OCT-AttenNet: Developing An Improved Deep Learning Framework for Multi-Class Eye Disease Detection

Ashikur Rahman Shad

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

SEPTEMBER 2025

APPROVAL

This thesis titled on "**OCT-AttenNet: Developing An Improved Deep Learning Framework for Multi-Class Eye Disease Detection**", submitted by **Ashikur Rahman Shad (ID: 212-35-724)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

---------- **Chairman**

Dr. S M Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---------- **Internal Examiner 1**

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---------- **Internal Examiner 2**

Khalid Been Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---------- **External Examiner**

Dr. Md. Sazzadur Rahman
Professor
Institute of Information Technology
Jahangirnagar University

Declaration

I hereby declare that this thesis has been completed under the supervision of **Mr. Khalid Been Md. Badruzzaman Biplob, Lecturer** (Senior Scale), Department of Software Engineering, Daffodil International University. I also affirm that this thesis is my original work, submitted for the degree of B.Sc. in Software Engineering, and neither the entire work nor any portion has been previously submitted for another degree at this or any other university.

Ashikur Rahman Shad
ID : 212-35-724
Department of Software Engineering
Daffodil International University

Certified By:

Mr. Khalid Been Md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Daffodil International University

ACKNOWLEDGEMENTS

First and foremost, I am profoundly grateful to Allah the Almighty for granting me the health, patience, and perseverance to complete this thesis. Without His countless blessings and guidance, none of this would have been possible.

I would like to express my deepest gratitude to my beloved parents, Mr. Mohammad Abu Sufian and Mrs. Mahbuba Akter Sheuli, for their unwavering support, encouragement, and for sponsoring my education. Their sacrifices, love, and belief in me have been the driving force behind my academic journey.

My heartfelt thanks go to my thesis supervisor, Mr. Khalid Been Badruzzaman Biplob, Lecturer (Senior Scale), Department of Software Engineering, Faculty of Science and Information Technology, for his continuous guidance, valuable feedback, and encouragement throughout this research. His expertise and mentorship have been instrumental in shaping this work.

I am also sincerely thankful to Abu Kowshir Bitto, a respected Master's senior at my university, for his generous support, insightful advice, and for always being willing to help whenever I needed guidance.

Special thanks to my university friends and family members who have stood by me through every challenge, offering moral support, laughter, and companionship along the way.

Lastly, I acknowledge the groundbreaking contributions of Yann LeCun, Yoshua Bengio, and Geoffrey Hinton for Convolutional Neural Networks (CNNs); Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin for the Transformer architecture; and Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio for the Attention Mechanism, without which this work would not have been possible.

ABSTRACT

It becomes really difficult to work without a sight. We have to save our sight before its too late. For this we need early detection of diseases. We developed a novel OCT-AttenNet model based on InceptionV3 and added BAM with ECA attention mechanism. We also applied several preprocessing and data enhancement techniques. Our proposed model OCT-AttenNet achieved an accuracy of 92% on a 10 class dataset collected from Bangladesh. It outperforms its backbone InceptionV3 by 2%. We did a comparative study of several CNN and transformers. Our proposed model outperforms all. We applied XAI like GradCAM++, IG to make it reliable to doctors and have a better understand of how the model is thinking. The model performed well with all diseases except early glaucoma and non-pathological myopia. The paper also covers how to improve this prediction.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$F \in \mathbb{R}C \times H \times W$	Feature map tensor with C channels, height H, and width W
Ms	Spatial attention mask in Bottleneck Attention Module (BAM)
σ	Sigmoid activation function, maps values between 0 and 1
ReLU	Rectified Linear Unit activation function
Conv d=4	Convolution operation with dilation rate of 4
$GAP(F)$	Global Average Pooling applied to feature map F
$W1$, $W2$	Weight matrices of the Multi-Layer Perceptron (MLP)
r	Reduction ratio in channel attention
Mc	Channel attention mask
$k = \varphi(C)$	Adaptive kernel size in Efficient Channel Attention, based on number of channels C
$\gamma, b \in \mathbb{R}$	Constants used in kernel size calculation for ECA
F′	Refined feature map after applying attention mechanisms

LIST OF ABBREVIATIONS

OCT	Optical Coherence Tomography
BAM	Bottleneck Attention Module
ECA	Efficient Channel Attention
AI	Artificial Intelligence
XAI	Explainable Artificial Intelligence
Grad-CAM	Gradient-weighted Class Activation Mapping
IG	Integrated Gradients
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
DMBO	Discrete Migratory Bird Optimizer
SS-OCT	Swept-Source Optical Coherence Tomography
PCA	Principal Component Analysis
YOLO	You Only Look Once (object detection algorithm)
ORID	Ocular Disease Intelligent Recognition Dataset
SVM	Support Vector Machine
CLAHE	Contrast Limited Adaptive Histogram Equalization
AHE	Adaptive Histogram Equalization
SMOTE	Synthetic Minority Oversampling Technique
GAP	Global Average Pooling
MLP	Multi-layer Perceptron
GPU	Graphics Processing Unit
CUDA	Compute Unified Device Architecture
ADAM	Adaptive Moment Estimation (optimizer)
RMSprop	Root Mean Square Propagation (optimizer)
ViT	Vision Transformer
SwinTransformer	Shifted Window Transformer
FLOPs	Floating Point Operations per Second (used in convolution cost)

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

We often see blind people in the walkway. In 2003, a study found that in Bangladesh, over 650,000 people are blind [1]. Blindness is usually caused by diseases that slowly lead to permanent blindness. These can be prevented if detected early. Central serous chorioretinopathy affects about 1.97 million people worldwide in 2025, and while men have a higher risk, it usually does not cause permanent blindness [2]. Glaucoma, on the other hand, often leads to permanent blindness, with the number of patients projected to be 111.8 million worldwide by 2040. [3] Myopia is predicted to affect about 4.76 billion people by 2050, and nearly 10% will have high myopia that can cause blindness through complications [4]. Diabetic retinopathy again is a significant cause of permanent blindness, affecting around 27% of people with diabetes worldwide. Retinal detachment occurs in about 10 to 18 people per 100,000 each year, and it can cause permanent blindness if not treated quickly [5]. Disc Edema is often seen in women of childbearing age with high body weight, and it too can lead to permanent blindness [6]. Macular scars are common in older adults with age-related macular degeneration, typically resulting in permanent vision loss [7]. Finally, retinitis pigmentosa affects about one in every 4,000 people worldwide, and it is a progressive disease that often causes permanent blindness [9]. Researchers have tested different CNN and Transformer-based Deep learning models. But these models are not used anywhere. There is still room for improvement.

1.2 Problem Statement

Losing sight affects a person's entire life. People in rural areas often have access to advanced medical facilities. There aren't enough medical experts in Bangladesh. There are five doctors for every 10,000 people in Bangladesh, and in rural areas, it's even worse. Doctors can't keep up with this ratio. Again, there are AI models for detecting fundus disease. But most of them are not trained for multi-class detections. Few are available who are trained on the Bangladeshi dataset. And many of these models are just black boxes, and don't show how they predicted what they predicted. This makes these models unreliable for doctors. Also, there is room for improvement for existing models.

1.3 Motivation

We need to find a way to ensure medical care for everyone, especially regarding eye diseases, because it's hard to work without sight. It can affect employment, the economy, everything. We have to figure out a way to increase the productivity of doctors so that even with this ratio, most people can get access to medical care. And today we have AI that can do so. We can use explainable AI (XAI) to make AI models reliable for doctors. For Bangladesh, we need models trained on a local dataset. So, different research claims different models as superior. We can compare these models and decide which one works best for our case. To tackle all of that, we develop a multi-class OCT model trained on 8types of eye disease datasets collected from Bangladeshi hospitals. And we are going to apply different explainable AI (XAI) to ensure the model is reliable for clinical use. And we also try to improve a baseline model for our dataset. CNN models don't come with an attention mechanism out of the box, unlike transformers. So we are going to try an attention mechanism on top of our best-performing CNN.

1.4 Significance of the Study

This study compares multiple top performing CNN and Transformer models. Different research claims different models as superior. This study helps researchers decide which model can be best for the diseases we tested. We worked on a dataset collected from Bangladeshi hospitals. The dataset reflects Bangladeshi scenario. Our custom model, OCT-AttenNet with XAI is suitable for this datasets. It can help in everyday diagnosis of eye disease and increase productivity of doctors. This can leverage eye disease detection in an early stage and prevent permanent blindness which is affecting employment, economy, etc.

1.5 Research Questions

This research tries to find answers to the following questions:

- Which CNN or Transformer model works best for classifying OCT images into nine classes (Central Serous Chorioretinopathy, Macular Scar, Diabetic Retinopathy, Myopia, Disc Edema, Retinal Detachment, Glaucoma, Retinitis Pigmentosa, and Healthy)?
- Can we improve the accuracy of the best model by adding spatial and channel attention like Bottleneck Attention Module (BAM) and Efficient Channel Attention (ECA)?
- Can explainability methods show the OCT image areas the model looks at when making a diagnosis?

1.6 Research Objective

- Create a new deep learning architecture, OCT-AttenNet, designed for multi-class classification of common diseases using Optical Coherence Tomography (OCT) images.
- Comparing leading Convolutional Neural Network (CNN) and Transformer-based models to evaluate their performance and determine the most effective baseline for a dataset collected from two hospitals in Bangladesh.
- Improving the baseline model (InceptionV3)
- An Attention Mechanism was added to the best CNN model, and it was compared with Transformer models, which already have an Attention Mechanism.
- Making the model reliable by applying XAI, highlighting the area of the image based on which the prediction was made.

1.7 Research Scope and Limitations

The following section details the boundaries of the study along with the constraints arising from the dataset, chosen techniques, and evaluation process.

1.7.1 Scope

- This Research has been conducted on nine categories of OCT images: Central Serous Chorioretinopathy, Macular Scar, Diabetic Retinopathy, Myopia, Disc Edema, Retinal Detachment, Glaucoma, Retinitis Pigmentosa, and Healthy.
- Experiments were performed on over 5000 fundus OCT images collected from Anwara Hamida Eye Hospital and BNS Zahrul Haque Eye Hospital in Faridpur, Bangladesh.
- Both well-established CNN architectures (InceptionV3, ResNet50, MobileNetV2, VGG16, VGG19) and modern Transformer architectures (Vision Transformer B/16, Swin Transformer) were tested for comparison.
- The proposed OCT-AttenNet model is based on InceptionV3. Bottleneck Attention Module (BAM) and Efficient Channel Attention (ECA) to improve feature representation.

- Explainability techniques (Grad-CAM, Grad-CAM++, Integrated Gradients) were applied to highlight the regions in OCT and fundus images that influenced the model's predictions, with the goal of improving transparency and clinical trust.

1.7.2 Limitations

- Data has been collected from two hospitals in Bangladesh. It may not represent conditions worldwide.
- The model is limited to 8 diseases. It can't detect diseases other than that.
- It can't work with external eye images or external eye diseases
- Support Vector Machines (SVM) were not included in comparisons, as prior evidence consistently shows CNNs and Transformers perform better for similar image classification tasks.
- Due to the limited size of the dataset, ImageNet-pretrained weights were used for transfer learning instead of training the models from scratch.
- Only 3 XAI (Grad-CAM, Grad-CAM++, and Integrated Gradients) were used. Shap, Lime, etc, were not used in this experiment.

1.8 Thesis Organization

This thesis has five main chapters. The first chapter describes the background and problem, then explains the research goals, importance, scope, and limits. The second chapter reviews related works on CNNs, Transformers, attention mechanisms, and explainability in medical imaging. The third chapter explains the research process. It starts with collecting OCT images from two hospitals in Bangladesh, then shows the preprocessing steps, model choice, and design of the proposed OCT-AttenNet. It also explains the training settings and the metrics for evaluation. The fourth chapter shows the results, compares models, and discusses the explainability outputs. The final chapter summarizes the main findings, contributions, and suggestions for further improvements.

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

In 2021, Li et al. conducted research on ocular disease intelligent recognition [10]. They found that feature fusion improves classification, and Inception-v4 performs the best. The results were an AUC of 86.91% and an F1 score of 87.93%. However, a limitation was that it struggles with multi-label cases and lacks explainability. Similarly, Gour and Khanna researched multi-class and multi-label ophthalmic disease detection using CNN transfer learning [11]. They found that transfer learning enhanced detection performance. The results showed strong classification scores, but a limitation was the imbalance issues and a lack of explainability. Moreover, Sarki et al. researched multi-class diabetic eye disease classification [12]. They found that deep learning has strong potential in DR screening. The results were an accuracy of 81.33%, sensitivity of 100%, and specificity of 100%. However, a limitation was class imbalance and no explainability.

In 2022, Saini and Susan conducted research on diabetic retinopathy screening with deep learning [13]. They found that DenseNet121 performed the best. The results showed the highest accuracy among the tested models. However, the limitation was that explainability was not used, and imbalance still affected the results. Similarly, Rodríguez et al. researched multi-label retinal disease classification with transformers [14]. They found that transformers outperform CNNs. The results were mAP 84.7% and AUC 92.1%. However, the limitation was that the dataset was small, and the imbalance persisted.

In 2023, Bhati et al. conducted research on ophthalmic disease detection using DKCNet [15]. They found that attention mechanisms improved performance. The results were an AUC of 96.08% and an F1 score of 94.28%. However, the limitations included the risk of overfitting and the lack of comparison with transformers.

In 2024, Golam Mohiuddin Niloy et al. conducted research on MobileNet-Eye for disease classification [16]. They found that MobileNetV2 was the most accurate. The results were 96% for MobileNetV2, 95% for EfficientNetB7, and 94% for ResNet50. However, the limitation was a small dataset with only three classes. Furthermore, Al-Fahdawi et al. conducted research on Fundus-DeepNet for multi-label disease classification [17]. They found that advanced CNNs and data fusion improved the results. The results were an AUC of 99.76% and an F1 score of 88.56%. But the limitation was the risk of overfitting and the need for more diverse datasets. Additionally, AlBalawi et al. conducted research on IoT-Opthom-CAD with Swin Transformers and XAI [18]. They found that transformers improved disease classification with explainability. The results showed strong performance in multiclass retinal disease classification. However, the limitation was the lack of very large-scale validation.

In 2025, Hiroki Maehara et al. conducted research on AI support for anterior segment disease diagnosis [19]. They found that AI improved human diagnostic accuracy, with results showing an increase from 79.2% to 88.8%. However, the limitation was the small number of cases and narrow testing scope. Similarly, Gülcan Gencer et al. researched retinal disease detection with SE-hybrid models and found that SE-hybrid was highly effective [20]. The results were 99.18% accuracy for Duke and 99.58% for UCSD. The limitation was the need for broader validation. Likewise, Ahmed Aizaldeen Abdullah et al. researched hybrid deep features with BEL and found that BEL achieved state-of-the-art performance, with results showing 99.77% accuracy [21]. The limitation was the complex pipeline and limited testing. Moreover, Vidivelli et al. researched deep learning for ophthalmological disorders and found that MobileNet with Adam optimizer was optimal, achieving 89.64% accuracy [22]. The limitation was the low overall accuracy and data imbalance. In addition, Md Najib Hasan et al. researched DIA-VXNET for diabetic eye disease detection and found that fusion with VGG16 and XceptionNet was highly accurate, with 99.76% accuracy [23]. The limitation was small dataset. Similarly, Arwa Albelaihi et al. researched RNN for diabetic eye disease detection and found that Bi-GRU with ResNet152V2 delivered strong performance, achieving 99.8% accuracy [24]. Lastly, Ahmed Aizaldeen Abdullah et al. conducted further research on hybrid BEL for the ODIR dataset and found that ensemble learning provided robust results, with 99.17% accuracy [21]. The limitation was limited generalization.

Table 2.1 Existing Models

Year	Dataset	Methods	Results Accuracy	XAI	Cite
2021	Kaggle RP Fundus	CNN Baseline	85.3%	No	23
		CNN+DCGAN	93.7%		
2022	Customized dataset consisting of 2,250 eye images collected from Shutterstock and Google. The dataset is comprised of 750 images each for normal eyes, conjunctivitis eyes, and cataract eyes. It was split into 1,689 training images and 561 validation images.	InceptionV3	97.08%	No	1
		ResNet50	95.68%		
		VGG16	95.48%		
2024	Customized eye disease dataset consisting of 3,000 eye images. The dataset is comprised of 1,000 images each for healthy eyes, eyes with uveitis, and eyes with Lid problems. The data was split into a training set of 2,400 images and a testing set of 600 images.	MobileNetV2	96%	No	2
		EfficientNetB7	95%		
		ResNet50	94%		
2025	Duke Dataset: NORMAL, AMD (Age-related Macular Degeneration), and DME.	SE-EfficientNetB0	93.44%	No	25
		SE-Xception	98.36%		
		SE-Hybrid	99.18%		
	UCSD Dataset: 84,495 OCT images classified into four categories: NORMAL, CNV (Choroidal Neovascularization), DME (Diabetic Macular Edema), and DRUSEN	SE-EfficientNetB0	98.96%		
		SE-Xception	99.38%		
		SE-Hybrid	99.58%		
2025	Kaggle Eye Disease dataset – 4217 images, expanded with augmentation and balancing (≈4392 images). Classes: Cataracts, Glaucoma, Diabetic Retinopathy, Normal	Blending Ensemble Learning (BEL)	99.77%	No	24
2025	Ocular Disease Intelligent Recognition (ODIR)	DenseNet, ResNet, Lenet	-	No	16
		MobileNet	89.64%		
2025	IDRID, HRF	DIA-VXNET	99.76%	No	15
2025	DIARETDBO, DIARETDB1, messidor, HEI-MED, ocular, and retina datasets containing 4 classes: DR, DME, glaucoma, and cataracts	Bi-GRU + ResNet152V2	99.8%	No	14
2025	ODIR dataset – 5000 patient images (left & right eyes). Classes used: Cataracts, Glaucoma, Diabetic Retinopathy, Normal	Blending Ensemble Learning (BEL)	99.17%	No	24

2.2 Research Gap

Most of the studies used CNNs, transformers, or hybrid models to detect eye diseases with high accuracy. However, most datasets were small or limited. Many studies did not include explainability. So doctors can't fully trust those. Some models are also very complex and difficult to implement in real hospitals. Few studies tested on diverse populations or real clinical images. This creates an opportunity for simple, interpretable, and clinically validated models that perform well on larger and more diverse datasets.

CHAPTER 3

METHODOLOGY

After gathering all the data, we start by preprocessing. We remove all blurry and ambiguous images. Since images were taken from different devices, we edit them to normalize their appearance. We stretch and crop to fix their shape, and adjust saturation to match images from various machines, ensuring that image formats do not interfere with the model training. Then, we split the data into training, testing, and validation sets, which will be discussed in the upcoming sections. We applied the CLAHE image enhancement technique to improve image details. Next, we balanced the classes by augmenting images through random flipping, adjusting brightness, contrast, hue, saturation, and other parameters to simulate real-world scenarios. Later, we cropped all images to 224x224 (except for InceptionV3, which uses 299x299).

We start the process of model training. We selected CNN models (InceptionV3, ResNet50, MobileNetV2, VGG16, VGG19) and Transformer models (SwinTransformer, VisionTransformerB16) for our training. We trained and validated them. We adjusted hyperparameters to achieve the best performance. We calculated evaluation metrics for each model and compared their results.

We begin developing our proposed model. We select the best performing CNN. We add a BAM layer within its final block. We replace the channel attention layer of BAM with ECA. We train the model and update it to find the best parameters. After validation, we finally obtain our desired OCT-AttenNet model. We apply XAI to it and compare its performance with the rest of the models.

Procedure

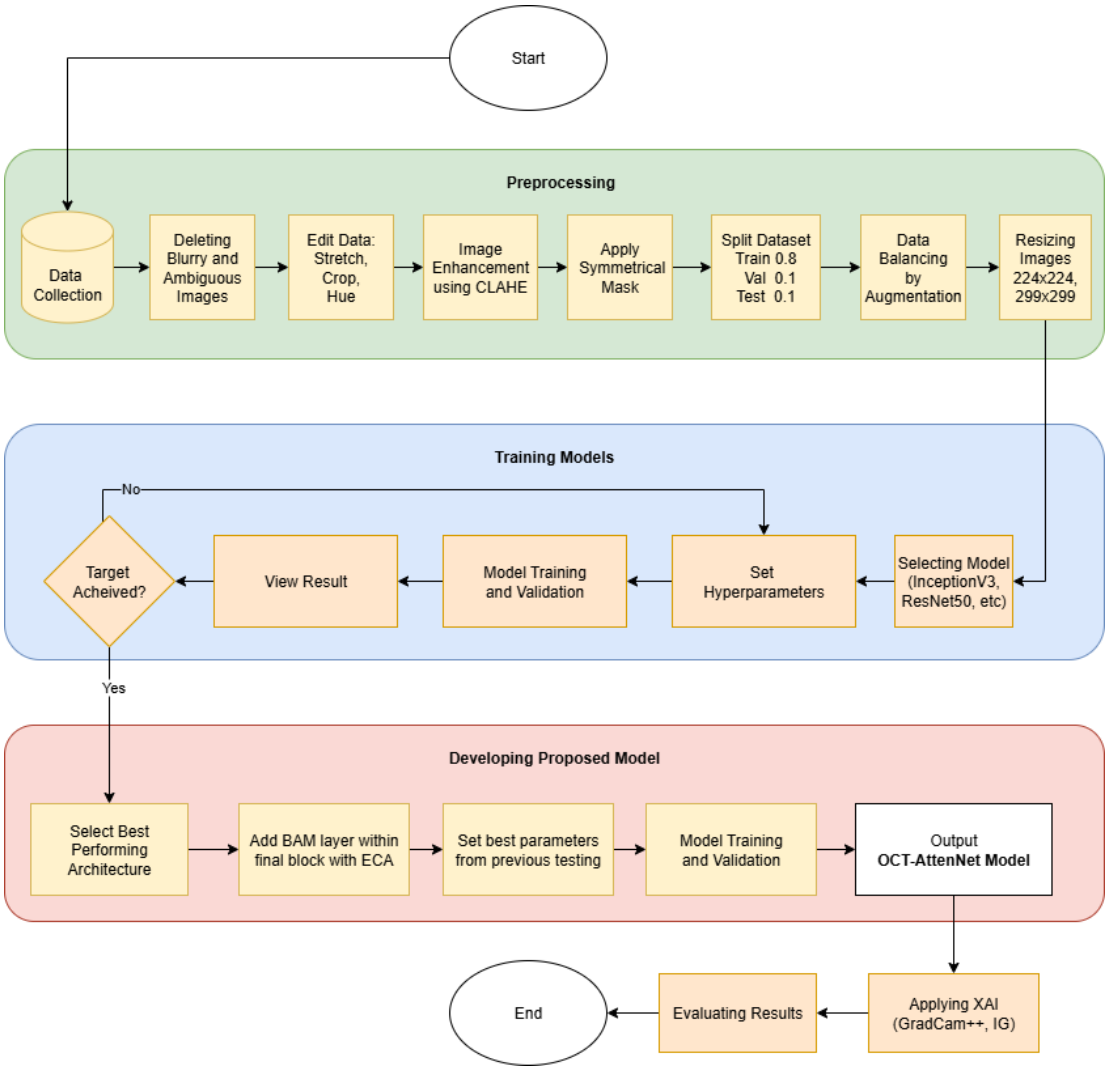


Figure 3.1 Step by step procedure diagram

3.1 Data Collection

Data were collected from Anwara Hamida Eye Hospital and BNS Zahrul Haque Eye Hospital in Faridpur, Bangladesh. The "Eye Disease Image Dataset" was published in Mendeley by Daffodil International University and Jahangirnagar University of Bangladesh in 2024. It contains raw 5335 images, of which 5318 are color fundus images. and 17 anterior segment images. Only fundus images (Central Serous Chorioretinopathy, Glaucoma, Myopia, Diabetic Retinopathy, Healthy, Retinal Detachment, Disc Edema, Macular Scar, Retinitis Pigmentosa) have been selected for the OCT model, and anterior segment images have been rejected (Pterygium).

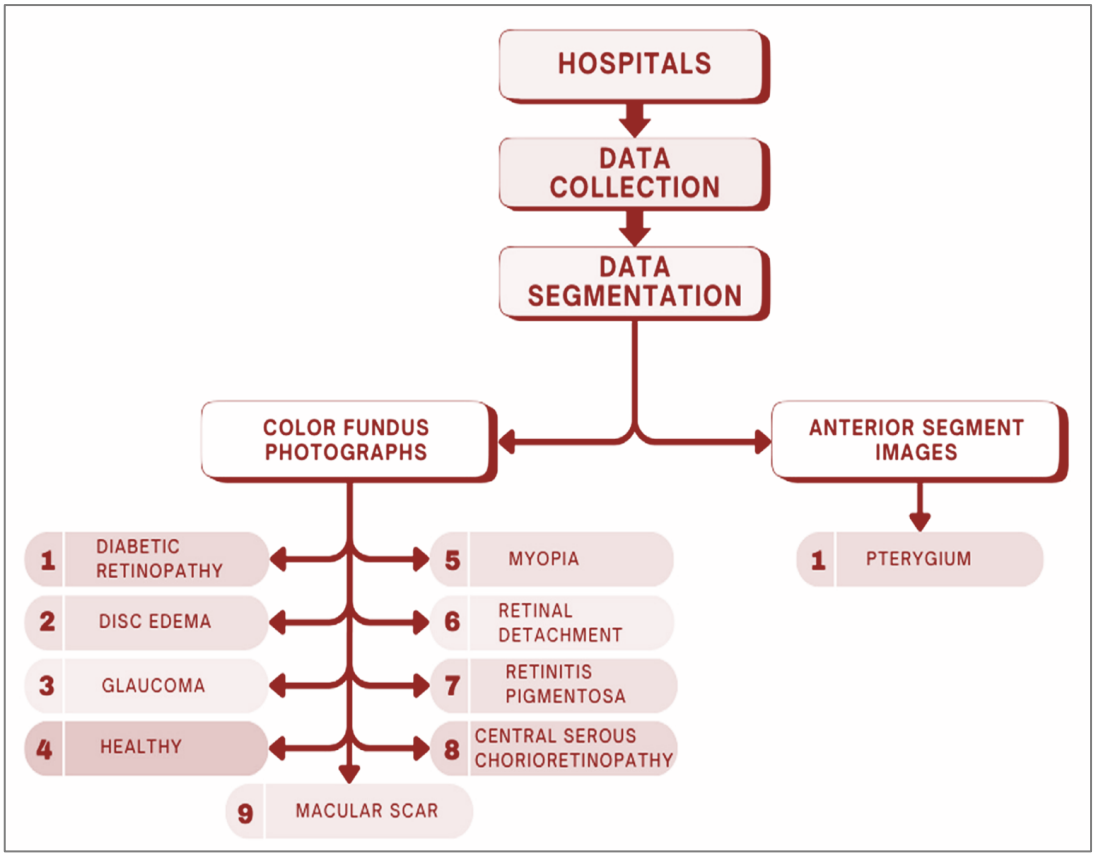


Figure 3.2 Structure of the dataset

3.2 Data Preprocessing

Data Preprocessing is the most crucial step in Machine Learning. Data Scientists spend most of their time preprocessing data. It involves data cleaning, data balancing, formatting data to fit the model, etc. The following are the preprocessing steps we have taken.

3.2.1 Deleting blurry and ambiguous images

Blurry and ambiguous images may confuse the model. That's why all blurry and ambiguous images were deleted from the dataset.

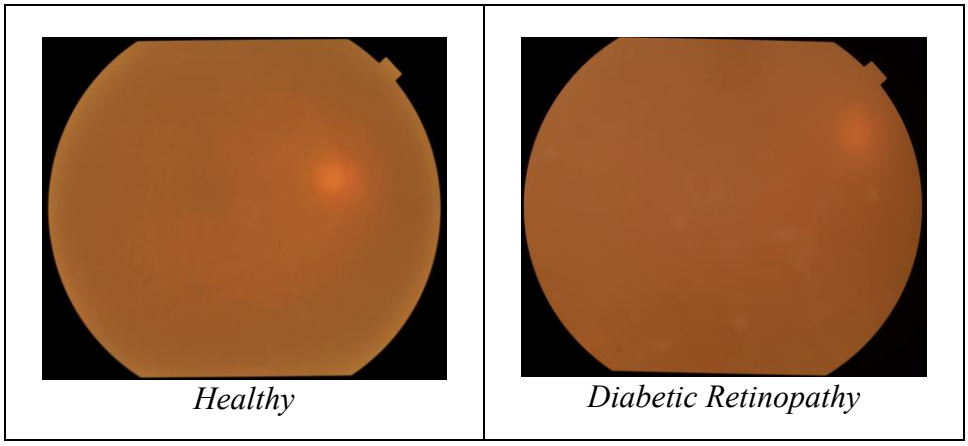


Figure 3.3 Example of blurry images

3.2.2 Edit Data - Stretch, Crop, and Tint

Different images were taken from different fundus cameras. In the dataset, the data were stretched to a square shape without cropping, which changed the fundus shape. So the images were stretched, cropped, and resized to match normal fundus images. And so that they are in alignment with other images of their respective class and don't get distracted by irrelevant features.

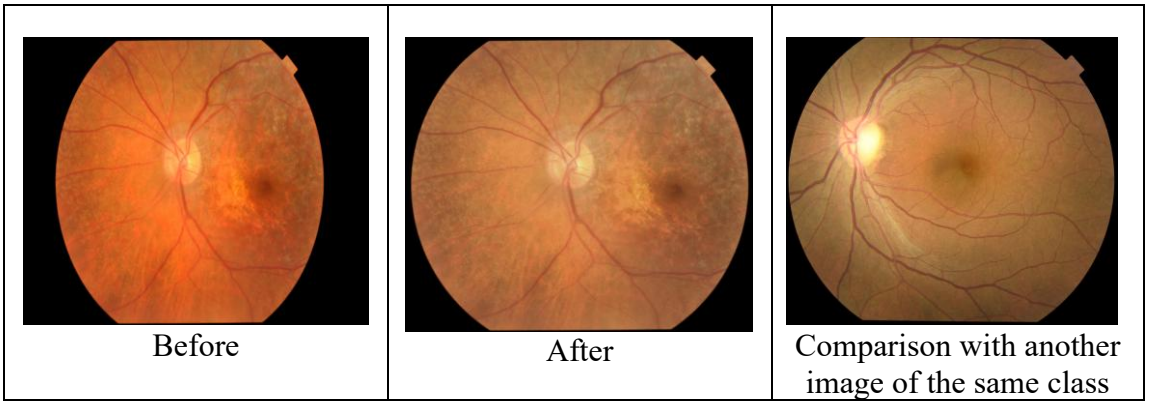


Figure 3.4 Editing image by stretching, cropping, changing hue, and saturation to match the existing data.

3.2.3 Image Enhancement - CLAHE

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image processing technique. It is used to enhance the contrast of images. Unlike Adaptive Histogram Equalization (AHE), CLAHE divides the image into small tiles and applies histogram equalization. Contrast Limit helps to prevent over-amplification of noise in homogeneous areas. Thus, CLAHE can improve low-contrast regions of an image and increase its details, making it popular in medical imaging. Here, we used CLAHE to improve the fundus images of our dataset. We set the clip limit to 2.0 and the tile grid to 8×8 size.

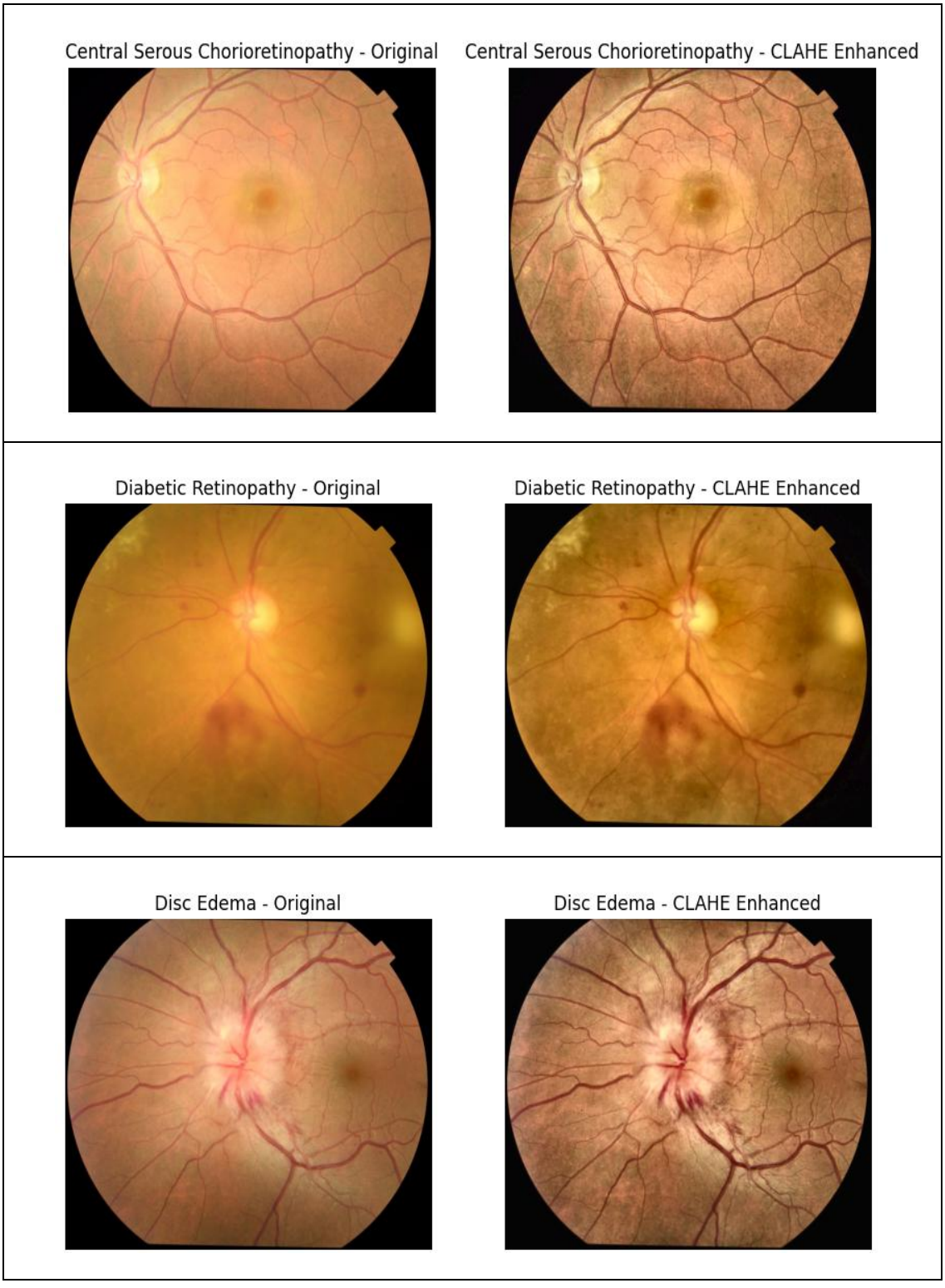


Figure 3.5 Examples of CLAHE Enhanced images (1)

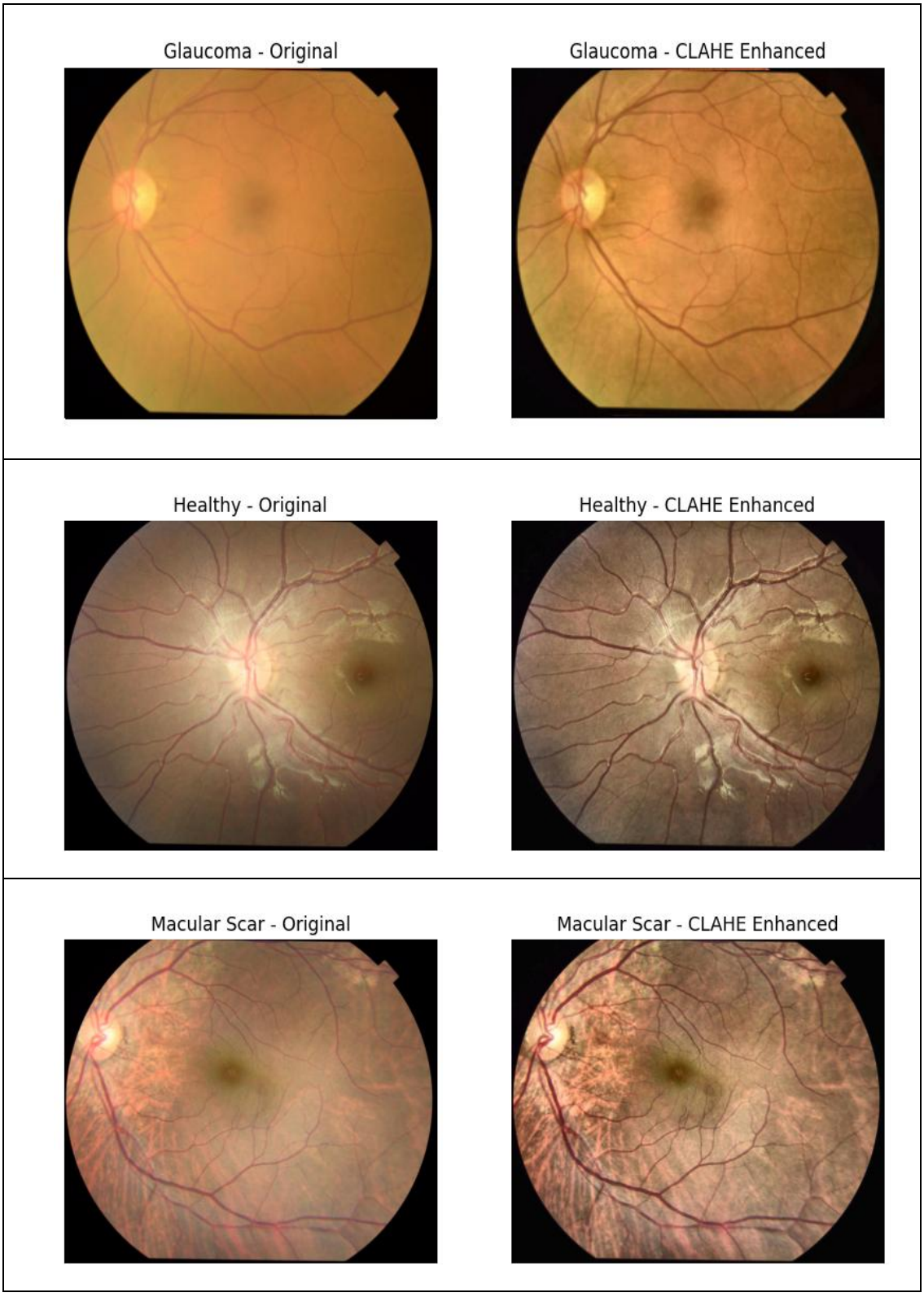


Figure 3.6 Examples of CLAHE Enhanced images (2)

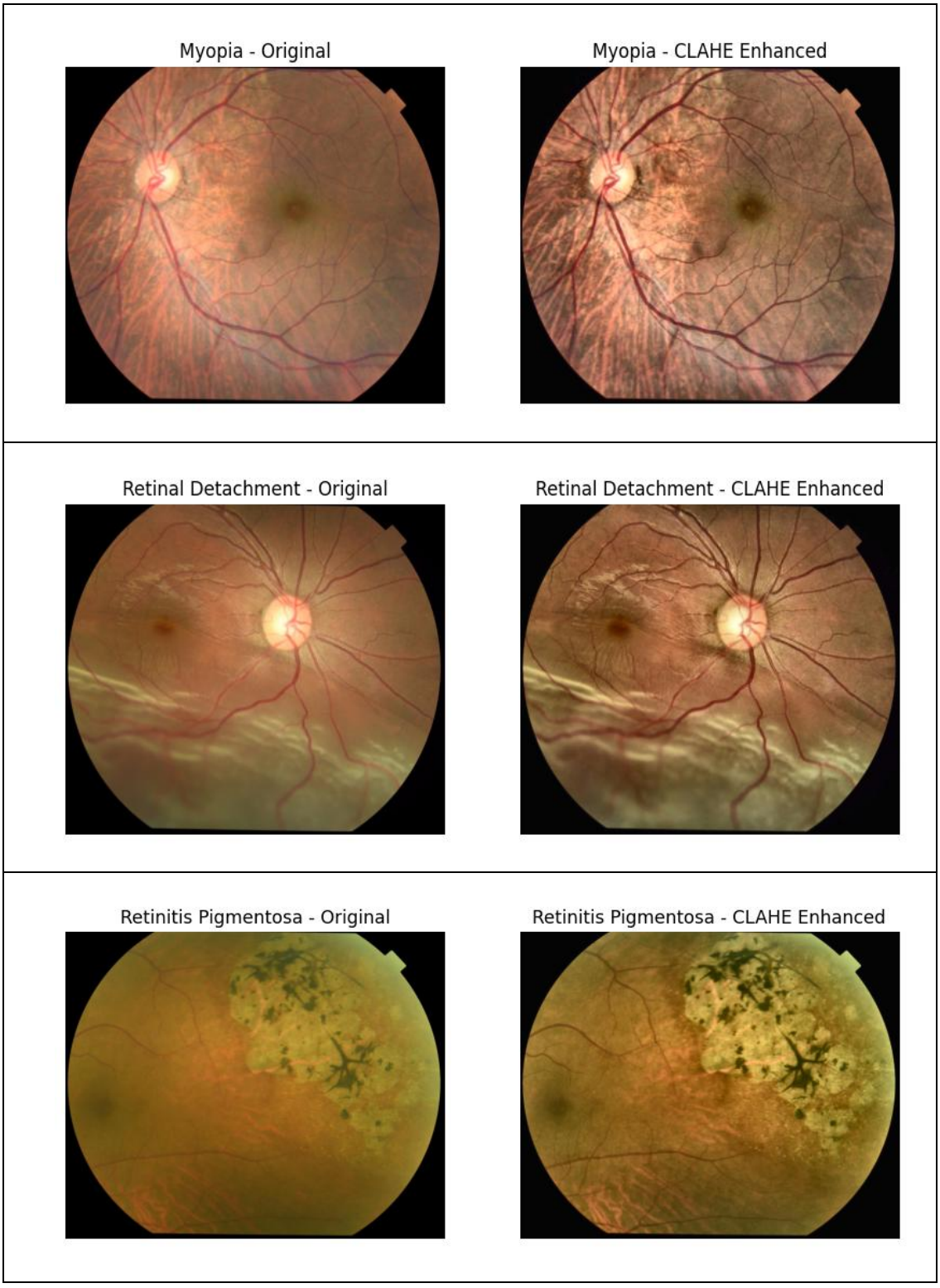
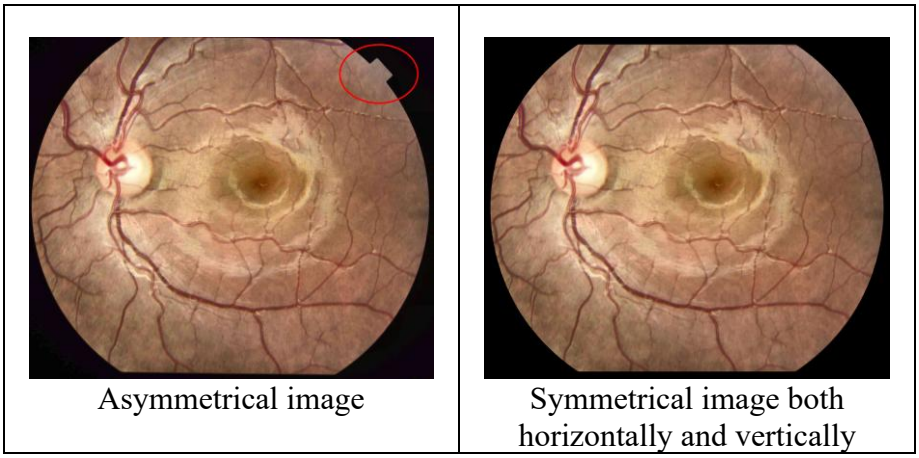


Figure 3.7 Examples of CLAHE Enhanced images (3)

3.2.4 Applying Symmetrical Mask

We applied a symmetrical mask above each image, so that the model doesn't confuse the shape of the camera sensor as a feature. Also, this lets us augment the image by flipping it too.



Asymmetrical image | Symmetrical image both horizontally and vertically

Figure 3.8 Applying symmetrical masks over images to make it symmetrical.

3.2.5 Data Splitting

The dataset has been split into three subsets: training, validation, and testing. Some of the classes in the dataset have a tiny amount of data. So the training dataset is set to 80%, and the other two splits are 10% each, instead of the conventional 70:20:10 split.

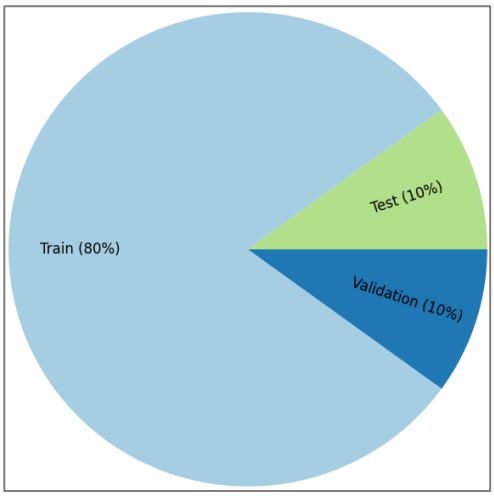


Figure 3.9 Train-Test Split pie-chart

3.2.6 Data Balancing

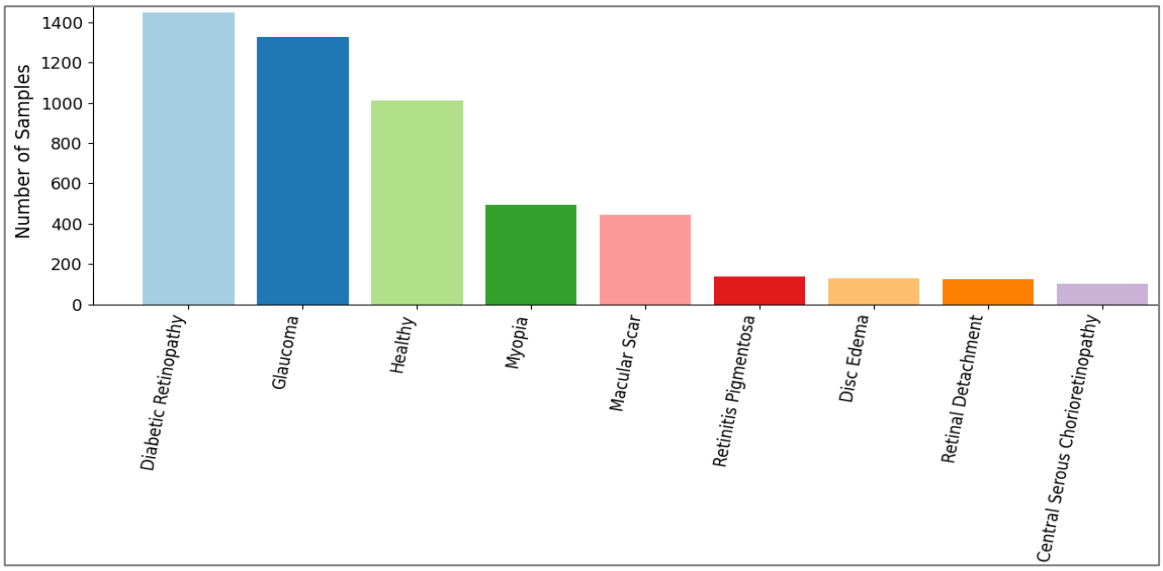


Figure 3.10 Class Distribution Bar Chart before Balancing

The raw dataset is highly imbalanced. The figure shows a considerable ratio of majority to minority classes. This imbalance increases the chance of bias towards courses with data. To solve this problem, the dataset needs to be balanced. Balancing can be done by increasing the size of the minority class (Oversampling) or by reducing the data in the majority classes (undersampling). Since the data in minority classes is tiny, undersampling would result in losing maximum data, which is crucial for model development. Thus, oversampling has been selected for data balancing.

There are several oversampling techniques. Random oversampling creates duplicate data of minority classes, which increases the chance of overfitting. SMOTE, or the Synthetic Minority Oversampling Technique, creates synthetic samples by combining a minority sample and its nearest neighbours in the feature space. However, the synthetically generated images often appear blurry and non-realistic, particularly for complex images such as fundus photographs. Data of minority classes has been augmented to balance with the number of data (1509) in class “Diabetic Retinopathy,” which is the majority class.

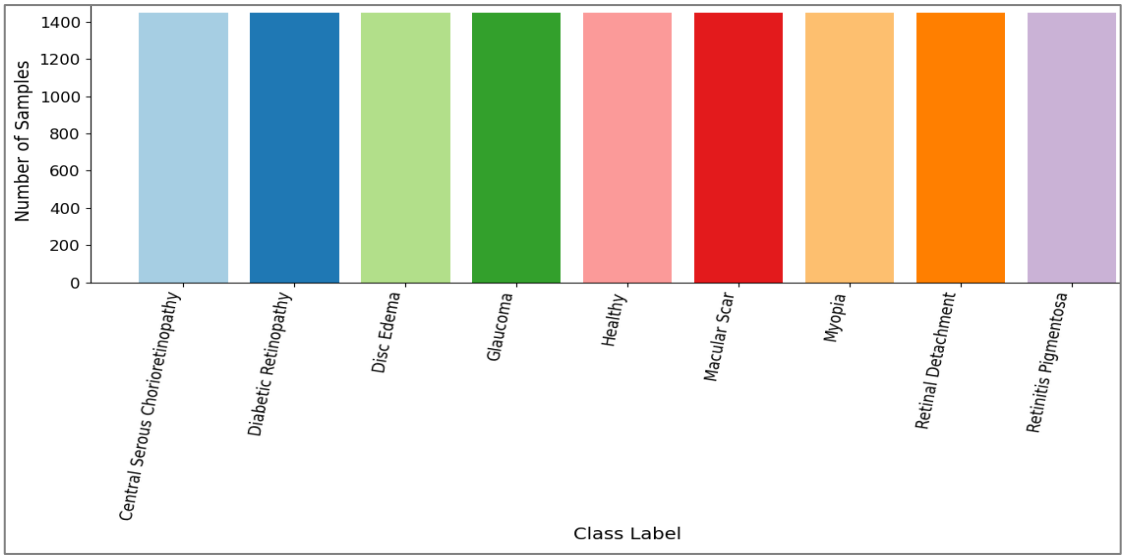


Figure 3.11 Class Distribution Bar Chart after Balancing

All augmented images have been flipped horizontally. Some of them have randomly been flipped vertically with a probability of 40%. Image brightness, Contrast, and Saturation have been shifted more or less within a range of ±0.05, and hue within a range of ±0.03. The following figures are examples of data augmentation. Features of images have shifted slightly, maintaining realism, which is crucial for medical imaging.

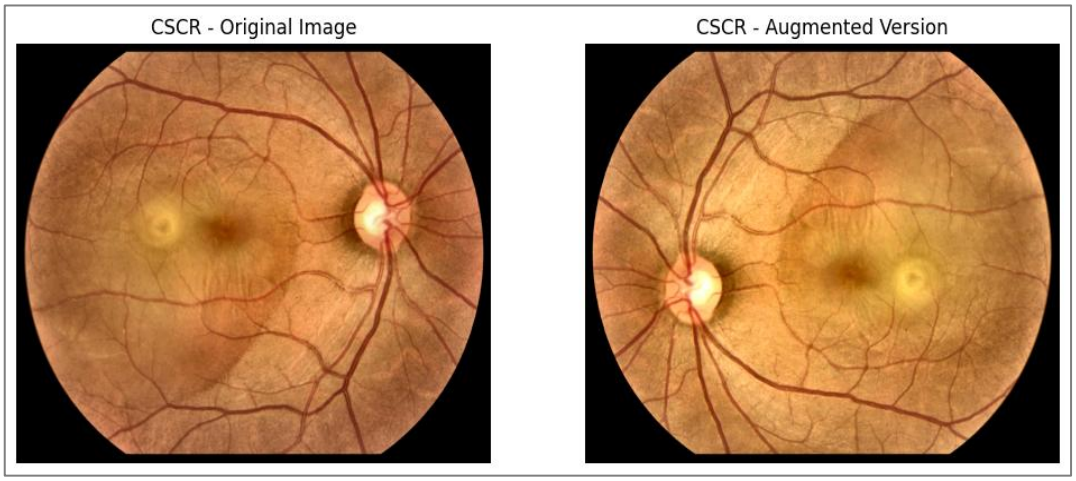


Figure 3.12 Examples of Augmented Images (1)

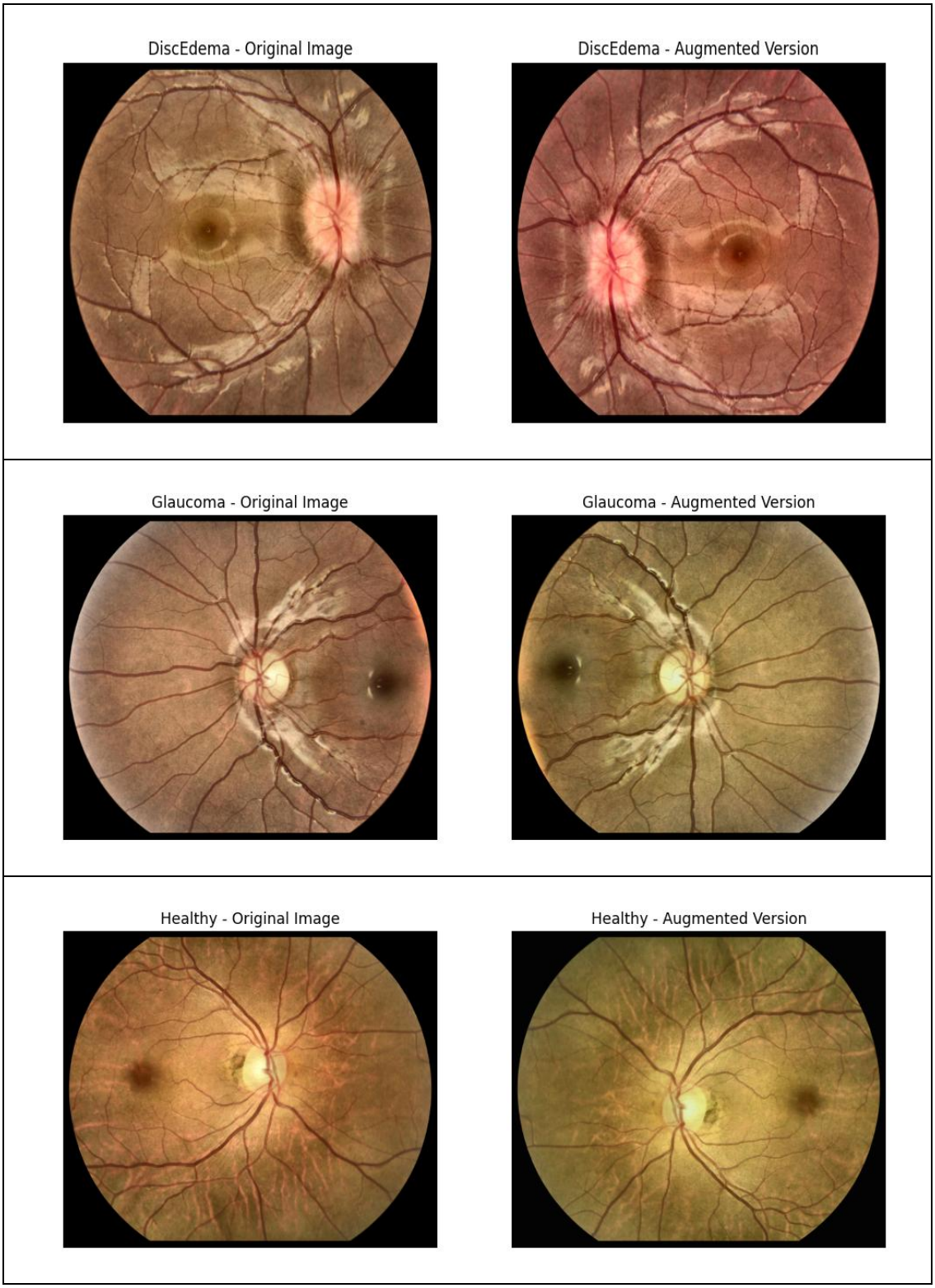


Figure 3.13 Examples of Augmented Images (2)

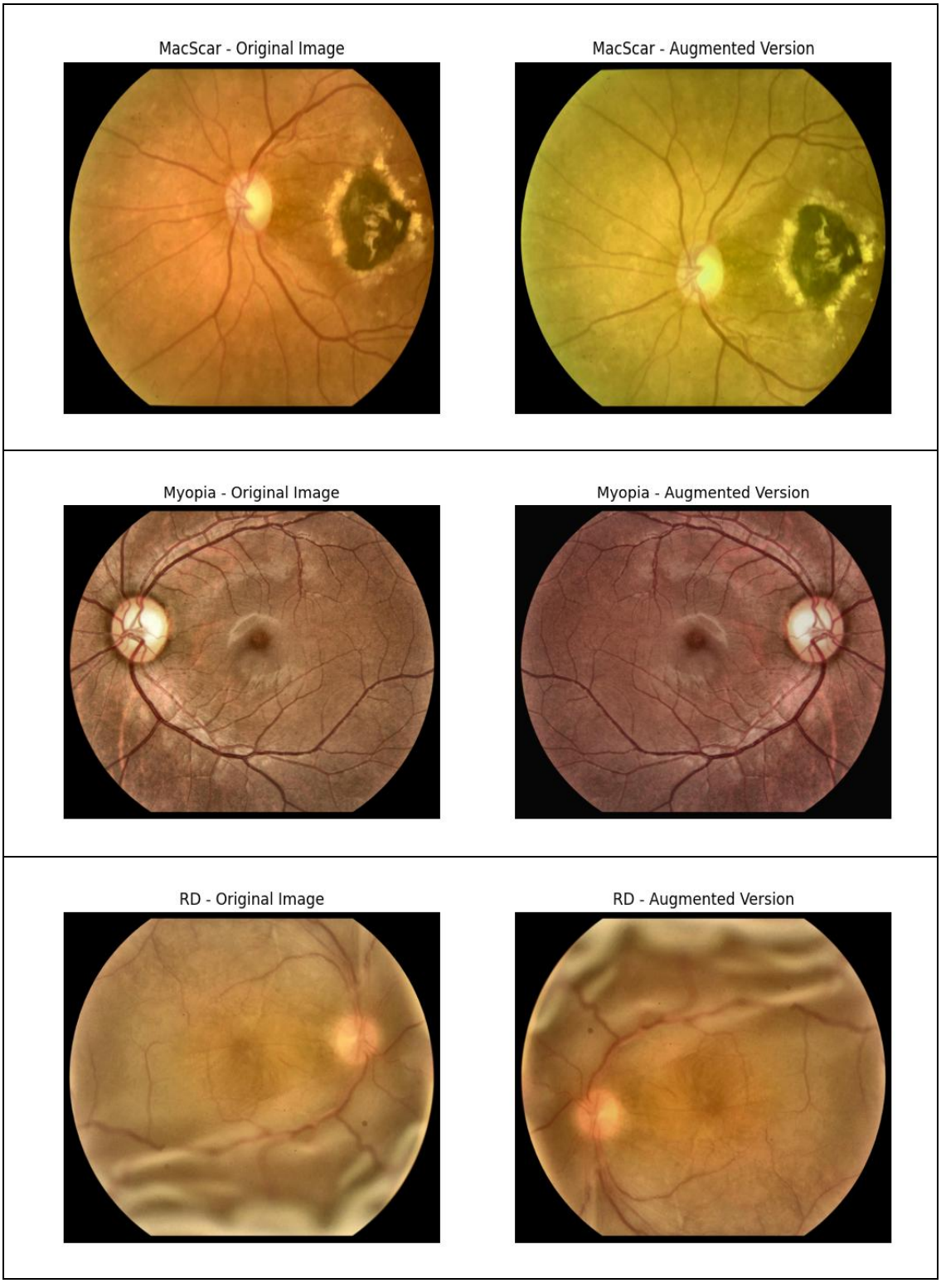


Figure 3.14 Examples of Augmented Images (3)

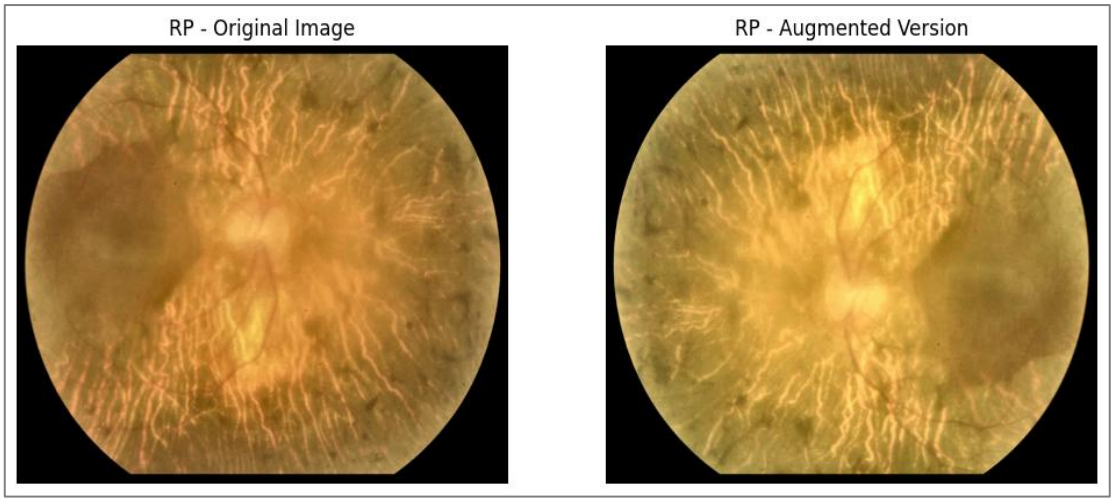


Figure 3.15 Examples of Augmented Images (4)

3.2.7 Resizing:

Images in the dataset are over 2000×1690 pixels. Each pixel is treated as a feature in Image Classification. Too many features result in the Curse of Dimensionality. It's called a curse because adding more features increases the model's complexity and affects its performance.

We used many pretrained models here, and they were trained to perform classification on datasets of specific sizes. These models work with square-shaped images. And it's part of how the model architecture works.

For InceptionV3 and models based on it, like OCT-AttenNet, images have been resized to 299×299 pixels. As for Resnet50, MobileNetV2, EfficientNetB0, DenseNet121, VGG19, VGG16, SwinTransformer, and VisionTransformer-B16, the images have been cropped to 224×224 pixels.

3.3 Parameters and hyperparameters:

We trained all models with the Adam optimizer (initial learning rate 1e-4), batch size 128, for 50 epochs. The learning rate was reduced on plateau using ReduceLROnPlateau (factor 0.1, monitored on validation loss). We used CrossEntropyLoss for multiclass training. Models were initialized with ImageNet-1K pretrained weights (transfer learning). The classifier head included dropout = 0.5; no additional L2 weight decay was applied beyond framework defaults. Early stopping was not used. Training ran on GPU (CUDA). Data augmentation details are described in the Preprocessing section.

3.4 Models

We selected 7 base models for our experiments. Among them, 5 are CNN including InceptionV3, ResNet50, MobileNetV2, VGG16, and VGG19. Then we have 2 Transformer models: VisionTransformerB16, SwinTransformer. We tested one model without preprocessing to test the impact of preprocessing. Lastly, we built the proposed model by taking the best performing CNN and adding attention mechanism to it.

3.5 Proposed Model: OCT-AttenNet

Our proposed model is built upon InceptionV3, which is the best performing model. We added Custom BAM with ECA module.

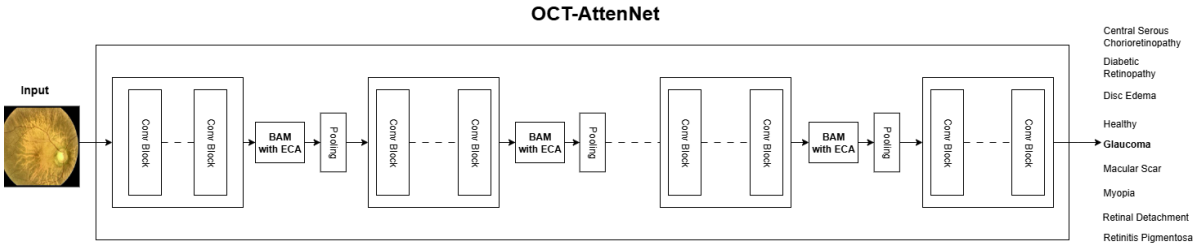


Figure 3.16 Proposed Model OCT-AttenNet Architecture Diagram

3.6 Custom BAM with ECA Module

BAM stands for Bottleneck Attention Module. It combines channel and spatial attention to refine features. We used Efficient Channel Attention instead of the default Channel Attention Module.

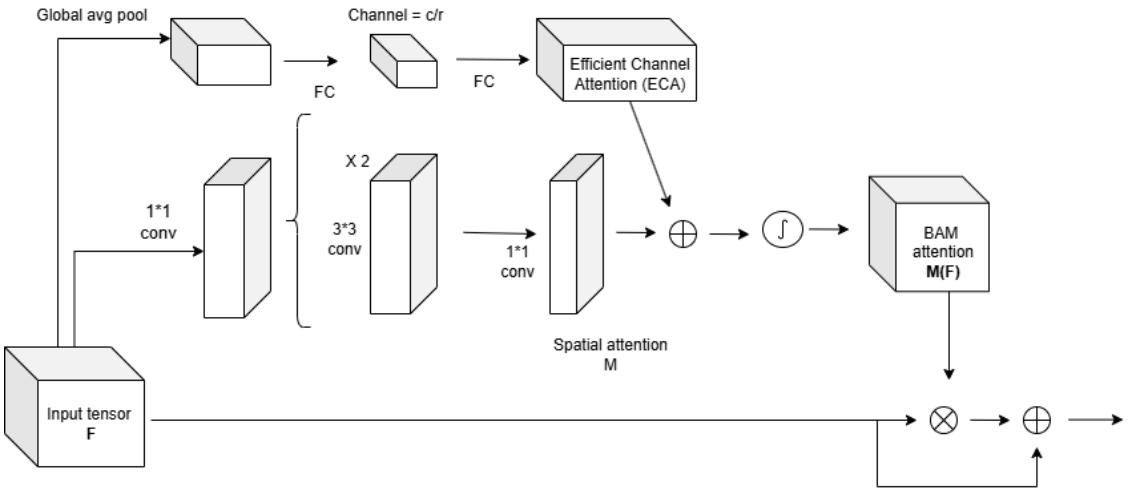


Figure 3.17 Proposed Model Architecture Diagram

We start with a feature map

$$F \in \mathbb{R}^{(C \times H \times W)}. \quad (1)$$

BAM makes an attention map

$$M(F) \in \mathbb{R}^{(C \times H \times W)} \quad (2)$$

The refined feature map is written as:

$$F' = F + F \otimes M(F) \quad (3)$$

The attention map is made by adding channel attention and spatial attention:

$$M(F) = \sigma(Mc(F) + Ms(F)) \quad (4)$$

Here $\otimes$ means element-wise multiplication. A skip connection helps the gradient flow. σ is the sigmoid function. Mc(F) is channel attention. Ms(F) is spatial attention.

3.7 Evaluation Matrix

Most papers we reviewed used Accuracy as their primary metric. We used all of the following matrices, which we calculated from the confusion matrix. But we will compare models and classes based on their F1 score.

Here, TP means true positive, TN means True negative, FP means false positive, and FN means false Negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} \tag{9}$$

$$True\ Negative\ Rate\ (TNR) = \frac{TN}{TN+FP} \tag{10}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \tag{11}$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{FN+TP} \tag{12}$$

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{14}$$

CHAPTER 4

RESULTS

4.1 Result Analysis

4.1.1 Model without Preprocessing (Raw InceptionV3)

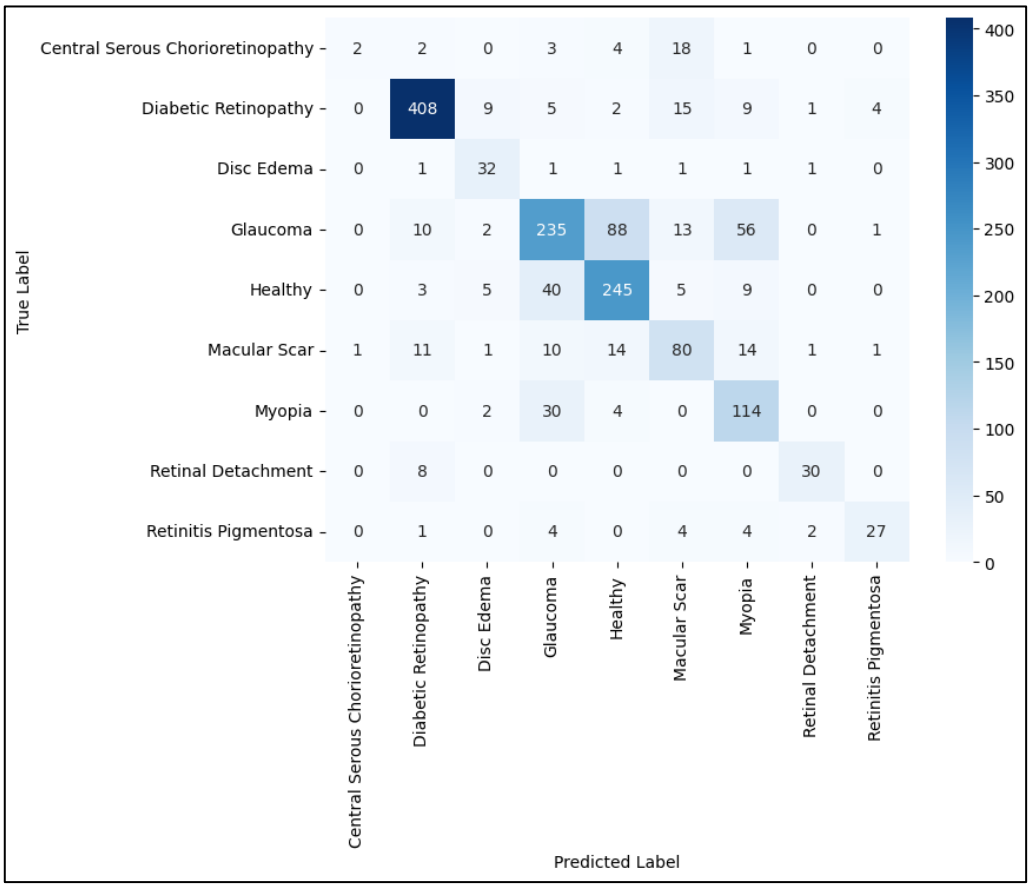


Figure 4.1 Model without Preprocessing (Raw InceptionV3)

In the confusion matrix, we can see that Glaucoma and Myopia both often confuse themselves with the class healthy and vice versa. And Macular Scar confused itself with almost all classes.

Table 4.1 Classification Report of Model without Preprocessing (Raw InceptionV3)

Class	TP	TN	FP	FN	TPR Recall	FPR	TNR Specificity	F1 Score
Central Serous Chorioretinopathy	2	1565	1	28	6.67%	0.06%	99.94%	12.12%
Diabetic Retinopathy	408	1107	36	45	90.07%	3.15%	96.85%	90.97%
Disc Edema	32	1539	19	6	84.21%	1.22%	98.78%	71.91%
Glaucoma	235	1098	93	170	58.02%	7.81%	92.19%	64.12%
Healthy	245	1176	113	62	79.80%	8.77%	91.23%	73.68%
Macular Scar	80	1407	56	53	60.15%	3.83%	96.17%	59.48%
Myopia	114	1352	94	36	76.00%	6.50%	93.50%	63.69%
Retinal Detachment	30	1553	5	8	78.95%	0.32%	99.68%	82.19%
Retinitis Pigmentosa	27	1548	6	15	64.29%	0.39%	99.61%	72.00%

We tested InceptionV3 on the raw model. Performance was not good. It showed a moderate performance in classifying Diabetic Retinopathy with an F1 score of 90%. Retinal Detachment performed worse than that with an F1 score of 82%. Rest of the classifications were even worse. The worst one turned out to be Central Serous Chorioretinopathy with an F1 Score of 12% only.

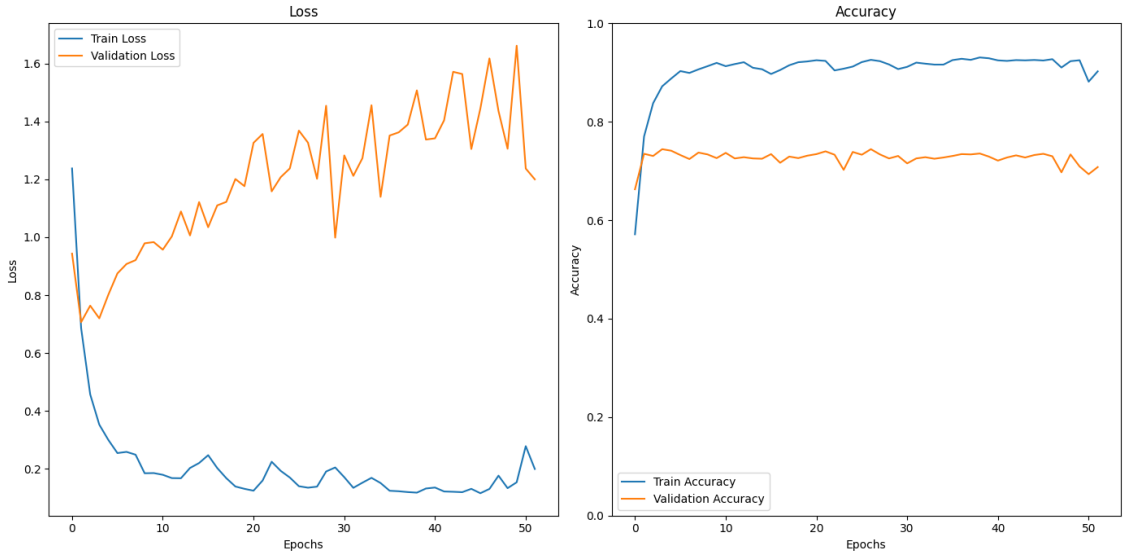


Figure 4.2 Model without Preprocessing (Raw InceptionV3) - Accuracy and Loss for Training and Validation per Epoch

4.1.2 InceptionV3

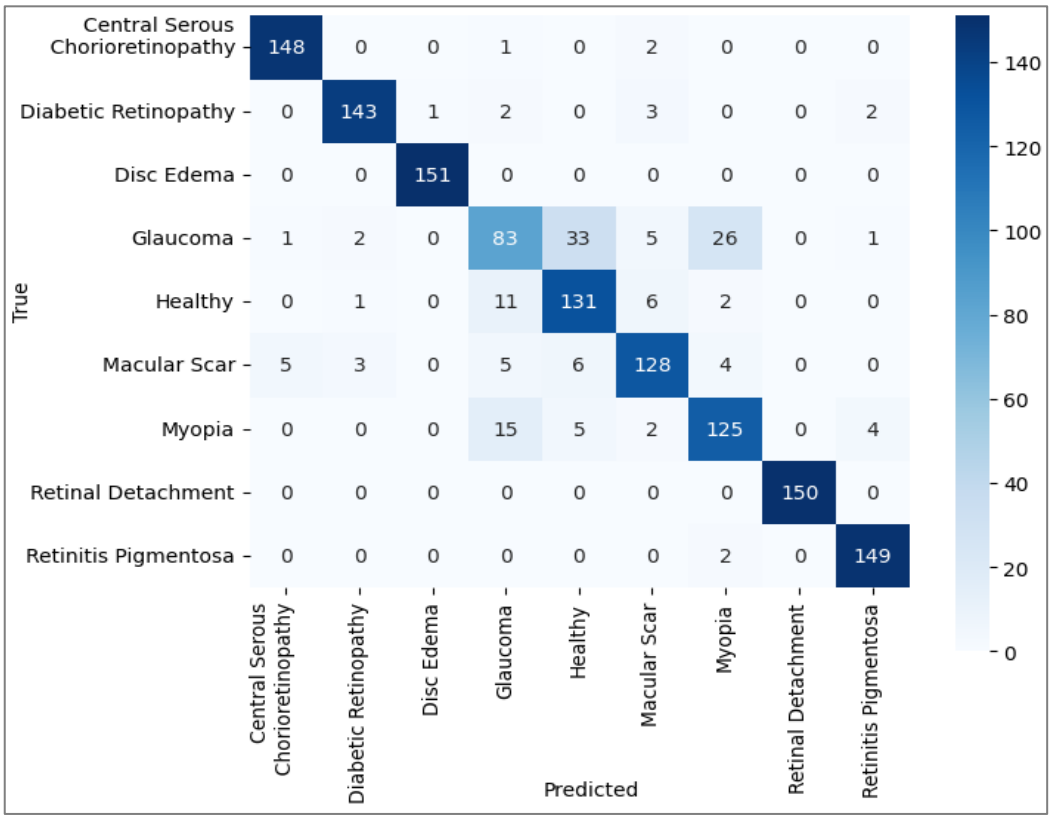


Figure 4.3 InceptionV3 Confusion Matrix

Table 4.2 InceptionV3 Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR Recall**	**FPR**	**TNR Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	148	1352	6	3	98.01%	0.44%	99.56%	1.99%	97.05%
Diabetic Retinopathy	143	1352	6	8	94.70%	0.44%	99.56%	5.30%	95.33%
Disc Edema	151	1357	1	0	100.00%	0.07%	99.93%	0.00%	99.67%
Glaucoma	83	1324	34	68	54.97%	2.50%	97.50%	45.03%	61.94%
Healthy	131	1314	44	20	86.75%	3.24%	96.76%	13.25%	80.37%
Macular Scar	128	1340	18	23	84.77%	1.33%	98.67%	15.23%	86.20%
Myopia	125	1324	34	26	82.78%	2.50%	97.50%	17.22%	80.65%
Retinal Detachment	150	1359	0	0	100.00%	0.00%	100.00%	0.00%	100.00%
Retinitis Pigmentosa	149	1351	7	2	98.68%	0.52%	99.48%	1.32%	97.07%

In InceptionV3, Retinal Detachment performed flawlessly with an F1 score of 100%. The model also successfully classified Central Serous Chorioretinopathy, Diabetic Retinopathy, Disc Edema, and Retinitis Pigmentosa with an F1 Score between 95% and 99%. The model showed an average performance while classifying Healthy, Macular Scar, and Myopia with an F1 score between 80% to 86%. The worst performing one was Glaucoma with an F1 Score of 61%.

Even in this case, the confusion matrix shows that Glaucoma and Myopia are often mixed up with Healthy and vice versa. Macular Scar was also confused with some classes, but only in small numbers.

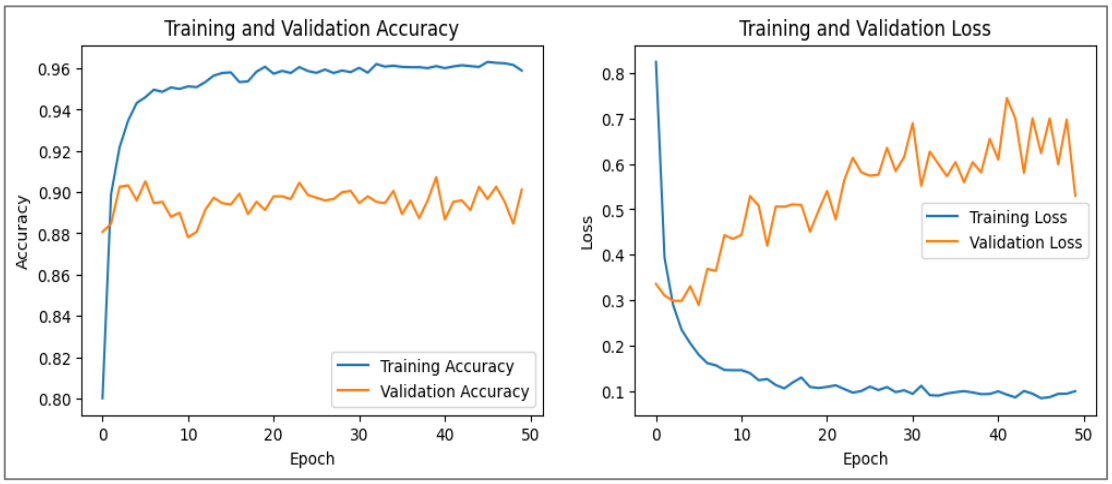


Figure 4.4 InceptionV3 - Accuracy and Loss for Training and Validation per Epoch

4.1.3 Resnet50

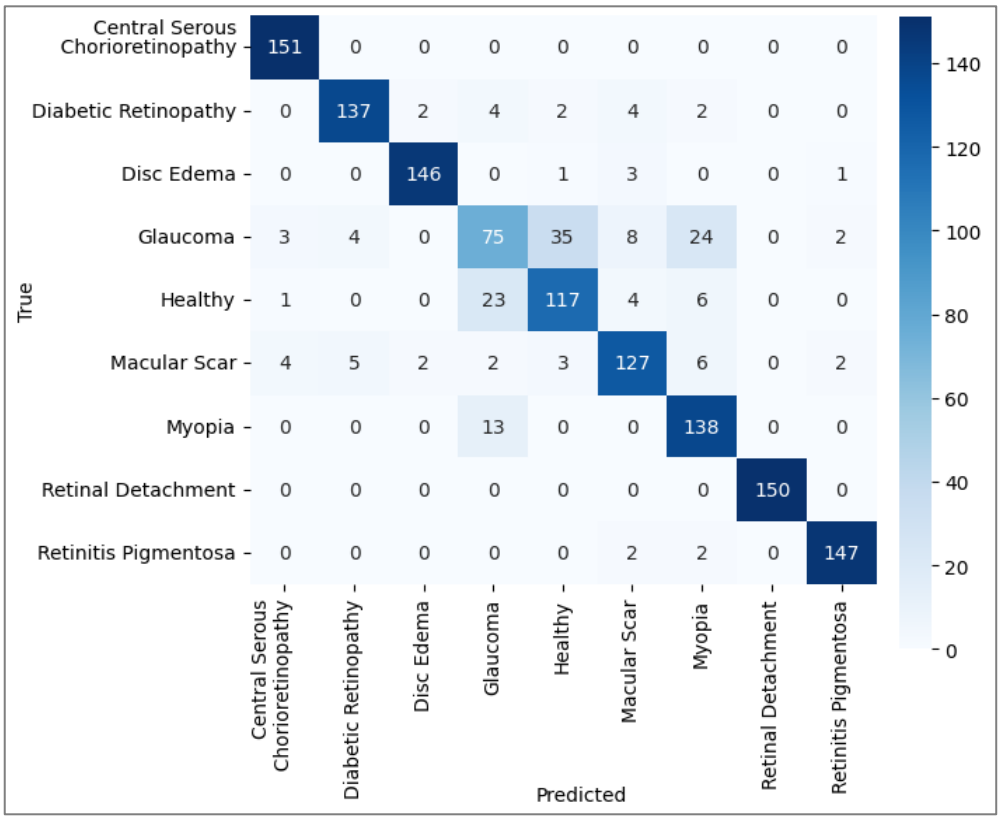


Figure 4.5 Resnet50 Confusion Matrix

Table 4.3 ResNet50 Classification Report

Class	TP	TN	FP	FN	TPR Recall	FPR	TNR Specificity	FNR	F1 Score
Central Serous Chorioretinopathy	151	1350	8	0	100%	0.59%	99.41%	0%	97.42%
Diabetic Retinopathy	137	1349	9	14	90.73%	0.66%	99.34%	9.27%	92.26%
Disc Edema	146	1354	4	5	96.69%	0.29%	99.71%	3.31%	97.01%
Glaucoma	75	1316	42	76	49.67%	3.09%	96.91%	50.33%	55.97%
Healthy	117	1317	41	34	77.48%	3.02%	96.98%	22.52%	75.73%
Macular Scar	127	1337	21	24	84.11%	1.55%	98.45%	15.89%	84.95%
Myopia	138	1318	40	13	91.39%	2.95%	97.05%	8.61%	83.89%
Retinal Detachment	150	1359	0	0	100%	0%	100%	0%	100.00%
Retinitis Pigmentosa	147	1353	5	4	97.35%	0.37%	99.63%	2.65%	97.03%

Even in ResNet50, Retinal Detachment performed flawlessly with an F1 score of 100%. The model also successfully classified Central Serous Chorioretinopathy, Disc Edema, and Retinitis Pigmentosa with an F1 score of about 97%. Diabetic Retinopathy and Myopia showed promising results with F1 scores around 91–92%. The model showed an average performance while classifying Macular Scar and Healthy with F1 scores between 75% and 85%. The worst-performing one was Glaucoma, with an F1 score of 56%.

Here too, Glaucoma and Myopia were frequently mistaken for Healthy, and Healthy was confused with them as well. Macular Scar caused a few errors, but the count stayed low.

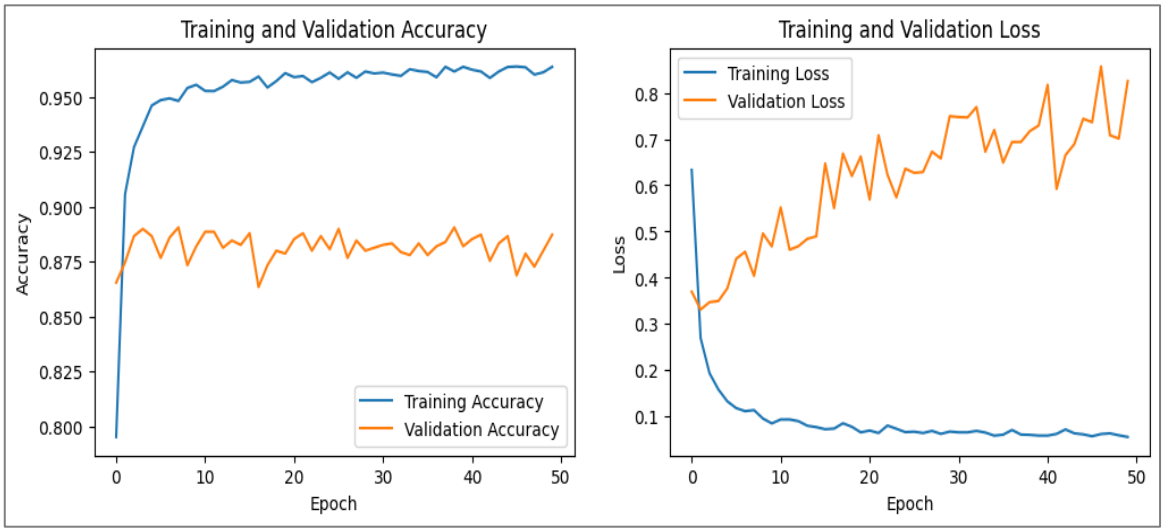


Figure 4.6 Resnet50 - Accuracy and Loss for Training and Validation per Epoch

4.1.4 MobileNetV2

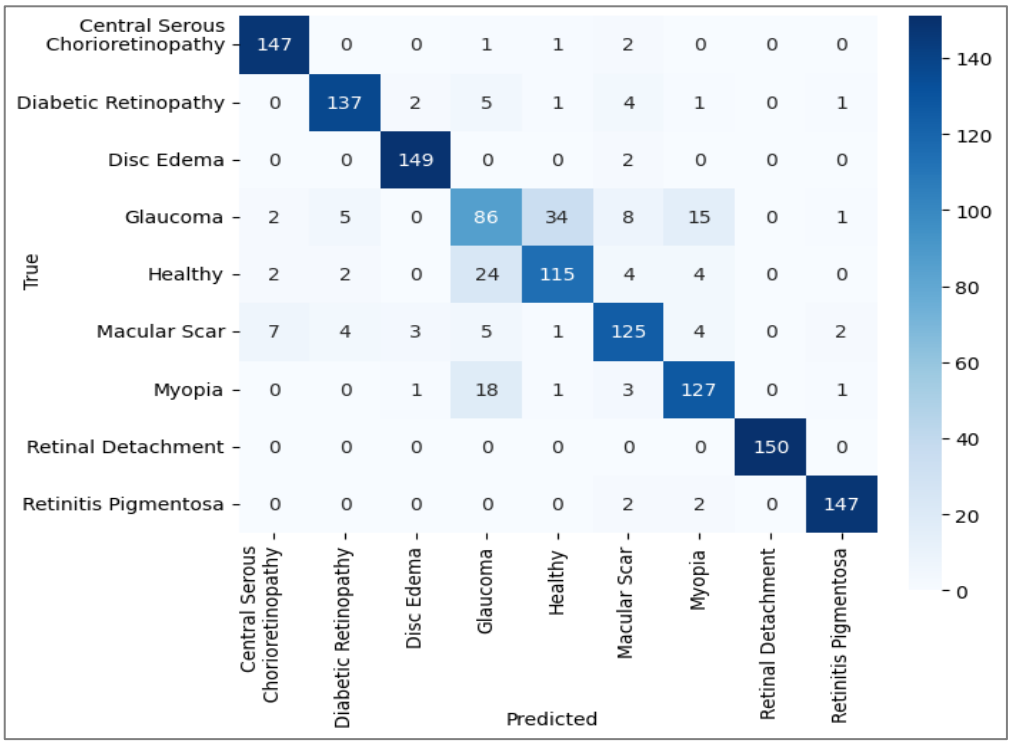


Figure 4.7 MobileNetV2 Confusion Matrix

Table 4.4 MobileNetV2 Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR Recall**	**FPR**	**TNR Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	147	1347	11	4	97.35%	0.81%	99.19%	2.65%	95.15%
Diabetic Retinopathy	137	1347	11	14	90.73%	0.81%	99.19%	9.27%	91.64%
Disc Edema	149	1352	6	2	98.68%	0.44%	99.56%	1.32%	97.39%
Glaucoma	86	1305	53	65	56.95%	3.90%	96.10%	43.05%	59.31%
Healthy	115	1320	38	36	76.16%	2.80%	97.20%	23.84%	75.66%
Macular Scar	125	1333	25	26	82.78%	1.84%	98.16%	17.22%	83.06%
Myopia	127	1332	26	24	84.11%	1.91%	98.09%	15.89%	83.55%
Retinal Detachment	150	1359	0	0	100%	0%	100%	0%	100.00%
Retinitis Pigmentosa	147	1353	5	4	97.35%	0.37%	99.63%	2.65%	97.03%

In MobileNetV2, Retinal Detachment also performed flawlessly with an F1 score of 100%. The model also successfully classified Disc Edema, Retinitis Pigmentosa, and Central Serous Chorioretinopathy with F1 scores between 95% and 97%. Diabetic Retinopathy showed good results with an F1 score of 91%. Myopia and Macular Scar were average with F1 scores around 83%. Healthy was also in the average range with a score of 76%. The worst performing one was Glaucoma with an F1 score of 59%.

Again, the confusion matrix highlights that Glaucoma and Myopia tend to overlap with Healthy in both directions. Macular Scar also made some misclassifications, though they were not many.

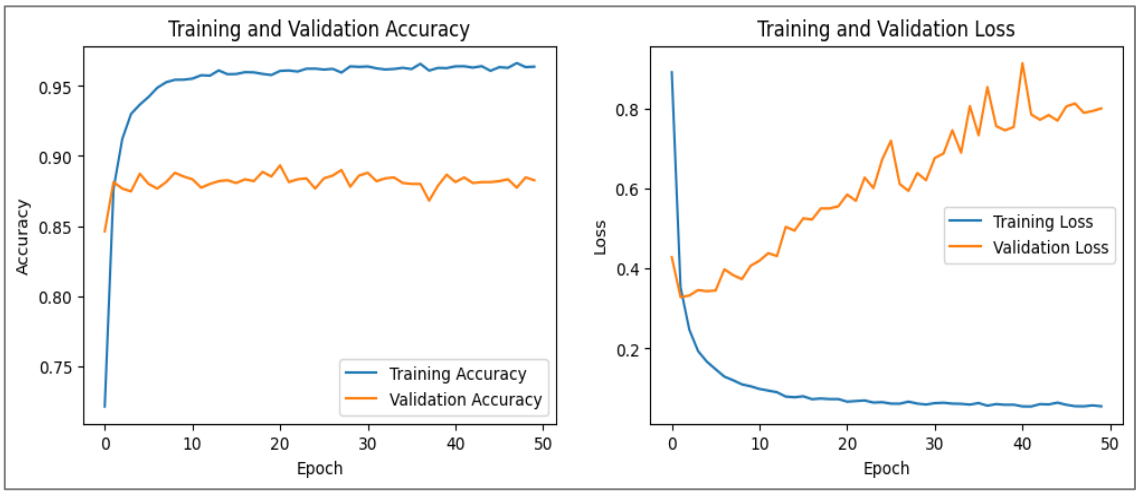


Figure 4.8 MobileNetV2 - Accuracy and Loss for Training and Validation per Epoch

4.1.5 VGG16

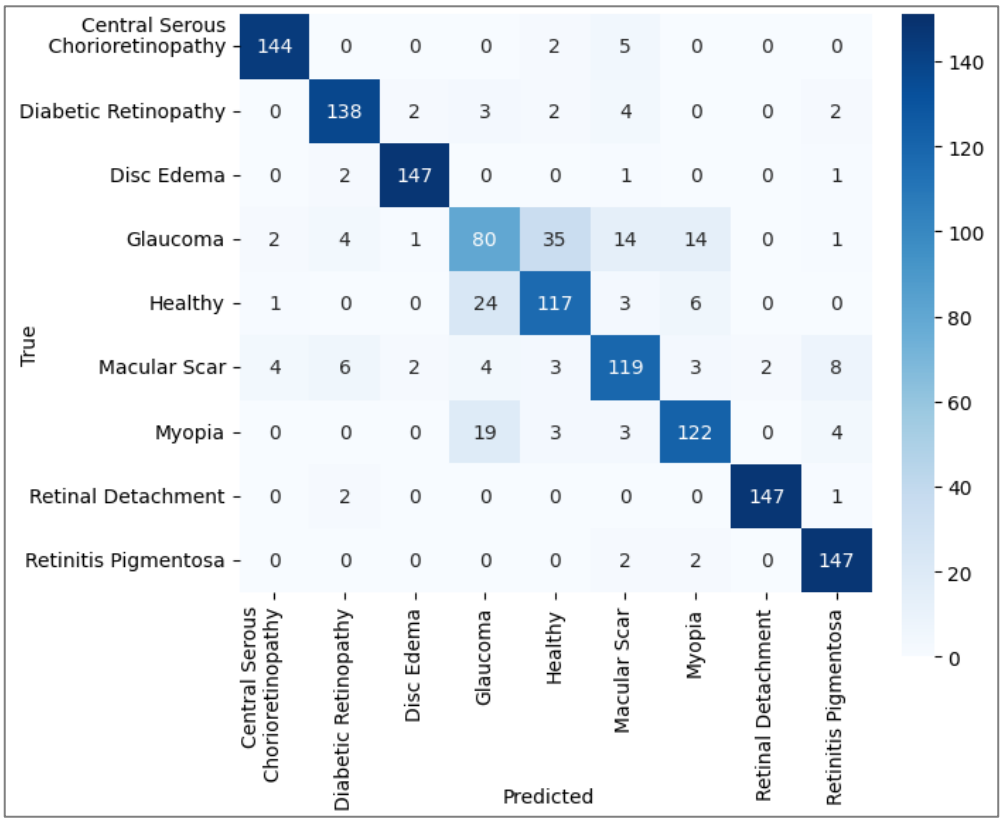


Figure 4.9 VGG16 Confusion Matrix

Table 4.5 VGG16 Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR Recall**	**FPR**	**TNR Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	144	1351	7	7	95.36%	0.52%	99.48%	4.64%	95.36%
Diabetic Retinopathy	138	1344	14	13	91.39%	1.03%	98.97%	8.61%	91.09%
Disc Edema	147	1353	5	4	97.35%	0.37%	99.63%	2.65%	97.03%
Glaucoma	80	1308	50	71	52.98%	3.68%	96.32%	47.02%	56.94%
Healthy	117	1313	45	34	77.48%	3.31%	96.69%	22.52%	74.76%
Macular Scar	119	1326	32	32	78.81%	2.36%	97.64%	21.19%	78.81%
Myopia	122	1333	25	29	80.79%	1.84%	98.16%	19.21%	81.88%
Retinal Detachment	147	1357	2	3	98.00%	0.15%	99.85%	2.00%	98.33%
Retinitis Pigmentosa	147	1341	17	4	97.35%	1.25%	98.75%	2.65%	93.33%

In VGG16, Retinal Detachment performed best with an F1 score of 98%. The model also successfully classified Disc Edema, Retinitis Pigmentosa, and Central Serous Chorioretinopathy with F1 scores between 95% and 97%. Diabetic Retinopathy showed good performance with a score of 91%. Myopia and Macular Scar were average with F1 scores around 79–82%. Healthy was also in the same range with 75%. The worst performing one was Glaucoma with an F1 score of 57%.

Here again, the confusion matrix shows a strong overlap between Glaucoma, Myopia, and Healthy. Macular Scar also appeared in wrong predictions, but very few.

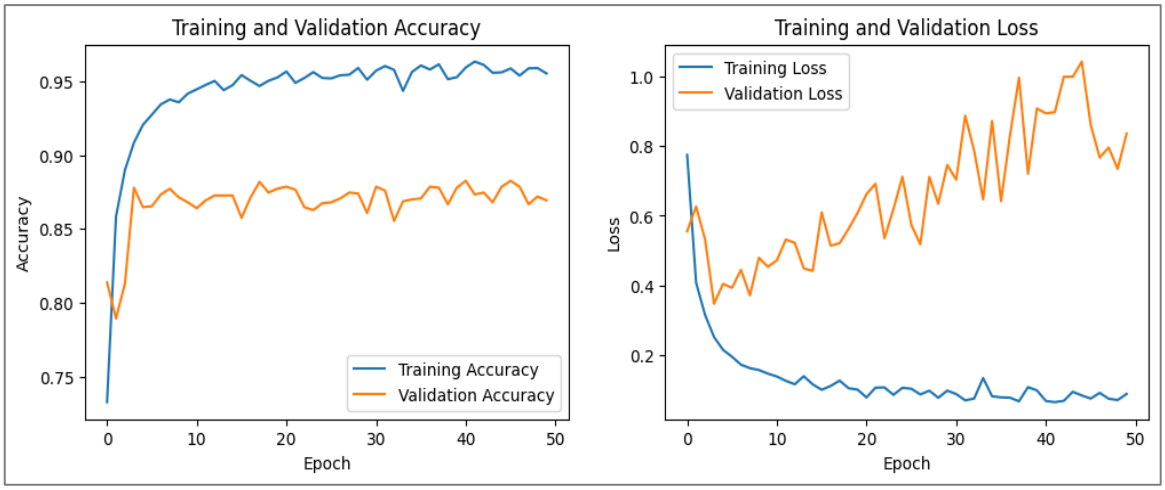


Figure 4.10 VGG16 - Accuracy and Loss for Training and Validation per Epoch

4.1.6 VGG19

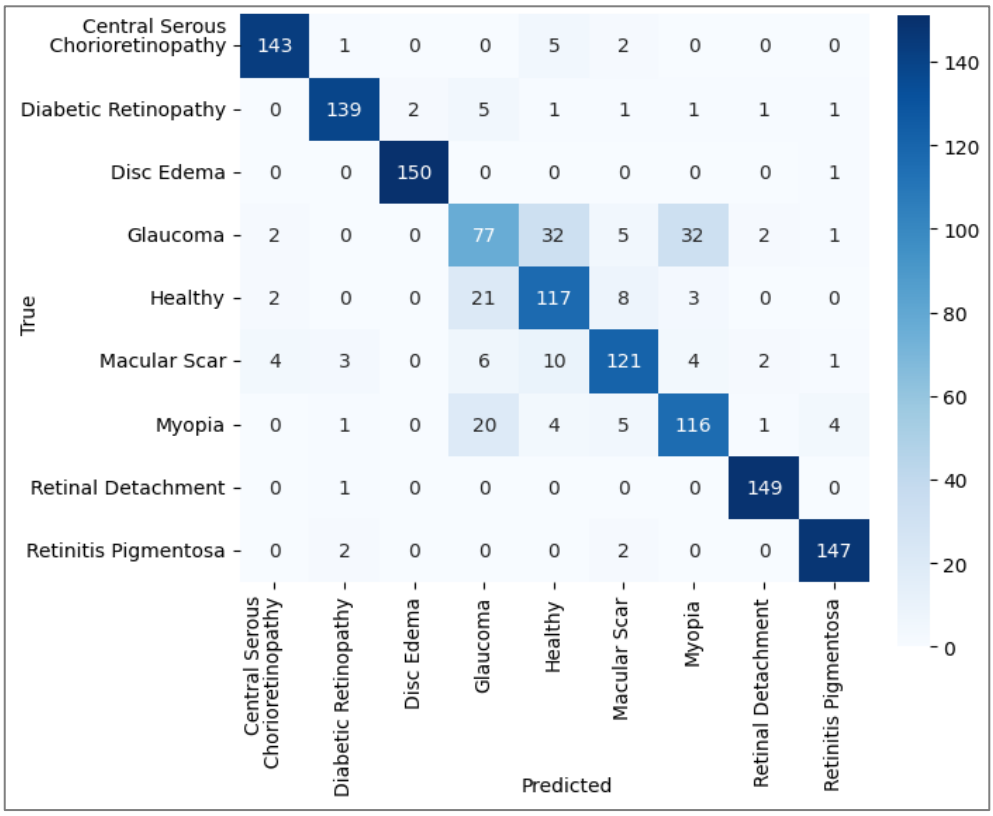


Figure 4.11 VGG19 Confusion Matrix

Table 4.6 VGG19 Classification Report

Class	TP	TN	FP	FN	TPR Recall	FPR	TNR Specificity	FNR	F1 Score
Central Serous Chorioretinopathy	143	1350	8	8	94.70%	0.59%	99.41%	5.30%	94.70%
Diabetic Retinopathy	139	1350	8	12	92.05%	0.59%	99.41%	7.95%	93.29%
Disc Edema	150	1356	2	1	99.34%	0.15%	99.85%	0.66%	99.01%
Glaucoma	77	1306	52	74	50.99%	3.83%	96.17%	49.01%	55.00%
Healthy	117	1306	52	34	77.48%	3.83%	96.17%	22.52%	73.12%
Macular Scar	121	1335	23	30	80.13%	1.69%	98.31%	19.87%	82.03%
Myopia	116	1318	40	35	76.82%	2.95%	97.05%	23.18%	75.57%
Retinal Detachment	149	1353	6	1	99.33%	0.44%	99.56%	0.67%	97.70%
Retinitis Pigmentosa	147	1350	8	4	97.35%	0.59%	99.41%	2.65%	96.08%

In VGG19, Disc Edema performed best with an F1 score of 99%. Retinal Detachment and Retinitis Pigmentosa also showed strong performance with F1 scores between 96% and 98%. Central Serous Chorioretinopathy and Diabetic Retinopathy followed with good results around 93–95%. Macular Scar, Myopia, and Healthy were average with F1 scores between 73% and 82%. The worst performing one was Glaucoma with an F1 score of 55%.

Even here, Healthy was often predicted instead of Glaucoma and Myopia, and sometimes the opposite happened. Macular Scar was also confused with other classes, but only in a few cases.

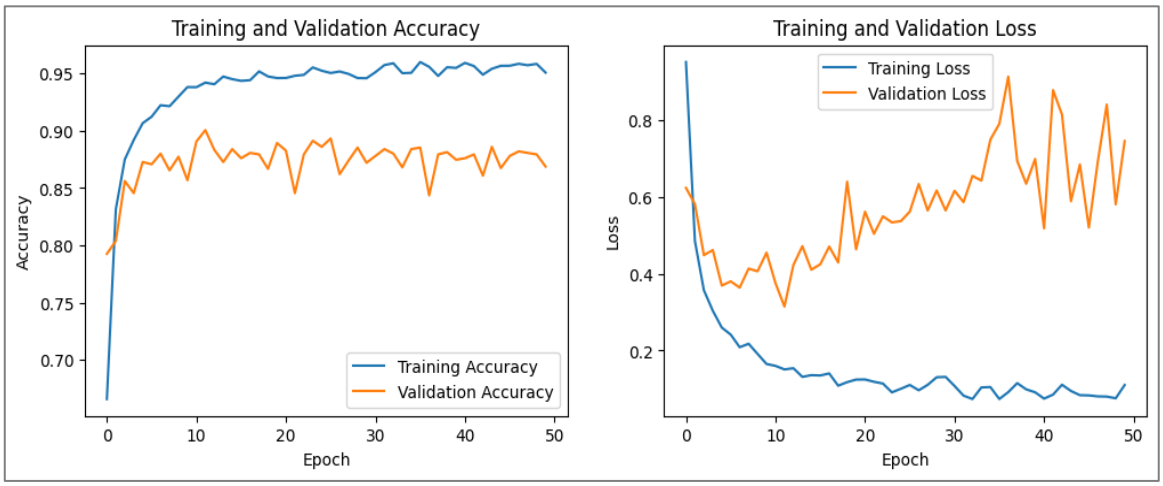


Figure 4.12 VGG19 - Accuracy and Loss for Training and Validation per Epoch

4.1.7 SwinTransformer

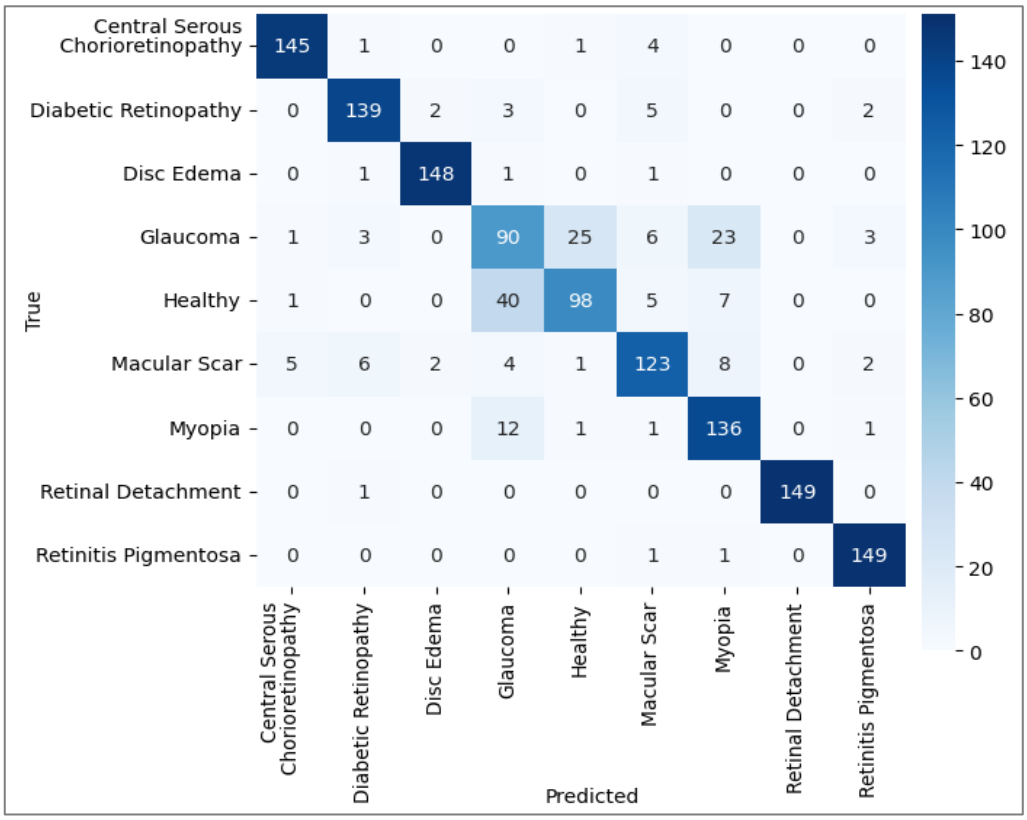


Figure 4.13 SwinTransformer Confusion Matrix

Table 4.7 SwinTransformer Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR Recall**	**FPR**	**TNR Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	145	1351	7	6	96.03%	0.52%	99.48%	3.97%	95.71%
Diabetic Retinopathy	139	1346	12	12	92.05%	0.88%	99.12%	7.95%	92.05%
Disc Edema	148	1354	4	3	98.01%	0.29%	99.71%	1.99%	97.69%
Glaucoma	90	1298	60	61	59.60%	4.42%	95.58%	40.40%	59.80%
Healthy	98	1330	28	53	64.90%	2.06%	97.94%	35.10%	70.76%
Macular Scar	123	1335	23	28	81.46%	1.69%	98.31%	18.54%	82.83%
Myopia	136	1319	39	15	90.07%	2.87%	97.13%	9.93%	83.44%
Retinal Detachment	149	1359	0	1	99.33%	0%	100%	0.67%	99.67%
Retinitis Pigmentosa	149	1350	8	2	98.68%	0.59%	99.41%	1.32%	96.75%

In SwinTransformer, Retinal Detachment performed best with an F1 score of 99%. The model also successfully classified Disc Edema, Retinitis Pigmentosa, and Central Serous Chorioretinopathy with F1 scores between 95% and 98%. Diabetic Retinopathy followed with a good score of 92%. Myopia and Macular Scar were average performers with F1 scores around 82–83%. Healthy was weaker with 71%. The worst performing one was Glaucoma with an F1 score of 60%.

As we see here, Glaucoma and Myopia repeatedly confused themselves with Healthy and vice versa. Macular Scar created some confusion too, but in small numbers.

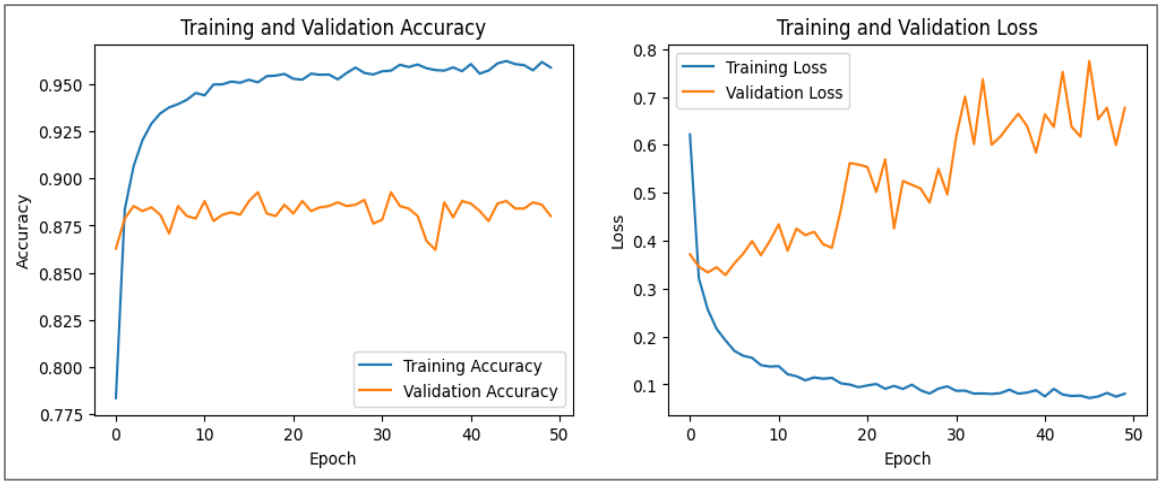


Figure 4.14 SwinTransformer - Accuracy and Loss for Training and Validation per Epoch

4.1.8 VisionTransformerB16

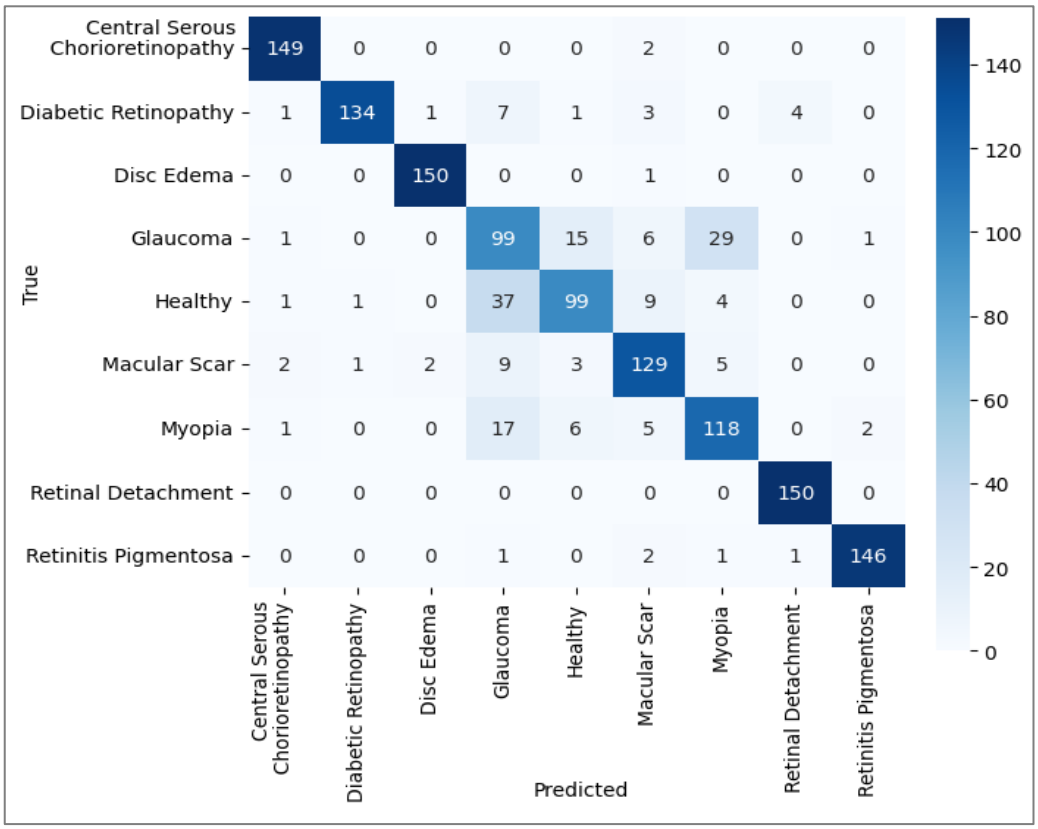


Figure 4.15 VisionTransformerB16 Confusion Matrix

Table 4.8 VisionTransformerB16 Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR Recall**	**FPR**	**TN /Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	149	1352	6	2	98.68%	0.44%	99.56%	1.32%	97.39%
Diabetic Retinopathy	134	1356	2	17	88.74%	0.15%	99.85%	11.26%	93.38%
Disc Edema	150	1355	3	1	99.34%	0.22%	99.78%	0.66%	98.68%
Glaucoma	99	1287	71	52	65.56%	5.23%	94.77%	34.44%	61.68%
Healthy	99	1333	25	52	65.56%	1.84%	98.16%	34.44%	72.00%
Macular Scar	129	1330	28	22	85.43%	2.06%	97.94%	14.57%	83.77%
Myopia	118	1319	39	33	78.15%	2.87%	97.13%	21.85%	76.62%
Retinal Detachment	150	1354	5	0	100%	0.37%	99.63%	0%	98.36%
Retinitis Pigmentosa	146	1355	3	5	96.69%	0.22%	99.78%	3.31%	97.33%

In VisionTransformerB16, Retinal Detachment performed flawlessly with an F1 score of 100%. Disc Edema also showed excellent performance with 99%. Central Serous Chorioretinopathy and Retinitis Pigmentosa followed closely with F1 scores around 97%. Diabetic Retinopathy showed good results with 93%. Macular Scar was average with 84%. Myopia and Healthy had weaker results with F1 scores between 72% and 77%. The worst performing one was Glaucoma with an F1 score of 62%.

Even in this confusion matrix, Glaucoma and Myopia are often classified as Healthy, and Healthy is misclassified as them. Macular Scar was also involved, but its errors were limited.

Figure 4.16 VisionTransformerB16 - Accuracy and Loss for Training and Validation per Epoch

4.1.9 OCT-AttenNet

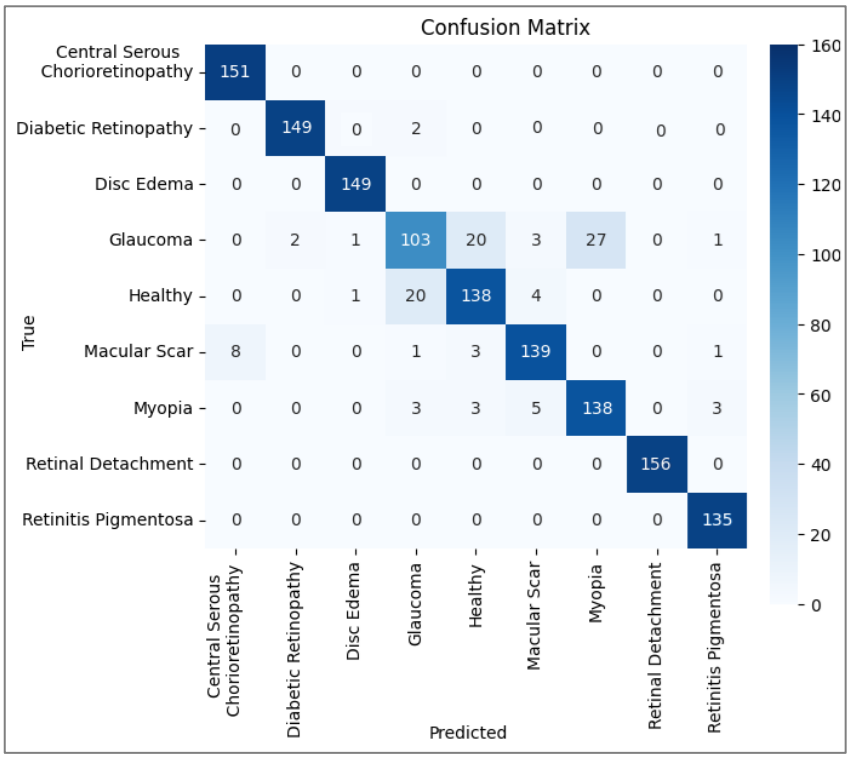


Figure 4.17 OCT-AttenNet Confusion Matrix

Table 4.9 OCT-AttenNet Classification Report

Class	**TP**	**TN**	**FP**	**FN**	**TPR/ Recall**	**FPR**	**TNR Specificity**	**FNR**	**F1 Score**
Central Serous Chorioretinopathy	149	1352	6	2	98.68%	0.44%	99.56%	1.32%	97.39%
Diabetic Retinopathy	134	1356	2	17	88.74%	0.15%	99.85%	11.26%	93.38%
Disc Edema	150	1355	3	1	99.34%	0.22%	99.78%	0.66%	98.68%
Glaucoma	99	1287	71	52	65.56%	5.23%	94.77%	34.44%	61.68%
Healthy	99	1333	25	52	65.56%	1.84%	98.16%	34.44%	72.00%
Macular Scar	129	1330	28	22	85.43%	2.06%	97.94%	14.57%	83.77%
Myopia	118	1319	39	33	78.15%	2.87%	97.13%	21.85%	76.62%
Retinal Detachment	150	1354	5	0	100.00%	0.37%	99.63%	0.00%	98.36%
Retinitis Pigmentosa	146	1355	3	5	96.69%	0.22%	99.78%	3.31%	97.33%

In OCT-AttenNet, Retinal Detachment performed flawlessly with an F1 score of 100%. Disc Edema also showed excellent performance with 99%. Central Serous

Chorioretinopathy and Retinitis Pigmentosa followed with strong F1 scores around 97%. Diabetic Retinopathy showed good results with 93%. Macular Scar was average with 84%. Myopia and Healthy had weaker results with F1 scores between 72% and 77%. The worst performing one was Glaucoma with an F1 score of 62%.

Once again, Glaucoma and Myopia were often confused with Healthy in the confusion matrix. But this is the best so far.

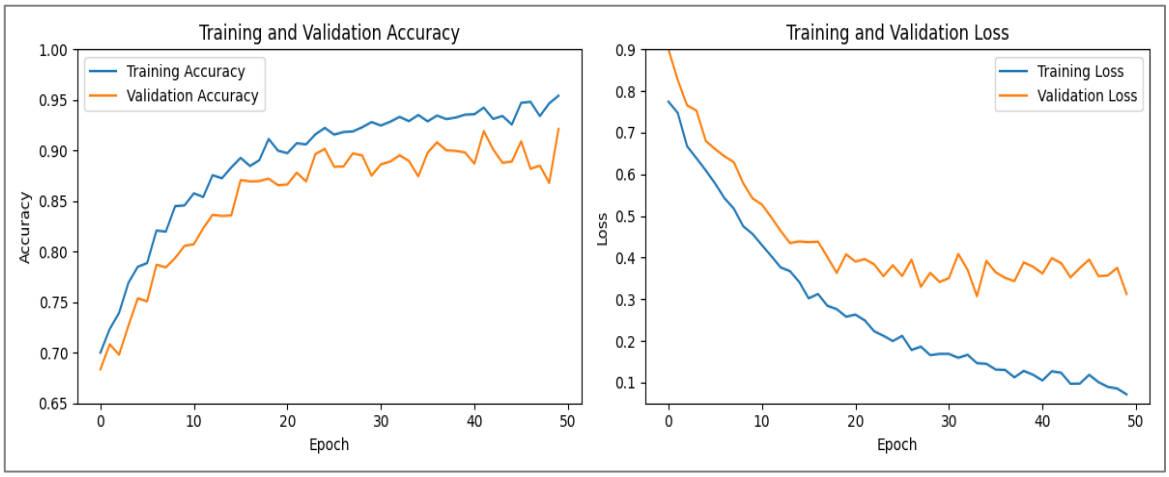
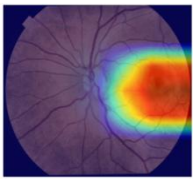
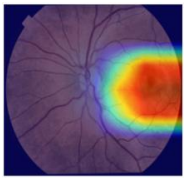
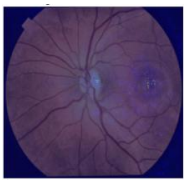
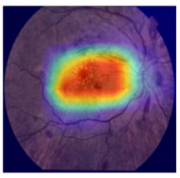
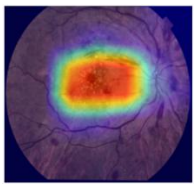
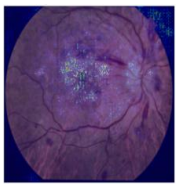
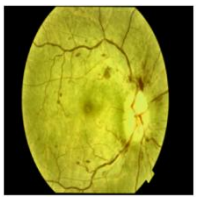
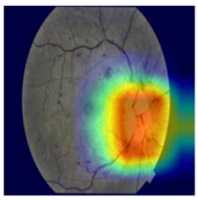
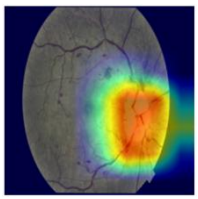
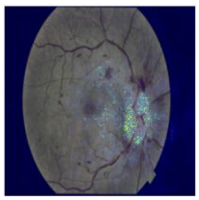


Figure 4.18 OCT-AttenNet - Accuracy and Loss for Training and Validation per Epoch

4.2 Explainable AI Analysis

We applied GradCAM (Gradient-weighted Class Activation Mapping), GradCAM++, and Integrated Gradients (IG) on our test images. We tested these XAI on our proposed model OCT-AttenNet. The outputs are promising. The outputs of both GradCAM and GradCAM++ are very similar. These XAI boldly highlighted the identified regions. IG, on the other hand, placed dots on the detected regions. Its output is also aligned with the GradCAM outputs. But since GradCAM covers a big region, IG outputs look more precise sometimes. However, their outputs differed sometimes, too. But slightly off the region. Not any major differences. We can see that the model has been able to properly identify right regions for Diabetic Retinopathy, Disc Edema, Retinal Detachment and Retinitis Pigmentosa, Central Serous Chorioretinopathy, and Macular Scar. However, with Myopia, and healthy, the prediction looks random. There reason for classification is not clear. And Glaucoma's heatmap also doesn't touch the right region, which is the optic disc.

Table 4.10 Preview of XAI applied on Models

Class	Original	GradCAM	GradCAM++	IG
Central Serous Chorioretinopathy				
Diabetic Retinopathy				
Disc Edema				

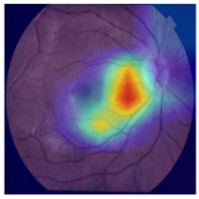
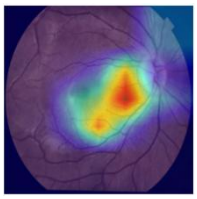
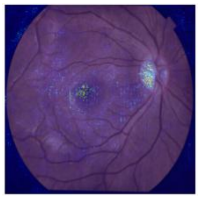
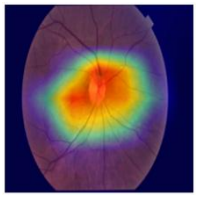
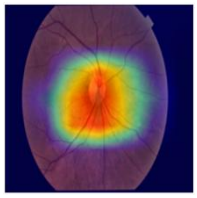
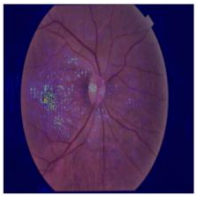
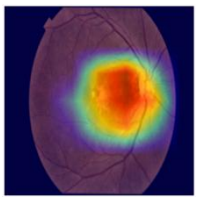
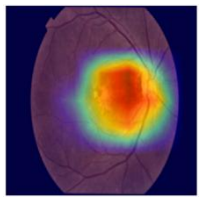
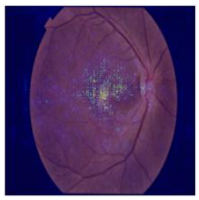
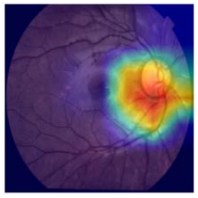
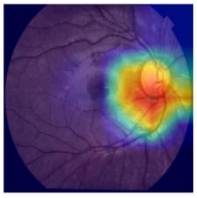
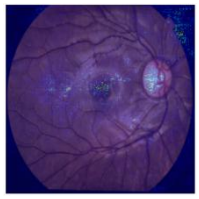
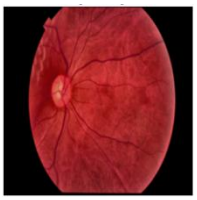
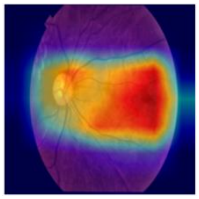
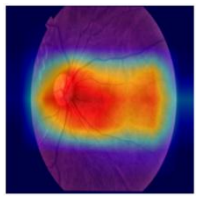
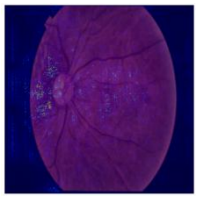
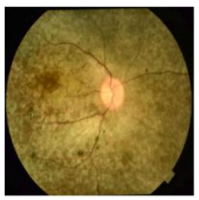
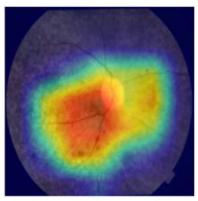
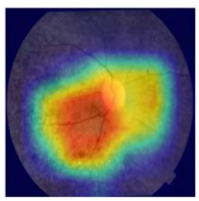
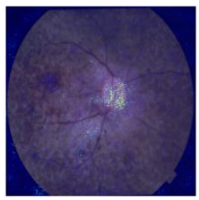
Glaucoma				
Healthy				
Macular Scar				
Myopia				
Retinal Detachment				
Retinitis Pigmentosa				

Figure 4.19 Preview of XAI including GradCam, GradCam++, Integrated Gradients (IG) applied on all classes.

4.3 Discussion

Table 4.11 All models comparison summary

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Model without Preprocessing – InceptionV3	73.5	74.4	73.5	73
InceptionV3	90	90	90	90
ResNet50	89	88	89	88
MobileNetV2	88	88	88	88
VGG16	87	87	87	87
VGG19	87	87	87	87
SwinTransformer	88	88	88	88
VisionTransformerB16	88	88	88	88
OCT-AttenNet	92	92.3	92	92.15

From the comparison of all models, we can clearly see that OCT-AttenNet outperformed all other models with an F1 Score and accuracy of 92%. OCT-AttenNet has clearly a superiority over the 2nd best model, InceptionV3. InceptionV3 has an accuracy of 90% which is also the base model for OCT-AttenNet. It outperforms the base model by 2% in accuracy. ResNet50 comes 3rd in terms of an accuracy of 89%. Both Transformers, VisionTransformerB16, and SwinTransformer, achieved an accuracy of 88% which is the same as MobileNetV2. The worst performing ones turned out to be VGG16 and VGG19 with an 87% accuracy. And also, the comparison shows the impact of preprocessing on the performance of the models. Without preprocessing, InceptionV3 achieved an accuracy of 73%, which is 17% less than InceptionV3 trained with a preprocessed dataset. In each model, the accuracy, precision, recall, and F1 score are almost the same. This shows a stable performance across all models.

OCT-AttenNet performed better than its core InceptionV3. Because OCT-AttenNet knows which region is more critical and what channel to look at, thanks to its BAM module, which contains both spatial attention and channel attention. They helped to differentiate between similar diseases.

We can see that the model perfectly identified Retinal Detachments with no mistakes. It also did great in detecting Diabetic Retinopathy, Disc Edema, and Retinitis Pigmentosa. Central Serous Chorioretinopathy and Macular Scar also did well. However, the performance dropped when it comes to Glaucoma and Myopia. The model often confuses both diseases with Healthy.

The primary reason behind the incorrect classifications of glaucoma is that only the optic disc is considered when detecting glaucoma. And when the size of the image is reduced to 299x299 or 224x224, the size of the optic disc decreases significantly. Thus, a lot of features are lost. Especially with glaucoma images of the early stages. Hence, the model fails to classify early-stage glaucoma images accurately.

The case with myopia is a bit different. Sometimes myopia is caused by the shape of the eyeball or lens. This can be identified by an autorefractor machine and not by OCT. Again, there is another type of myopia called Pathological Myopia. Different diseases or conditions inside the eye cause it. It can be identified through OCT images. Here, our dataset does not mention whether an image represents refractive myopia or pathological myopia. So the model is confusing refractive myopia with healthy, which is the main cause behind the model giving false positives and false negatives to some of the myopia images.

CHAPTER 5

CONCLUSION

5.1 Findings & Contributions

In this research, we experimented various CNN and transformers on a local 9 class eye disease dataset. We found that InceptionV3 performed the best on our dataset. We developed a Hybrid CNN with Attention Mechanism based on Inception, and BAM and ECA for Multiclass Eye Disease Detection with 92% F1 Score. It managed to beat the original InceptionV3 by 2% improvement in accuracy and F1 score. Our model was trained with eye diseases found in Bangladesh. So that model appropriate for deployment in developing country like Bangladesh. Our Model consist of 8 disease, mostly retinal disease. All of them are internal eye diseases which most people won't be able to recognize. We also augmented to increase the weight of rare disease for AI detection. And 3 XAI including GradCAM, GradCAM++, Integrated Gradients (IG) has been implemented, so that the model becomes reliable by clinicians. This AI model will help to prevent permanent blindness through early detection.

5.2 Limitations

This dataset for this research only covers images collected from two hospitals in Faridpur District, Bangladesh, which may not reflect the entire scenario of Bangladesh. The dataset has a class name "Myopia" instead of separating "Pathological Myopia" and "Refractive Myopia". Since OCT images can't distinguish between Refractive myopia and healthy, it will confuse the model. The main bottleneck of our model is in the early stages of glaucoma. Only the optic disc is relevant to glaucoma detection. Therefore, the remaining features also confuse the model, causing it to pay less attention to the optic disc and its features. This research compares traditionally best models, and not state-of-the-art models. The study also doesn't show how it can be implemented in hospitals or in rural areas.

5.3 Recommendations for Future Work

There is room for improvement everywhere, including models, datasets, approaches, etc. The model can work better if pathological myopia is separated from the rest of the myopia. Refractive myopia can't be detected through OCT, so it should be removed. Feature extractions can be done for the optic disc and other parts of images, and train them separately. This makes the model pay more attention to details. It will also enhance the detection of diseases such as glaucoma. Also, instead of limiting the dataset to just two hospitals, data from all over the country should be tested. State-of-the-art models can be tested on this OCT dataset. Further research work can be done on the implementation of OCT models in Bangladesh

CHAPTER 6

REFERENCES

1. Khan, A. J., Islam, M. E., & Bari, M. A. (2003). Prevalence and causes of blindness and visual impairment in Bangladeshi adults: Results of the National Blindness and Low Vision Survey of Bangladesh. British Journal of Ophthalmology, 87(7), 808-812.

2. Frederiksen, I. N., Arnold-Vangsted, A., Anguita, R., Boberg-Ans, L. C., Skovgaard Eriksen, N., Ferro Desideri, L., Huemer, J., Iovino, C., Künzel, S. E., Ørskov, M., Pauleikhoff, L. J. B., & Roed Rasmussen, M. L. (2025). Global incidence of central serous chorioretinopathy: A systematic review, meta-analysis, and forecasting study. Ophthalmology and Therapy, 14(10), 2443-2467.

3. Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C.-Y. (2014). Global prevalence of glaucoma and projections to 2040. Ophthalmology, 121(11), 2081-2090.

4. Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K. S., Sankaridurg, P., Wong, T. Y., Naduvilath, T., & Resnikoff, S. (2016). Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology, 123(5), 1036-1042.

5. Teo, Z. L., Tham, Y.-C., Yu, M., & Wong, T. Y. (2021). Global prevalence of diabetic retinopathy and projection. Ophthalmology, 128(5), 1580-1591.

6. Saraf, S. S., Patel, P. J., & Wong, E. Y. L. (2022). Demographics and seasonality of retinal detachment: A review. Ophthalmology Retina, 6(1), 5-10.

7. Reier, L., Smith, K., & Jones, A. (2022). Optic disc edema and elevated intracranial pressure: A review. Cureus, 14(5), e25423.

8. Daniel, E., Martin, D. F., Maguire, M. G., Fine, S. L., Jaffe, G. J., Grunwald, J. E., Toth, C. A., Huang, J., Ying, G.-S., Hagstrom, S. A., & Ferris, F. L. (2014). Risk of scar in the Comparison of Age-Related Macular Degeneration Treatments Trials. Ophthalmology, 121(1), 150-157.

9. Suleman, N., Khan, H., Ahmed, M., & Yousuf, M. (2025). Current understanding on retinitis pigmentosa: A literature review. Frontiers in Ophthalmology, 5, Article 1600283.

10. Li, N., Li, T., Hu, C., Wang, K., & Kang, H. (2021). A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Revised Selected Papers (pp. 177–193). Springer International Publishing.

11. Gour, N., & Khanna, P. (2021). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. Biomedical Signal Processing and Control, 66, 102329.

12. Sarki, F. A., Ahmed, K., Wang, H., & Zhang, Y. (2021). Convolutional neural network for multi-class classification of diabetic eye disease. Medical & Biological Engineering & Computing, 59(1), 103–115.

13. Saini, M., & Susan, S. (2022). Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets. Computers in Biology and Medicine, 149, 105989.

14. Rodríguez, M. A., AlMarzouqi, H., & Liatsis, P. (2022). Multi-label retinal disease classification using transformers. IEEE Journal of Biomedical and Health Informatics, 27(6), 2739–2750.

15. Bhati, A., Gour, N., Khanna, P., & Ojha, A. (2023). Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. Computers in Biology and Medicine, 153, 106519.

16. Niloy, G. M., Sammak, M. H., Bitto, A. K., Das, A., Biplob, K. B. M., & Hridoy, G. G. (2024, March). MobileNet-Eye: An efficient transfer learning for eye disease classification. In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) (pp. 1–6). IEEE.

17. Al-Fahdawi, H. M., Al-Timemy, A. H., Escudero, J., & Al-Mousa, A. (2024). Fundus-DeepNet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases. Computers in Biology and Medicine, 161, 106887.

18. AlBalawi, T., Aldajani, M. B., Abbas, Q., & Daadaa, Y. (2024). IoT-Opthom-CAD: IoT-enabled classification system of multiclass retinal eye diseases using dynamic Swin Transformers and explainable artificial intelligence. International Journal of Advanced Computer Science and Applications, 15(7), 453–461.

19. Maehara, H., Ueno, Y., Yamaguchi, T., Kitaguchi, Y., Miyazaki, D., Nejima, R., ... & Oshika, T. (2025). Artificial intelligence support improves diagnosis accuracy in anterior segment eye diseases. NPJ Digital Medicine, 8, 55.

20. Gencer, G., & Gencer, K. (2025). Advanced retinal disease detection from OCT images using a hybrid squeeze and excitation enhanced model. PLOS ONE, 20(2), e0318657.

21. Abdullah, A. A., Aldhahab, A., & Al Abboodi, H. M. (2025). Eye disease classification based on hybrid deep features with principal component analysis and blending ensemble learning. International Journal of Intelligent Engineering and Systems, 18(6), 179–191.

22. Vidivelli, S., Padmakumari, P., Parthiban, C., DharunBalaji, A., Manikandan, R., & Gandomi, A. H. (2025). Optimising deep learning models for ophthalmological disorder classification. Computers in Biology and Medicine, 163, 107100.

23. Hasan, M. N., Rabbi, M. E., Das, S., Siddique, N., & Wang, H. (2025). DIA-VXNET: A framework for automated diabetic eye disease detection using transfer learning with feature fusion network. Biomedical Signal Processing and Control, 87, 105500.

24. Albelaihi, A., Ibrahim, D. M., & Hussein, D. (2025). RNN diabetic framework for identifying diabetic eye diseases. Indonesian Journal of Electrical Engineering and Computer Science, 37(3), 1830–1844.

25. Sharmin, S., Rashid, M. R., Khatun, T., Hasan, M. Z., & Uddin, M. S. (2024). A dataset of color fundus images for the detection and classification of eye diseases. Data in Brief, 57, 110979.

26. Park, J., Woo, S., Lee, J. Y., & Kweon, I. S. (2018). Bam: Bottleneck attention module. arXiv preprint arXiv:1807.06514.

27. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11534-11542).

APPENDICES

Appendix A: Dataset Availability

The publicly available retinal OCT image dataset used in this study can be accessed from the Mendeley Data repository at the following link:

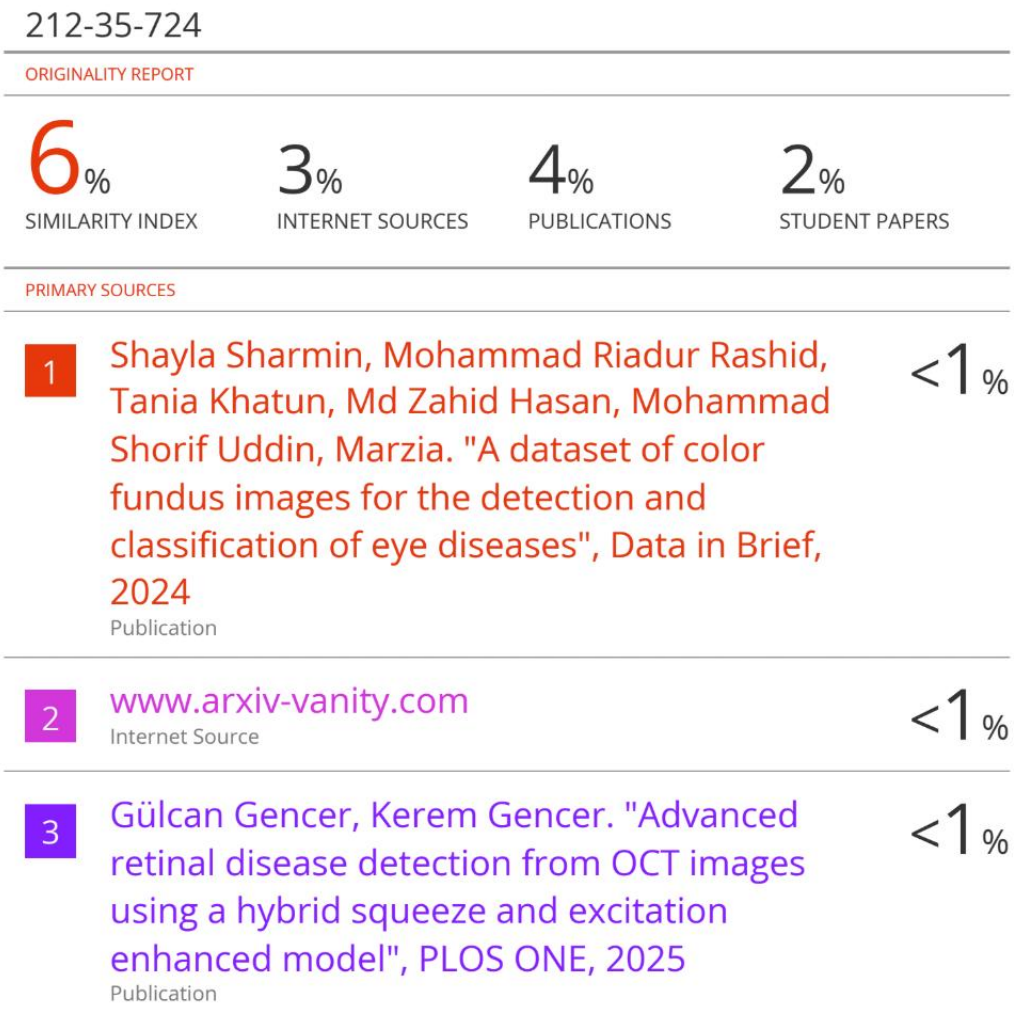
https://data.mendeley.com/datasets/s9bfhswzjb/1

Appendix B: Code Availability

The source code used to implement the OCT-AttnNet model and conduct experiments is openly available on GitHub:

https://github.com/ashikur-rahman-shad/eye-disease-detection/

PLAGIARISM REPORT



ACCOUNT CLEARANCE