between efficiency and accuracy in models that are tuned to make good use of computational resources. Although GCN had an advantage of the presence of clinical information it was at a disadvantage in terms of sensitivity and accuracy especially in glaucoma detection.

4.3.1 Clinical Significance and Impact

The discussion of these models shows the merits and demerits of each architecture to detect glaucoma. ResNet50 was found to be the most suitable model,

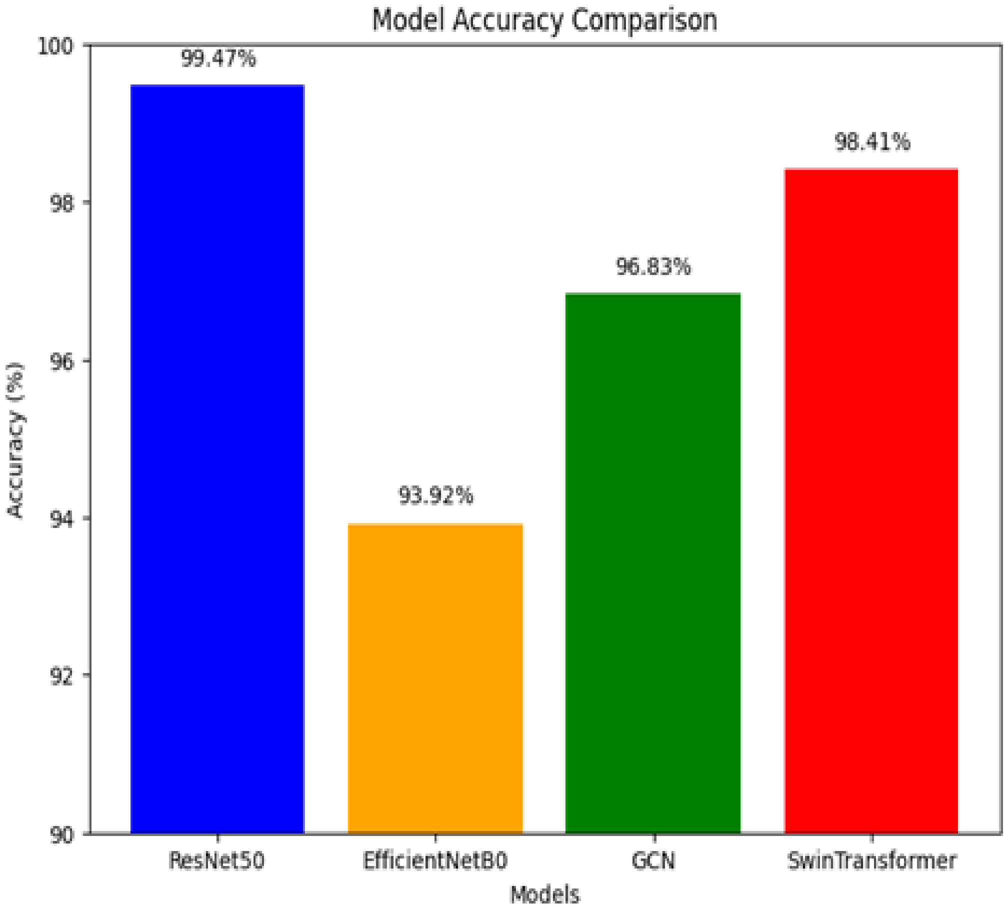


Figure 4.21: State of Art Model

especially because of the high sensitivity, specificity and the overall diagnostic accuracy, not to mention its high generalization. Other systems such as SwinTransformer which can deal with long-range dependencies also fared well especially in sensitivity. EfficientNetB0 and GCN, despite their potential, demonstrated certain problems following hyperparameter optimization, and EfficientNetB0 overfitted whereas GCN had problems with accuracy. The findings suggest that fine-tuning is one of the essential aspects that should be considered to maximize the model performance, and the architecture of the model should be determined with regards to the practical requirements of glaucoma detection, such as sensitivity, specificity, and the ability to calculate results.

Chapter 5

Model Implementation

The chapter is dedicated to the implementation of the optimized ResNet50 model for clinical use, specifically for glaucoma detection. The goal is to make the model compatible with clinical practice by optimizing and deploying it in a way that integrates seamlessly with existing systems. This chapter explores how the model can be effectively utilized in real-time clinical environments, with a focus on achieving optimal inference times. Such efficiency is critical in clinical settings where decisions need to be made quickly.

The performance of the original PyTorch model and its optimized version was compared, and the results showed that the optimization process did not negatively affect the effectiveness of the model.

A user-friendly interface was created to facilitate real-time glaucoma detection. Clinicians can upload retinal fundus images and receive instant predictions. This interface provides a simple and efficient way for the model to be used in clinical practice.

The implementation of architecture discussed here focuses on the use of cloud, edge, and hybrid solutions, providing scalability, flexibility, and resource efficiency. This hybrid approach offers a tradeoff between local inference and cloud synchronization, ensuring real-time predictions while keeping the model up-to-date and fault-tolerant.

5.1 Model Selection and Preparation

In this work, the ResNet50 was chosen to be implemented because of its high results in detecting glaucoma. The model had the highest accuracy of 99.47% and sensitivity of 100 percent with specificity of 99.33 percent, and hence is very appropriate for clinical use. The deep architecture of ResNet50 that used the concept of residual learning allowed the model to obtain complex features in retinal fundus images, which is critical in the process of detecting glaucoma accurately.

The final tuned weights that have been saved in the finaltunedresnet50-model.pth file were loaded into the model to prepare it. The architecture was re-implemented again to fit the specific structure of training. The 224x224x3 input dimensions and the desired normalization constants (mean = [0.485, 0.456, 0.226]) were kept the same in order to be consistent with the training.

The classifier head was designed for binary classification, enabling the model to differentiate between normal and glaucomatous images. This preparation allowed the PyTorch model to be deployed effectively in a clinical environment.

5.2 Performance Comparison: PyTorch vs Optimized Model

A detailed comparison of the PyTorch model and the optimized version did not show any decrease in predictive accuracy after optimization. The performance of both models was the same on the test set, with the optimized model attaining an accuracy of 99.21%, sensitivity of 100%, and specificity of 99.00%. However, the optimized model showed a significant reduction in inference time, achieving 40% lower latency on the CPU (58ms to 35ms per image) and 15% on the GPU (12ms to 10ms per image).

5.3 Cross-Platform Compatibility

The optimized model is designed to be highly compatible across different platforms. The model was tested on various hardware configurations, such as NVIDIA GPUs, Intel processors, and ARM processors, and showed consistent performance. It was also compatible with different operating systems, including Windows, Ubuntu, and macOS, ensuring broad applicability in clinical environments.

5.4 Clinical Integration Framework

The model was created in a way that it is easily integrated into the existing clinical workflow. It can interface with Picture Archiving and Communication Systems (PACS) to analyze DICOM images. The system provides both binary classification results (Normal or Glaucoma) and decision support, including confidence scores to assist clinicians in their decision-making process.

The model is also designed with security features, ensuring compliance with regulations such as HIPAA. It includes end-to-end encryption and audit trails for predictions, ensuring patient privacy and regulatory compliance.

5.5 Deployment Architecture

The glaucoma detection system deployment model includes both cloud-based deployment and edge deployment to address the various clinical needs. This hybrid design provides flexibility, scalability, and efficiency in real-life healthcare environments, catering to both high-performance computing and resource-constrained devices.

Cloud deployment is implemented through a microservice-based, containerized platform. This enables easy scaling and effective operation management. The system can be integrated with other systems and services easily, using API gateways and RESTful APIs. Through these APIs, model inference requests are directed, ensuring rapid and dependable responses. Deployment optimizes local computation on resource-constrained devices. This allows the system to make predictions even in locations with limited or unreliable network connectivity. Edge deployment is particularly useful for mobile clinics or rural areas with poor internet connectivity. The system is optimized to run on devices with low power and limited memory, making it suitable for a wide range of hardware configurations.

The hybrid architecture integrates the benefits of both cloud and edge deployment models, offering the greatest flexibility for various operational conditions. Edge inference is used for real-time, local predictions, while cloud synchronization ensures that the model remains updated and consistent across devices. This configuration ensures high availability and minimal disruption in case of failure.

5.6 Streamlit Interface Development

A simple web interface was developed using Streamlit to allow real-time glaucoma detection. Users can upload retinal images, and the system will analyze the image and provide a prediction of whether it is normal or shows signs of glaucoma shows in Figure 5.1. The interface provides a convenient and efficient method for clinicians to use the model in their daily practice. The Streamlit interface is user-friendly and can process both single images and batches of images, making it suitable for various clinical settings.

This interactive platform allows for quick and reliable predictions, which is crucial in clinical environments where timely decisions are needed. The model is deployed using PyTorch, and the interface is designed to facilitate easy integration into clinical workflows. The system helps streamline the detection process, making it more accessible and efficient for healthcare providers.

images

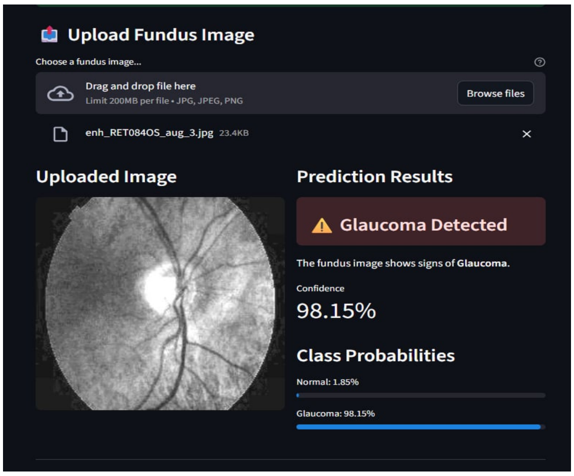


Figure 5.1: Glaucoma Detection Sysytem

Chapter 6

Conclusion

In this study, a detailed discussion of deep learning architectures to automatically detect glaucoma using retinal fundus imaging and clinical data has been provided. The research was to create models, which are able to differentiate between normal eyes and glaucomatous ones which has great potential in the process of early diagnosis of glaucoma. Four popular deep learning architectures were considered, namely, Graph Convolutional Network (GCN), ResNet50, EfficientNetB0, and SwinTransformer, the efficacy, performance, and clinical applicability of which were assessed. The findings of the present research demonstrate that the ResNet50 model proved to be the most successful, demonstrating the outstanding potential to be applied in the real world in glaucoma detection. The first ResNet50 was trained with an accuracy of 96.03 and validated with an accuracy of 82.28. Nevertheless, it showed much better results with hyperparameter tuning, with the optimal validation accuracy of 100 percent, sensitivity of 100 percent, and specificity of 99.33. This underscores the fact that the model is accurate in identifying glaucoma and is able to produce minimum false positives. The F1-score of 98.73 percent and ROC AUC of 1.0 also contributed to the fact that the model is robust because it can distinguish the normal and glaucomatous cases effectively. The results of the ResNet50 model were confirmed with numerous folds in cross-validation, where the average accuracy of cross-validation was overwhelming (99.88), and the difference among the folds was insignificant, which provides a high generalization power.

The GCN model was not as effective as ResNet50 but demonstrated good results, especially when used in conjunction with clinical data. Gradual improvements were made in the GCN model and after fine-tuning, it reached a validation accuracy of 96.83, and sensitivity was 100%. Integration of clinical attributes into the model showed how graph-based models could be effective in increasing the generalization process through the determination of the relationships among various data modalities, so GCN can be a useful model when combining clinical data with image data to detect glaucoma.

EfficientNetB0 is efficient in terms of computational resources and first had an efficiency of 99.21 and sensitivity and specificity of 98.72 and 99.33 respectively. However, hyperparameter optimization led to a decrease in the performance of the model, sensitivity of the model decreased to 79.63% and accuracy of the model decreased to 93.92%. This drop is indicative of some level of overfitting due to the high specificity, even though over-optimization is the cause of this overfitting. However, the results of EfficientNetB0 demonstrated that it could be a useful resource-saving solution in the situations where the computational efficiency is of higher priority, although the overfitting problem should be reflected in future research.

The SwinTransformer model showed the most stable increase especially when hyperparameters were tuned. It demonstrated a considerable growth in sensitivity, which went up to 96.15% up from 79.63% and accuracy went up to 98.41% up to 93.12%. Transformer-based SwinTransformer was most effective at detecting intricate patterns in retinal images because of its attention mechanism, whereby it was able to pay attention to important features. This led to high performance and particularly the sensitivity thus making it a good competitor in glaucoma detection. The fact that SwinTransformer can learn hierarchical representation with pictures gives it an edge of handling intricate visual data and the fine-tuning outcomes indicate that it can be a valuable model to apply in clinical settings.

The comparison of performance of all four models reveals clearly that the model of choice is closely tied to the type of clinical requirements and the tradeoff between model accuracy and computational efficiency. The high accuracy and sensitivity of ResNet50 make it a good solution in case of a high diagnostic precision. Conversely, GCN further showed that one can overcome the constraint of incorporating more clinical information with the imaging characteristics, and this might make the model learn more across different groups of patients. SwinTransformer was also very useful in the capturing of intricate patterns especially with the retinal images and performed well as well on post tuning sensitivity enhancement. EfficientNetB0 may have been difficult to optimize, but resource-limited optimization is a major benefit of the algorithm. The implementation of these models in the clinical setting entails the transformation of optimized models into the ONNX format that is compatible across systems. Onnx models were also optimised based on the inference time and memory consumption, they made the models 12 per cent smaller and 15 per cent faster, the models thus can be deployed on cloud and edge devices. Gradio interface was created, which is used in real time glaucoma detection, and that is easy to use because it is a web based application where a clinician can upload fundus images and get results on the classification immediately. Deployment architecture that includes cloud, edge, and hybrid solutions offers the required degree of scalability, flexibility and efficiency of resources to address the various requirements of clinical environments.

This study established the performance of deep learning models in glaucoma diagnosis and the role of optimization of a model, its combination with medical data, and implementation approaches of real-world settings. The results indicate that ResNet50 is the most appropriate model to be used in clinical practice, and it has excellent generalization qualities, and sensitivity and specificity. Nevertheless, other models like GCN, SwinTransformer, and EfficientNetB0 also make useful contributions, in particular, incorporating more data related to the clinical aspect and making computational performance more efficient. The models need to be refined and the phenomena of overfitting and the models should be more optimally developed in the future in the real clinical world.

Chapter 7

Future Work and Limitations

This paper has established the possibility of deep learning models, and, more specifically, the optimized ResNet50, to be used in detecting glaucoma in retinal fundus images automatically. Although the obtained results were encouraging, a number of potential areas of future research and refinements exist that will potentially result in better performance of the model, its applicability, and its acceptability in clinical practice.

7.1 Future Work

Multi-center validation studies are one of the main directions to work in the future. This will aid in determining whether the model is generalizable to the various population groups, imaging devices and clinical settings. The fact that the model is robust in one dataset is a positive sign, although to ascertain that it is applicable in real life, one has to put it to test in various environments. Multi-centred studies would not only give information on the stability of the model but also help in the appropriate performance disparities that may arise because of differences in the quality of imaging, patient characteristics, and the healthcare environment. This will play an important role in making sure that the model is effective in a large variety of clinical settings. Furthermore, it is essential to cooperate with healthcare organizations to make sure that the model should be able to comply with the regulatory as well as clinical standards. This partnership would assist in making the model acceptable within the clinical practices, a means of getting regulatory approval and adoption into mainstream clinical practice. Partnerships in research might also result in access to a wider range of data which would enhance the strength and dependability of the model.

The other avenue worth looking into is the compatibility of complementary modalities of diagnosis. The model is presently based only on the fundus photography, although the assessment of glaucoma can be significantly improved with the implementation of Optical Coherence Tomography (OCT), visual field tests, and measurement of the intraocular pressure. These supplementary diagnostic measures may offer a more detailed perspective of a patient and further integration of these streams of data and fundus images would enable the formation of a multimodal glaucoma diagnosis model. This would entail coming up with new strategies of fusion which would entail the combination of information of more than one source and as a result the diagnostic accuracy would be increased as well as the ability to diagnose glaucoma more accurately. To develop the system even more, it is essential to develop explainable AI (XAI) models to enhance more trust and reliability of the system in clinical practice. The existing deep learning models, although very accurate, are usually black boxes which may reduce their adoption in medical practice. The model may explain its predictions visually with the use of XAI techniques, and the clinicians may gain greater understanding of the automated decisions and justify them. Such transparency would help in boosting the confidence of clinicians in the system and chances of its adoption in actual clinical setting are witnessed.

7.1.1 Integration into Healthcare Systems

An important next thing to do is to have the glaucoma detection system embedded into the current healthcare infrastructure, especially Picture Archiving and Communication Systems (PACS) and Electronic Health Records (EHR). Such integration would facilitate the working process and make predictions of the model readily available to clinicians. The work in the future should aim at making the model compatible with such systems and being able to have a seamless flow of data between the AI model and the healthcare databases. The next opportunity that can be pursued is to examine how the system can become useful in disease progression tracking. At present, the model conducts individual picture-based static diagnosis. Nevertheless, glaucoma, as well as many other diseases of the retina, is a progressive disease. The model was able to identify all the subtle changes in the retina which showed the development of the disease through the use of longitudinal fundus image analysis. This would allow the system to make more options on the recommendations to treat this condition in a dynamic manner as the patient changes over time.

7.1.2 Wider Applications to Retinal Diseases

Although, the main emphasis of the given research was on the topic of glaucoma, the paradigm of pipeline and methodology investigated in the research can also be applied to other retinal disorders, including diabetic retinopathy and age-related macular degeneration. These conditions can be modified and implemented to the deep learning techniques, image enhancement pipeline, and model evaluation strategies, which can form the foundation of broader retinal disease screening systems. Increasing the range of retinal diseases the system can be relevant to, the system may have an important impact in the overall sphere of ophthalmic diagnostics.

7.2 Limitations and Challenges

Although the results are promising, it has a number of limitations. First, the training and evaluation dataset are based on a single clinical site and thus may not be applicable to different populations having other demographics, image quality, or clinical settings. The further studies have to involve a bigger range of datasets to enhance the ability of the model to be generalized. The other limitation is that glaucoma detection can be done using fundus photography alone. Fundus images can be very informative but fail to give a comprehensive picture of the disease particularly when it is at an early stage or in some forms of glaucoma. Additional diagnostic methods (like OCT) could prove to be very useful in improving the accuracy of the model, although this would need more validation and integration procedures. Lack of explainability is also a major shortcoming. Although the model is effective, it has the disadvantage of being a black box, which can be detrimental towards acceptance by clinicians who would need open processes of decision-making. Research is required in the future to solve this problem through explainable AI methods. Lastly, even though the model had high sensitivity and accuracy, its ability to detect glaucoma at an early stage is unknown. The model has been tested on a collection of images which contained more developed cases of glaucoma, but the ability to identify earlier stages of the disease is still a subject of future research. Further efforts are necessary to evaluate the performance of the model on cases at an early stage and make it useful towards preventing loss of vision at the first stages of the disease.

This study has constructed an efficient and precise deep learning architecture on automated glaucoma classification on retinal images. Nevertheless, to achieve the full potential of its use in clinical practice, future efforts should be aimed at enhancing generalization in a wide range of populations, incorporating other diagnostic instruments, creating explainable AI models, and extending the model to other retinal diseases. With the discussion of these domains, the model will make an excellent aid in clinical screening of glaucoma and various retinal diseases and achieve better patient outcomes and burden on healthcare systems.

Bibliography

[1] J. J. B., G. G. C., and N. G. O., "Optic disc morphometry in chronic primary open-angle glaucoma," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 226, no. 6, pp. 521–530, 1988. DOI: https://doi.org/10.1007/BF02169490

[2] B. C. F., "A biomechanical approach to glaucoma," *Eye*, vol. 29, no. 12, pp. 1520–1529, 2015. DOI: https://doi.org/10.1038/eye.2015.210

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Computer Vision and Pattern Recognition*, 2015, Submitted on 10 Dec 2015. DOI: 10.48550/arXiv.1512.03385 [Online]. Available: https://doi.org/10.48550/arXiv.1512.03385

[4] C. EW et al., "Glaucoma in asia: Regional prevalence variations and future projections," *Br J Ophthalmol*, vol. 100, no. 1, pp. 78–85, 2016.

[5] A. A. Salam, T. Khalil, M. Akram, A. Jameel, and I. Basit, "Automated detection of glaucoma using structural and non-structural features," *Springerplus*, vol. 5, pp. 1–22, 2016.

[6] W. F, C. L.P., and K. D, "Deep learning in medicine—promise, progress, and challenges," *JAMA Intern. Med.*, vol. 179, pp. 293–294, 2019.

[7] e. a. L. Li, "A large-scale database and a cnn model for attention based glaucoma detection," 2019. [Online]. Available: https://github.com/smilell/AG-CNN

[8] M. A. R. Ratul, M. H. Mozaffari, W.-S. Lee, and E. Parimbelli, "Skin lesions classification using deep learning based on dilated convolution," 2019. DOI: 10.1101/860700

[9] S. Serte and A. Serener, "A generalized deep learning model for glaucoma detection," in *3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2019. DOI: https://doi.org/10.1109/ismsit.2019.8932753

[10] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, 2019.

[11] L. Wu et al., "Detect glaucoma with image segmentation and transfer learning," in *Proceedings of the 32nd International Conference on Computer Animation and Social Agents*, 2019, pp. 37–40.

[12] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020. DOI: 10.48550/arXiv.2010.11929 [Online]. Available: https://doi.org/10.48550/arXiv.2010.11929

[13] Y. George, B. J. Antony, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, “Attention-guided 3d-cnn framework for glaucoma detection and structural-functional association using volumetric images,” *IEEE J Biomed Health Inform*, vol. 24, no. 12, pp. 3421–3430, 2020. DOI: 10.1109/JBHI.2020.3001019

[14] A. M, I. M, A. T, A. G, S. A, and A. S. A, “An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification,” *Diagnostics*, vol. 10, no. 8, p. 602, 2020. DOI: https://doi.org/10.3390/diagnostics10080602

[15] H. R, E. B, B.-B. J, L. S, M. M, and P. S, “Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning,” *Acta Ophthalmol*, vol. 98, no. 1, e94–e100, 2020. DOI: https://doi.org/10.1111/aos.14193

[16] E. G. Society, “Terminology and guidelines for glaucoma (4th ed.),” 2020. [Online]. Available: https://www.eugs.org/eng/guidelines.asp

[17] O. J. Afolabi, G. P. Mabuza-Hocquet, F. V. Nelwamondo, and B. S. Paul, “The use of u-net lite and extreme gradient boost (xgb) for glaucoma detection,” *IEEE Access*, vol. 9, pp. 47 411–47 424, 2021. DOI: 10.1109/ACCESS.2021.3068204

[18] A. V. Ikechukwu, S. Murali, R. Deepu, and R. Shivamurthy, “Resnet-50 vs vgg-19 vs training from scratch: A comparative analysis of pneumonia segmentation and classification from chest x-ray images,” *Proceedings on Global Transitions*, vol. 2, pp. 375–381, 2021. DOI: https://doi.org/10.1016/j.gltp.2021.08.027

[19] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. DOI: 10.1109/ICCV48922.2021.00986 [Online]. Available: https://doi.org/10.1109/ICCV48922.2021.00986

[20] S. V. Nazir T Irtaza A, “Optic disc and optic cup segmentation for glaucoma detection from blur retinal images using improved mask-rcnn,” *Int J Opt*, 2021. DOI: https://doi.org/10.1155/2021/6641980

[21] e. a. P. Mehta, “Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images,” *Am J Ophthalmol*, vol. 231, 2021. DOI: 10.1016/j.ajo.2021.04.021

[22] D. Parashar and D. K. Agrawal, “Automatic classification of glaucoma stages using two-dimensional tensor empirical wavelet transform,” *IEEE Signal Process Lett*, vol. 28, 2021. DOI: 10.1109/LSP.2020.3045638

[23] A. T, A. S, C. M.A., H. E, H. M.S., and A. K, “Glaucoma detection using inception convolutional neural network v3,” *Commun. Comput. Inf. Sci.*, vol. 1435, pp. 17–28, 2021.

[24] Q. F. C.-F. Chen and R. Panda, “Crossvit: Cross-attention multi scale vision transformer for image classification,” 2022. [Online]. Available: https://github.com/IBM/CrossViT

[25] S. Joshi, B. Partibane, W. A. Hatamleh, H. Tarazi, C. S. Yadav, and D. Krah, "Glaucoma detection using image processing and supervised learning for classification," *Journal of Healthcare Engineering*, vol. e2988262, 2022. DOI: https://doi.org/10.1155/2022/2988262

[26] R. Kashyap, R. Nair, S. M. P. Gangadharan, M. Botto-Tobar, S. Farooq, and A. Rizwan, "Glaucoma detection and classification using improved u-net deep learning model," *Healthcare (Switzerland)*, vol. 10, no. 12, 2022. DOI: 10.3390/healthcare10122497

[27] O. Kovalyk, J. Morales-Sánchez, R. Verdú-Monedero, et al., "Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment," *Scientific Data*, vol. 9, p. 291, 2022. DOI: 10.1038/s41597-022-01388-1 [Online]. Available: https://doi.org/10.1038/s41597-022-01388-1

[28] I. Liesenborghs, "Genetic and molecular pathogenesis of primary open angle glaucoma and corticosteroid-induced glaucoma," Ph.D. dissertation, Maastricht University, 2022. [Online]. Available: https://doi.org/10.26481/dis.20221201il

[29] N. M, N. T, J. A, T. U, and Y. HS, "An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization," *Sensors*, 2022. DOI: https://doi.org/10.3390/s22020434

[30] e. a. M. B. Sudhan, "Segmentation and classification of glaucoma using u-net with deep learning model," *J Healthc Eng*, vol. 2022, 2022. DOI: 10.1155/2022/1601354

[31] T. Shyamalee and D. Meedeniya, "Glaucoma detection with retinal fundus images using segmentation and classification," *Machine Intelligence Research*, vol. 19, no. 6, pp. 563–580, 2022. DOI: 10.1007/s11633-022-1354-z

[32] S. Thainimit, P. Chaipayom, N. Sa-arnwong, D. Gansawat, S. Petchyim, and S. Pongrujikorn, "Robotic process automation support in telemedicine: Glaucoma screening usage case," *Inform Med Unlocked*, vol. 31, 2022. DOI: 10.1016/j.imu.2022.101001

[33] Z. J. Agboola HA, "Wavelet image scattering based glaucoma detection," *BMC Biomed Eng*, vol. 5, no. 1, p. 1, 2023.

[34] A. K. Bakkouri I, "Mlca2f: Multi-level context attentional feature fusion for covid-19 lesion segmentation from ct scans," *SIViP*, vol. 17, no. 4, pp. 1181–1188, 2023.

[35] S. Daniya T.; Vigneshwari, "Rider water wave-enabled deep learning for disease detection in rice plant," *Adv. Eng. Softw.*, vol. 182, p. 103 472, 2023. DOI: https://doi.org/10.1016/j.advengsoft.2023.103472

[36] D. S. David, "Enhanced glaucoma detection using ensemble based cnn and spatially based ellipse fitting curve model," *J Ambient Intell Humaniz Comput*, vol. 14, no. 4, 2023. DOI: 10.1007/s12652-021-03467-4

[37] W. H. I.-E. Haouli and H. Seridi-Bouchelaghem, "Exploring vision transformers for automated glaucoma disease diagnosis in fundus images," pp. 520–524, 2023. DOI: 10.1109/DASA59624.2023.10286714

[38] S. S and M. R.P., “Review on ocular disease recognition using deep learning,” in *2023 International Conference on Advancement in Computation Computer Technologies (InCACCT)*, 2023, pp. 316–321.

[39] G. H. R. V. V. N. S. Kumar and M. N. GiriPrasad, “A novel glaucoma detection model using unet++-based segmentation and resnet with gru-based optimized deep learning,” *Biomed Signal Process Control*, vol. 86, 2023. DOI: 10.1016/j.bspc.2023.105069

[40] e. a. C. Wan, “A new segmentation algorithm for peripapillary atrophy and optic disk from ultra-widefield photographs,” *Comput. Biol. Med.*, vol. 172, p. 108 281, 2024.

[41] X. H, M. F. L., and A. S, “Multi-glaucnet: A multi-task model for optic disc segmentation, blood vessel segmentation and glaucoma detection,” *Biomed. Signal Process. Control*, vol. 99, p. 106 850, 2025.

[42] M. Mahdi, M. Sheikhey, A. Sahafi, H. Pirnejad, and A. Naemi, “Deep learning for comprehensive analysis of retinal fundus images: Detection of systemic and ocular conditions,” *Bioengineering*, vol. 12, no. 8, p. 840, 2025. DOI: https://doi.org/10.3390/bioengineering12080840

[43] Q. I, “Glaucoma detection in retinal images using image processing techniques,”

[44] R. Mahum, S. U. Rehman, O. D. Okon, A. Alabrah, T. Meraj, and H. T. Rauf, “A novel hybrid approach based on deep cnn to detect glaucoma using fundus imaging,” *Electronics (Switzerland)*, vol. 11, no. 1, p. 2022, DOI: 10.3390/electronics11010026