Faculty of Computer Science and Information Technology

Improved Multiclass Classification of Eye Diseases using a Feature-Augmented Enhanced Deep Learning Approach

Alvin Choo Ming Siang

Master of Science
2026

Improved Multiclass Classification of Eye Diseases using a Feature-Augmented Enhanced Deep Learning Approach

Alvin Choo Ming Siang

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Visual Image Processing)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2026

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

……

Signature

Name: Alvin Choo Ming Siang

Matric No.: 23020166

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date: 30.09.2025

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest appreciation to my research supervisor, Associate Professor Dr Stephanie Chua Hui Li, for her continuous guidance, encouragement, motivation, and insightful feedback that she has shared with me throughout the course of this research. I am equally thankful to my co-supervisors, Professor Dr Dayang Nurfatimah Awang Iskandar and Professor Dr Lim Lik Thai, for their valuable support, constructive suggestions, and academic mentorship, which have been instrumental in shaping the direction and quality of this thesis.

I would also like to sincerely deliver my appreciation to Ministry of Higher Education, Malaysia (MOHE), for funding this research under the Fundamental Research Grant [FRGS/1/2023/ICT02/UNIMAS/02/1]. Besides, I would like to thank Universiti Malaysia Sarawak (UNIMAS), the Centre for Graduate Studies, and Faculty of Computer Science and Information Technology, for providing the facilities, resources, and academic environment necessary to carry out this work.

Last but not least, I am forever grateful to my beloved family, especially my parents and friends for their unwavering support, patience and encouragement throughout this journey. Their belief in me has been a constant source of strength and motivation in completing this research.

ABSTRACT

Vision impairment is a global health issue often caused by conditions such as cataracts, diabetic retinopathy, and glaucoma. Early detection is critical to prevent irreversible vision loss. Retinal fundus imaging plays a central role in diagnosis, and deep learning offers promising automation for disease detection. However, multiclass eye disease classification remains challenging due to limited annotated datasets, overlapping clinical features and intra-class heterogeneity. This research proposes an enhanced deep learning pipeline for classifying retinal fundus images into four classes: cataracts, diabetic retinopathy, glaucoma, and normal. A dataset, named CDGN, was constructed by integrating and standardizing images from eight publicly available sources, offering improved diversity, resolution, and patient demographics to enhance model robustness and generalizability. Image enhancement was applied to make disease-specific feature more pronounced, while attention mechanisms improved focus on relevant regions, and ensemble learning further boosted performance across heterogeneous data. Multiple convolutional neural network (CNN) architectures were explored through transfer learning. An ablation study quantified the individual contributions of image enhancement, attention, and ensemble learning. Experimental results demonstrate progressive improvements in accuracy, recall, precision, F1-score and AUC across enhancement stages, with the ensemble model achieving the highest performance. These findings indicate that the feature-enhanced deep learning pipeline effectively addresses challenges in multiclass eye disease classification, supporting clinical decision-making and advancing automated diagnostic systems in ophthalmology.

Keywords: Multiclass eye disease classification, retinal fundus images, convolutional neural networks, attention mechanism, ensemble learning

Penambahbaikan Pengelasan Berbilang Kelas Penyakit Mata Menggunakan Pendekatan Pembelajaran Mendalam yang Dipertingkatkan melalui Pengayaan Ciri

ABSTRAK

Kecacatan penglihatan merupakan isu kesihatan global yang sering disebabkan oleh keadaan seperti katarak, retinopati diabetik, dan glaukoma. Pengesanan awal adalah penting bagi mengelakkan kehilangan penglihatan yang tidak boleh dipulihkan. Imejan fundus retina memainkan peranan penting dalam diagnosis, dan pembelajaran mendalam menunjukkan potensi dalam pengesanan penyakit secara automatik. Walau bagaimanapun, klasifikasi penyakit mata pelbagai kelas masih mencabar disebabkan oleh set data beranotasi yang terhad, ciri klinikal yang serupa atau bertindih, serta heterogeniti dalam kelas. Kajian ini mencadangkan satu rangka kerja pembelajaran mendalam yang dipertingkatkan bagi mengklasifikasikan imej fundus retina kepada empat kelas: katarak, retinopati diabetik, glaukoma, dan normal. Satu set data baharu yang dinamakan CDGN telah dibina dengan menggabungkan dan menyeragamkan imej daripada lapan sumber awam yang tersedia, memberikan kepelbagaian, resolusi, dan demografi pesakit yang lebih baik untuk meningkatkan ketahanan dan keupayaan generalisasi model. Penambahbaikan imej telah digunakan untuk menonjolkan ciri-ciri penyakit yang spesifik, manakala mekanisme perhatian meningkatkan fokus pada kawasan imej yang relevan, dan pembelajaran ensembel seterusnya meningkatkan prestasi data yang heterogen. Pelbagai seni bina rangkaian neural konvolusi (CNN) telah diterokai melalui pemindahan pembelajaran. Kajian ablasi dijalankan untuk menilai sumbangan individu penambahbaikan imej, mekanisme perhatian, dan pembelajaran ensembel. Keputusan eksperimen menunjukkan peningkatan progresif dalam ketepatan, kepekaan (recall), ketepatan ramalan (precision), skor F1 dan AUC pada setiap peringkat penambahbaikan,

dengan model ensembel mencatatkan prestasi tertinggi. Hasil kajian ini menyerlahkan keberkesanan rangka kerja pembelajaran mendalam yang dipertingkatkan dengan ciri tambahan dalam menangani cabaran klasifikasi pelbagai kelas penyakit mata, menyokong proses menjana keputusan klinikal dan memacu pembangunan sistem diagnostik automatik dalam bidang oftalmologi.

Kata kunci: *Pengelasan berbilang kelas penyakit mata, imej fundus retina, rangkaian neural konvolusi (CNN), mekanisme perhatian, pembelajaran ensembel*

TABLE OF CONTENTS

		Page
DECLARATION		i
ACKNOWLEDGEMENT		ii
ABSTRACT		iii
ABSTRAK		iv
TABLE OF CONTENTS		vi
LIST OF TABLES		xii
LIST OF FIGURES		xiv
LIST OF ABBREVIATIONS		xix
CHAPTER 1: INTRODUCTION		1
1.1	Background	1
1.2	Motivation	3
1.3	Problem Statement	4
1.4	Research Questions	7
1.5	Research Objectives	7
1.6	Research Scope	8
1.7	Expected Outcomes	9
1.8	Significance of Research	9

1.9 Thesis Outline 11

1.10 Chapter Summary 12

CHAPTER 2: LITERATURE REVIEW 13

2.1 Overview 13

2.2 Anatomy of the Eye and Its Related Diseases 13

2.2.1 Cataracts 15

2.2.2 Diabetic Retinopathy 16

2.2.3 Glaucoma 18

2.3 Conventional Diagnostic Methods for Eye Diseases 20

2.3.1 Retinal Fundus Imaging 23

2.4 Machine Learning 24

2.4.1 Deep Learning 27

2.4.2 Convolutional Neural Network 30

2.5 Review of Related Work 52

2.5.1 Binary Classification of Eye Diseases 53

2.5.2 Multiclass Classification of Eye Diseases 59

2.6 Publicly Available Retinal Fundus Image Datasets 72

2.6.1 Eye Disease Retinal Images Dataset 73

2.6.2 Ocular Disease Intelligence Recognition (ODIR-5K) Dataset 73

2.6.3 Retinal Fundus Multi-Disease Image Dataset (RFMiD) 74

2.6.4 Glaucoma Fundus Imaging Dataset 75

2.6.5 PAPILA Dataset 76

2.6.6 DRISHTI-GS Dataset 76

2.6.7 Eye Disease Diagnosis and Fundus Synthesis (EDDFS) Dataset 77

2.6.8 1000 Fundus Images with 39 Categories Dataset 77

2.7 Comparative Analysis of Related Work 78

2.8 Chapter Summary 94

CHAPTER 3: METHODOLOGY 95

3.1 Overview 95

3.2 Research Methodology 95

3.3 Execution Environment 97

3.4 Data Collection 97

3.4.1 Dataset Selection and Integration 99

3.5 Data Preprocessing 105

3.6 Image Enhancement 106

3.7 Model Training 110

3.7.1 Transfer Learning 110

3.7.2 Attention Mechanism 116

3.7.3 Ensemble Learning 119

3.8 Modal Evaluation 121

3.8.1 Confusion Matrix and Derived Metrics 121

3.8.2 Area Under the Receiver Operating Characteristic Curve (AUC-ROC) 124

3.9 Chapter Summary 126

CHAPTER 4: RESULTS AND DISCUSSION 127

4.1 Overview 127

4.2 Experimental Setup 127

4.2.1 Dataset Description 128

4.2.2 Image Preprocessing and Enhancement 128

4.2.3 Model Architectures 128

4.2.4 Training Configuration 129

4.2.5 Evaluation Metrics 129

4.3 Model Performance on Original CDGN Dataset (Baseline) 129

4.3.1 VGG16 130

4.3.2 Inception-v3 134

4.3.3 ResNet50 138

4.3.4 DenseNet121 142

4.3.5 EfficientNet-B0 146

4.3.6 Summary and Discussion 150

4.4 Model Performance on Feature-Enhanced CDGN Dataset 152

4.4.1 VGG16 152

4.4.2 Inception-v3 156

4.4.3 ResNet50 160

4.4.4 DenseNet121 164

4.4.5 EfficientNet-B0 168

4.4.6 Summary and Discussion 172

4.5 Attention-Based Model Performance on Feature-Enhanced CDGN Dataset 174

4.5.1 Attention-Based VGG16 174

4.5.2 Attention-Based Inception-v3 178

4.5.3 Attention-Based ResNet50 182

4.5.4 Attention-Based DenseNet121 186

4.5.5 Attention-Based EfficientNet-B0 191

4.5.6 Summary and Discussion 195

4.6 Ensemble Model Performance on Feature-Enhanced CDGN Dataset 198

4.7 Comparative Evaluation of Model Configurations 203

4.7.1 Macro-Performance Summary 204

4.7.2 The Impact of Attention Mechanisms (Ablation Study) 205

4.7.3 Class-Specific Discriminative Analysis 206

4.8 Chapter Summary 208

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS 210

5.1 Overview 210

5.2 Contribution 210

5.3 Limitations 212

5.4 Recommendations and Future Works 213

5.5 Conclusion 215

REFERENCES 216

LIST OF TABLES

		Page
Table 2.1	Commonly Used Activation Functions in Deep Learning	35
Table 2.2	Summary of Reviewed Studies on Eye Disease Classification Using CNNs	81
Table 3.1	Hardware and Software Specifications	98
Table 3.2	Summary of Publicly Available Datasets Prior to Filtering	100
Table 3.3	Final Distribution of Retinal Fundus Images by Class in CDGN Dataset	103
Table 3.4	Overview of Selected CNN Architectures	111
Table 3.5	Model Size and Number of Parameters (Trainable and Non-Trainable) for Each Selected Pre-trained CNN Models	112
Table 3.6	Training Hyperparameter used for all Pre-trained CNN Models	114
Table 3.7	Confusion Matrix for Binary Classification	122
Table 3.8	Confusion Matrix for Multiclass Classification (Class C_k) (Krüger, 2016)	122
Table 4.1	Classification Metrics of VGG16	132
Table 4.2	Classification Metrics of Inception-v3	136
Table 4.3	Classification Metrics of ResNet50	140
Table 4.4	Classification Metrics of DenseNet121	145
Table 4.5	Classification Metrics of EfficientNet-B0	149
Table 4.6	Performance Metrics of Baseline CNN Models on Original CDGN Dataset	151
Table 4.7	Classification Metrics of VGG16 on Feature-Enhanced CGDN Dataset	154
Table 4.8	Classification Metrics of Inception-v3 on Feature-Enhanced CDGN Dataset.	158
Table 4.9	Classification Metrics of ResNet50 on Feature-Enhanced CDGN Dataset	162

Table 4.10 Classification Metrics of DenseNet121 on Feature-Enhanced CDGN Dataset 166

Table 4.11 Classification Metrics of EfficientNet-B0 on Feature-Enhanced CDGN Dataset 170

Table 4.12 Performance Metrics of CNN Models on Feature-Enhanced CDGN Dataset 173

Table 4.13 Classification Metrics of Attention-Based VGG16 on Feature-Enhanced CDGN Dataset 176

Table 4.14 Classification Metrics of Attention-Based Inception-v3 on Feature-Enhanced CDGN Dataset 180

Table 4.15 Classification Metrics of Attention-Based ResNet50 on Feature-Enhanced CDGN Dataset 185

Table 4.16 Classification Metrics of Attention-Based DenseNet121 on Feature-Enhanced CDGN Dataset 189

Table 4.17 Classification Metrics of Attention-Based EfficientNet-B0 on Feature-Enhanced CDGN Dataset 193

Table 4.18 Performance Metrics of Attention-Based CNN Models on Feature-Enhanced CDGN Dataset 195

Table 4.19 Classification Metrics of Ensemble Model on Feature-Enhanced CDGN Dataset 200

Table 4.20 Comparative Performance Metrics of Best-Performing Models from Each Experimental Configuration 204

Table 4.21 Class-Specific AUC Performance Comparison (Non-Attention vs. Attention-based) 207

LIST OF FIGURES

		Page
Figure 2.1	Basic structure of eye (Galloway & Amoaku, 1999)	14
Figure 2.2	Clouding of the lens in a cataractous right eye (Khazaeni, 2023b)	16
Figure 2.3	Stages of diabetic retinopathy: (a) Mild non-proliferative, (b) Moderate non-proliferative, (c) Severe non-proliferative, (d) Proliferative (Qummar et al., 2019).	18
Figure 2.4	Snellen Chart (Turbert, 2022)	21
Figure 2.5	Slit-lamp examination (Porter, 2018)	22
Figure 2.6	Hand-held ophthalmoscope with labelled parts (Al-Zubaidy, 2020)	22
Figure 2.7	Labelled retinal fundus image showing the optic disc, optic cup, macula, fovea, and retinal blood vessels (Al-Zubaidy, 2020)	23
Figure 2.8	Illustration showing the hierarchical relationship between artificial intelligence, machine learning, and deep learning (Sarker, 2021).	29
Figure 2.9	Block diagram of deep neural network (Choudhary & Kesswani, 2020).	29
Figure 2.10	Common architecture of CNN for image classification (Taye, 2023)	31
Figure 2.11	Illustratiion of the convolution operation (Purwono et al., 2023)	33
Figure 2.12	Illustration of max-pooling for dimension reduction (Purwono et al., 2023)	34
Figure 2.13	AlexNet architecture (Han et al., 2017)	38
Figure 2.14	VGG16 architecture (Purwono et al., 2023)	39
Figure 2.15	Naïve version of the Inception module (Szegedy, Liu, et al., 2015)	41
Figure 2.16	Dimension reduction in Inception module (Szegedy, Liu, et al., 2015)	41
Figure 2.17	GoogLeNet architecture (Szegedy, Liu, et al., 2015)	43
Figure 2.18	Spatial factorization in Inception modules: (a) Original Inception module in Inception-v1, (b) Inception module with factorised convolutions (Szegedy, Vanhoucke, et al., 2015).	44
Figure 2.19	Illustration of skip connection (He et al., 2016)	45

Figure 2.20 Comparison of the architectures of VGG19, Plain ResNet34 and ResNet34 with residual blocks (He et al., 2016). 47

Figure 2.21 Direct connection in DenseNet, each layer takes all preceeding feature maps as input (Huang et al., 2017). 48

Figure 2.22 Transition layers between two dense blocks in DenseNet (Huang et al., 2017) 48

Figure 2.23 Illustration of conventional scaling and proposed compound scaling (Tan & Le, 2019) 51

Figure 2.24 Comparison of EfficientNets and other exisiting CNNs on ImageNet Top-1 Accuracy and Parameters Count (Tan & Le, 2019) 52

Figure 3.1 Research flowchart illustrating the sequential stages of the proposed methodology 96

Figure 3.2 Examples of poor image quality 102

Figure 3.3 Examples of incomplete central image where optic disc is not photographically visible 102

Figure 3.4 Examples of image offset: (a) Fundus region offset, (b) offset with background distortion 103

Figure 3.5 Cataracts 104

Figure 3.6 Diabetic retinopathy 104

Figure 3.7 Glaucoma 104

Figure 3.8 Normal 104

Figure 3.9 Image enhancement process 109

Figure 3.10 Comparison of original and enhanced retinal fundus images across four classes: (a) displays the original images from the CDGN dataset, (b) shows the corresponding enhanced images. 109

Figure 3.11 Architecture of the CNN model used in this research: a pre-trained base model, followed by custom fully connected layers and a final softmax output layer. 113

Figure 3.12 Architecture of CNN model with an integrated spatial attention module. The attention block, highlighted in red, is positioned between the base model and the custom classification head. 118

Figure 3.13 Overview of soft voting ensemble mechanism where the class probabilities predicted by five attention-based CNN models are averaged to generate the final prediction. 121

Figure 4.1 Training and validation accuracy and loss plots of VGG16 130

Figure 4.2 Confusion matrix of VGG16 131

Figure 4.3 ROC curves and corresponding AUC scores for each class of VGG16 133

Figure 4.4 Training and validation accuracy and loss plots of Inception-v3 134

Figure 4.5 Confusion matrix of Inception-v3 135

Figure 4.6 ROC curves and corresponding AUC scores for each class of Inception-v3 137

Figure 4.7 Training and validation accuracy and loss plots of ResNet50 138

Figure 4.8 Confusion matrix of ResNet50 139

Figure 4.9 ROC curves and corresponding AUC scores for each class of ResNet50 141

Figure 4.10 Training and validation accuracy and loss plots of DenseNet121 143

Figure 4.11 Confusion matrix of DenseNet121 144

Figure 4.12 ROC curves and corresponding AUC scores for each class of DenseNet121 145

Figure 4.13 Training and validation accuracy and loss plots of EfficientNet-B0 147

Figure 4.14 Confusion matrix of EfficientNet-B0 148

Figure 4.15 ROC curves and corresponding AUC scores for each class of EfficientNet-B0 149

Figure 4.16 Training and validation accuracy and loss plots of VGG16 on feature-enhanced CDGN dataset 153

Figure 4.17 Confusion matrix of VGG16 153

Figure 4.18 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of VGG16 155

Figure 4.19 Training and validation accuracy and loss plots of Inception-v3 on feature-enhanced CGDN dataset 156

Figure 4.20 Confusion matrix of Inception-v3 157

Figure 4.21 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of Inception-v3 159

Figure 4.22 Training and validation accuracy and loss plots of ResNet50 on feature-enhanced CDGN dataset 160

Figure 4.23 Confusion matrix of ResNet50 161

Figure 4.24 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of ResNet50 163

Figure 4.25 Training and validation accuracy and loss plots of DenseNet121 on feature-enhanced CDGN dataset 164

Figure 4.26 Confusion matrix of DenseNet121 165

Figure 4.27 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of DenseNet121 167

Figure 4.28 Training and validation accuracy and loss plots of EfficientNet-B0 on feature-enhanced CDGN dataset 168

Figure 4.29 Confusion matrix of EfficientNet-B0 169

Figure 4.30 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of EfficientNet-B0 171

Figure 4.31 Training and validation accuracy and loss plots of attention-based VGG16 on feature-enhanced CDGN dataset 174

Figure 4.32 Confusion matrix of attention-based VGG16 175

Figure 4.33 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based VGG16 177

Figure 4.34 Training and validation accuracy and loss plots of attention-based Inception-v3 on feature-enhanced CDGN dataset 179

Figure 4.35 Confusion matrix of attention-based Inception-v3 179

Figure 4.36 ROC curves and corresponding AUC scores for each class in enhanced CDGN dataset of attention-based Inception-v3 181

Figure 4.37 Training and validation accuracy and loss plots of attention-based ResNet50 on feature-enhanced CDGN dataset 183

Figure 4.38 Confusion matrix of attention-based ResNet50 184

Figure 4.39 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based ResNet50 185

Figure 4.40 Training and validation accuracy and loss plots of attention-based DenseNet121 on feature-enhanced CDGN dataset 187

Figure 4.41 Confusion matrix of attention-based DenseNet121 188

Figure 4.42 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based DenseNet121 190

Figure 4.43 Training and validation accuracy and loss plots of attention-based EfficientNet-B0 on feature-enhanced CDGN dataset 191

Figure 4.44 Confusion matrix of attention-based EfficientNet-B0 192

Figure 4.45 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based EfficientNet-B0 194

Figure 4.46 Confusion matrix of ensemble model 199

Figure 4.47 ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of ensemble model 201

Figure 4.48 Comparison of Performance (Accuracy) of Non-Attention Model versus Its Attention Counterpart on Feature-Enhanced Dataset 206

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CDR	Cup-to-Disk Ratio
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolution Neural Network
COCO	Common Objects in Context
DenseNet	Densely Connected Convolutional Network
DR	Diabetic Retinopathy
FLOPS	Floating-Point Operation Per Second
FN	False Negative
FP	False Positive
ILSVRC	ImageNet Large-Scale Visual Recognition Challenges
ISNT	Inferior, Superior, Nasal, Temporal
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
NRR	Neuroretinal Rim
ODIR-5K	Ocular Disease Intelligence Recognition
ONH	Optic Nerve Head
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
ResNet	Residual Network
RFMiD	Retinal Fundus Multi-Disease Image Dataset

RGB	Red, Green, Blue
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SMOTE	Synthetic Minority Oversampling Technique
SNN	Simulated Neural Network
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VGG	Visual Geometry Group
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background

Vision is one of the most important and dominant senses among the five human senses. It is essentially needed in every aspect of life, giving individuals the ability to see and engage with their surroundings through their eyes. Vision plays an indispensable role in daily activities, such as guiding movement, helping people avoid obstacles, recognising objects, and coordinating actions with remarkable precision. Beyond these fundamental functions, vision opens a world of limitless possibilities. It allows individuals to connect with others on a deeper level, fostering empathy and understanding through subtle cues like facial expressions and body language.

For some individuals, however, this essential sense is either partially or completely diminished, resulting in a condition known as visual impairment. Visual impairment is the term used to describe any eye condition affects the visual system and its function (World Health Organization [WHO], 2023). It can have lifelong adverse effects, making it challenging for individuals to learn, walk, read, and carry out daily activities without vision. Visual impairment covers a range of conditions that affect the visual system, with a spectrum of severity, ranging from mild visual disturbances that can be corrected with spectacles or contact lenses to complete blindness.

According to the World Report on Vision, the first report by World Health Organization (WHO) on global vision health, it is estimated that at least 2.2 billion people worldwide are suffering from some form of visual impairment, and that in at least 1 billion—

nearly half—could have been avoided or are still unresolved (WHO, 2019). Additionally, the report estimates that at least 650 million people globally are living with moderate to severe vision loss. The report identified age-related macular degeneration, cataracts, diabetic retinopathy, glaucoma, and refractive errors as among the leading causes of visual impairment and blindness (WHO, 2019).

Age-related macular degeneration is brought on by the macula's deterioration with aging and is prevalent among individuals aged 50 years and above (Muchuchuti & Viriri, 2023). According to WHO (2019), cataracts is caused by clouding in the eye lens that is due to excess blood sugar, which can happen at any age but are more prevalent in older people. Diabetic retinopathy is complication of uncontrolled, long-term high blood glucose levels, which harms the retina's blood vessels. These blood vessels may become swell, leak, or blocked, leading to vision loss due to swelling in the central part of the retinal. Glaucoma, on the other hand, is an eye disease led by higher-than-normal intraocular pressure which progressively damages the optic nerve (WHO, 2023). Refractive errors are vision disorders that caused by irregularities in the eye's shape or the curvature of the cornea, which interfere with eye's ability to concentrate on objects at varying distances. These errors can be corrected using eyeglasses, contact lenses, or refractive surgical procedures.

In 2020, among 206 million adults aged 50 years and older with moderate to severe vision impairment, the predominant causes included cataract (78.8 million), diabetic retinopathy (2.9 million), glaucoma (4.1 million), refractive errors (86.1 million), and age-related macular degeneration (6.2 million). For the estimated 33.6 million blind adults in the same age group, the primary causes of blindness were cataracts (15.2 million), diabetic retinopathy (0.9 million), glaucoma (3.6 million), refractive errors (2.3 million), and age-

related macular degeneration (18.4 million) (GBD 2019 Blindness and Vision Impairment Collaborators & Vision Loss Expert Group of the Global Burden of Disease Study, 2021).

Vision impairment can pose significant challenges in daily life, affecting an individual's ability to perform day-to-day tasks, navigate their surroundings, and interact with others. While many cases of vision loss are preventable, this is not always possible, as many eye diseases often manifest gradually over time. Therefore, early detection and prompt treatment are crucial to mitigating or even preventing permanent vision loss. Traditionally, the detection of eye diseases relies on the expertise of ophthalmologists, who examine retinal fundus images to identify the presence of abnormalities. This process is tedious and may take longer time, which may be limited in many underserved areas due to a shortage of experienced ophthalmologists (Muchuchuti & Viriri, 2023). As a result, there has been a growing interest recently in using deep learning techniques in detection and classification of eye disease.

1.2 Motivation

Visual impairment and blindness that caused by various types of eye diseases affect millions of individuals globally. Therefore, early diagnosis and timely treatment are vital to halting disease progression and preventing blindness. While abnormalities on the external surface of the eye can be detected directly with the naked eye, a more-in-depth examination requires a fundus camera or ophthalmoscope to produce a retinal fundus image (Arif et al., 2023). A retinal fundus image captures the back surface of the eye, which comprises the fovea, retina, macula, optic disc, and blood vessels. Ophthalmologists or clinicians usually use these images to detect the presence of eye diseases.

However, eye disease detection is a challenging task that often requires years of medical experiences. According to Qummar et al. (2019), the diagnostic process has low repeatability due to its time-consuming, laborious, and arbitrary nature. Even professional ophthalmologists may face challenges in making accurate diagnoses based on fundus image. These challenges have motivated the development of deep learning networks for classification of eye diseases. With advancements in machine learning and computer vision, deep learning has become a viable solution for classifying various eye diseases (Muchuchuti & Viriri, 2023). Automated detection and classification systems should be useful as assistive technologies to lessen the workload on overworked ophthalmologists and address the shortage of experienced professionals.

The successful application of vision transformers and deep Convolutional Neural Networks (CNN) in computer-aided diagnosis has significantly advanced the efforts in automated detection of eye diseases (Qummar et al., 2019; Raj et al., 2019). Current research primarily focuses on detecting a single type of eye diseases, where the models are designed to classify fundus images as either healthy or showing abnormalities, or to distinguish between normal images and those with a specific disease. These models have shown high effectiveness, with many models developed in several studies achieving classification accuracies exceeding 90% (Tsiknakis et al., 2021; Muchuchuti & Viriri, 2023). These results establish a strong foundation for extending automated eye disease detection frameworks toward more comprehensive multiclass classification tasks.

1.3 Problem Statement

Automated eye disease detection and classification have gained increasing attention in the realm of medical imaging, showing high efficacy in binary classification tasks.

However, a significant gap remains in transitioning these models to real-world clinical settings, which require simultaneous multiclass classification. In actual clinical practice, ophthalmologists must distinguish between various eye conditions that often exhibit subtle, overlapping pathological patterns. As highlighted by Muchuchuti and Viriri (2023), there is a need to develop multiclass eye diseases classification models that better reflect real-world clinical practice and assist ophthalmologists in effectively diagnosing various eye conditions.

Current research in this area faces critical challenges that limit the diagnostic reliability of automated systems, primarily stemming from the inherent complexity of disease symptoms, overlapping clinical presentations of different eye diseases especially during the early stages of the diseases, and limitation of existing data (Tsiknakis et al., 2021; Muchuchuti & Viriri, 2023). The core challenge in multiclass eye classification lies in the subtlety of diagnostic features. Some eye diseases exhibit similar or overlapping clinical features, particularly in their early stages, which are often visually faint and easily obscured by the dominant global structures of the retina. Conventional deep learning models sometimes struggle to isolate these features, leading to high inter-class confusion where different diseases are misidentified due to overlapping visual presentations. Without enhancement of these subtle features, the model's discriminative power remains insufficient for clinical-grade diagnosis. resulting in high inter-class confusion.

In clinical environment, retinal fundus images are captured across various devices, resolutions, and patient demographics. This results in intra-class variability, meaning a single disease may present with a wide range of visual symptoms (e.g., microaneurysms, haemorrhages, exudates), further complicating classification. This complexity is

compounded by the limitation of large-scale annotated datasets. Most publicly available datasets contain only a few hundred to a few thousand images. These complexities hinder the models' ability to generalize across different cases within the same class and to effectively identify the most relevant features for correct prediction. Current research often uses small, homogeneous datasets that fail to reflect this diversity. While increasing dataset size through multiple source integration captures this reality, it also introduces significant noise and data heterogeneity. Furthermore, the diagnostic performance of deep learning models is often compromised by class imbalance. In most retinal fundus image datasets, certain diseases are significantly less represented than others. This imbalance typically biases models toward majority classes, leading to poor sensitivity for rarer conditions and inconsistent diagnostic outcomes. These technical challenges—ranging from feature obscurity to data bias—can lead to misdiagnosis and delayed treatment, ultimately preventing patients from receiving timely, life-saving care.

Most existing works focus on classification using raw images without addressing the inherent visibility gap of disease-specific features. There is a research gap in developing deep learning architectures that remain robust and generalizable across diverse, heterogeneous clinical data. This research proposes a feature-augmented enhanced deep learning approach that integrates image enhancement with attention-based ensemble learning. A suitably sized dataset will be constructed by combining publicly available datasets to introduce variability and diversity. By making the disease-specific features to be more pronounced through image enhancement and employing attention mechanisms to focus on the most relevant features within retinal fundus images, this methodology aims to mitigate issues related to inter-class and intra-class confusion and enhance the model performance.

1.4 Research Questions

According to the problem statement, the research questions are listed as follows:

i. How can eye disease-specific features be enhanced from fundus images?

ii. What is the impact of incorporating image enhancement, attention mechanisms, and ensemble learning into pre-trained deep learning models for multiclass classification of eye diseases?

iii. How will the feature-augmented enhanced deep learning models perform compared to conventional deep learning models?

1.5 Research Objectives

The primary objective of this research is to develop a deep learning model for multiclass eye disease classification that can classify cataracts, diabetic retinopathy, and glaucoma, and normal using retinal fundus images. The specific objectives of this research are as follows:

i. To formulate the enhancement of eye disease-specific features in retinal fundus images using image enhancement techniques.

ii. To investigate and compare the effectiveness of image enhancement, attention mechanisms, and ensemble learning within pre-trained deep learning models for multiclass eye disease classification.

iii. To evaluate the performance of the feature-augmented enhanced deep learning models against the conventional deep learning models for eye disease classification in classifying cataracts, diabetic retinopathy, glaucoma, and normal eyes.

1.6 Research Scope

This research focuses on developing an enhanced multiclass eye disease classification model using a feature-enhanced deep learning approach. The classification task will concentrate on the three most common types of eye diseases, namely cataracts, diabetic retinopathy, and glaucoma, as well as normal eyes. Although other eye diseases such as age-related macular degeneration, hypertensive retinopathy, and radiation retinopathy exist, they are beyond the scope of this research. The selection of these three diseases is based on findings from the Third National Eye Survey (NES III), conducted by Malaysian Ministry of Health in the Eastern Zone (Pahang, Kelantan, Terengganu) and the Sarawak Zone between July and October 2023. The survey identified cataracts, diabetic retinopathy, and glaucoma as the leading causes of blindness in these regions ("National Survey Shows Fewer Cases of Blindness in Those Over 50," 2023).

This research will utilize retinal fundus images as the primary data source, compiled by selecting and combining several publicly available datasets, such as RFMiD and those from Kaggle. These datasets contain retinal fundus images categorized according to specific eye diseases. The data will be pre-processed to ensure quality before being used to train, validate, and evaluate the deep learning model. The model will be trained using Convolutional Neural Network (CNN) architectures, which are ideal for image classification tasks. Evaluation will be based on standard performance metrics, including accuracy, precision, recall, F1-score, and area under the curve (AUC) scores to measure the model's effectiveness in multiclass eye disease classification.

However, the availability of high-quality, labelled fundus images is a potential limitation of this research. Additionally, the generalizability of the model may be affected

by factors such as image quality and inconsistencies in how the images are acquired. This research focuses solely on classifying fundus images into one of the four target classes: cataracts, diabetic retinopathy, glaucoma, and normal. As such, each image in the dataset is belonging to only one class, and multilabel classification is not considered. Retinal fundus images exhibiting more than one type of eye diseases are excluded from this research.

1.7 Expected Outcomes

The expected outcome of this research is the development of a multiclass eye diseases classification model capable of accurately classifying cataracts, diabetic retinopathy, and glaucoma using transfer learning, attention mechanisms, and ensemble learning techniques. Besides, this research aims to enhance the classification performance by incorporating image enhancement techniques. The model will learn to distinguish between the three eye diseases from normal eye by identifying relevant features from the retinal fundus images during the training process. This approach is intended to improve the reliability of the model and foster trust among ophthalmologists, clinicians, and patients.

The performance of the proposed enhanced deep learning models will be evaluated in comparison to conventional deep learning approaches. Through the integration of image enhancement, attention mechanisms, and ensemble learning, the model is expected to demonstrate improved performance. The accuracy, generalizability, and applicability of the model are expected to be enhanced, with the long-term goal of providing ophthalmologists with an assistive tool that enabling remote screening and early detection of eye diseases.

1.8 Significance of Research

This research on enhancing the multiclass eye disease classification through deep learning is significant in addressing the challenges faced in ophthalmic diagnosis. The

development of deep learning-based eye disease classification models has become a popular topic in medical image analysis. Deep learning approaches are able to identify the subtle abnormalities in retinal fundus images that may indicate the presence of eye diseases. Most existing studies have focused primarily on binary classification, where models distinguish between normal and diseased eye, such as identifying diabetic retinopathy or glaucoma. While these models have shown promising results, they do not fully demonstrate the actual clinical setup, where multiple types of eye diseases must be diagnosed simultaneously. This research aims to bridge that gap by developing a model that can accurately classify three most common eye diseases—cataracts, diabetic retinopathy, and glaucoma—using a feature-augmented enhanced deep learning approach.

Research on multiclass eye disease classification using deep learning has the potential to revolutionise eye care and improve global vision health. The proposed model can be deployed in areas with limited access to ophthalmologists, especially in areas with elevated rates of eye diseases, thereby increasing the accessibility and affordability of eye care services. The model is designed to emulate the diagnostic decision-making behaviours of professional ophthalmologists by analysing retinal fundus images, serving as a valuable tool that augments clinical expertise. It can facilitate more efficient diagnostic processes by providing second opinions and highlighting potential abnormalities. Early detection through these systems facilitates prompt intervention and treatment, helping to minimize the risk of vision loss and blindness.

Furthermore, this deep learning-based multiclass eye disease classification system can be integrated into public health initiatives to enhance screening programs, monitor disease trends, and inform policy decisions. This could contribute to lower the global burden

of vision impairment and blindness. It also aligns with the United Nations' third Sustainable Development Goal (SDG 3), which focuses on ensuring healthy lives and promoting well-being for all ages. Vision impairment is a significant public health issue worldwide that affects individuals across the life course, limiting their ability to work and engage socially. By enabling early detection, precise diagnosis, and timely intervention, this research can significantly reduce the impact of eye diseases and improve the standard of living for millions around the world.

1.9 Thesis Outline

This thesis includes five chapters that describe the research of enhancing multiclass classification of eye diseases using a feature-augmented enhanced deep learning approach.

Chapter 1 Introduction explains about the background of eye diseases and related classification technologies. Besides, this chapter depicts the motivation, problem statement, research questions, research objectives, research scope, significance of research, expected outcomes, and the thesis outline.

Chapter 2 Literature Review reviews and analyses existing studies on eye diseases classification using deep learning. This chapter delves more into the strengths and limitations between the past studies and identifies gaps that this research aims to address.

Chapter 3 Methodology outlines the development of the proposed multiclass eye diseases classification model using a feature-augmented enhanced deep learning approaches. This chapter describes the process of data collection, data preprocessing, image enhancement, model development, model training, and model evaluation to analyse the model effectiveness in classifying targeted eye diseases.

Chapter 4 Results and Discussion emphasizes on the performance results of the developed classification model. The model is evaluated using performance metrics such as accuracy, precision, recall, F1-score and AUC scores. The results are discussed in relation to the research objectives and compared with baseline performance.

Chapter 5 Conclusion and Recommendations summarises the research process and key findings. In addition, this chapter discusses the contributions of this research along with its objectives, outlines the research limitations and proposes recommendations for future research directions.

1.10 Chapter Summary

This chapter outlines the background of the research, focusing on vision impairment and eye diseases. Recent research has shown notable advancements in developing deep learning models for detecting and classifying eye diseases, highlighting this is a promising field of study. However, to enhance the generalizability and robustness of these models, several issues need to be addressed. These research gaps are detailed in the Problem Statement subsection. Based on the identified problems, this chapter also presented the research questions, objectives, scope, expected outcomes, and the significance of the research. These components offer a clear and comprehensive insight into the motivation and direction of this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter delves into the overview of the anatomy of the human eye and discusses three eye diseases that are central to this research, which are cataracts, diabetic retinopathy, and glaucoma. It then briefly outlines conventional diagnostic methods used in clinical practice to detect eye diseases. With the advancements in machine learning and deep learning, diagnostic capabilities have significantly improved, particularly in image classification. This chapter proceeds by introducing the core principles of machine learning and deep learning, providing a detailed explanation of their mechanisms. A specific focus is given to Convolution Neural Networks (CNNs), the most widely used deep learning technique for image classification tasks, including a review of prominent CNN architectures. These technological developments have propelled research in the classification of eye diseases. Finally, this chapter present a comprehensive review and comparison of existing studies that apply deep learning approaches to eye disease classification. The insights acquired in this chapter lays the groundwork for the subsequent analysis of eye disease classification methods.

2.2 Anatomy of the Eye and Its Related Diseases

The eyes, being among the most vital sensory organs in humans, play a central role in enabling vision, giving individuals the ability to observe the world and interact with their surrounding environment. The human eye has a complex structure, with easily identifiable components including the sclera, cornea, iris, and pupil, as shown in Figure 2.1. Generally, the human eye can be distinguished as three layers: the outer, middle, and inner layers. The

outermost layer of the eye comprises the cornea and sclera. The cornea functions to refract and transmits light toward the lens and retina, while the sclera serves as a protective barrier against infection and physical damage by forming an outer membrane of connective tissue and maintaining its shape. A transparent mucous membrane known as the conjunctiva covers the visible portion of the sclera (Willoughby et al., 2010).

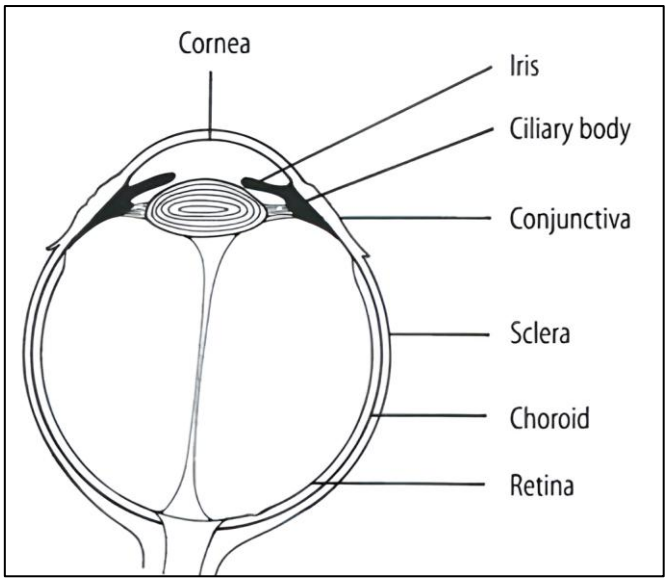


Figure 2.1: Basic structure of eye (Galloway & Amoaku, 1999)

The middle layer of the eye is composed of the iris, ciliary body and choroid. The iris regulates the amount of light entering the eye by adjusting the pupil's size. The ciliary body produces aqueous humour and controls the strength and shape of the lens. Additionally, the choroid is a layer rich in blood vessels that delivers oxygen and nutrients to the outer regions of the retina. The retina, which lines the inner surface of the eye, contains a complex, layered neuronal structure responsible for capturing and processing light (Willoughby et al., 2010). Most vertebrate eyes contain two distinct types of photoreceptors: rod cells and cone cells. These photoreceptors facilitate phototransduction—the process by which incoming light is converted into electrical signals transmitted to the brain for interpretation. In humans, rod cells outnumber cone cells by approximately 20 to 1 (McCaa, 1982).

Eye diseases and degenerative conditions are a major concern due to their impact on the functionality of this essential organ. Eye diseases encompass a wide variety of morbidities that affect various components of the visual system and their associated functions (World Health Organization [WHO], 2019). Although categorizing eye diseases can be challenging due to their diversity, one common approach is to differentiate between those that typically do not lead to vision impairment and those that do. Cataracts, diabetic retinopathy, and glaucoma are among the major eye diseases that can cause vision impairment and even blindness (WHO, 2019). The following sub-sections provide a brief overview of these three eye diseases, which are the focus of this research aimed at enhancing eye disease classification models.

2.2.1 Cataracts

Cataracts are among the most common eye diseases and represent a primary contributors to visual impairment and preventable blindness across the globe. Various subtypes of cataracts exist, including nuclear, cortical, posterior, congenital, senile, traumatic, and bilateral cataracts (Marouf et al., 2022). Cataracts are characterized by clouding of the lens of the eye, resulting in vision deterioration (WHO, 2019). Situated at the anterior part of the eye, the lens is a small, transparent structure that focuses incoming light onto the retina, allowing for sharp vision across varying distances. Figure 2.2 shows the cataractous eye of a patient, where white clouding is evident. Marouf et al. (2022) describe cataracts as the agglomeration of protein within the eye, forming a lump-like mass that obstructs the normal passage of light to the retina through the lens, thereby preventing a sharply defined image from forming. In addition, cataract formation typically begins with an increase in the lens's water content, followed by dehydration and a reduction in oxygen uptake (McCaa, 1982).

Figure 2.2: Clouding of the lens in a cataractous right eye (Khazaeni, 2023b)

According to the World Health Organization (WHO, 2019), cataracts are highly prevalent in countries with low to middle income levels. While cataracts are asymptomatic in the early stages, they can progressively impair vision over time. Major symptoms of cataracts include the gradual onset of painless blurred vision, degradation of night vision, fading of colours, and diplopia (Marouf et al., 2022). Patients with cataracts often describe their vision as resembling the view through a fogged or frosted window. This blurred vision can hinder daily activities and significantly reduce quality of life. The likelihood of developing cataracts increases with age, smoking, and exposure to ultraviolet light (WHO, 2019). Additional risk factors include diabetes, obesity, family history, previous eye surgery, and prolonged use of corticosteroid medications (Khazaeni, 2023b). In most cases, cataract-related blindness is preventable, as vision can be effectively restored through use of artificial intraocular lenses or surgical removal of the cataracts (Cicinelli et al., 2023).

2.2.2 Diabetic Retinopathy

Diabetic retinopathy is an eye condition triggered by blocked, leaking or damaged blood vessels in the retina. Swelling in the central part of the retina can cause vision impairment, and if left untreated, eventual vision loss (WHO, 2019). This condition

mutilates the retinal blood vessels in individuals with diabetes. According to the World Health Organization (WHO, 2019), lifestyle changes contributing to the rising prevalence of diabetes are expected to increase the number of individuals affected by diabetic retinopathy from 146 million in 2014 to 180.6 million by 2030. Moreover, Dai et al. (2021) estimated that by 2040, around 600 million people globally will have diabetes, with approximately one-third likely to develop diabetic retinopathy.

There are two primary types of diabetic retinopathy: non-proliferative and proliferative (Qummar et al., 2019). Non-proliferative diabetic retinopathy, the early stage of the disease, is further categorized into mild, moderate, and severe stages. In the mild stage, the main manifestation is the appearance of microaneurysms—tiny, round red spots located at the ends of blood vessels—as shown in Figure 2.3(a). In the moderate phase, these microaneurysms rupture and extended into deeper retinal layers, resulting in flame-shaped haemorrhages, as depicted in Figure 2.3(b). Severe non-proliferative diabetic retinopathy defined by the presence of over 20 intraretinal haemorrhages in each of the four quadrants, along with clear venous beading and noticeable intraretinal microvascular abnormalities, as shown in Figure 2.3(c). Proliferative diabetic retinopathy represents a more advanced stage of condition, marked by neovascularization—the abnormal formation of new blood vessels on the inner surface of the retina—creating a dysfunctional microvascular network (Figure 2.3(d)).

Patients are typically asymptomatic during the initial phase of diabetic retinopathy. However, in advanced stages, symptoms such as floaters, blurred vision, distorted vision, and progressive decline in visual acuity may occur (Qummar et al., 2019). Therefore, early detection is challenging but critically important to avoid adverse outcomes in later stages.

Effective diabetes management through regular physical activity, a balanced diet, and adherence to prescribed medications is crucial in preventing or postponing the development of diabetic retinopathy. Treatment options for diabetic retinopathy include laser therapy, intravitreal injections, and surgical procedures, all aimed at preventing further vision loss and improving visual outcomes (National Eye Institute, 2024).

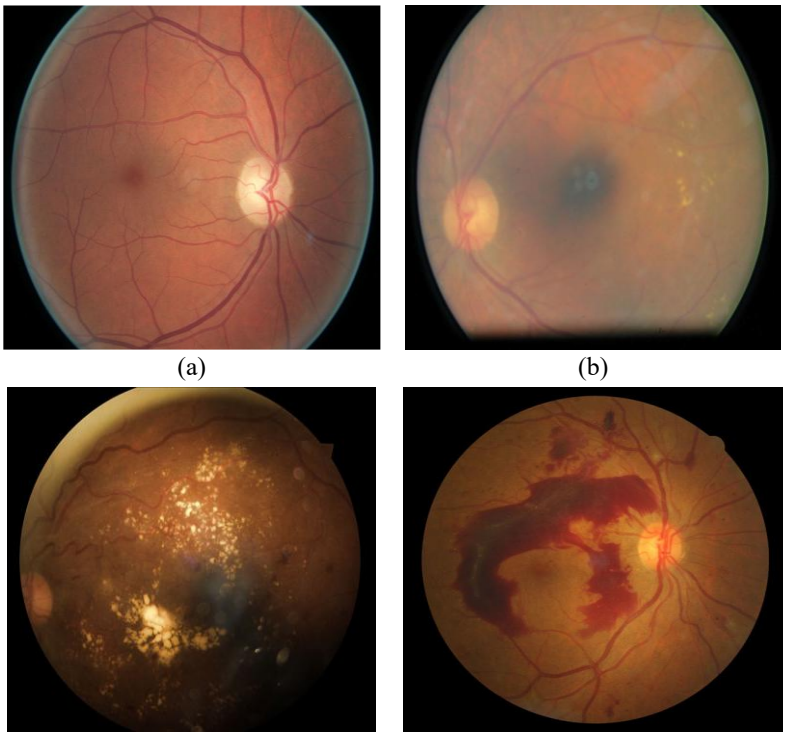


Figure 2.3: Stages of diabetic retinopathy: (a) Mild non-proliferative, (b) Moderate non-proliferative, (c) Severe non-proliferative, (d) Proliferative (Qummar et al., 2019).

2.2.3 Glaucoma

Glaucoma is an eye disease marked by the destruction of the optic nerve, leading to optic neuropathy with various potential aetiologies, and marked by the loss of retinal ganglion cells (Akter et al., 2022). Clinically, glaucoma is a silent disease that causes gradual and irreversible deterioration of the visual field, often remain asymptomatic until its advanced stages, when pronounced and irreversible vision loss occurs (Kovalyk et al., 2022a). WHO (2019) estimated that approximately 64 million people globally are affected

by glaucoma, with 6.9 million (10.9%) experiencing moderate or severely impaired vision or blindness. Noteworthy risk factors for glaucoma include advanced age, increased intraocular pressure, severe near-sightedness, and a family history of the disease (Schuster et al., 2020).

The absence of a cure for established glaucomatous damage underscores the significance of early diagnosis to avoid progression to total blindness. Currently, diagnosing and monitoring glaucoma are often challenging, as this disease may exhibit substantial overlap in ocular characteristics between healthy individuals and those in early stage of glaucoma (Akter et al., 2022). Accurate diagnosis typically requires a thorough eye assessment, additional testing, and interpretation of extensive clinical data. Ophthalmologists rely on measurements of intraocular pressure, assessment of functional impairment via visual field testing, and manual analysis of structural damage in the optic nerve and retinal nerve fibre layer using retinal fundus imagery (Bragança et al., 2022).

Clinically, glaucoma is characterised by anatomical alterations in the optic nerve head (ONH), notably thinning and backward displacement of the lamina cribrosa, which leads to ONH cupping (Hemelings et al., 2021). During fundus examination, ophthalmologists assess the ONH for characteristic changes, such as thinning of neuroretinal rim (NRR). This is typically quantified by the vertical cup-to-disc ratio (CDR). An elevated vertical CDR or an interocular asymmetry greater than 0.2 is considered suspicious for glaucoma (Hemelings et al., 2021). Additional pathological signs visible in fundus images that support glaucoma diagnosis include disc haemorrhages, NRR thickness deviations from the Inferior, Superior, Nasal, Temporal (ISNT) rule, nerve fibre layer defects, peripapillary atrophy, and NRR notches (Z. Zhang et al., 2010).

2.3 Conventional Diagnostic Methods for Eye Diseases

The diagnosis of eye diseases involves the identification of specific conditions that affects an individual's vision. This process is usually conducted by ophthalmologists or optometrists, using specialized equipment and diagnostic tools. Ophthalmologists are medical professionals specializing in the evaluation and treatment of eye diseases, including both surgically and non-surgical interventions. In contrast, optometrists are healthcare practitioners who focus primarily on diagnosing and managing vision problems and refractive errors (Khazaeni, 2023a). Early diagnosis of eye disease is essential for timely intervention and improved treatment outcomes. Initial assessments are generally based on symptoms reported by the patient, the physical appearance of the eyes, and preliminary examination findings. To assess various facets of vision and eye health, ophthalmologists may perform several diagnostic procedures, including visual field testing, refraction testing, slit-lamp examination, and ophthalmoscopy.

i. Visual Field Testing

The visual field, also referred to as peripheral vision, encompasses the entire area visible to each eye, including the corners or edges (Turbert, 2022). Visual field testing is a routine component of a comprehensive eye examination used to measure the extent of vision loss. This test identifies the appearance and location of blind spots and evaluates peripheral vision function. In patients with glaucoma, for instance, visual field testing can reveal any possible peripheral vision loss, aiding in diagnosis and monitoring of disease progression.

ii. Refraction Testing

Refraction testing is used to diagnose visual acuity problems caused by refractive errors, such as near-sightedness (myopia), far-sightedness (hyperopia), astigmatism,

and presbyopia (Khazaeni, 2023a). Figure 2.4 shows the Snellen chart—an eye chart displaying rows of letters in various sizes or the letter “E” in different orientation—is used for this purpose. This test determines whether an individual has 20/20 vision, which represents normal visual acuity measured at 20 feet. The smallest row of letters that the person can read accurately indicates their degree of visual acuity and informs the prescription needed for corrective lenses.

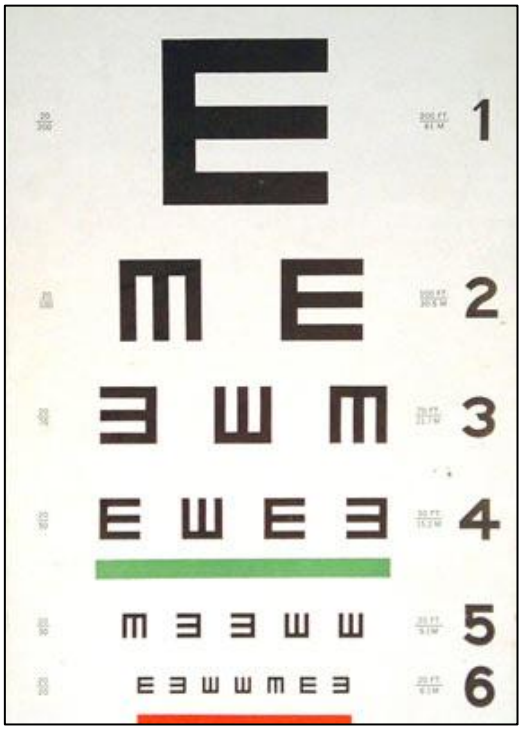


Figure 2.4: Snellen Chart (Turbert, 2022)

iii. Slit-Lamp Examination

The slit-lamp examination utilizes a table-mounted specialized microscope (Figure 2.5) equipped with a focused light beam under high magnification and depth-measurement capability to examine both the external and internal structure of the eye, including the cornea, lens, iris, and retina (Khazaeni, 2023a). Pupil-dilating eye drops are typically administrated to provide a better view at the back of the eye. (Porter, 2018).

Figure 2.5: Slit-lamp examination (Porter, 2018)

iv. Ophthalmoscopy

Ophthalmoscopy, also known as fundoscopy, is a diagnostic procedure that uses an ophthalmoscope (Figure 2.6) or specialised lenses to examine the retina, blood vessels, and optic nerve located at the back of the eye. It is a standard component of comprehensive eye examinations and is commonly performed to screen for eye diseases. As with the slit-lamp examination, pupil-dilating eye drops are often administered to enhance the field of view.

Figure 2.6: Hand-held ophthalmoscope with labelled parts (Al-Zubaidy, 2020)

2.3.1 Retinal Fundus Imaging

Ophthalmoscopy or fundoscopy, is a quick, non-invasive examination that facilitates the early identification of various eye conditions. It produces a retinal fundus image, a highly specialized type of medical image of the retina at approximately 10-15 times magnification captured using a fundus camera operated by ophthalmologists (Al-Zubaidy, 2020). This imaging technique is an important mean for documenting retinal health and is commonly utilized in diagnosing several eye diseases. Figure 2.7 illustrates a retinal fundus image, highlighting the key anatomical structures of the retina, including the blood vessels, fovea, macula, and optic disc. A healthy retina appears clear, with no signs of haemorrhages, cotton wool spots, or exudates. The optic disc, also known as the optic nerve head, is a circular region at the posterior pole of the eye where the retinal ganglion cell axons converge to form the optic nerve, which carrying visual signals to the brain. It also serves as the entry and exit point for the retinal artery and vein (Belden, 2023).

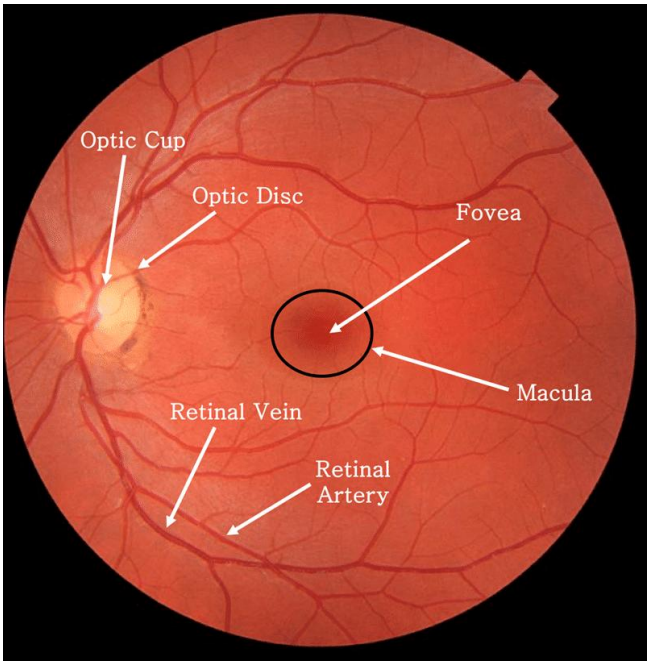


Figure 2.7: Labelled retinal fundus image showing the optic disc, optic cup, macula, fovea, and retinal blood vessels (Al-Zubaidy, 2020)

The distinguishable depression in the centre of the optic disc is called the optic cup, formed by the inward folding of the embryonic optic vesicle. According to Al-Zubaidy (2020), the cup-to-disc ratio (CDR)—defined as the proportion of optic cup's diameter relatively to the overall diameter of the optic disc—is a clinically important metric fundus examination. A CDR of 0.3 or lower is generally regarded as normal. The area between the temporal vascular arcades is referred to as the macula lutea, and the depression located near the centre of the macula is called the fovea. The fovea, which has roughly the same diameter as the optic disc, appears darker due to increased concentration of pigment. It contains the highest density of cone photoreceptors that plays a key role in producing sharp central vision (Al-Zubaidy, 2020).

Based on the information acquired from the various eye examinations, the ophthalmologist provides a diagnosis and recommends a suitable treatment plan. This may include medications, laser therapy surgery intervention, or other procedures, depending on the specific condition diagnosed. However, manual diagnosis is a tedious and error-prone process, requiring highly skilled experts or experienced ophthalmologists to interpret the result—especially in the analysis retinal fundus images for disease classification (Qummar et al., 2019). Moreover, essential ophthalmic diagnostic equipment such as slit lamps and fundus cameras is often unavailable in low-resource settings, even where trained health workers are present, especially in public healthcare sector of low- and middle-income countries (WHO, 2019).

2.4 Machine Learning

Over the past few years, artificial intelligence (AI) and machine learning have transformed a myriad of industries and disciplines, offering substantial improvements in

efficiency, accuracy, and decision-making. AI involves replicating human cognitive ability—such as learning and problem-solving—within machines using computer algorithms (Srivastava et al., 2023). Machine learning, a branch within the broader field of AI, involves encoding and modifying computational parameters based on data (Srivastava et al., 2023). It serves as a collective term for diverse algorithms capable of making intelligent predictions by learning from large datasets, often consist of millions of unique data (Nichols et al., 2018). Machine learning seeks to automate the creations of analytical models capable of performing cognitive functions, such as object detection or natural language translation, by employing algorithms that learn patterns and make predictions from problem-specific training data. As a result, machines can recognize intricate patterns and reveal subtle insights without being explicitly programmed for each tasks (Janiesch et al., 2021).

Machine learning has demonstrated good applicability in tasks involving high-dimensional data, such as clustering, classification, and regression. Recent developments in machine learning have enabled systems to approach, and in some cases surpass, human performance in semantic understanding and information extraction. Consequently, machine learning algorithms have been effectively applied across various domains, including speech and image recognition, natural language processing (NLP), next-best-offer analysis, fraud detection, and increasingly, medical diagnostics. Machine learning can be categorized into four main types: supervised, unsupervised, semi-supervised, and reinforcement learning (Chapelle et al., 2006; Nichols et al., 2018; Jhaveri et al., 2022).

i. Supervised Learning

In supervised learning, models are trained on datasets include both the input and corresponding correct output labels. Its primary goal is to learn the mapping between inputs and outputs, enabling the model to make accurate predictions on unseen data

(Chapelle et al., 2006; Nichols et al., 2018). This approach is more widely used in classification and regression tasks. Common algorithms used in supervised learning include support vector machines (SVM), random forest, and various neural networks such as long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) (Taye, 2023).

ii. Unsupervised Learning

In unsupervised learning, models are trained on unlabelled data, in contrast to supervised learning, to uncover hidden patterns, structures, trends, or correlations within the datasets (Chapelle et al., 2006). Common applications include clustering, anomaly detection, association rule mining, dimensionality reduction, feature extraction, and density estimation (Jhaveri et al., 2022). A notable example of an unsupervised learning system is Google's AlphaGo Zero, which learned to play the game of Go without prior labelled examples (Silver et al., 2016).

iii. Semi-Supervised Learning

Semi-supervised learning bridges the gap between supervised and unsupervised approaches, utilizing a combination of labelled and unlabelled data for training (Chapelle et al., 2006; Jhaveri et al., 2022). It is useful in real-time applications where large amounts of unlabelled data and very small amounts of labelled data are available. Semi-supervised learning achieves better prediction goals compared to using labelled data alone, particularly in tasks such as text classification, fraud detection, and machine translation (Jhaveri et al., 2022).

iv. Reinforcement Learning

Reinforcement learning is a strategy guided by environmental feedback, where machines or software agents independently learn to adopt optimal actions that

increase rewards while reducing penalties (Sutton & Barto, 1998; Jhaveri et al., 2022). Reinforcement learning operates on a system of rewards and penalties, where the agent is rewarded for desirable actions and penalized for undesirable ones. This method is frequently employed in field requiring sequential decision-making, such as robotics, self-driving automobile, and the gambling sector (Taye, 2023).

A range of effective interventions are now available to reduce the risk of developing eye diseases, prevent vision impairment, and mitigate their long-term impact. Several emerging technologies have been adopted in ophthalmology, such as mobile-based software applications for vision assessment and cataracts surgery benchmarking, as well as AI systems for the detection of various eye diseases (Muchuchuti & Viriri, 2023). Given the increasing reliance on digital imaging modalities and quantitative assessment metrics in ophthalmology, this field is particularly well-suited for AI integration. It is crucial to emphasize that AI is designed to support, not replace, ophthalmologists, by augmenting their diagnostic capabilities and clinical decision-making. Through the integration of AI technologies, healthcare professionals can deliver quicker and more precise diagnoses, tailor personalized treatment strategies, and enhance overall clinical outcomes.

2.4.1 Deep Learning

Neural networks, also referred to as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a specific type of machine learning algorithm inspired by the structure and functioning of the human brain (Goodfellow et al., 2016). They aim to emulate the learning capabilities of biological neurons in performing computational tasks. A typical neural network is composed of interconnected nodes, or neurons, arranged into an input layer, multiple hidden layers, and an output layer. Each neuron processes information by receiving, transforming, and transmitting signals using associated weights and threshold

values. When the computed output exceeds a defined threshold, the signal is transmitted to subsequent layers as determined by the activation function. Through iterative training on labelled data, neural networks adjust their internal parameters (weights and biases), thereby improving performance over time.

Deep learning, as illustrated in Figure 2.8, is a subset of machine learning and part of the broader AI field. While “deep learning” and “neural networks” are frequently used interchangeably, deep learning specifically denotes the application of neural networks with numerous layers, known as deep neural networks, where the term “deep” signifies the depth of layers in the network (IBM, n.d.). Basic neural network generally consists of two or three layers, whereas a deep neural network feature several hidden layers positioned between the input and output layers, as depicted in Figure 2.9 (Choudhary & Kesswani, 2020). Moreover, deep neural networks usually contain more advanced neurons compared to simple ANNs and may incorporate complex operations or utilize multiple activation functions within individual neuron. These capabilities allow them to handle raw input data and autonomously discover the necessary representations for specific learning task—a fundamental aspect known as deep learning.

Deep learning has found extensive use across various application and research field, including but not limited to visual recognition, object detection, text analytics, sentiment analysis, natural language processing (NLP), medical and healthcare systems, cybersecurity, business intelligence, and more (Sarker, 2021; Taye, 2023). Properly trained deep neural networks can achieve notable effectiveness in both classification and regression tasks. In the medical and healthcare fields, numerous deep learning models exhibiting high specificity and sensitivity have been developed to detect or classify specific medical conditions using

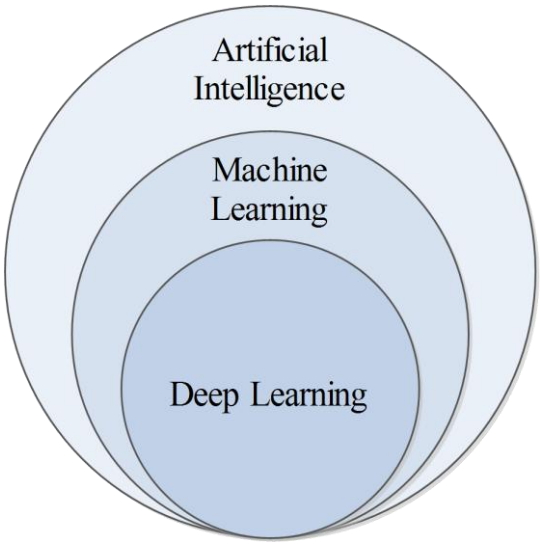


Figure 2.8: Illustration showing the hierarchical relationship between artificial intelligence, machine learning, and deep learning (Sarker, 2021).

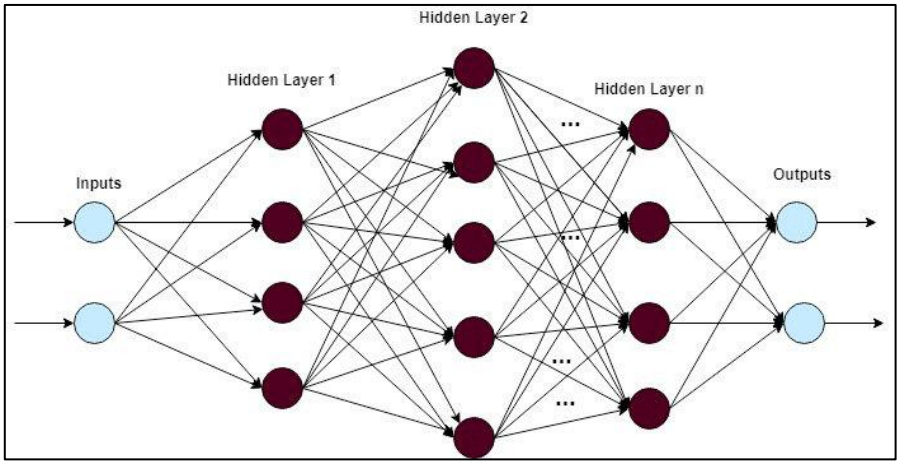


Figure 2.9: Block diagram of deep neural network (Choudhary & Kesswani, 2020).

imaging data (Dai et al., 2021). Deep learning architecture is designed for hierarchical feature extraction: early layers handle primary data processing or learn simple features, while deeper layers progressively capture more complex features (Taye, 2023). As a result, deep learning is especially advantageous in domains that handle large, high-dimensional and complex data types, such as text, image, video, speech, and audio (Janiesch et al., 2021; Taye, 2023). Compared to other machine learning algorithms, deep learning generally requires less time for execution during testing but demand a lengthy training period due to the large number of parameters involved (Sarker, 2021).

2.4.2 Convolutional Neural Network

Convolutional Neural Network (CNN or ConvNet) are among the most commonly employed deep learning architectures. Inspired by artificial neural networks (ANNs), CNNs are specifically designed to process two-dimensional data, making them particularly effective for image classification and object recognition tasks (LeCun et al., 1989; Sarker, 2021). Image classification refers to the process of assigning a label or class to an input image. It is a supervised learning approach where the model is trained on annotated datasets and subsequently used to predict new, unseen images. CNNs have become a state-of-the-art approach, demonstrating remarkable performance in a wide range of domains, such as computer vision, self-driving vehicles, medical image analysis, NLP, radiology and more (Ayeni, 2022; Sarker, 2021).

CNNs are structured with multiple layers of neurons that facilitate hierarchical feature learning, first popularized by LeCun et al. (1989) through their work on handwritten digit recognition using backpropagation. They are particularly effective to address tasks that involve dataset with spatial dependencies, such as images, where the arrangement of pixels is crucial (Janiesch et al., 2021). CNNs are capable of learning spatial feature hierarchies—such as edges, patterns, and shapes—that are essential for accurate object recognition in images. This process is conducted automatically and adaptively through backpropagation across multiple layers (Ayeni, 2022). The capabilities of automatically discover relevant features from raw input data, without requiring manual feature engineering makes CNNs more powerful than traditional networks (Sarker, 2021).

A standard CNN architecture comprises multiple convolution layers, followed by pooling layers, and concludes with fully connected layers responsible for final classification.

As described by Nichols et al. (2018), the first few layers in CNNs are the convolution layers that extract relevant features from the input image by applying a set of filters (or kernels). Each filter captures a small feature and the degree of similarity between the filter and a portion of the image determines its output. The resulting outputs are then processed through non-linear activation functions, such as sigmoid or Rectified Linear Unit (ReLU), to incorporate non-linearity into the network. The final decision-making process, typically classification, is handled by fully connected layers (Nichols et al., 2018). Figure 2.10 illustrates a common CNN architecture used for image classification, generally comprising convolutional layers, pooling layers, activation function, and fully connected layer.

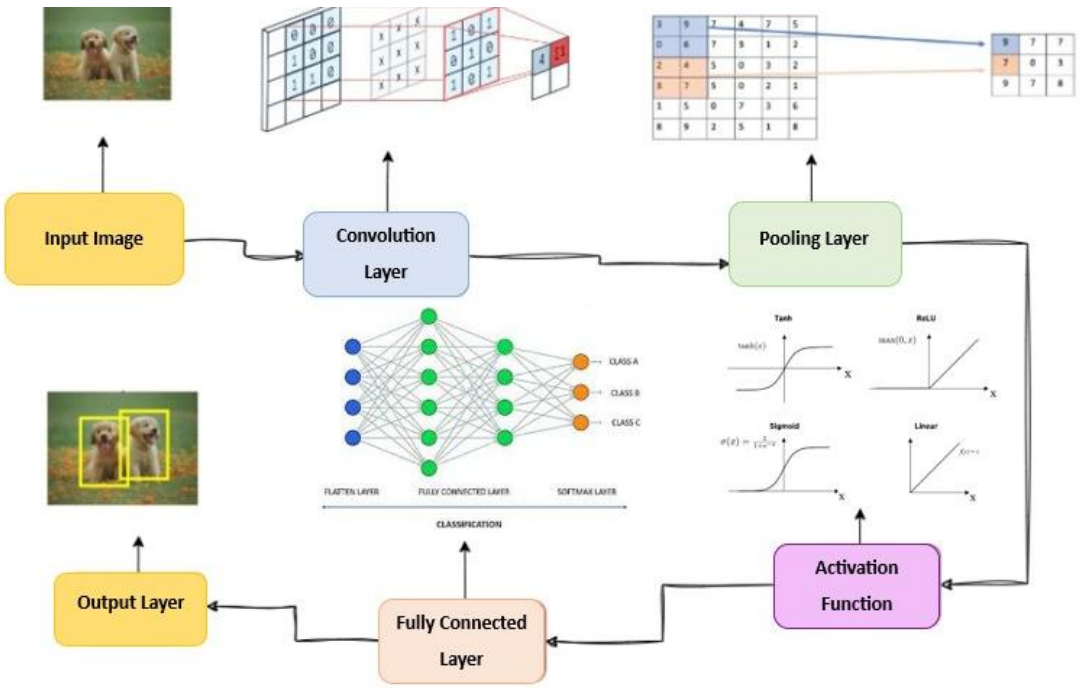


Figure 2.10: Common architecture of CNN for image classification (Taye, 2023)

a) Convolutional Layer

In image classification, an image is fed into the input layer of a CNN, where it is interpreted as a matrix of numerical values. Each number in this matrix represents the light intensity at a particular pixel. These numerical representations, commonly referred to as

matrices. A key concept in CNN architecture is the receptive field that describes local connectivity where neurons in one layer are linked to a subset of neurons in the preceding layer, capturing localized patterns from the input image (Indolia et al., 2018). The connection between neurons is done via a weight vector, commonly referred to as a filter or kernel. This filter is applied across the entire input matrix, and since it is shared across spatial locations, the network can recognize specific features irrespective of their position within the image (Indolia et al., 2018).

The convolution operation is performed by sliding the filter across the input image in a systematic manner to compute feature maps, which highlight specific features detected at various locations. When *N* filters are applied, the result is *N* corresponding feature maps, each capturing different visual characteristics of the input image (Indolia et al., 2018). The filter, although having the same dimensional size as the input, is smaller in spatial dimensions than the input image. As the filter slides across the input image, it computes the dot product between its weight and the corresponding image region, producing a two-dimensional feature map, as depicted in Figure 2.11. According to Indolia et al. (2018), the output value a_{ij} at location *(i, j)* in the feature map, resulting from the convolution operation, can be calculated using Equation 2.1,

$$a_{ij} = \sigma\big((W * X)_{ij} + b\big) \qquad \text{Equation 2.1}$$

where *X* is the input, *W* is the filter or kernel that slides over the input, σ is the activation function introducing non-linearity in the network, * represents the convolution operation, and *b* is the bias term. Each convolutional layer is defined by several parameters, including the kernel size, stride (the number of pixels the kernel slides at each step), and padding (the size of the zero-padding pixels added around the input feature map) (Purwono et al., 2023).

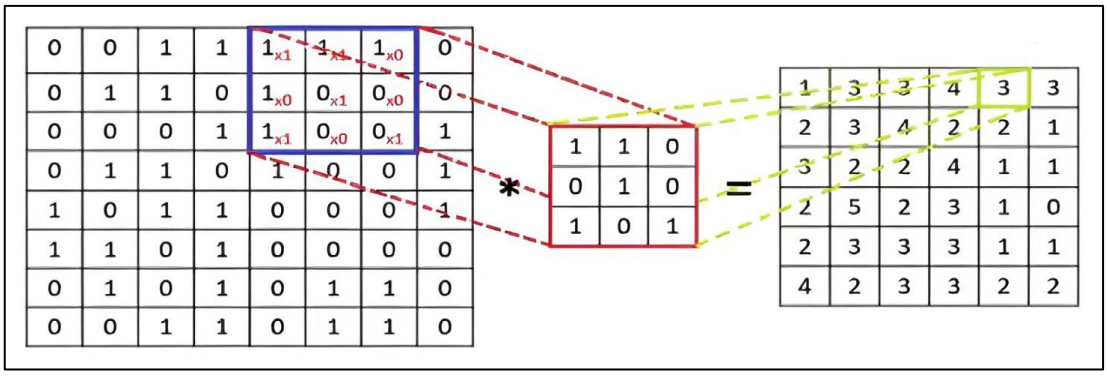


Figure 2.11: Illustratiion of the convolution operation (Purwono et al., 2023)

b) Pooling Layer

Once a feature is detected, its precise location becomes less significant. Consequently, pooling or subsampling layers are usually placed after convolutional layers to down-sample feature maps and reduce their spatial dimensions (Yamashita et al., 2018). This process significantly reduces the number of trainable parameters and computational loads while introducing translation invariance, enabling the model to detect features regardless of their location (Indolia et al., 2018; Purwono et al., 2023). Additionally, pooling layers helps mitigate the risk of overfitting. To perform a pooling operation, a filter is applied to regions of the input feature map, and the values are processed using a specific function.

The most prevalent technique is max pooling, which involves segmenting the input feature maps into small regions and selecting the highest value from each region, discarding the remaining values. Figure 2.12 shows an example of max pooling operation using a 2×2 filter with a stride of 2, a configuration commonly used in practical applications, which reduces the in-plane dimensions by a factor of two (Yamashita et al., 2018). Another technique is global average pooling, which compresses each feature map into a 1×1 array by averaging all its values. Typically applied once prior to the fully connected layer, pooling reduces the spatial dimensions of the feature map while retains its depth.

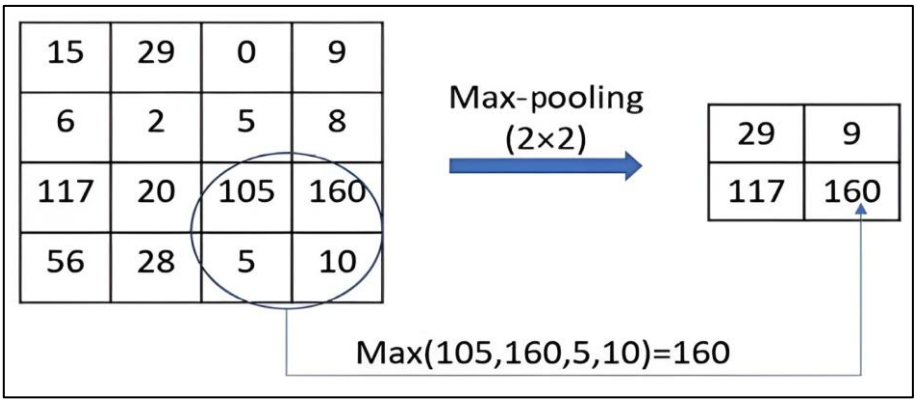


Figure 2.12: Illustration of max-pooling for dimension reduction (Purwono et al., 2023)

c) Fully Connected Layer

The fully connected layer—also referred to as dense layer or convolutional output layer, is usually positioned at the final stage of the network (Purwono et al., 2023). At this point, the feature maps produced by the final convolutional or pooling layer are flattened, transforming them into a one-dimensional vector. This flattened feature vector is fed into one or more fully connected layer, which are responsible for translating the learned features into final outputs, such as class probabilities in classification tasks. Typically, each fully connected layer is followed by a non-linear activation function, and the final layer includes one output node per class in the classification task. The final outputs are commonly processed through a softmax activation function to generate a probability distribution across the predicted classes (Purwono et al., 2023).

d) Activation Function (Non-Linearity Layer)

Activation functions or non-linearity layers play an essential role in CNNs. Its main objective is to determine the final output of a neural network. The activation function maps output values to a specific range, such as between -1 and 1 or 0 and 1, depending on the function used. According to Purwono et al. (2023), activation function can be broadly classified into two categories:

i. **Linear Activation Functions:** These functions can generally be expressed in a simplified form as $F(x) = CY$, where C is a constant representing the weight of each neuron. The output of these functions is proportional to their input. While linear activation functions provide a single definitive answer of yes or no, they are limited in expressiveness and generally not suitable for learning complex mapping.

ii. **Non-linear Activation Functions:** Modern neural networks commonly utilized these functions to establish complex relationship between inputs and outputs, which are essential for intricate learning and modelling tasks. Among them, the Rectified Linear Unit (ReLU) is the most widely used non-linear activation function in CNNs, while other common activation functions include sigmoid, softmax, and hyperbolic tangent (tanh), and Leaky ReLU (Balakrishnan et al., 2021; Purwono et al., 2023; Taye, 2023).

Table 2.1 summarizes several commonly used activation functions in CNNs and other neural networks architecture.

Table 2.1: Commonly Used Activation Functions in Deep Learning

Activation Function	Explanation
Sigmoid	The sigmoid function takes real numbers as inputs and maps the output to a range between 0 and 1, forming a S-shaped curve. It is mathematically represented as Equation 2.2. $f(x) = \frac{1}{1 + e^{-x}}$ Equation 2.2 It is usually used in binary classification problems.

Table 2.1 continued

Softmax	The softmax function is commonly applied in the final layer of neural networks for multiclass classification, producing a probability distribution over all possible classes, with each values ranging from 0 and 1. It is mathematically defined as Equation 2.3. $f(x_i) = \frac{e^x}{\sum_j e^{x_j}}$ Equation 2.3
Hyperbolic Tangent (tanh)	Like the sigmoid function, the hyperbolic tangent function accepts real-valued inputs but constrains its output within the range of -1 and 1. Its mathematical representation is presented as Equation 2.4. $f(x) = \frac{e^x + e^{-x}}{e^x + e^{-x}}$ Equation 2.4 This function is mostly employed for NLP tasks and recognition in RNNs.
Rectified Linear Unit (ReLU)	The Rectified Linear Unit (ReLU) function outputs zero for any negative input while retains positive values as they are. It is favoured for its relatively lower computational load and improved computational speed. Mathematically, ReLU can be represented as Equation 2.5. $f(x) = \max(0, x)$ Equation 2.5
Leaky ReLU	The Leaky ReLU function mitigate the dying ReLU issue by permitting a small, non-zero gradient for negative input, ensuring that all inputs contribute to learning. The mathematical representation for leaky ReLU is as in Equation 2.6. $f(x) = \begin{cases} x, & x > 0 \\ mx, & x \leq 0, m \text{ is a small constant} \end{cases}$ Equation 2.6

Note. Mathematical equations are adapted from Balakrishnan et al. (2021), Purwono et al. (2023), and Taye (2023).

With the foundational understanding of CNN architecture and its core components established, it is essential to explore how these principles have been implemented in real-world applications. Over the years, various CNN architectures have been developed, each incorporating specific innovations aimed at enhancing performance, minimizing computational complexity, or tackling specific challenges associated with computer vision tasks. Yann LeCun and his colleagues developed LeNet-5, the first CNN, specifically designed for handwritten digits recognition (LeCun et al., 1989). The model was trained on the Modified National Institute of Standards and Technology (MNIST) dataset, comprising thousands of grayscale images depicting handwritten digits. This work marked a significant breakthrough, showcasing the potential of CNNs in performing image classification.

The success of LeNet-5, combined with subsequent advancements in computational power and increased availability of data, catalysed the widespread adoption of CNNs across various domains. Since then, the field has seen the development of a wide variety of CNN architectures, each tailored to specific application requirements and learning capabilities. The diversity in CNN designs often lies in the number of layers, depth, and architectural configurations. Notable examples include Visual Geometry Group (VGG), Inception, and Residual Network (ResNet), each of which has consistently achieved high accuracy on benchmark computer vision tasks. The following section introduces several prominent CNN architectures, including AlexNet, VGGNet, ResNet, GoogLeNet/Inception, DenseNet, and EfficientNet, each of which has significantly contributed to advancements in computer vision and deep learning for image-based analysis.

2.4.2.1 AlexNet

AlexNet, a CNN architecture implemented using GPU acceleration, was introduced by Alex Krizheysky, Ilya Sutskever and Geoffrey E. Hinton. It gained prominence by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 (Krizhevsky et al., 2017). AlexNet comprises eight layers: five convolutional layers and three fully connected layers, as illustrated in Figure 2.13. A defining technical innovation was the replacement of traditional tanh activations with the ReLU function, which significantly accelerated training convergence (Krizhevsky et al., 2017). Architectural down-sampling is achieved through max pooling following the first, second, and fifth convolutional layers, culminating in a 1000-class softmax output. To manage the overfitting risks inherent in its 60 million parameters, AlexNet utilizes data augmentation—including image translations, horizontal reflections, and Principal Component Analysis (PCA) for colour intensity adjustment—alongside dropout. By randomly omitting 50% of neurons in the first two fully connected layers during training, dropout reduces complex neuron co-adaptation, thereby substantially enhancing model's generalizability.

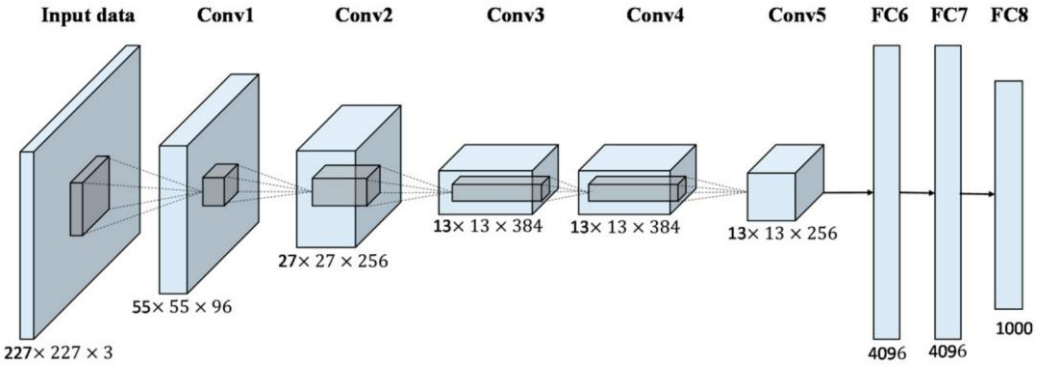


Figure 2.13: AlexNet architecture (Han et al., 2017)

2.4.2.2 Visual Geometry Group Network (VGG)

The Visual Geometry Group (VGG) architecture was developed by Karen Simonyan and Andrew Zisserman from the University of Oxford, and it achieved prominence in

ILSVRC 2014 by demonstrating that increasing network depth by adding more convolutional layers is a primary driver of performance (Simonyan & Zisserman, 2014). The design's core innovation is the consistent use of small 3×3 convolutional filters to capture spatial information in all directions—left, right, up, down, and centre—while preserving spatial resolution after convolution operation (Simonyan & Zisserman, 2014).

VGG16, the most widely used variant, consists of 13 convolutional layers and three fully connected layers, as illustrated in Figure 2.14. It employs 2×2 max pooling across five stages to systematically down-sample the feature map size by 50% (Purwono et al., 2023). The initial two fully connected layers in VGG16 each contain 4,096 neurons, while the third layer includes 1,000 neurons, aligning with the number of output classes. The final classification layer uses the softmax activation function, and ReLU is applied across all hidden layers (Simonyan & Zisserman, 2014). VGG16 requires an input image with a fixed resolution of 224×224 pixels. Notably, VGG16 maintains a minimalist preprocessing approach, involving only the subtraction of the mean RGB values calculated on the training dataset.

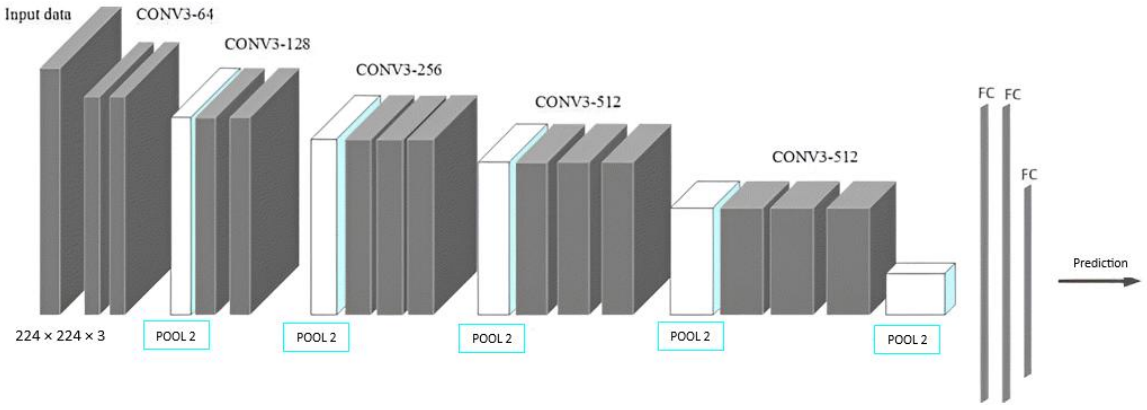


Figure 2.14: VGG16 architecture (Purwono et al., 2023)

2.4.2.3 Inception

Since LeNet, the fundamental structure of CNNs has included stacked convolutional layers, optional normalization, max pooling, and fully connected layers. Variations of this structure have become standard in the image classification research and have consistently delivered top performance in challenges like the ILSVRC. One straightforward approach to enhancing the performance of the deep neural networks is to increase their size—by adding more layers or increasing the size of existing layers—while using dropout to reduce overfitting. This approach, however, has two significant drawbacks: a heightened risk of overfitting due to increase parameters and a sharp rise in required computational resources.

To address these limitations, Szegedy, Liu, et al. (2015) introduced sparsity into deep networks, by replacing fully connected layers with sparse layers. In 2014, Christian Szegedy and colleagues at Google introduced GoogLeNet, the first iteration of the Inception architecture. The name "GoogLeNet" combines "Google" and "LeNet", reflecting both its origin and inspiration. This architecture aimed to increase deep neural networks' efficiency by reducing the number of parameters without compromising accuracy, thereby tackling the challenges of computational limitations in deeper models. GoogLeNet utilizes 22 layers, broadly organized into three parts: a stem of initial convolutional layers, ReLU activations and max pooling; the output classifier using dense layer and softmax activation function as last part; and the core inception module in the middle part.

Departing from purely sequential arrangement, GoogLeNet employs parallel structures within its inception models. These modules perform multiple convolutions and pooling operations simultaneously, concatenating the results into a single output vector. This design enables the network to capture both local and global features while minimizing

parameters counts. (Szegedy, Liu, et al., 2015). In the naïve version (Figure 2.15), 1×1, 3×3, and 5×5 filters are applied in parallel to extract features at different scales. To manage the computational complexity of increasing depth, the authors added a 1×1 "bottleneck" convolutional layer prior to the 3×3 and 5×5 convolutions, as well as following the max pooling layer (Figure 2.16). These modules serve as dimensionality reduction layers, allowing for deeper and wider networks without noticeably sacrificing performance (Szegedy, Liu, et al., 2015).

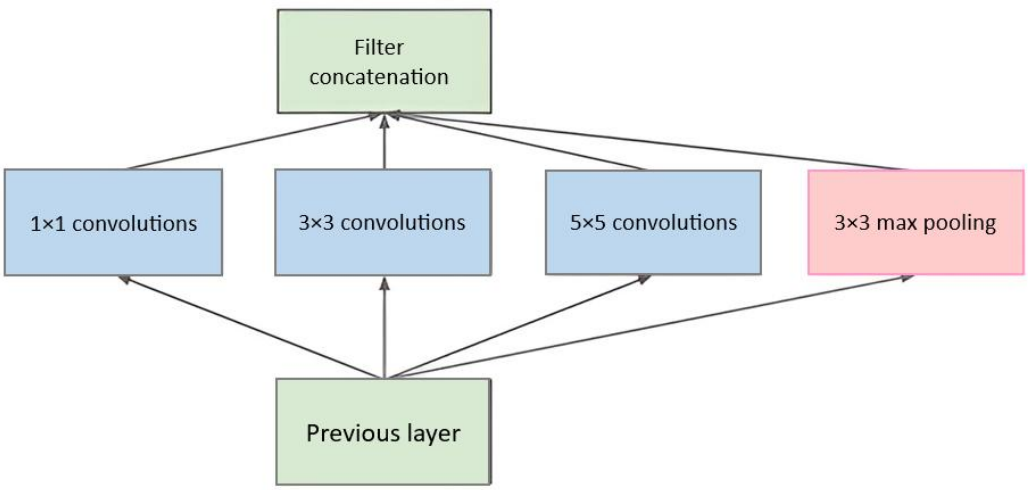


Figure 2.15: Naïve version of the Inception module (Szegedy, Liu, et al., 2015)

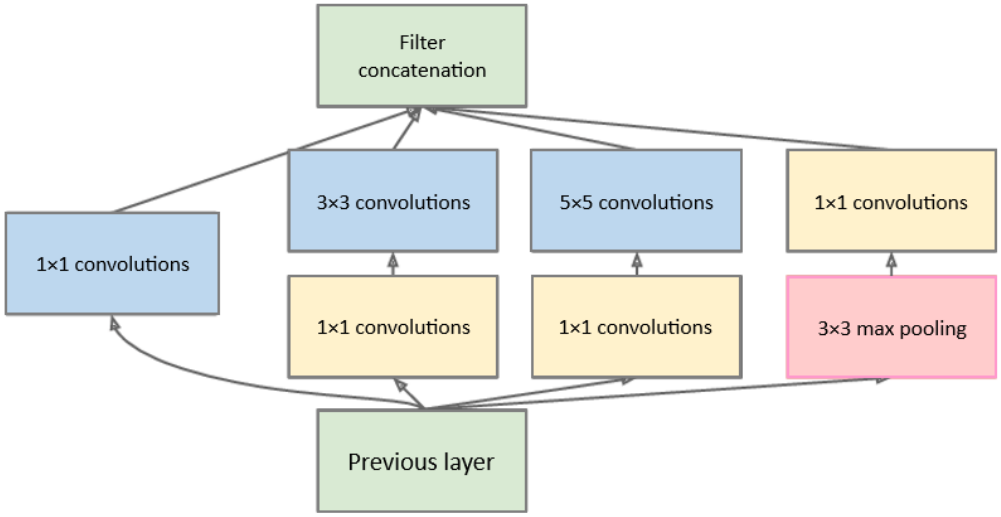


Figure 2.16: Dimension reduction in Inception module (Szegedy, Liu, et al., 2015)

To mitigate the vanishing gradient problem in its 22-layer depth, GoogLeNet incorporates two auxiliary classifiers. These compute intermediate losses over the same class labels, applying softmax to the output of the middle inception modules to provide additional supervision (Szegedy, Liu, et al., 2015). A schematic view of the complete GoogLeNet architecture is shown in Figure 2.17. Later renamed Inception-v1, this model set a precedent for a series of improved architectures. To further improve efficiency, Inception-v2 introduced batch normalization to reduce internal covariate shifts—a change in the network activations' distribution caused by changes in network parameters during training (Ioffe & Szegedy, 2015; Szegedy, Vanhoucke, et al., 2015). By stabilizing these distributions, batch normalization permits higher learning rates and less meticulous parameter tuning, significantly accelerating the training process (Ioffe & Szegedy, 2015).

Another major improvement in Inception-v2 was the use of spatial factorization, where larger convolutions are factorized into smaller convolutions. For example, 5×5 convolutional filters (Figure 2.18(a)) were replaced by two successive 3×3 filters (Figure 2.18(b)), reducing parameters and computational cost without degrading performance. Similarly, $n \times n$ convolutions were replaced by combinations of $1 \times n$ and $n \times 1$ convolutions, which significantly reduces computational cost as n grows. The initial 7×7 convolutional layer was factorized into a series of 3×3 convolutional layers. Inception-v2 also introduced label smoothing to regularize the output of the network and improve generalization performance (Szegedy, Vanhoucke, et al., 2015).

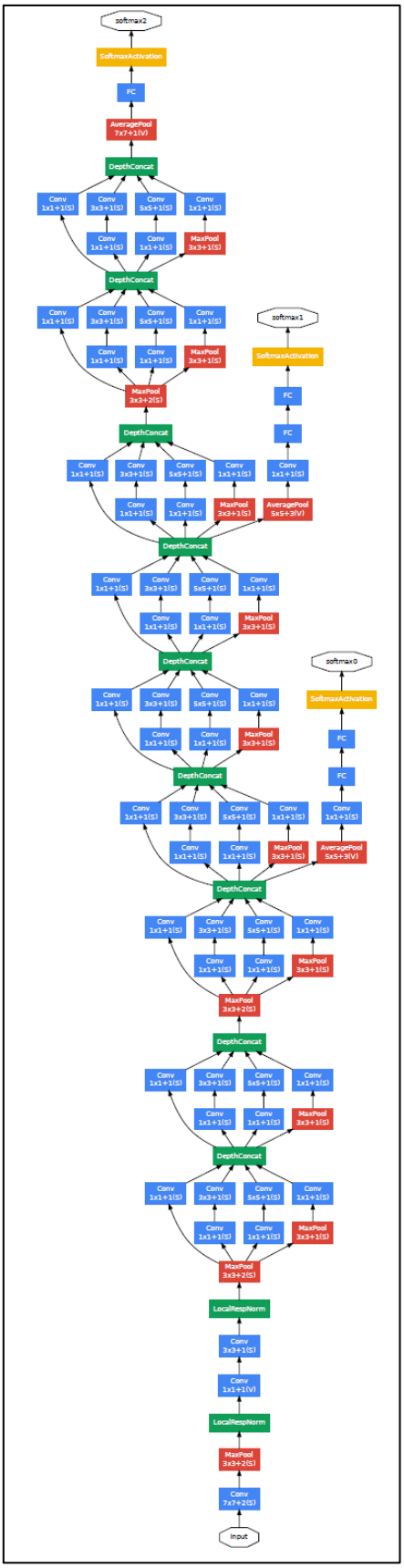
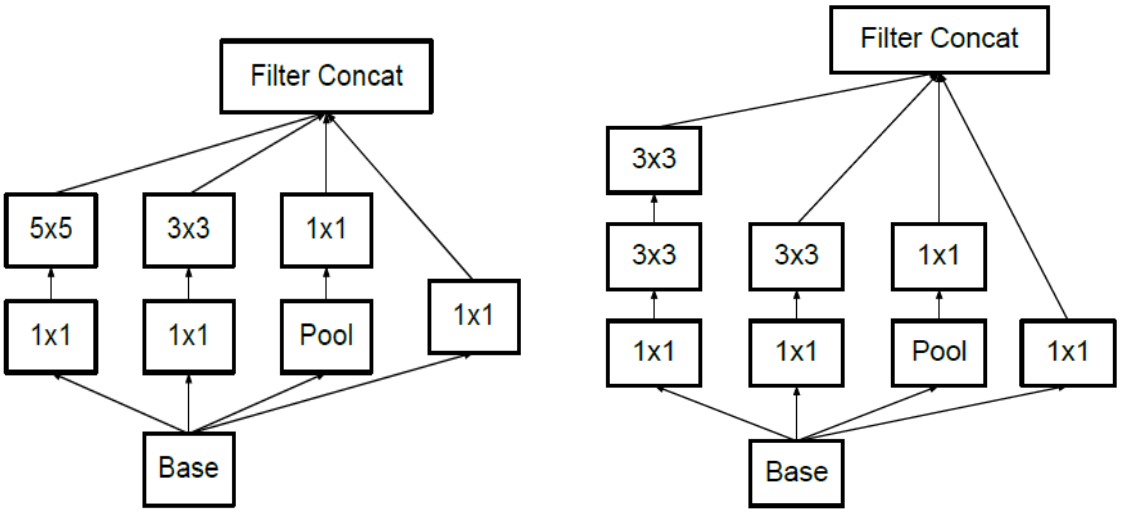


Figure 2.17: GoogLeNet architecture (Szegedy, Liu, et al., 2015)



(a) (b)

Figure 2.18: Spatial factorization in Inception modules: (a) Original Inception module in Inception-v1, (b) Inception module with factorised convolutions (Szegedy, Vanhoucke, et al., 2015).

Inception-v3 retained all features from Inception-v2 while adding batch normalization to the auxiliary classifiers—normalizing both convolutional and fully connected layer—further stabilize training and boost performance (Szegedy, Vanhoucke, et al., 2015). This iteration achieved state-of-the-art performance on ImageNet with superior computational efficiency. Subsequently, Inception-v4 incorporated residual connections into the inception modules, resulting in hybrid models such as Inception-ResNet-v1 and Inception-ResNet-v2 (Szegedy et al., 2016). These models combined the advantages of residual learning with the efficiency of Inception modules, achieving enhanced performance across various image recognition benchmarks.

2.4.2.4 Residual Network (ResNet)

Following the success of AlexNet, architectural trends shifted toward increasing the depth to reduce error rates, progressing from eight layers in AlexNet to 16 layers in VGG16. However, beyond a specific threshold, deeper networks suffer from accuracy saturation and degradation. This is primarily caused by the vanishing or exploding gradient problem, where

gradients become too small or excessively large during backpropagation, worsening both, training and testing error. To address this problem, He et al. (2016) introduced skip connections, which allow the network to learn residual functions rather than direct mapping. This innovation paved the way for creation of the Residual Network (ResNet) architecture by Microsoft Research in 2015.

ResNet variants, such as ResNet50, ResNet101 and ResNet152, which consists of 50, 101, and 152 layers, utilize these connections to maintain performance across deeper layers. In 2015, a ResNet ensemble won the ILSVRC with a top-5 error rate of 3.57%, while also leading in ImageNet detection and localization as well as in Common Objects in Context (COCO) detection and segmentation tasks (He et al., 2016). Skip connections provide alternative paths for gradients by connecting the activations of a layer to layers further downstream, bypassing intermediate layers. This structure forms what is known as a residual block, as illustrated in Figure 2.19. In this structure, the network learns a residual mapping defined as $F(x) = H(x) - x$, which can be rewritten as $H(x) = F(x) + x$. This formulation ensures that deeper layers do not degrade the model performance relative to their shallower counterparts.

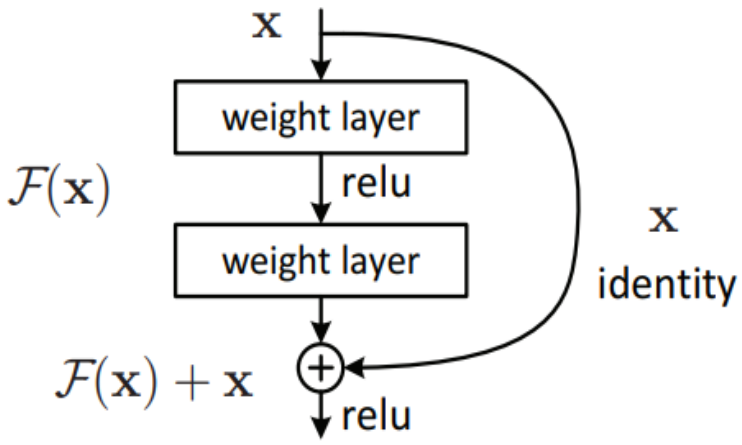


Figure 2.19: Illustration of skip connection (He et al., 2016)

The initial ResNet34 architecture was developed by converting a VGG-inspired plain network into a residual version through the insertion of skip connections (He et al., 2016). While both architectures utilize 3×3 filters, ResNets maintains lower complexity and fewer filters compared to VGG. Figure 2.20 illustrates the architecture of VGG19, the plain version of ResNet34 and its residual counterpart, showing the incorporation of skip connection through stacked two-layer residual blocks. The ResNet50 architecture further optimized the design by replacing the standard two-layer residual block of ResNet34 with three-layer bottleneck blocks to manage computational demands in deeper configurations.

2.4.2.5 Densely Connected Convolutional Network (DenseNet)

In deep CNNs, effective gradient propagation is often hampered as layers increases. This issue is particularly pronounced in early layers, where vanishing gradients during backpropagation impede the updating of weights and biases, diminishing learning capacity. While ResNet addressed this using skip connections to produce identity mappings via summation, this approach can still limit information flow throughout the model. To further improve information flow, Huang et al. (2017) proposed Densely Connected Convolutional Network (DenseNet), an extension of traditional CNN architectures that builds upon and refines the ResNet framework. While a conventional CNN with L layers has only L connections (each layer connected to its immediate successor), and ResNet has $2L$ connections, DenseNet introduces direct connections from each layer to every succeeding layers, as illustrated schematically in Figure 2.21, resulting in a $\frac{L(L+1)}{2}$ total connections (Huang et al., 2017).

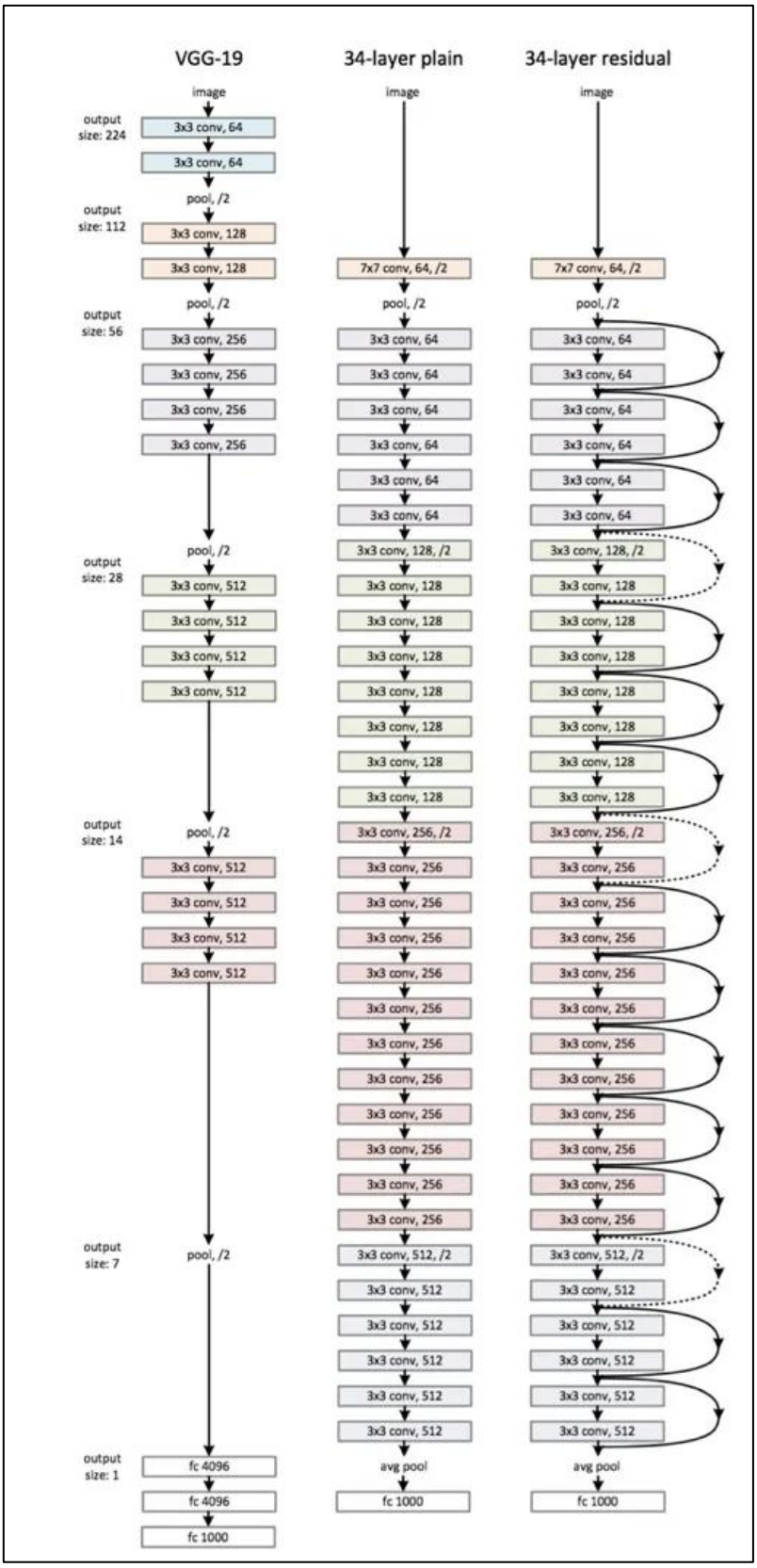


Figure 2.20: Comparison of the architectures of VGG19, Plain ResNet34 and ResNet34 with residual blocks (He et al., 2016).

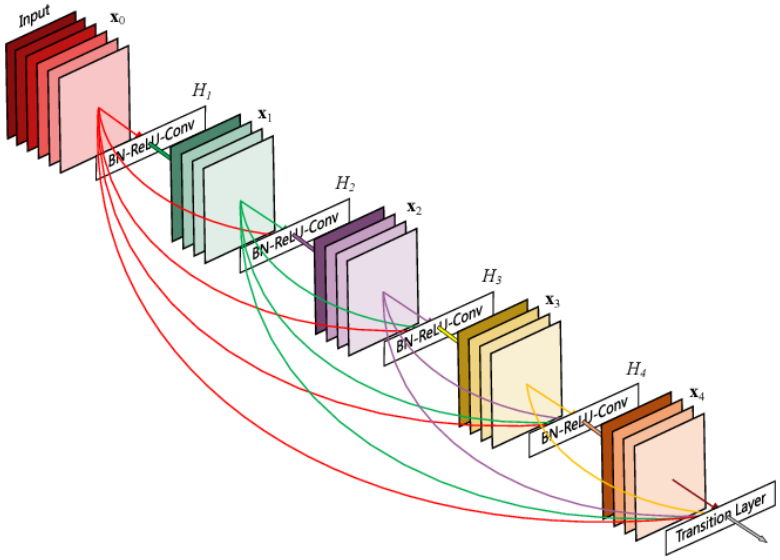


Figure 2.21: Direct connection in DenseNet, each layer takes all preceeding feature maps as input (Huang et al., 2017).

Within a DenseNet, layers are organized into dense block, as visualized in Figure 2.22. Each layer ℓ performs a composite non-linear transformation $H_\ell(\bullet)$ comprising batch normalization, ReLU activation, and a 3×3 convolution. If each transformation produces k feature maps—a value known as the growth rate—then the input to the ℓ^{th} layer comprises $k_0+k\times(\ell-1)$ feature maps, where k_0 is the number of channels in the input layer. This architecture promotes parameter efficiency and reduces redundancy, as each layer has access to the global state of the network through preceding feature maps. However, the number of parameters can still increase as each layer receives growing number of input feature maps. Therefore, the authors inserted a 1×1 convolutional bottleneck layer before each 3×3 convolution within the dense block (Huang et al., 2017).

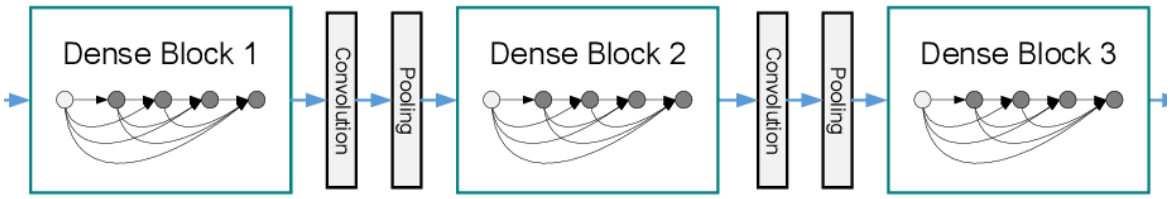


Figure 2.22: Transition layers between two dense blocks in DenseNet (Huang et al., 2017)

In addition to direction connection, Huang et al. (2017) alleviated vanishing gradients and performance degradation by reusing preceding feature maps, resulting in a more compact DenseNet representation. Unlike traditional CNNs that passed only the most recent output, DenseNet concatenates the current output with all preceding feature maps, allowing for more efficient information preservation. However, as convolutional operations reduce spatial resolution, compatibility issues arise when concatenating feature maps from different layers. To resolve this, Huang et al. (2017) introduced transition layer between dense blocks. These layers consist of batch normalization, a 1×1 convolutional layer, and a 2×2 average pooling (Figure 2.22). This architectural design enhances computational efficiency while reducing the number of parameters.

DenseNet is available in multiple configurations, such as DenseNet-121, DenseNet-169, and DenseNet-201 (each with a growth rate of $k = 32$), and DenseNet-161 ($k = 48$). These models were assessed on the ImageNet validation set using both single-crop and 10-crop testing methods, these models achieved top-1 error rates of 25.02% (23.61%), 23.80% (22.08%), 22.58% (21.46%), and 22.33% (20.85%), respectively. DenseNets achieve performance level comparative to leading ResNet architectures, while utilizing substantially fewer parameters and computational resources. For instance, DenseNet-201, with only 20 million parameters, achieves similar performance to 101-layer ResNet, which has over 40 million parameters.

2.4.2.6 EfficientNet

EfficientNet, developed by Tan and Le (2019) at Google, introduced a new approach to model scaling called compound scaling. Traditionally, CNNs were scaled up to improve accuracy by increasing a single dimension: the depth of the network (as seen in ResNet),

widening the network or—less commonly—increasing the image resolution. In this context, depth refers to the number of layers, width to the number of channels per layer, and resolution to the spatial dimension of the input image. Although arbitrarily scale two or all three dimension is possible, doing so often lead to suboptimal accuracy and efficiency and requires laborious manual tuning (Tan & Le, 2019).

Tan and Le (2019) found that improved performance could be achieved by jointly scaling the width, depth, and resolution in a balance way. They proposed a compound scaling method using a fixed set of coefficients to uniformly scale all three dimensions in a simple yet highly effective manner. This concept is illustrated in Figure 2.23: (a) displays a baseline network; (b), (c), (d) depicts conventional scaling individual dimensions; and (e) shows the compound scaling approach. As explained by Tan and Le (2019), if 2^{ϕ} times more computational resources are used, then the width, depth, and resolution of the network can be increased by α^{ϕ}, β^{ϕ}, and γ^{ϕ}, respectively. The compound scaling is governed mathematically by Equation 2.7:

$$\text{Depth } d = \alpha^{\phi}, \text{Width } w = \beta^{\phi}, \text{Resolution } r = \gamma^{\phi} \qquad \text{Equation 2.7}$$

such that

$$\alpha \bullet \beta^{2} \bullet \gamma^{2} \approx 2 \text{ and } \alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

In this formulation, α, β, γ are constants determined via a small grid search conduction on a baseline model, while ϕ represents a user-defined scaling coefficient controlling available resources. Since the computational cost of a convolutional operation, measured in Floating-Point Operations Per Second (FLOPS), is proportional to the square of the input resolution, width, and depth, scaling the network using Equation 2.7 increases FLOPS by a factor of

$(\alpha \bullet \beta^2 \bullet \gamma^2)^{\phi}$. Tan and Le (2019) limited $\alpha \bullet \beta^2 \bullet \gamma^2$ to be approximately 2, so the total FLOPS increases by 2^{ϕ} for any new ϕ.

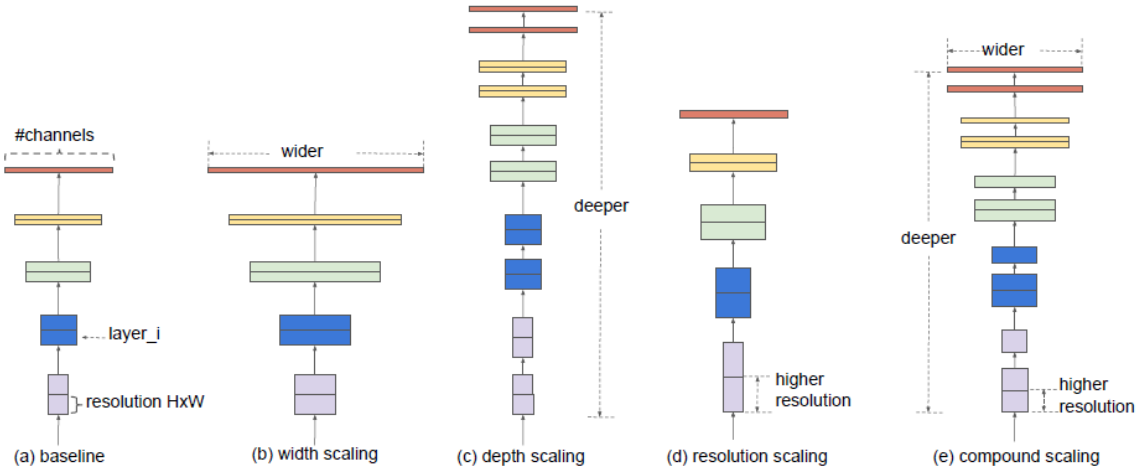


Figure 2.23: Illustration of conventional scaling and proposed compound scaling (Tan & Le, 2019)

Tan and Le (2019) initially applied compound scaling to existing architectures such as MobileNet and ResNet, where they observed improved accuracy compared to single-dimension scaling. They then deployed neural architecture search to design a new baseline model, EfffficientNetB0, which adopts the Mobile Inverted Bottleneck Convolution (MBConv)—technique originally introduced in MobileNetV2—as the primary building block and squeeze-and-excitation optimization. By fixing the derived optimal coefficients (α = 1.20, β = 1.10, γ = 1.15.) and varying ϕ, the authors produced a family of models from EfficientNet-B1 to EfficientNet-B7 (Tan & Le, 2019). These models outperformed earlier CNN architectures in terms of both accuracy and efficiency, as shown in Figure 2.24, which compares their top-1 accuracy and number of parameters. For instance, EfficientNet-B7 attained a top-1 accuracy of 84.3% on the ImageNet dataset, while being 8.4 times more compact and 6.1 times faster during inference than earlier high-performing CNNs.

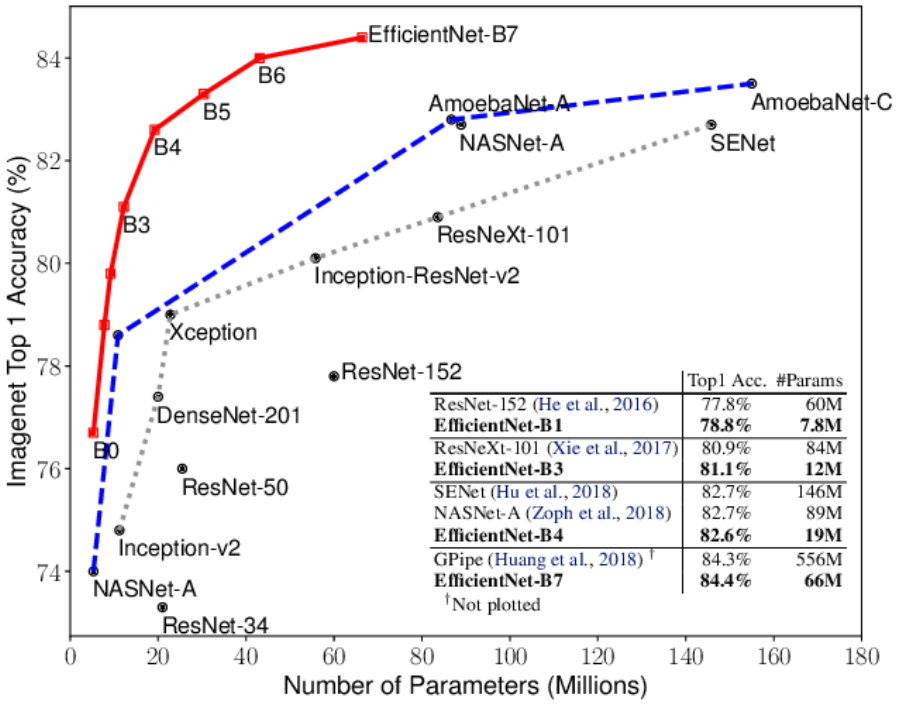


Figure 2.24: Comparison of EfficientNets and other exisiting CNNs on ImageNet Top-1 Accuracy and Parameters Count (Tan & Le, 2019)

The evolution of CNNs has been marked by continuous innovation aimed at improving accuracy, efficiency, and scalability. Beginning with AlexNet, which demonstrated the potential of deep learning on large scale image datasets, subsequent architectures such as VGG and Inception refined network design through deeper layers and multi-path processing. ResNet addressed the degradation problem in deep network through residual connections, enabling the training of extremely deep models. DenseNet further enhanced information flow by introducing dense connectivity, while EfficientNet introduced a principled compound scaling method to balance depth, width, and resolution for optimal performance. Together, these architectures form the foundation of modern CNNs and have significantly advanced the field of computer vision.

2.5 Review of Related Work

In recent year, deep learning models, particularly CNNs, have been increasingly applied in the medical domain. These models are establishing new benchmarks in medical

image analysis due to their capacity to automatically learn and extract intricate features. This makes them well-suited for detecting subtle pattern in medical imagery. In ophthalmology, CNNs have played a pivotal role in advancing automated eye disease diagnosis. There have been notable advancements in the detection and grading of individual eye diseases through classification methods, segmentation techniques, or a combination of both. Recent studies has primarily explored the diagnostic capabilities of CNNs for four major eye diseases—age-related macular degeneration, cataracts, diabetic retinopathy, and glaucoma—using widely available colour retinal fundus images (Hemelings et al., 2021). This section reviews studies published between 2021 and the present that employ CNN-based architectures for eye disease classification using retinal fundus images, encompassing both binary and multiclass classification approaches. This review aims to identify prevailing trends, methodological innovations, and existing challenges in the development of robust deep learning models for eye disease classification.

2.5.1 Binary Classification of Eye Diseases

Binary classification of eye diseases typically involves distinguishing between the presence and absence of a specific eye disease. This approach simplifies the diagnostic process by focusing on whether a condition exists, such as identifying diabetic retinopathy versus a healthy retina. Numerous studies have employed deep learning models, particularly CNNs, to achieve high accuracy in binary classification tasks using retinal fundus images. This section reviews key research efforts that utilize binary classification techniques to detect individual eye diseases, highlighting model architectures, datasets and performance metrics.

Emir and Çolak (2023) evaluated the performance of three pre-trained models—VGG16, Inceptionv3, and ResNet50—in classifying various ophthalmological diseases

using the ODIR-5K dataset. The classification task was binary, assigning each patient's fundus images to either disease-positive or disease-negative category based on diagnostic keywords corresponding to eight disease classes. VGG16 achieved the highest scores for normal (AUC = 0.717, Final = 0.557) and diabetic retinopathy (AUC = 0.677, Final = 0.528). Inception-v3 performed best for age-related macular degeneration (AUC = 0.748, Final = 0.663) and other unspecified conditions (AUC = 0.621, Final = 0.481), while ResNet50 yielded the highest score for glaucoma (AUC = 0.762, Final = 0.664), cataract (AUC = 0.964, Final = 0.903), hypertension (AUC = 0.764, Final = 0.626) and myopia (AUC = 0.959, Final = 0.836). These results highlight the varying strengths of different architectures across the disease types.

Bakir and Yilmaz (2022) proposed a deep learning-based framework for the early diagnosis of cataract using retinal fundus images. Their approach utilized transfer learning for feature extraction, employing four pre-trained CNN architectures: VGG-16, ResNet, InceptionV3, and MobileNet. Features extracted from these models were classified using a simple neural network. The study focused on binary classification between normal and cataract-affected eyes using a subset of the Ocular Disease Intelligence Recognition (ODIR-5K) dataset. Among the tested models, ResNet achieved the highest accuracy of 95.51%, showing the effectiveness of transfer learning in cataracts fundus images classification. Furthermore, Singh et al. (2023) employed the VGG16 models for cataracts detection using retinal fundus images, achieving a notable classification accuracy of 96.10%.

Bernabe et al. (2021) proposed a CNN model for classifying eye diseases, specifically diabetic retinopathy (negative class) and glaucoma (positive class), using pre-processed retinography images. The dataset comprised 650 images from the ORIGA dataset and 397 from the DIARETDB0 dataset. The CNN architecture consists of two convolutional

layers, followed by max-pooling and a fully connected multilayer perceptron. The model achieved an accuracy of 99.89%, with performance metrics such as recall, specificity, precision, and F1-score all approaching 1. These results underscore the potential of CNN-based approaches in effectively distinguishing between major eye diseases.

Dai et al. (2021) introduced DeepDR, a ResNet-based deep learning architecture designed for real-time evaluation of image quality, identification of retinal lesion, and grading of diabetic retinopathy from early to advanced stages. The model was pre-trained on ImageNet and fine-tuned using a real-world dataset of fundus images from diabetic patients. DeepDR achieved high performance in detecting retinopathy-related abnormalities, with area under the receiver operating characteristic curve (AUC) values of 0.901 for microaneurysms, 0.941 for cotton-wool spots, 0.954 for hard exudates, and 0.967 for hemorrhages. For severity grading, the system attained AUC values ranging from 0.943 to 0.972. Its robustness was further validated on three external datasets—EyePACS-1, MESSIDOR 2, and the Kaggle dataset, where AUC values for grading ranged from 0.916 to 0.970.

Sakri et al. (2021) introduced an automated classification system for diabetic eye disease, integrating image enhancement methods including illumination correction and contrast-limited adaptive histogram equalization (CLAHE), along with image segmentation techniques to identify blood vessels, the macular region and the optic nerve. Their study employed several pre-trained models (VGG-16, Xception, ResNet50) and a custom CNN, all of which achieved classification accuracies exceeding 90%. The findings demonstrated that transfer learning, leveraging pre-trained models, played a significant role in improving diagnostic performance.

Gheisari et al. (2021) investigated the effectiveness of deep learning in glaucoma detection by combining spatial and temporal feature extraction. Spatial features were extracted from fundus images using CNN, specifically VGG16 and ResNet50, while temporal features were derived from fundus video using recurrent neural networks (RNNs), particularly LSTM units. The integrated CNN-RNN architecture significantly enhanced glaucoma detection performance, achieving higher specificity and sensitivity compared to standalone VGG16 or ResNet50 models.

Hemelings et al. (2021) proposed a methodology to enhance explainable deep learning for detection of glaucoma and analysis of vertical cup-disc ratio. Their model was trained on retinal fundus images subjected to a cropping strategy, where the crop radius—defined as a percentage of the image dimensions—was centred on the optic nerve head (ONH) and applied across uniformly spaced intervals from 10% to 60%. The model trained on original uncropped images achieved an AUC of 0.94 for glaucoma detection, while models trained on images excluding the ONH still performed well (AUC = 0.88). These findings indicates that the deep learning models can utilize peripheral retinal features beyond the ONH region for effective glaucoma detection.

Sandoval-Cuellar et al. (2021) developed a 15-layer CNN for glaucoma detection inspired by existing architectures such as AlexNet and VGG16. Using the ORIGA dataset, the model achieved an accuracy of 93.22%, sensitivity of 94.14%, and AUC of 93.98%, demonstrating its effectiveness in identifying glaucomatous features. Similarly, Shoukat et al. (2022) proposed an automated deep learning approach for glaucoma diagnosis using retinal fundus images. Their method involved grayscale conversion and data augmentation to enhance image quality and increase dataset diversity. The study employed transfer learning with three pre-trained models—VGG16, ResNet50, and EfficientNetB7—trained

and evaluated on three datasets: RIM-ONE, G1020, and REFUGE. Among these, EfficientNet-B7 demonstrated the most robust performance, achieving accuracies of 97% (RIM-ONE), 99.2% (G1020), and 99% (REFUGE).

Bragança et al. (2022) acquired a new dataset named Brazil Glaucoma, comprising 2,000 retinal fundus images evenly divided into glaucomatous and non-glaucomatous classes. The images were acquired using a smartphone paired with a Welch Allyn panoptic direct ophthalmoscope, highlighting the practicality of cost-effective data acquisition. To evaluate the dataset, the authors employed an ensemble of CNN model, including DenseNet, MobileNet, Inception-v3, InceptionResnet, Resnet50v2, Resnet101, and Xception. The ensemble model achieved an accuracy of 90.0% in distinguishing glaucoma from normal fundus images.

Bulut et al. (2022) proposed a deep learning-based approach for automatic classification of eye diseases using colour fundus images. The study employed EfficientNet-B6, fine-tuned with various hyperparameter configurations. The dataset included over 101,000 images, comprising 21,842 clinical images from Akdeniz University Hospital and several public sources. The model was trained to distinguish between normal and abnormal retinal conditions, covering more than ten types of retinal disorders. Pre-processing steps included resizing, normalization, and data augmentation. The best-performing model achieved a sensitivity of 94.39%, specificity of 86.04%, and an overall accuracy of 86% on the test set. The study emphasizes the importance of distinguishing “referrable” from “non-referrable” retinal disorders to support clinical decision-making and reduce unnecessary specialist referrals.

Shamia et al. (2022) developed an online graphical user interface (GUI) platform integrated with deep CNN-based expert system for diagnosing three eye diseases: diabetic retinopathy, cataracts, and glaucoma. The system achieved classification accuracies of 91% for diabetic retinopathy, 90% for cataracts, 86% for glaucoma. Similarly, Arslan and Erdaş (2023) studied the binary classification of retinal fundus images targeting the same three eye diseases using a deep learning framework. Using a dataset of 2,748 images evenly split between normal and diseased cases, compiled from multiple public sources, five CNN architectures—EfficientNet, VGG, ResNet, DenseNet, and Xception—were assessed using 10-fold cross-validation. Among them, EfficientNet delivered the best performance, achieving 94.88% accuracy, 94.88%, recall, 95.02% precision, 94.88% F1-score, and a Matthews correlation coefficient of 89.89%. The study highlights EfficientNet's scalability and efficiency, making it particularly suitable for early diagnosis using limited data.

Thanki (2023) introduced a dual learning-based approach that integrates deep neural networks with traditional machine learning classifiers for glaucoma classification. The system utilizes SqueezeNet, a lightweight CNN, to extract deep features from colour retinal fundus images, which are subsequently classified using six machine learning algorithms: decision tree, k-nearest neighbour (kNN), logistic regression, Naive Bayes, random forest, and support vector machine (SVM). Among these, the combination of deep neural network and logistic regression outperformed others in accuracy, sensitivity, and precision across multiple datasets, including DRISTHI-GS and ORIGA. The study demonstrates the effectiveness of combining deep feature extraction with classical classifiers, achieving a maximum training accuracy of 100% and outperforming several existing glaucoma screening systems. This hybrid approach underscores the potential of ensemble-like

strategies that leverage both deep learning and machine learning for robust medical image classification.

The reviewed studies demonstrate growing effectiveness of deep learning techniques in the binary classification of major eye diseases, particularly cataracts, diabetic retinopathy, and glaucoma. Techniques like transfer learning and preprocessing such as image enhancement, grayscale conversion, and segmentation, consistently achieve high accuracy across diverse datasets. Models like VGG, ResNet, and EfficientNet have shown strong performance. Several studies also highlight the potential of hybrid frameworks that integrate deep learning with traditional machine learning classifiers, further improving diagnostic capability. Collectively, these findings underscore the promise of CNN-based systems in supporting early, accurate, and scalable screening of eye diseases using retinal fundus images. Although binary classification provides a foundational understanding of deep learning applications in ophthalmology, it is inherently limited in addressing the complexity of real-world clinical scenarios where multiple diseases may coexist or present overlapping features. Therefore, the next section will explore multiclass classification approaches, which form the core focus of this research, aiming to develop more comprehensive and scalable diagnostic solutions.

2.5.2 Multiclass Classification of Eye Diseases

Multiclass classification extends the diagnostic ability of deep learning models by enabling the simultaneous identification of multiple eye diseases or different stages of a single disease within a unified framework. This approach is particularly valuable in clinical settings where comprehensive and efficient screening is essential. Recent advancements in CNN architectures and training strategies have facilitated the development of robust multiclass classifiers capable of distinguishing between eye diseases such as age-related

macular degeneration, cataracts, diabetic retinopathy, and glaucoma. This section explores the studies that implement multiclass classification methods, emphasizing their contributions to automated ophthalmic diagnosis, disease grading, and the broader goal of scalable, real-world clinical deployment.

i. Classification of Three Eye Diseases/Conditions

Chea and Nam (2021) studied on the multiclass classification of three prevalent eye diseases—age-related macular degeneration, diabetic retinopathy, and glaucoma—using several deep learning architectures including ResNet (ResNet50, ResNet101, ResNet152) and VGG (VGG16, VGG19). To enhance model performance, they applied image preprocessing methods were employed, including shrinking the region of interest, applying CLAHE on iso-luminance plane, and performing data augmentation. Their models achieved a peak accuracy of 91.19% and an average accuracy of 85.79%. The findings demonstrated that deep learning-based multiclass classification, when applied to publicly available datasets, can yield strong performance and underscore the value of publicly accessible retinal fundus image datasets as valuable data for computer-aided detection of eye diseases.

Arif (2023) employed EfficientNet-B0 architecture to classify cataracts, glaucoma, and normal retinal fundus images. The dataset, sourced from Kaggle, initially consisted of 300 images and was expanded to 3,600 images through data augmentation. Several dataset configurations were tested, including the original, augmented, grayscale pre-processed augmented, and threshold pre-processed augmented dataset. The optimal results were obtained using the Adam optimiser with a learning rate of 0.00001, batch size of 32, and 20 training epochs. Notably, the grayscale pre-processed augmented dataset achieved an accuracy of 79.22%, precision of 80.30%, recall of 79.22%, and F1-score of 78.87%.

Pan et al. (2023) utilized deep learning architectures—Inception-v3 and ResNet-50—to categorise retinal fundus images into three classes: macular degeneration, tessellated fundus, and normal. The study leveraged transfer learning and fine-tuning of pre-trained CNNs on a dataset of 1,032 fundus images. ResNet-50 model achieved a classification accuracy of 93.81%, outperforming Inception-v3, which attained 91.76%.

Shamsan et al. (2023) proposed a hybrid deep learning framework for the automatic classification of colour fundus photographs to detect multiple eye diseases, including cataract, diabetic retinopathy, and glaucoma. The framework introduced three strategies involving feature extraction and fusion. The first strategy applied MobileNet and DenseNet-121 separately with PCA for dimensionality reduction, followed by classification using an ANN. The second strategy fused features from both CNN models before and after PCA, while the third strategy combined CNN features with handcrafted features. Handcrafted features were extracted using techniques such as Gray Level Co-occurrence Matrix, Fuzzy Colour Histogram, Local Binary Pattern, and Discrete Wavelet Transform. The dataset, comprising 4,217 colour fundus images from multiple sources, was enhanced using averaging and Laplacian filters and augmented to address class imbalance. The best performance was achieved by fusing MobileNet features with handcrafted features, resulting in an AUC of 99.23%, accuracy of 98.5%, sensitivity of 98.75%, precision of 98.45%, and specificity of 99.4%. The study demonstrates the efficacy of integrating deep learning features and handcrafted descriptors for robust multi-disease classification in retinal imaging.

Deepak and Bhat (2024) studied the multiclass classification of eye diseases—specifically cataract, glaucoma, and normal conditions—using transfer learning. The study

evaluated three CNN architectures: SqueezeNet, EfficientNet-b0, and DarkNet-53, optimizing them with respect to batch size (6, 8, 10) and optimizer type (SGDM, RMSProp, Adam). The models were trained and evaluated on the ODIR-5K dataset, with data augmentation applied to improve generalization. DarkNet-53 achieved the highest classification accuracy of 99.4%, along with a precision of 99.39%, recall of 99.40%, and a false negative rate of 0.60%.

Vardhan et al. (2024) proposed a deep learning-based framework for automated classification of major eye diseases—cataract, diabetic retinopathy, and glaucoma—using transfer learning techniques. The study evaluated three prominent CNN architectures: VGG19, InceptionV3, and ResNet50, both in their original forms and with transfer learning applied. The modified ResNet50 model outperformed both VGG19 and InceptionV3, achieving accuracy of 99.94%, precision of 96.28%, and F1-score of 96.24%.

ii. Classification of Four Eye Diseases/Conditions

Hemalakshmi et al. (2021) proposed a hybrid classification approach for retinal fundus images to detect age-related macular degeneration, choroidal neovascularization, diabetic retinopathy, and normal cases. Their approach integrates Multi-Scale Discriminative Robust Local Binary Pattern (MS-DRLBP) features with a hybrid convolutional neural network and radial basis function classifier (CNN-RBF). The preprocessing pipeline includes adaptive median filtering, contrast enhancement using top-hat transformation, and retina region expansion. The feature extraction process involves RGB component separation, gradient-based edge detection, and texture analysis using local binary pattern (LBP) descriptors, complemented by statistical metrics such as mean, standard deviation, kurtosis, skewness, and entropy. The CNN-RBF classifier outperformed

conventional classifiers (CNN, RBF, SVM, Naïve Bayes, Nearest Neighbour, Adaptive Neuro-Fuzzy Inference System/ANFIS), achieving 97.22% accuracy, 96.49% sensitivity, and 100% specificity on the STARE dataset.

Toki et al. (2022) developed a custom CNN model named RetinalNet-500 for classification of normal eye, cataracts, diabetic retinopathy, and glaucoma. To benchmark its performance, they also trained several pre-trained models including Inception-v3, MobileNetV2, and Xception. All models had achieved promising accuracy: Inception-v3 and MobileNetV2 both achieved 97.30%, Xception reached 96.45%, and RetinalNet-500 attained 95.15%. These results affirm the reliability of both custom and pre-trained architectures in multiclass eye disease classification.

Babaqi et al. (2023) explored the effectiveness of transfer learning in multiclass eye diseases classification. Their study compared a custom CNN model with a transfer learning-based EfficientNet model for distinguishing healthy eyes and those exhibited signs of cataracts, diabetic retinopathy, or glaucoma. The use of a pre-trained EfficientNet model significantly enhanced classification performance, with accuracy increasing from 84% (custom CNN) to 94% (EfficientNet), demonstrating the advantages of transfer learning in medical image analysis.

Saini et al. (2023) trained an EfficientNet-B3 model to classify fundus images into cataracts, diabetic retinopathy, glaucoma and normal. The study also investigated the impact of hyperparameter tuning on model performance. The best performance was achieved with a batch size of 32 and 20 training epochs, yielding an impressive accuracy of 99.85%. Similarly, Sharma et al. (2023) developed a modified EfficientNetB3-based deep learning model to classify cataract, diabetic retinopathy, glaucoma, and healthy eyes. The dataset was expanded from 3,400 to 11,400 images using data augmentation, including CLAHE

enhancement, flipping, rotation, and grid distortion. The model achieved a classification accuracy of 97.5%.

Cui et al. (2023) proposed a transfer learning-based approach for eye disease classification using several pre-trained models, including VGG19, ResNet50, EfficientNet-B0, and DenseNet. These models were employed to classify fundus images into four classes: cataracts, diabetic retinopathy, glaucoma, and normal. Among the evaluated models, ResNet50 obtained the highest test accuracy of 0.9291. Likewise, Erdas and Arslan (2024) conducted a comparative study to evaluate the performance of deep learning architectures—EfficientNet, DenseNet, VGG, ResNet, and Xception—for the classification of cataracts, diabetic retinopathy, glaucoma, and health classes. Utilizing a dataset of 4,217 retinal fundus images and 10-fold cross-validation, EfficientNet outperformed other architectures, achieving an accuracy of 87.84% and an F1-score of 93.53%, demonstrating its balance between computational efficiency and classification performance.

On the other hand, Kaur et al. (2023) developed a customized ResNet-18 model for classifying retinal fundus images into four classes: cataracts, diabetic retinopathy, glaucoma, and normal. The model attained a notable classification accuracy of 94%, indicating the effectiveness of tailored deep learning architecture in multiclass eye disease detection. In a comparable manner, Imaduddin et al. (2024) employed a transfer learning strategy through fine-tuning a pre-trained ResNet-50 model. Fine-tuning involved hyperparameter adjustments such as learning rate and early stopping. Training on a dataset comprised 4,217 images, the model achieved an accuracy of 92%.

Focusing on the same four eye diseases, Varghese and Pandian (2023) presented an effective eye disease classification model based on the InceptionResNetV2 architecture with

a fine-tuning mechanism. The model demonstrated substantial performance improvements after fine-tuning, with accuracy increasing from 81.00% to 94.74%, precision from 79.25% to 93.99%, recall from 82.49% to 94.24% and F1-score from 81.49% to 94.11%. Kumar et al. (2023) implemented an ensemble learning approach using majority voting to combine the probabilistic outputs of two top-performing models: Xception and DenseNet169. This ensemble approach surpassed multiple state-of-the-art methods, achieving 99.80% accuracy, 97.00% recall, 97.10% precision, and an F1-score of 97.05%, emphasizing the strength of ensemble techniques in improving eye diseases classification.

Ejaz et al. (2024) utilized the RFMiD and RFMiD 2.0 datasets, focusing on four classes: diabetic retinopathy, media haze, optic disc cupping, and normal. Three CNN architectures with 12, 14, and 20 layers were implemented, alongside with extensive preprocessing techniques including data augmentation, cropping, resizing, and one-hot encoding to address class imbalance. The 12-layer CNN achieved the best balance between accuracy and overfitting, with a validation accuracy of 89.81% and a testing accuracy of 88.72%.

Ejaz et al. (2025) proposed a hybrid deep learning approach for early diagnosis of retinal diseases, combining two custom CNN architectures for feature extraction and Canonical Correlation Analysis (CCA) for feature fusion. The model was trained and evaluated on the RFMiD and RFMiD 2.0 datasets, focusing on four classes: diabetic retinopathy, media haze, optic disc cupping, and healthy. After preprocessing and data augmentation, features extracted from both CNNs were fused and classified using various machine learning algorithms. Among the evaluated models, the ensemble learning classifier delivered the highest accuracy of 93.39%, outperforming random forest, SVM, and KNN.

The study demonstrated that combining deep feature extraction with traditional machine learning classifiers significantly enhances classification performance.

iii. Classification of More Than Four Eye Diseases/Conditions

Guergueb and Akhloufi (2021) evaluated several pre-trained deep learning models—EfficientNet-B5, EfficientNet-B6, EfficientNet-B7, DenseNet121, DenseNet169, and DenseNet201—for the classification of eight eye diseases: Age-Related Macular Degeneration, Cataract, Diabetic Retinopathy, Glaucoma, Hypertension, Pathological Myopia, Normal, and Others. Utilizing ODIR-5K dataset, EfficientNet-B7 emerged as the best-performing model, achieving an AUC score of 96.04%. This research highlights the effectiveness of pre-trained deep learning model, especially EfficientNet in handling complex multiclass classification tasks in ophthalmology. Raza et al. (2021) developed a deep learning-based approach for eye disease classification using digital fundus imaging and the Inception-V4 model. Their method achieved a classification accuracy of 96% on a dataset consisting of 602 digital retinal images.

Helen and Gokila (2023) developed a CNN-based model named EYENET to classify bulging eye, cataracts, crossed eye, glaucoma, and uveitis. The model achieved an overall accuracy of 92.3%, demonstrating the potential of CNN model for automated diagnosis across a diverse range of eye diseases. Similarly, Jain and Patidar (2023) utilized transfer learning with ResNet50 to classify bulging eye, crossed eye, cataracts, glaucoma, and uveitis, with their model achieved an accuracy of 90.55%.

Wahab Sait (2023) proposed a lightweight, AI-driven eye disease classification model designed to detect and classify eye diseases from fundus images with high accuracy and low computational cost. The model integrates several advanced techniques: denoising

autoencoders for image enhancement, single-shot detection for multi-scale feature extraction, and a whale optimization algorithm enhanced with Levy Flight and wavelet mutation for feature selection. Classification is performed using a fine-tuned ShuffleNet V2 architecture. Evaluated on two datasets—ODIR-5K and Eye Disease Classification dataset—the model attained accuracy scores of 99.1% and 99.4%, respectively. The study highlights the potential of the model to be deployed in resource-constrained clinical settings and mobile applications, while also addressing challenges such as data imbalance and image variability through preprocessing and augmentation strategies.

Aslam et al. (2024) employed transfer learning with five pre-trained CNN architectures—VGG-16, VGG-19, ResNet-50, ResNet-152, and DenseNet-121—on the ODIR-5K dataset, which includes eight categories of eye diseases. Among the tested models, VGG-19 achieved the best performance, with 95% accuracy, recall, precision, and F1-score. The authors emphasized the significance of fine-tuning and hyperparameter optimization in enhancing model performance. The study by Chaudhari et al. (2024) presented a cost-effective and user-friendly system for real-time eye disease classification, specifically targeting children in underprivileged areas. Using the ODIR-5K dataset, the system combined a simple CNN model with a pre-trained VGG-19 model, achieving an accuracy of 98.10% in categorizing fundus images into classes such as cataract, diabetes, glaucoma, normal, and others. The system features a ReactJS-based GUI for seamless user interaction and provides personalized dietary recommendations based on the detected eye diseases.

Chavan and Pete (2024) proposed a novel deep learning framework, Multilevel Glowworm Swarm Convolutional Neural Network (MGSCNN), for automated retinal fundus image classification. The model first performs a binary classification to distinguish

between normal and abnormal images. If an image is identified as abnormal, it is then further classified into one of 39 specific retinal disease categories using an optimized CNN architecture. MGSCNN leverages Glowworm Swarm Optimization (GSO) to simultaneously tune the CNN's structure and hyperparameters. Evaluated on the RFMiD dataset, the model achieved a classification accuracy of 95.09%, It also demonstrated strong performance across other metrics, including sensitivity (96.59%), specificity (93.59%), precision (93.53%), recall (96.67%), and F1-score (95.07%), highlighting its effectiveness in large-scale, multi-disease retinal screening applications. outperforming traditional models such as CNN, SVM, and LSTM.

Khalid et al. (2024) introduced CAD-EYE, a novel deep learning-based system for multiclass eye disease classification, targeting contrast-related disorders, diabetic retinopathy, glaucoma, hypertensive retinopathy, and normal cases. The system integrated feature fusion from two pre-trained CNNs—MobileNetV2 and EfficientNetB0—enhanced by a Fluorescence Imaging Simulation preprocessing technique to improve interpretability. The fused features are processed through dense layers and classified using an XGBoost classifier. Trained on a custom-curated dataset of 65,871 fundus images, CAD-EYE achieved 98% accuracy and outperformed several state-of-the-art models such as DenseNet-169, Inception-v3, and ResNet50. The system introduced fluorescence-based pre-processing for the first time in this domain, enhancing both performance and explainability.

Mannepalli et al. (2024) employed fine-tuned DenseNet-121 model, focusing on five classes—glaucoma, maculopathy, myopia, retinitis pigmentosa, and healthy eyes—using a dataset of 250 images. Preprocessing steps included resizing and applying CLAHE specifically to the green channel to improve the visibility of retinal blood vessels. The DenseNet-121 model was adapted by freezing pretrained layers and adding custom

classification layers. The model outperformed baseline models, achieving 97% accuracy, 94% recall, 91% precision, 92% F1-score, and 96% AUC.

Acevedo et al. (2025) developed a custom CNN for classification of five eye conditions—cataract, diabetic retinopathy, glaucoma, retina diseases, and normal—using pre-processed retinal images. The architecture comprised three convolutional layers, three max pooling layers, one batch normalization layer, two fully connected layers, and a final output layer. Preprocessing steps involved grayscale conversion, blur filtering, and Canny edge detection to enhance feature visibility. The model was trained and validated on a balanced dataset of 1,000 images (200 per class) using an 80-20 hold-out validation split. The proposed CNN achieved 97% accuracy, precision, recall, and F1-score, demonstrating strong performance in distinguishing both healthy and diseased eyes. Compared to more complex pre-trained models, this lightweight architecture offers competitive accuracy with lower computational demands.

Tashkandi (2025) studied the classification of age-related macular degeneration, cataracts, diabetic retinopathy, glaucoma, and high myopia retinal images. The study integrated six publicly available datasets and applied extensive preprocessing methods including resizing, normalization, augmentation, and grayscale conversion. Several models were evaluated, including traditional machine learning algorithms (SVM, Random Forest) and deep learning architectures (VGG16, MobileNetV1, and a hybrid CNN-RNN model). Among these, MobileNetV1 achieved the highest accuracy (98%), outperforming other models in recall, precision, and F1-score across most disease classes.

iv. Classification of Diabetic Retinopathy Stages

Topaloglu (2023) introduced a novel deep learning-based image classification framework for eye disease identification, featuring a custom CNN architecture enhanced by a new "Care model" approach. This model improves upon traditional CNNs by incorporating pixel-level rescaling and retraining strategies to improve feature extraction and classification accuracy. The study focused on diabetic retinopathy stages grading (Non-Diabetic Retinopathy, Mild, Moderate, Severe, and Proliferative Diabetic Retinopathy) using a dataset of over 45,000 retinal images sourced from Kaggle. The model architecture was based on VGG19 and included preprocessing techniques such as image resizing, rotation, and noise removal. The model achieved 88% training accuracy, 87% testing accuracy, 93% precision, and 83% recall.

In a comparable fashion, Kallel and Echtioui (2024) proposed a transfer learning-based approach for grading diabetic retinopathy stages using retinal fundus images. The study evaluated four pre-trained CNN models—VGG16, VGG19, Inception-v3, and DenseNet169—on the APTOS 2019 dataset, which includes five diabetic retinopathy severity classes. Preprocessing techniques such as Gaussian filtering, circular cropping, and pixel scaling were applied, along with data augmentation (horizontal flipping) to address class imbalance. Each model was fine-tuned by replacing the dense layers with global average pooling, dropout, and softmax layers. Inception-v3 achieved the highest classification accuracy of 96.88%, outperforming others in both accuracy and F1-score. VGG19 also delivered strong performance in precision and recall while slightly lower in accuracy.

Shamrat et al. (2024) introduced DRNet13, a novel deep learning model for automated classification of diabetic retinopathy stages using fundus images. The study compared DRNet13 against fifteen pre-trained CNN models, including ResNet, VGG, Inception, and DenseNet variants. Using a Kaggle-sourced dataset comprising 3,662 images, expanded to 7,500 through augmentation, the authors applied preprocessing techniques such as median filtering and gamma correction to enhance image quality. DRNet13, a 13-layer CNN architecture, achieved a test accuracy of 96%, precision of 0.97, sensitivity of 0.98, and an AUC of 0.99, outperforming all baseline models. The model also demonstrated strong generalization across five diabetic retinopathy classes, with detailed analysis using confusion matrices, ROC curves, and feature maps.

I. K. Gupta et al. (2025) introduced a novel computer-aided diagnosis framework, FIDRC-DLFFO for the automated diabetic retinopathy stages grading using retinal fundus images. The model integrates median filtering for noise suppression, utilizes Inception-ResNet-v2 for deep feature extraction, and incorporates a Gated Recurrent Unit (GRU) optimized using the Fennec Fox Optimization (FFO) algorithm. Validated on a balanced subset of the Kaggle DR dataset, the model achieved an average accuracy of 98.00%, outperforming several benchmark models such as ResNet-50, VGG16, and Inception-V3. This work distinguishes itself by combining transfer learning, recurrent neural networks, and bio-inspired optimization, offering a comprehensive and scalable approach for the screening of diabetic retinopathy, particularly suited for deployment in settings with limited resources.

Sushith et al. (2025) proposed a Temporal Aware Hybrid Deep Learning (TAHDL) framework for early detection and progression monitoring of diabetic retinopathy. The model integrates CNN for spatial feature extraction and recurrent neural network (RNN),

specifically LSTM with attention mechanisms, to capture temporal dependencies across sequential retinal scans. The framework incorporates multiscale convolutional paths and contrast enhancement (CLAHE) preprocessing to improve feature visibility. Evaluated on DRIVE, Kaggle DR, and EyePACS datasets, TAHDL achieved high accuracy (up to 97.5%), outperforming traditional models such as CNN, RNN, and Vision Transformers.

The reviewed studies collectively demonstrate the rapid advancement and effectiveness of deep learning approaches in multiclass classification of eye diseases using retinal fundus images. A wide range of architectures—including ResNet, EfficientNet, DenseNet, VGG, Inception, and custom CNN models—have been explored, often enhanced through transfer learning, ensemble methods, and hybrid frameworks. These approaches have consistently yieled high accuracy, recall, precision, and F1-score across diverse datasets and disease categories such as age-related macular degeneration, cataracts, diabetic retinopathy, and glaucoma. Overall, the body of work affirms the potential of deep learning as a reliable and efficient tool for automated multiclass eye disease classification, laying a solid foundation for future research and real-world applications.

2.6 Publicly Available Retinal Fundus Image Datasets

Publicly available retinal fundus image datasets play a crucial role in the development, evaluation, and benchmarking of automated eye disease diagnosis systems. These datasets provide standardized and well-annotated retinal fundus images that enable researchers to train and compare deep learning models. Over the years, several retinal fundus image datasets have been introduced, varying in terms of image resolution, disease categories, annotation quality and clinical scope. This section provides an overview of commonly used publicly available retinal fundus image datasets reported in the literature,

which form the foundation for many existing studies and support comparative analysis in eye disease classification research.

2.6.1 Eye Disease Retinal Images Dataset

The Eye Disease Retinal Images Dataset, uploaded to Kaggle by Doddi (2020), contains retinal fundus images categorised into four classes: cataracts, diabetic retinopathy, glaucoma, and normal eyes. The dataset was compiled from multiple publicly available sources, including Indian Diabetic Retinopathy Image Dataset (IDRiD), Ocular Diseases Intelligence Recognition (ODIR), High-Resolution Fundus (HRF) dataset, retinal_dataset and Digital Retinal Images for Vessel Extraction (DRIVE). Due to the dataset's diverse sources, the images vary in both resolution (including 256×256, 512×512, and 2592×1728 pixels) and file format (JPG, JPEG, and PNG). This variation contributes to a more heterogeneous and representative dataset, which can enhance the model's generalization ability across varying imaging conditions.

2.6.2 Ocular Disease Intelligence Recognition (ODIR-5K) Dataset

The Ocular Disease Intelligence Recognition dataset, commonly referred to as ODIR-5K, is a structured ophthalmic dataset containing colour retinal fundus images from 5,000 patients, along with associated metadata such as age and diagnostic keywords provided by ophthalmologists (Peking University, 2019). The dataset was released as part of the Peking University International Competition on Ocular Disease Intelligent Recognition, organised in 2019. Images were acquired from patients undergoing ocular examinations at various hospitals and medical institutions across China. The dataset was compiled by Shanggong Medical Technology Co. Ltd, with fundus images captured using a variety of

commercially available cameras, including Canon, Zeiss, and Kowa devices. All images were pre-processed to standardized resolution of 512×512 pixels.

The ODIR-5K dataset is designed for multi-label classification, meaning each image may be associated with multiple diagnostic labels. Images are categorized into eight classes based on diagnostic keywords from ophthalmologists: Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age-related Macular Degeneration (A), Hypertension (H), Myopia (M), and Other diseases/abnormalities (O).

2.6.3 Retinal Fundus Multi-Disease Image Dataset (RFMiD)

The Retinal Fundus Multi-Disease Image Dataset (RFMiD) was released as part of the Retinal Image Analysis for Multi-Disease Detection Challenge and consists of 3,200 retinal fundus images annotated for 46 distinct ocular conditions. Annotations were determined through the adjudicated consensus of two senior retinal experts (Pachade et al., 2021b). The authors highlight that RFMiD represents the sole publicly available dataset encompassing a comprehensive spectrum of eye diseases commonly encountered in clinical practice. Each image is labelled by two expert ophthalmologists, ensuring reliable annotations.

Images were captured using three different digital fundus camera models, resulting in varied resolutions of 2048×1536, 2144×1424, and 4288×2848 pixels (Pachade et al., 2021b). Excluding normal eyes, the dataset includes 45 disease categories, such as Diabetic Retinopathy (DR), Age-Related Macular Degeneration (ARMD), Media Haze (MH), Drusen (DN), Myopia (MYA), Optic Disc Cupping (ODC), and many others (Pachade et al., 2021b). Images with multiple ocular conditions are assigned multi-label annotations, supporting the development of generalizable models for screening a wide variety of eye diseases.

2.6.4 Glaucoma Fundus Imaging Dataset

The Glaucoma Fundus Imaging Dataset, compiled by Jain (2022), consists of glaucomatous retinal fundus images. This dataset integrates samples from three publicly available datasets: G1020, ORIGA, and REFUGE. The combined dataset aims to provide a set of glaucomatous and non-glaucomatous fundus images to facilitate robust model training. The component datasets are described below:

(a) G1020 Dataset

The G1020 dataset, introduced by Bajwa et al. (2020), contains 1,020 high-resolution colour fundus images, with resolutions ranging from 1944×2108 to 2426×3007 pixels. It provides verified annotations pertinent to glaucoma diagnosis, as well as segmentation of the optic disc and optic cup. Of the total images, 296 are labelled as glaucoma.

(b) ORIGA Dataset

The Online Retinal Fundus Image database for Glaucoma Analysis and Research (ORIGA) dataset, acquired by Zhang et al. (2010), aims to disseminate retinal images accompanied by clinically validated ground truth annotations to facilitate research in glaucoma detection and retinal image segmentation. It consists of 650 fundus images, annotated by professionals at the Singapore Eye Research Institute, of which 168 images are labelled as glaucomatous.

(c) REFUGE Dataset

The Retinal Fundus Glaucoma Challenge (REFUGE) dataset was released as part of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2018. It comprises 1,200 fundus images, each accompanied by ground truth segmentations masks for the optic disc and optic cup,

along with clinical labels of glaucoma (Orlando et al., 2020). This dataset is highly unbalanced, containing only 120 glaucomatous images.

By combining these datasets, the Glaucoma Fundus Imaging Dataset provides a diverse and representative sample of fundus images for glaucoma classification. These datasets are widely used in the literature for glaucoma detection and optic disc/cup analysis.

2.6.5 PAPILA Dataset

The PAPILA dataset was acquired between 2018 and 2020 at the Department of Ophthalmology, Hospital General Univesitario Reina Sofía (HGURS) in Murcia, Spain, by Kovalyk et al. (2022a). Its objective is to advance the early glaucoma diagnosis by providing joint information from both eyes of each patient, thus enabling a more comprehensive diagnostic approach. The dataset comprises 488 retinal fundus images from 244 patients. All images are of high resolution (2576×1934 pixels) and are accompanied by relevant clinical data. Additionally, the dataset includes expert-annotated segmentation of the optic disc and optic cup, and diagnostic labels based on thorough clinical assessment. Each image is categorized into one of three diagnostic classes: Healthy, Glaucoma or Suspect.

2.6.6 DRISHTI-GS Dataset

The DRISHTI-GS dataset is publicly accessible that curated to support study in automated glaucoma assessment, with a focus on segmentation of optic disc and optic cup (Sivaswamy et al., 2014). It consists of 101 monocular retinal fundus images with a resolution of 2896×1944 pixels. Of the 101 images, 70 are labelled as glaucomatous and 31 as normal, divided into training and test subsets. The images were acquired at the Aravind Eye Hospital in Madurai, India. Each image includes ground truth annotations, provided by

four glaucoma experts with varying clinical experience (three, five, nine, and twenty years, respectively) (Sivaswamy et al., 2015a).

2.6.7 Eye Disease Diagnosis and Fundus Synthesis (EDDFS) Dataset

The Eye Disease Diagnosis and Fundus Synthesis (EDDFS) dataset is a vast, high-quality retinal fundus image dataset acquired by Xia et al. (2022b) to support research in both data-driven approaches to eye disease diagnosis and the generation of synthetic fundus images. The dataset comprises a total of 28,877 colour retinal fundus images, including 15,000 images of normal eyes and images representing eight different eye diseases: diabetic retinopathy, age-related macular degeneration, glaucoma, pathological myopia, hypertension, retinal vein occlusion, Laser-Assisted In Situ Keratomileusis (LASIK) spot and others. Some images are multi-labelled, indicating the presence of more than one disease in a single eye. The annotation process was conducted by an ophthalmologist and six domain experts, with assistance from three trained annotators. To ensure label accuracy and consistency, multiple rounds of annotation were performed.

2.6.8 1000 Fundus Images with 39 Categories Dataset

The 1000 Fundus Images with 39 Categories dataset is a curated subset of a much larger dataset comprising 249,620 fundus images, which was originally used to train a deep learning platform for fundus image analysis developed by Cen et al. (2021). This subset, made publicly available via Kaggle, includes images representing 39 distinct fundus diseases and conditions, encompassing a wide range of common retinal abnormalities. These categories include conditions such as large optic disc, disc swelling and elevation, cotton-wool spots, diabetic retinopathy, possible glaucoma, optic atrophy, severe hypertensive retinopathy, tessellated fundus, and others. The images were collected from three major

sources: Joint Shantou International Eye Centre (JSIEC) in China, Lifeline Express Diabetic Retinopathy Screening System (LEDRS) in China, and Eye Picture Archive Communication System (EyePACS) in the United States. A rigorous multi-stage annotation process was applied. Initial labels were assigned by general retina specialists, then verified by senior ophthalmologists with over seven years of experience and further reviewed by a retina expert panel in cases of ambiguity, ensuring high label accuracy.

2.7 Comparative Analysis of Related Work

The previous section reviewed a range of studies employing CNNs for eye disease classification with the use retinal fundus images, covering both binary and multiclass approaches. These works vary in terms of model architecture, types of eye diseases, number of disease classes, preprocessing techniques, and dataset characteristics. Table 2.2 provided a structured summary of these published works, tabulating the types of eye disease, datasets and data sizes, best architectures, and performance metrics. While the recorded performances in many studies appear high, a critical analysis of the methodologies, datasets, and classification scopes reveals several systemic gaps regarding task complexity, data homogeneity, and feature visibility that this research aims to address.

A significant portion of the literatures focus on binary classification (e.g., Thanki et al., 2023; Gupta et al., 2025). While these models achieve high accuracy (often > 95%), they fail to account for clinical reality where multiple diseases may coexist or present similar symptoms. Studies that attempt broader multiclass classification, such as studies by Chea and Nam (2021), Guergueb and Akhloufi (2021), show a notable drop in performance (accuracies ranging from 74.98% to 76.71%). This highlights a primary challenge: the

difficulty in maintaining discriminative power as the number of class increases and clinical boundaries become less distinct.

A recurring challenge in existing research is the reliance on small-scale or localized datasets. CNNs typically require large volumes of labelled data to effectively learn complex patterns and generalize across diverse cases. However, many studies are constrained by datasets with relatively few samples (Raza et al., 2021; Arif et al., 2023; Acevedo et al., 2025). This scarcity of data often leads to overfitting and compromises the robustness and generalization ability of the model to diverse patient demographics. To address these issues, some researchers have employed data augmentation and transfer learning from large-scale datasets. Nonetheless, the fundamental challenges of limited data size which may not adequately capture the inter-class and intra-class variability in disease presentation, as well as the complex visual nature of retinal fundus images where multiple disease classes are present in the classification tasks, further limiting the robustness and generalizability of the trained models.

Furthermore, many researcher rely on single-source repositories like EyePACS or MESSIDOR (Hemalakshmi et al.; 2021; Jain & Patidar, 2023; Pan et al, 2023; Ejaz et al., 2024), which lack the visual diversity—such as disease manifestations, resolution differences, lighting variations and demographic heterogeneity—found in real-world clinical environments. Conversely, larger-scale data integration efforts often introduce significant noise that standard CNNs struggle to process without specific refinement (Sushith et al., 2025). Models trained on single, clean datasets do not generalize well to the variability of integrated data sources, as retinal fundus images can be affected by various factors such as illumination and acquisition methods (Meedeniya et al., 2025).

There is a notable lack of focus on addressing the visibility gap of disease-specific features. Standard models often tend to prioritize dominant global structure of the retina, while fail to isolate minute pathological features (e.g. microaneurysms). Enhancement processes are essential to amplify these subtle features within the retinal fundus images (Priyadharsini & Jagadeesh, 2023). In addition, inter-class confusion rises, particularly in early disease stage where the disease symptoms overlapping is more pronounced. Many current works do not employ mechanisms in the models to differentiate between these subtle clinical overlaps.

In conclusion, while the literature demonstrates the transformative potential of CNN in eye disease classification, there is a clear need for a methodology that address the limitations of feature obscurity (inter-class variability) and data heterogeneity (inter-class variability). This research is positioned to fill these gaps by proposing a pipeline that integrates feature enhancement (to make disease-specific/diagnostically clinical features), attention mechanisms (to focus the model on spatially salient regions), and ensemble learning (to aggregate diverse architectural strengths and improve generalizability across heterogenous data sources).

Table 2.2: Summary of Reviewed Studies on Eye Disease Classification Using CNNs

Reference	Eye Conditions	Dataset	Number of Images	Best Architecture	Performances
Bernabe et al. (2021)	Diabetic retinopathy, Glaucoma	ORIGA	168	Custom CNN	Acc=0.9989, Sen=0.9999, Spe=1, Pre=1, F1=0.9976
		DIARETDB0	397		
Dai et al. (2021)	Diabetic retinopathy, Normal	Local Dataset	666383	Custom CNN	AUC=0.943-0.972
Gheisari et al. (2021)	Glaucoma, Normal	Local Dataset	1810 + 295 video	VGG16 + LSTM	Sen=0.96, Spe=0.96, AUC=0.99, F1=0.9619
Hemelings et al. (2021)	Glaucoma, Normal	Local Dataset, REFUGE	13551	ResNet50	AUC=0.9400
Sandoval-Cuellar et al. (2021)	Glaucoma, Normal	ORIGA	650	Custom CNN	Acc=0.9322, Sen=0.9414, AUC=0.9398
Sarki et al. (2021)	Mild diabetic retinopathy, Normal	MESSIDOR, MESSIDOR-2, DRISHTI-GS, Retinal Dataset (Github)	100 per class	Custom CNN	Acc=0.9333, Sen=1.0000, Spe=0.8667, Pre=1.0000
	Mild diabetic macular edema, Normal				Acc=0.9143, Sen=0.9444, Spe=0.8824, Pre=0.9400

Table 2.2 continued

	Glaucoma, Normal				Acc=1.0000, Sen=1.0000, Spe=1.0000, Pre=1.0000
Shoukat et al. (2021)	Glaucoma, Normal	G1020	1020	EfficientNet-B7	Acc=0.992, Sen=0.980, Spe=0.970
		REFUGE	4366		Acc=0.990, Sen=0.975, Spe=0.980
		RIM-ONE	169		Acc=0.960, Sen=0.970, Spe=0.980
Bakır and Yılmaz (2022)	Cataracts, Normal	ODIR	1094	ResNet	Acc=0.9551, Sen=0.9500, Pre=0.9600, F1=0.9500
Bragança et al. (2022)	Glaucoma, Normal	Brazil Glaucoma (BrG)	2000	Ensemble model	Acc=0.9050, Sen=0.8500, Spe=0.9600, Pre=0.9550, AUC=0.9650, F1=0.8990
Bulut et al. (2022)	Abnormal, Normal	EyePACS, DIARETB0, IDRiD, MESSIDOR, MESSIDOR2, APTOS2019	21842	EfficientNet-B6	Acc=0.86, Sen=0.9439, Spe=0.8604
Shamia et al. (2022)	Diabetic retinopathy, Normal	Local dataset	Not specified	Custom CNN	Acc=0.91

Table 2.2 continued

	Cataracts, Normal				Acc=0.90
	Glaucoma, Normal				Acc=0.86
Arslan and Erdaş (2023)	Diseased, Normal	Eye Disease Classification (Kaggle), 1000 Fundus images with 39 categories, Mendeley dataset	2748	EfficientNet	Acc=0.9488, Sen=0.9488, Pre=0.9502, F1=0.9488, MCC=0.8989
Emir and Çolak (2023)	Health	ODIR	5000	VGG16	Acc=0.667, Sen=0.646, Spe=0.675, Pre=0.454, AUC=0.717, F1=0.667
	Cataracts			ResNet50	Acc=0.971, Sen=0.785, Spe=0.985, Pre=0.797, AUC=0.964, F1=0.971
	Diabetic retinopathy			VGG16	Acc=0.653, Sen=0.561, Spe=0.700, Pre=0.494, AUC=0.677, F1=0.686

Table 2.2 continued

	Glaucoma			ResNet50	Acc=0.934, Sen=0.294, Spe=0.972, Pre=0.375, AUC=0.762, F1=0.934
	Age-related macular degeneration			Inception-v3	Acc=0.940, Sen=0.292, Spe=0.975, Pre=0.389, AUC=0.748, F1=0.940
	Hypertension			ResNet50	Acc=0.916, Sen=0.400, Spe=0.933, Pre=0.167, AUC=0.764, F1=0.916
	Myopia			ResNet50	Acc=0.957, Sen=0.696, Spe=0.971, Pre=0.552, AUC=0.957, F1=0.959
	Other			Inception-v3	Acc=0.719, Sen=0.125, Spe=0.954, Pre=0.516, AUC=0.621, F1=0.719
Shah et al. (2023)	Cataracts, Normal	Eye Disease Classification (Kaggle)	1112	VGG19	Acc=0.97
Singh et al. (2023)	Cataracts, Normal	ODIR	1309	VGG16	Acc=0.9610

Table 2.2 continued

Thanki et al. (2023)	Glaucoma, Normal	DRISTHI-GS, HRF, ORIGA	Not specified	SqueezeNet + Logistic Regression	Acc=0.653-0.990, Sen=0.543-1.000, Pre=0.722-0.986, AUC=0.836-1.000, F1=0.838-0.993
Chavan and Pete (2024)	Abnormal, Normal	RFMiD	Not specified	Multilevel Glowworm Swarm CNN	Acc=0.9402, Sen=0.9518, Spe=0.9298, Pre=0.9250, F1=0.9382
Gupta et al. (2025)	Cataracts, Normal	Eye Disease Classification (Kaggle)	1112	VGG19	Acc=0.981
Guergueb and Akhloufi (2021)	Age-related macular degeneration, Cataracts, Diabetic retinopathy, Glaucoma, Hypertension, Pathological myopia, Normal, Others	ODIR	5000	EfficientNet-B7	Acc=0.7498, Sen=0.7357, Spe=0.8523, AUC=0.9604
Chea and Nam (2021)	Age-related macular degeneration, Diabetic retinopathy, Glaucoma, Normal	NOISE-STRESS	2335	ResNet50	Acc=0.7671

Table 2.2 continued

Hemalakshmi et al. (2021)	Age-related macular degeneration, Choroidal neovascularization, Diabetic retinopathy, Normal	STARE	180	CNN + Radial Basis Function model	Acc=0.9722, Sen=0.9649, Spe=1.0000, Pre=1.0000 F1=0.9821
Raza et al. (2021)	Cataracts, Glaucoma, Retinal disease, Normal	Cataract Dataset (Kaggle)	601	Inception-v4	Acc=0.9666
Toki et al. (2022)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Local dataset	6000	Inception-v3 MobileNetV2	Acc=0.9730
Arif et al. (2023)	Cataracts, Glaucoma, Normal	Cataract Dataset (Kaggle)	300	EfficientNet-B0	Acc=0.7922, Sen=0.7922, Pre=0.8030, F1=0.7887
Babaqi et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4200	EfficientNet	Acc=0.94
Cui et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	ResNet50	Acc=0.9291

Table 2.2 continued

Helen and Gokila (2023)	Bulging eyes, Crossed eyes, Cataracts, Glaucoma, Uveitis	Kaggle dataset	383	Custom CNN	Acc=0.923
Jain and Patidar (2023)	Bulging eyes, Crossed eyes, Cataracts, Glaucoma, Uveitis	Local dataset	375	ResNet50	Acc=0.9055
Kaur et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	ResNet-18	Acc=0.94
Kumar et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	Ensemble of Xception and DenseNet169	Acc=0.9980, Sen=0.9700, Pre=0.9710, F1=0.9705
Nagpal et al. (2023)	Mild diabetic retinopathy, Moderate diabetic retinopathy, Severe diabetic retinopathy, Hypertensive, Normal	MESSIDOR ODIR	6000	ResNet101	Acc=0.98139

Table 2.2 continued

Pan et al. (2023)	Normal, Macular degeneration, Tessellated fundus	Local dataset	1032	ResNet50	Acc=0.9381
Saini et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	EfficientNet-B3	Acc=0.9985
Shamsan et al. (2023)	Cataracts, Diabetic, retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	MobileNet + ANN	Acc=0.985
Sharma et al. (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	EfficientNet-B3	Acc=0.975
Topaloglu (2023)	Diabetic retinopathy stages: None, Mild, Moderate, Severe, Proliferative	Kaggle dataset	45321	VGG19	Acc=0.87, Sen=0.83, Pre=0.93, F1=0.83

Table 2.2 continued

Wahab Sait (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	ShuffleNetV2	Acc=0.991, Sen=0.98.9, Spe=0.963, Pre=0.989, F1=0.989, Kappa=0.964
Varghese and Pandian (2023)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	ODIR	4217	Inception-ResNet-V2	Acc=0.994, Sen=0.987, Spe=0.963, Pre=0.987, F1=0.987, Kappa=0.965
		Eye Disease Classification (Kaggle)			Acc=0.9474, Sen=0.9424, Pre=0.9399, F1=0.9411
Aslam et al. (2024)	Age-related macular degeneration, Cataracts, Diabetic retinopathy, Glaucoma, Hypertensive, Myopia, Normal, Others	ODIR	5000	VGG19	Acc=0.95, Sen=0.95, Pre=0.95, F1=0.95
Chaudhari et al. (2024)	Age-related macular degeneration, Cataracts, Diabetic retinopathy, Glaucoma, Hypertensive, Myopia, Normal, Others	ODIR	5000	VGG19	Acc=0.9435

Table 2.2 continued

Deepak and Bhat (2024)	Cataracts, Glaucoma, Normal	ODIR	5000	Darknet-53	Acc=0.9940, Sen=0.9940, Spe=0.9939, F1=0.9969
Ejaz et al. (2024)	Diabetic retinopathy, Media haze, Optic disc cupping, Healthy	RFMiD RFMiD 2.0	1908	Custom CNN	Acc=0.9047
Erdaş and Arslan (2024)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	EfficientNet	Acc=0.8784, Sen=0.9284, Pre=0.9441, F1=0.9353, MCC=0.8387
Kallel and Echtioui (2024)	Diabetic retinopathy stages: None, Mild, Moderate, Severe, Proliferative	APTOS2019	5990	VGG19	Acc=0.9688, Sen=0.8540, Pre=0.9200, F1=0.8660
Khalid et al. (2024)	Contrast-related condition, Diabetic retinopathy, Glaucoma, Hypertensive retinopathy, Normal	Local dataset	65871	CAD-EYE	Acc=0.980, Sen=0.973, Spe=0.979, F1=0.980

Table 2.2 continued

Imaduddin et al. (2024)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	ResNet50	Acc=0.92
Mannepalli et al. (2024)	Glaucoma, Maculopathy, Myopia, Retinitis pigmentosa, Normal	Github dataset	250	DenseNet121	Acc=0.92, Sen=0.94, Pre=0.91, F1=0.92
Ryan et al. (2024)	Age-related macular degeneration, Cataracts, Diabetic retinopathy, Glaucoma, Pathological myopia, Hypertension	ODIR, Eye Disease Classification (Kaggle)	Not specified	ResNet152-v2	Acc=0.7250
Shamrat et al. (2024)	Diabetic retinopathy stages: None, Mild, Moderate, Severe, Proliferative	Diabetic Retinopathy 224×224 Gaussian Filtered (Kaggle)	3662	DRNet13	Acc=0.96, Sen=0.98, Spe=0.99, Pre=0.97, F1=0.97, AUC=0.99, MSE=0.03, FPR=0.04

Table 2.2 continued

Vardhan et al. (2024)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	4217	Inception-v3	Acc=0.9994, Pre=0.9628, F1=0.9624
Acevedo et al. (2025)	Cataracts, Diabetic retinopathy, Glaucoma, Normal	Eye Disease Classification (Kaggle)	200 per class	Custom CNN	Acc=0.97, Sen=0.97, Pre=0.9712, F1=0.97
I. K. Gupta et al. (2025)	Diabetic retinopathy stages: None, Mild, Moderate, Severe, Proliferative	Diabetic Retinopathy Detection (Kaggle)	400 per class	Inception-ResNet-v2	Acc=0.9800, Sen=0.9498, Pre=0.9494, F1=0.9491, MCC=0.9371
Sushith et al. (2025)	Diabetic retinopathy stages: None, Mild, Moderate, Severe, Proliferative	DRIVE	200	Temporal Aware Hybrid CNN + RNN + Attention Mechanism	Acc=0.9404-0.9750
		Diabetic Retinopathy Detection (Kaggle)	35126		
		EyePACS DR	104000		

Table 2.2 continued

Tashkandi (2025)	Age-related macular degeneration, Cataracts, Diabetic retinopathy, Glaucoma, Myopia	ODIR Eye Disease Classification (Kaggle), Cataract dataset (Kaggle), ARMD, SMDG-19, HPMI	Around 10000	MobileNetV1	Acc=0.98, Sen=0.99, Pre=0.99, F1=0.99

Note. Acc = Accuracy, Pre = Precision, Sen = Sensitivity, Spe = Specificity, F1 = F1-score, AUC = Area Under the Curve, MSE = Mean Squared Error, MCC = Matthews Correlation Coefficient.

2.8 Chapter Summary

This chapter began with a briefly discussion about the anatomy of the human eye, followed by an introduction of the three targeted eye diseases for this research: cataracts, diabetic retinopathy, and glaucoma. The causes, symptoms and risk factors associated with these eye diseases were outlined. The chapter then discussed convention eye diagnosis methods, with a particular emphasis on retinal fundus imaging. Following this, recent advancements in machine learning and deep learning were introduced, leading to a detailed explanation of Convolutional Neural Network (CNN) architectures, including key models such as AlexNet, VGG, Inception, ResNet, DenseNet, and EfficientNet. This chapter also reviewed related research on classification of eye disease using deep learning and outlined several publicly available retinal fundus images datasets. This chapter concluded with a comparative analysis of existing approaches to acquire a deeper understanding of eye disease classification and identify the research gaps that this study aims to address.

CHAPTER 3

METHODOLOGY

3.1 Overview

This chapter presents the methodology employed in this research, detailing the systematic approach undertaken to develop and evaluate an enhanced deep learning model for multiclass classification of eye disease. The methodology is organized into five phases: data collection, data preprocessing, image enhancement, model training, and model evaluation. Each phase is described in detail to provide clear understanding of the overall research design and implementation.

3.2 Research Methodology

The research flowchart provides a visual representation of the overall methodology adopted in the study. The proposed methodology follows a structured feature-augmented deep learning pipeline for multiclass eye disease classification. As illustrated in Figure 3.1, the methodology consists of five main stages: data collection, data preprocessing, image enhancement, model training, and model evaluation. The process commences with the combination of a comprehensive dataset from publicly available sources comprising retinal fundus images categorised into four classes: cataracts, diabetic retinopathy, glaucoma, and normal. The images are then pre-processed to standardized resolution and quality, followed by enhancement procedures designed to improve the visibility of disease-specific and clinically relevant retinal features. Subsequently, multiple convolutional neural network (CNN) architectures are trained using transfer learning, enabling efficient feature extraction from limited labelled data. Attention mechanisms are then be integrated to direct the model's focus on the most relevant image regions. The probabilistic outputs of these attention-based

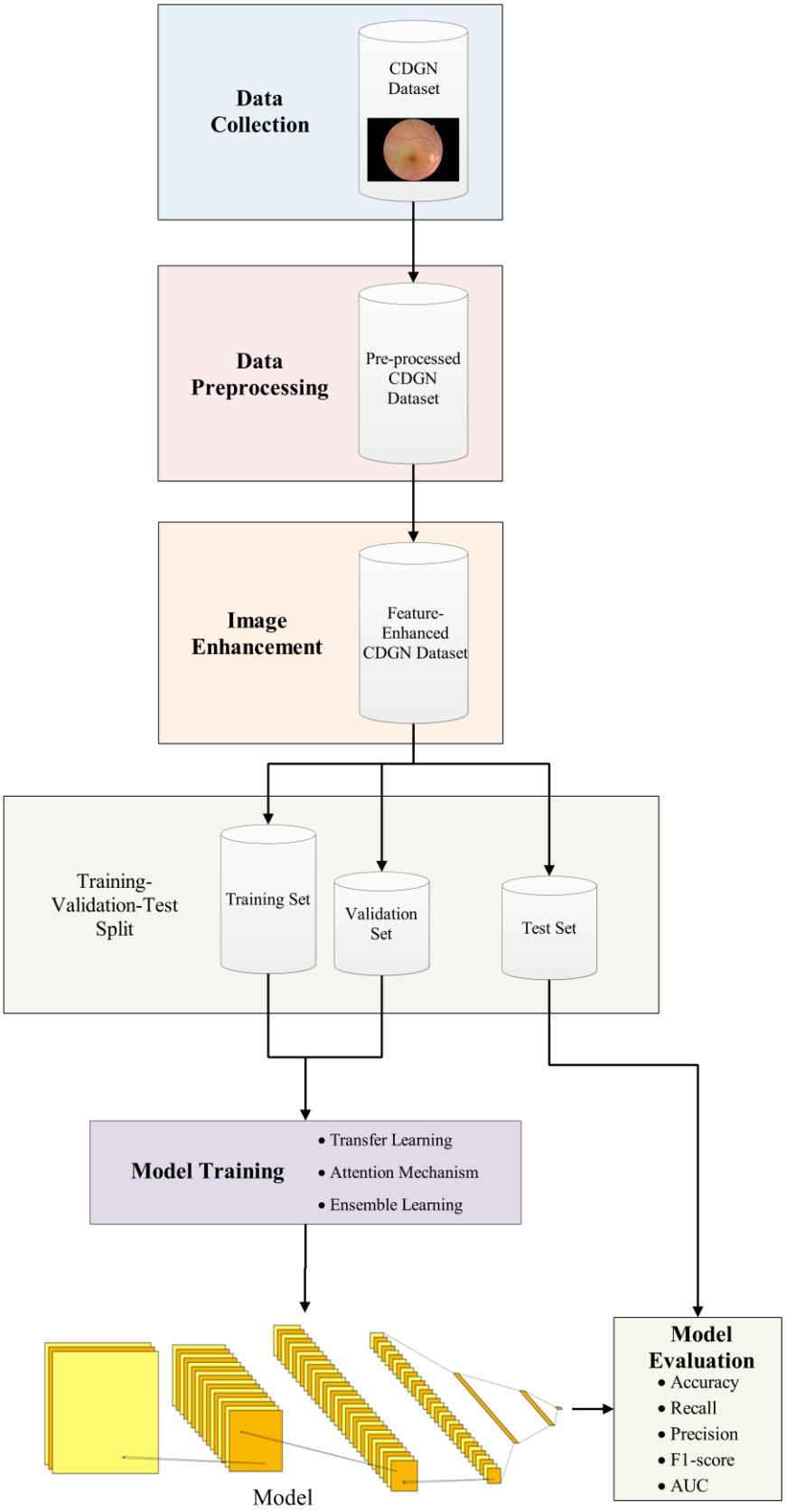


Figure 3.1: Research flowchart illustrating the sequential stages of the proposed methodology

models are subsequently combined using ensemble learning approach to enhance robustness and overall classification performance. Finally, model performance is assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. This structure pipeline ensures a systematic and effective approach in developing a robust deep learning eye disease classification model.

3.3 Execution Environment

The implementation of the proposed methodology was conducted locally within a Python virtual environment. Jupyter Notebook served as the primary development interface for code development, execution, and testing. This environment was configured with Python 3.12.2 and included the following key libraries and frameworks: TensorFlow 2.18.0, Keras 3.8.0, OpenCV 4.9.0.90, and Scikit-Image. These libraries were utilized for tasks such as image preprocessing, image enhancement, and convolutional neural network (CNN) model development. All CNN models were developed using the Keras library, an open-source neural network application programming interface (API) written in Python, running on the TensorFlow backend.

All experiments were executed on a personal laptop with the hardware and software specifications as outlined in Table 3.1. Although TensorFlow was installed with GPU support, GPU acceleration was not utilized, as no compatible GPU was detected during execution. Consequently, all training and inference operations were performed using the CPU.

3.4 Data Collection

The dataset used in this research consists of retinal fundus images across four classes: cataracts, diabetic retinopathy, glaucoma, and normal eyes. Manual collection and expert

Table 3.1: Hardware and Software Specifications

Hardware/Software	Specification
Device	Acer Laptop
Processor	8th generation Intel Core i5-8250U (1.6GHz, up to 3.4GHz with Turbo Boost)
Graphics	NVIDIA GeForce MX150
Memory	12 GB Random Access Memory (RAM)
Operating System	64-bit Windows 11 Home

annotation of such medical images are time-consuming and labour-intensive, making publicly available datasets a practical alternative. These datasets—commonly hosted on platforms like Kaggle or provided by other researchers—are typically well-annotated and have been widely adopted in existing CNN-based eye disease classification studies. Moreover, ethical concerns related to patient consent and data privacy are generally addressed by the dataset providers, thereby simplifying the ethical review process and facilitating smoother research approval.

The use of publicly available datasets streamlines the research workflow, reduces acquisition costs, and provides access to high-quality, well-curated retinal fundus images. Recognizing their established utility, this research combined multiple publicly available datasets into a single, comprehensive dataset. Each dataset contributes images corresponding to one or more of the targeted eye conditions. By combining these datasets, the overall sample size is increased, and the diversity of the image data is enhanced. This approach not only simplifies data collection but also strengthen the generalizability and robustness of the deep learning models developed in this research. The following subsections outline the

selection criteria and filtering procedure applied to the publicly available datasets incorporated in this research.

3.4.1 Dataset Selection and Integration

Retinal fundus images representing the cataracts, diabetic retinopathy, glaucoma, and normal classes were selected from the publicly available datasets described in Chapter 2 (Section 2.6). For the ODIR-5K dataset, only images labelled with the codes C (cataract), D (diabetes), G (glaucoma), and N (normal) were selected. Specifically, C represents cataracts, D refers to diabetic retinopathy, which includes mild, moderate, severe non-proliferative retinopathy as well as proliferative and severe proliferative retinopathy, G denotes glaucoma, and N indicates a normal fundus image.

For the RFMID, images were selected based on single-label annotations to ensure unambiguous classification. The following labels were included: MH = 1 for cataracts (approximated via Media Haze); DR = 1 for diabetic retinopathy; ODC = 1 for glaucoma through optic disc cupping; and Disease_risk = 0 for normal eyes. Images with multiple disease labels were excluded. Additional glaucomatous images were sourced from the Glaucoma Fundus Imaging Dataset, PAPILA, DRISHTI-GS, EDDFS, and 1000 Fundus Images with 39 Categories Dataset. Only images explicitly labelled as glaucomatous were included, while suspect cases were excluded to reduce label ambiguity and ensure dataset reliability.

Table 3.2 summarizes the dataset used, including the number of the selected images for each class, as well as associated metadata such as image resolution and file format. The table reports the total number of images per target class in each dataset before filtering, providing an overview of the data sources and their characteristics for model development.

Table 3.2: Summary of Publicly Available Datasets Prior to Filtering

Dataset	Source URL	Number of Images				Image Format	Image Resolution (pixels)	Acquisition Method
		C	DR	G	N			
Eye Disease Retinal Images Dataset	https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data	1038	1098	1007	1074	JPG, JPEG, PNG	256×256, 512×512, 2592×1728	Not specified
Ocular Disease Intelligence Recognition (ODIR-5K)	https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k	281	1383	203	2818	JPG	512×512	CANON, ZEISS, and Kowa fundus cameras
Retinal Fundus Multi-disease Image Dataset (RFMiD)	https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid	523	632	445	669	PNG	2048×1536, 2144×1424, 4288×2848	TOPCON 3D OCT-2000, Kowa VX-10, and TOPCON TRC-NW300
Glaucoma Fundus Imaging Dataset	https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets	–	–	504	–	JPG	Vary	Different models and types of fundus cameras

Table 3.2 continued

PAPILA Dataset	https://figshare.com/articles/dataset/PAPILA/14798004/1?file=28454352	–	–	87	–	JPG	2576×1934	TOPCON TRC-NW400 non-mydriatic retinal camera
DRISHTI-GS Dataset	http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php	–	–	70	31	PNG	2047×1760	Image centred on optic disc with a 30-degree field of view
Eye Disease Diagnosis and Fundus Synthesis (EDDFS)	https://github.com/xia-xx-cv/EDDFS_dataset	–	–	633	–	JPG	1024×1024	Captured by five fundus cameras
1000 Fundus Images with 39 Categories	https://www.kaggle.com/datasets/linchundan/fundusimage1000	–	–	13	–	JPG	Vary	ZEISS FF450 Plus IR and TOPCON TRC_50DX mydriatic retinal camera with 35 to 50-degree field of view

Note. C = Cataracts, DR = Diabetic retinopathy, G = Glaucoma, N = Normal. The number of images refers to the original count for each targeted class before filtering. Final totals may be lesser following the filtering process.

Following selection, a filtering process was carried out to improve data quality. Fundus images were excluded if they exhibited the problem of poor image quality, incomplete central image (such as the optic disc photographically invisible), misaligned or offset fundus capture, and background offset. Examples of these exclusion cases are displayed in Figure 3.2, Figure 3.3, and Figure 3.4, respectively. These quality issues were found to adversely affect the predictive performance of deep learning models. To establish a robust and representative dataset, only high-quality images were retained.

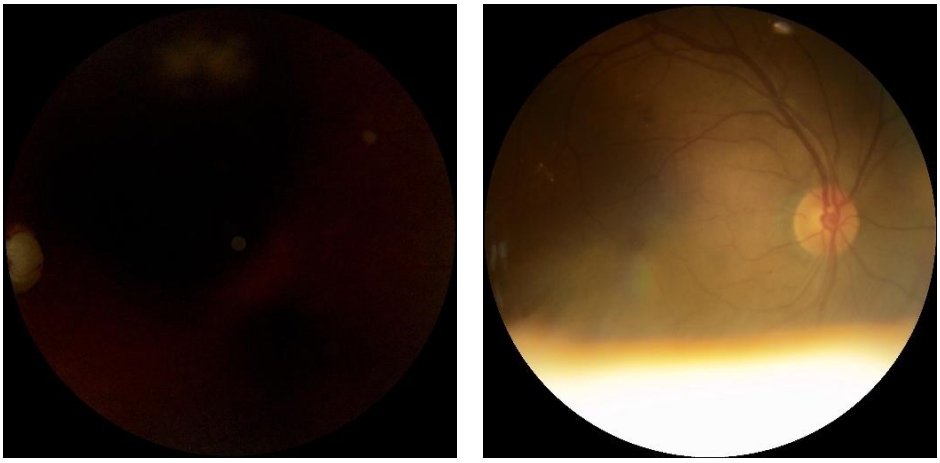


Figure 3.2: Examples of poor image quality

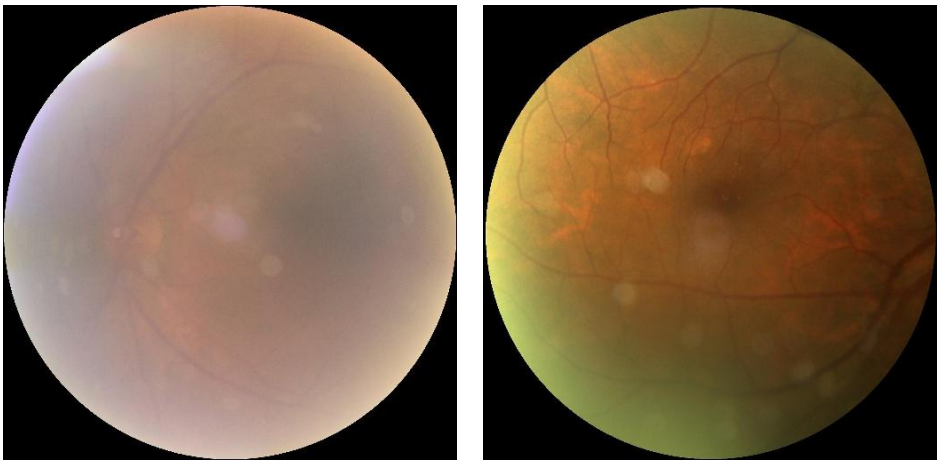
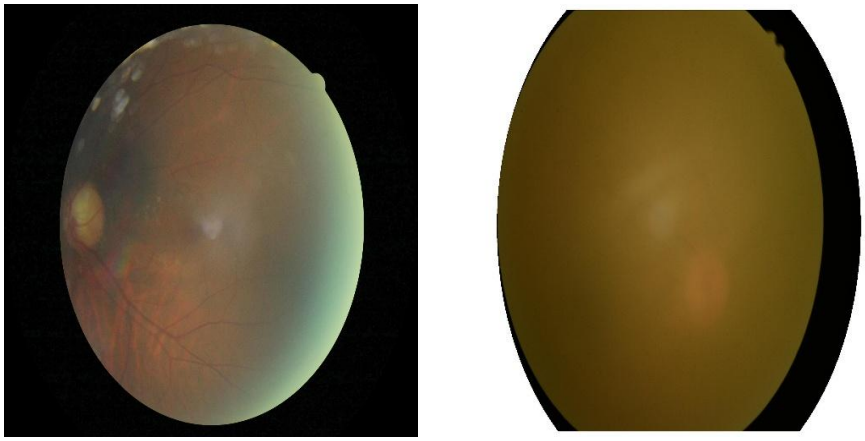


Figure 3.3: Examples of incomplete central image where optic disc is not photographically visible



(a) (b)

Figure 3.4: Examples of image offset: (a) Fundus region offset, (b) offset with background distortion

After the selection, filtration and combination of retinal fundus images from these publicly available datasets, a combined dataset was obtained. To streamline references to this dataset throughout the research, it is designated as the **CDGN dataset**, an acronym derived from the initial letters of the four represented classes: **C**ataracts, **D**iabetic retinopathy, **G**laucoma, and **N**ormal. Representative examples of input images for each class are shown in Figure 3.5, Figure 3.6, Figure 3.7, and Figure 3.8, respectively. The distribution of images across these classes is summarized in Table 3.3. The CDGN dataset serves as the primary dataset for all subsequent stages in the proposed methodology.

Table 3.3: Final Distribution of Retinal Fundus Images by Class in CDGN Dataset

Class	Number of Images	Image Resolution	Image Format
Cataracts	1,305	Vary	PNG, JPG, JPEG
Diabetic retinopathy	2,875	Vary	PNG, JPG, JPEG
Glaucoma	2,280	Vary	PNG, JPG, JPEG
Normal	3,251	Vary	PNG, JPG, JPEG
Total Images	**9,711**	–	–

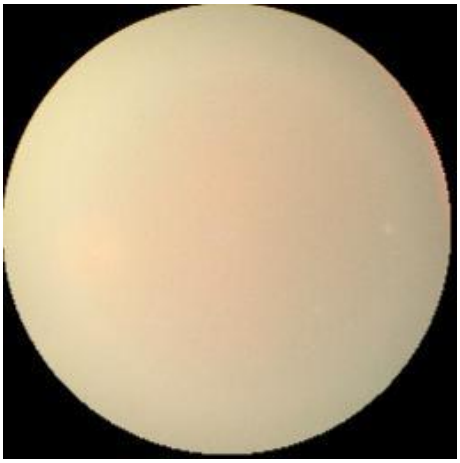


Figure 3.5: Cataracts

Figure 3.6: Diabetic retinopathy

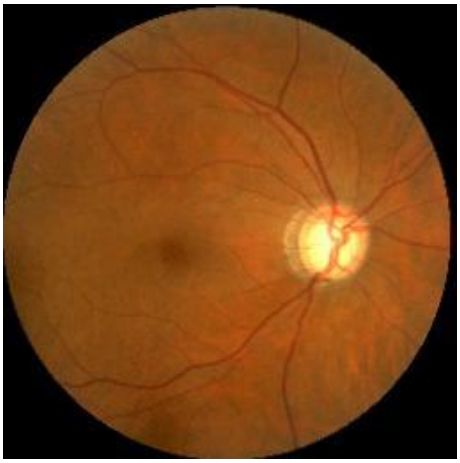


Figure 3.7: Glaucoma

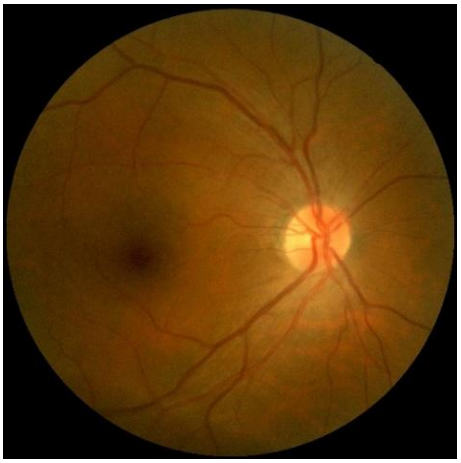


Figure 3.8: Normal

The CDGN dataset was constructed through systematic integration and rigorous filtering of eight distinct publicly available datasets, resulting in a comprehensive multiclass corpus of 9,711 retinal fundus images. While the use of multi-source data is common in contemporary research, the novelty of this dataset lies in its high-fidelity clinical diversity and targeted class density. By combining data across eight separate sources, the model is exposed to a significantly broader spectrum of real-world noise, including variations in camera sensor specifications, illumination conditions, and patient demographics, which smaller or single-source datasets hard to capture. This study ensures that only images with clear, representative disease features for the four targeted classes—cataracts, diabetic retinopathy, glaucoma, and normal—were retained.

As detailed in Table 3.3, the final distribution of each class shows high concentration of data specifically for these four high-prevalence conditions allows the model to develop a deeper, more robust understanding of intra-class variability. Furthermore, by standardizing images from eight sources with varying original format, resolutions, and acquisition methods, the dataset serves as a benchmark for evaluating the effectiveness of the proposed approach. This curated synthesis effectively bridges the gap between controlled academic benchmarks and the heterogeneous nature of actual clinical imaging, thereby providing a more reliable foundation for multiclass eye disease classification.

3.5 Data Preprocessing

Since the CDGN dataset was collected through combining images from multiple publicly available sources, the retinal fundus images vary significantly in resolution and quality. To ensure consistency and improve computational efficiency, several preprocessing techniques were applied:

i. Cropping

Each image was cropped to remove the black background, preserving only the fundus region. This step reduces visual noise and helps focus the model on relevant anatomical structures.

ii. Resizing

After cropping, all images were resized to a standard resolution of 224×224 pixels, as required by the input specifications of the pre-trained models used in this research.

The standardized CDGN dataset was split into three subsets: training set (80%), validation set (10%), and test set (10%). The training and validation set were used during model training, whereas the test set was reserved exclusively to evaluate the final

performance of model. To ensure that each subset of the dataset maintained a representative distribution of the classes, stratified splitting was employed. This is essential in classification tasks where class imbalance exists, as it helps to prevent biased model training and evaluation, thereby improving its ability to generalize to unseen data. This approach also ensures that the testing set provides a fair and reliable estimate of the model's performance.

3.6 Image Enhancement

Image enhancement is an essential step that improves image quality and facilitates the extraction of subtle visual features. Image enhancement aimed at highlighting the hidden or important details in an image. This is especially important in medical imaging tasks, such as eye disease classification, where small abnormalities can be diagnostically significant. By enhancing contrast and local image details such as subtle features and abnormalities, enhancement techniques help enhance deep learning models' performance in recognizing key pathological features.

In this research, Contrast Limited Adaptive Histogram Equalization (CLAHE) is employed to enhance the contrast of retinal fundus images before model training. CLAHE is an advanced form of adaptive histogram equalization that enhances local contrast while minimizing the risk of noise amplification (Pisano et al., 1998). Unlike traditional histogram equalization methods, CLAHE operates on small tiles (local regions) within the image and applies contrast enhancement individually, followed by bilinear interpolation to eliminate artificially induced boundaries. This localized approach makes CLAHE particularly suitable for medical imaging applications where fine details are crucial.

The enhancement process was performed on the L* channel of the CIELAB colour space, as this channel represents the luminance information (lightness) of the image, which

is ideal for contrast enhancement without affecting colour fidelity. The procedure for applying CLAHE comprises the following steps:

Step 1: Input Image

The original retinal fundus image, in RGB format, was used as the input.

Step 2: Colour Space Conversion

The image was converted from RGB to CIELAB colour space. This colour space separates the luminance component (L*) from the chromatic components (a*, b*), facilitating targeted contrast enhancement.

Step 3: Channel Separation

The L* channel was extracted from the CIELAB image.

Step 4: CLAHE Application

CLAHE was applied to the L* channel with the following parameters: Clip Limit: 2.0 and Tile Grid Size: (8, 8).

Step 5: Image Reconstruction

The enhanced L* channel was merged back with the original a* and b* channels to reconstruct the CIELAB image, which was then converted back to RGB colour space to obtain the final enhanced fundus image.

Following the CLAHE enhancement on the L*channel, additional feature-specific enhancement techniques were applied to further improved the visibility of diagnostically relevant structures in the retinal fundus images. Two separate morphological operations were implemented to enhance blood vessels and optic disc, respectively:

i. Blood Vessel Enhancement

The L* channel-enhanced images were first converted to grayscale. Morphological top-hat and black-hat transformation using a rectangular kernel (15×15) were applied

to emphasize fine vessel structures. The resulting image were normalized and converted back to colour before being blended with the original image using weighted averaging (85% original, 15% enhanced) to preserve anatomical context while improving vessel contrast.

ii. Optic Disc Enhancement

The grayscale version of the enhanced images underwent a closing operation using a large elliptical kernel (30×30) to suppress background noise and isolate the optic disc region. The difference between the original and closed image was normalized, converted back to colour, and similarly blended with the input image to highlight the optic disc without distorting surrounding features.

These enhancement steps, as summarize in Figure 3.9, were applied uniformly across all images in the dataset to ensure consistency and avoid introducing class-specific biases. The resulting set of enhanced images formed an alternative version of the CDGN dataset, referred to as the **feature-enhanced CGDN dataset**. This version was used in parallel with the original dataset to investigate the effectiveness of image enhancement on classification performance. Figure 3.10 presents side-by-side comparison of the original and enhanced images for each class.

The proposed enhancement framework constitutes a targeted approach to bridge the visibility gap between raw retina fundus data and clinical-grade diagnostic features. First, CLAHE application on L* channel is justified as a mean to standardize local contrast across eight heterogeneous data sources without inducing the colour shifts that often occur in RGB-based equalization. Second, the Top-hat and Black-hat transformations are integrated to specifically amplify the high-frequency components of the retinal vasculature and micro-lesions, addressing the inter-class similarity between early-stage diseases. Finally, the

elliptical morphological closing provides a localized enhancement of the optic disc, a critical feature for glaucoma differentiation. By employing a weighted blending strategy, this framework achieves a novel balance: it amplifies faint, diagnostically features while preventing the loss of global context, thereby providing a more discriminative input the subsequent multiclass classification model.

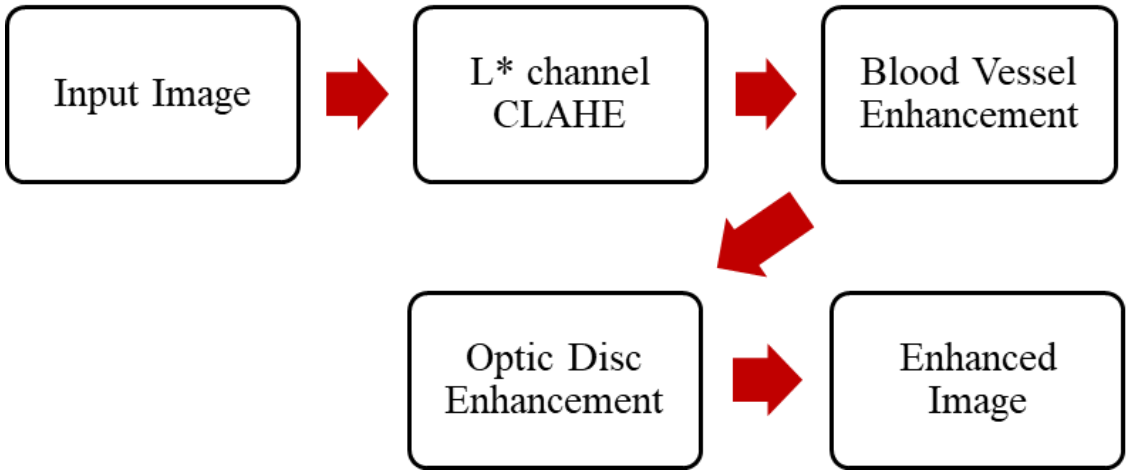


Figure 3.9: Image enhancement process

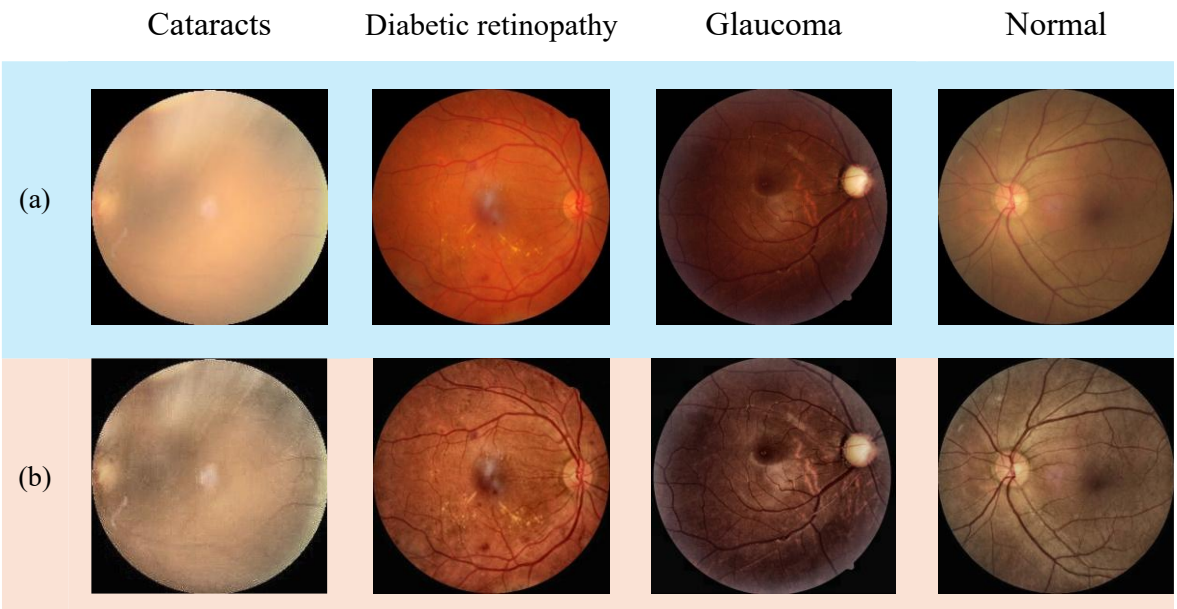


Figure 3.10: Comparison of original and enhanced retinal fundus images across four classes: (a) displays the original images from the CDGN dataset, (b) shows the corresponding enhanced images.

3.7 Model Training

In this research, convolutional neural network (CNN) architectures are employed to develop deep learning models for eye disease classification from retinal fundus images. As this task is framed as a multiclass image classification task, CNNs are chosen due to their demonstrated superiority in handling such tasks, particularly in medical image analysis. To enhance the performance of the models, this research incorporates transfer learning, attention mechanisms, and ensemble learning as key strategies within the training framework.

3.7.1 Transfer Learning

CNNs have achieved notable success in the domain of image classification and are widely adopted across various domains, including medical imaging. However, training deep CNN architectures from scratch generally necessitate extensive annotated datasets to achieve optimal predictive performance. In medical context, the collection of such datasets presents significant challenges due to ethical concerns surrounding patient privacy, the scarcity of labelled medical images, and the substantial cost and specialized expertise required for accurate annotation. Crowdsourcing alternatives are often impractical, as non-expert annotators may compromise label accuracy.

To overcome this constraint, transfer learning is employed. Transfer learning involves reusing knowledge obtained from solving one problem (typically on a large dataset) and applying it to a different but related task. This is analogous to a teacher transferring subject matter expertise to students: the pre-trained model (teacher) can guide the new model (student) to learn faster and more efficiently. Specifically, CNNs pre-trained on large-scale datasets such as ImageNet—comprising millions of labelled images—can be adapted for new tasks by reusing their learned feature representations, thereby eliminating the need to

train a model from scratch. Employing a pre-trained model is advantageous because it reduces training time, lowers computational cost, and improves performance, especially when training data is limited (Alzubaidi et al., 2021).

In this research, five CNN architectures—VGG16, Inception-v3, ResNet-50, DenseNet121, and EfficientNet-B0—were adopted. Table 3.4 provides an overview of the selected pre-trained models, including their top-1 accuracy on the ImageNet validation set. These models are chosen based on their proven performance and widespread use in prior studies. By leveraging these pre-trained weights and biases from ImageNet, the models are retrained on the CDGN dataset rather than being trained from scratch.

Table 3.4: Overview of Selected CNN Architectures

CNN Architecture	Author(s)	Number of Parameters	Dataset Trained	Top-1 Accuracy
VGG16	Simonyan and Zisserman (2014)	138.4 million	ImageNet	71.3%
Inception-v3	Szegedy, Vanhoucke, et al. (2015)	23.9 million	ImageNet	74.9%
ResNet-50	He et al. (2016)	25.6 million	ImageNet	77.9%
DenseNet121	Huang et al. (2017)	8.06 million	ImageNet	75.0%
EfficientNet-B0	Tan and Le (2019)	5.3 million	ImageNet	77.1%

All models were implemented using the Keras deep learning framework with TensorFlow as the backend. The base CNNs were loaded without their top classification layers (include_top = False), allowing them to function as feature extractors. Custom classification layers were then added on top to tailor the model to the specific four-class task in this research. The architecture for each adapted model includes:

- **Base Model**: A frozen pre-trained CNN (with ImageNet weights), preserving general feature representations.
- **Custom Fully Connected Layers**:
    - Dense Layer with 1024 units, followed by Batch Normalization and Dropout (rate = 0.5)
    - Dense Layer with 512 units, followed by Batch Normalization and Dropout (rate = 0.5)
    - Dense Layer with 256 units, followed by Batch Normalization and Dropout (rate = 0.5)
- **Output Layer**: Dense layer with 4 units and softmax activation for multiclass classification.

The overall model architecture—consisting of the pre-trained base model and custom fully classification head—is illustrated in Figure 3.11. Table 3.5 presents the model size and parameter counts for each pre-trained CNN architecture.

Table 3.5: Model Size and Number of Parameters (Trainable and Non-Trainable) for Each Selected Pre-trained CNN Models

Pre-trained models	Size (MB)	Number of Trainable Parameters	Number of Non-trainable Parameters
VGG16	262.60	54,116,988	14,721,936
Inception-v3	493.22	107,488,772	21,806,880
ResNet50	886.06	208,155,140	23,594,880
DenseNet121	51.40	6,435,844	7,038,016
EfficientNet-B0	16.86	370,244	4,050,467

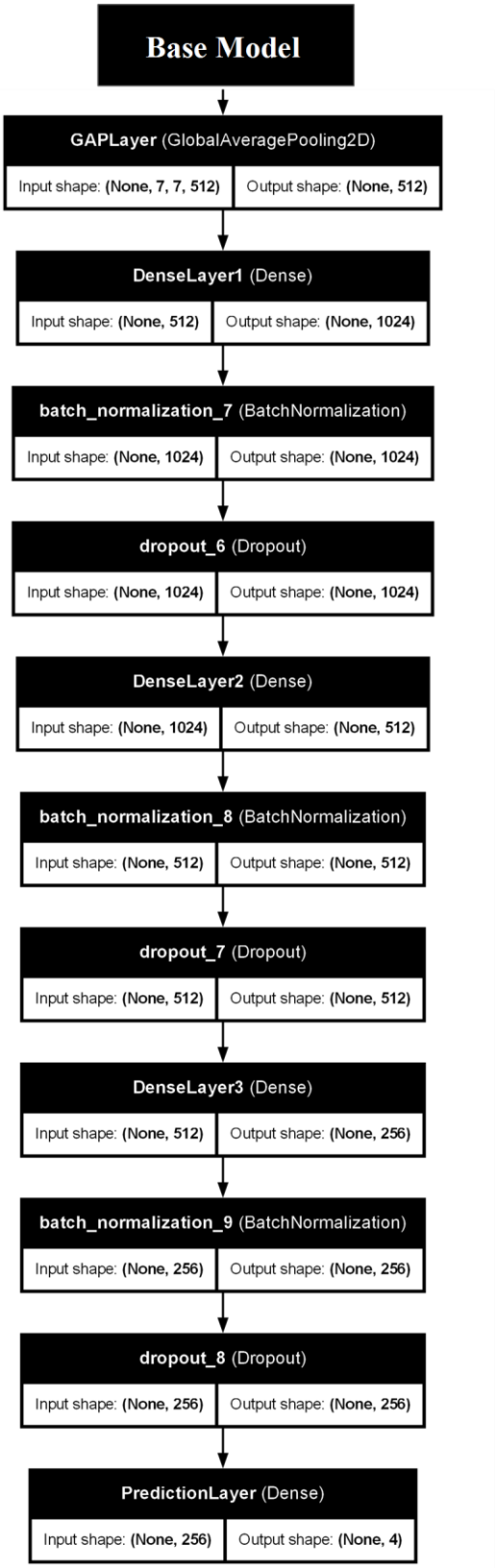


Figure 3.11: Architecture of the CNN model used in this research: a pre-trained base model, followed by custom fully connected layers and a final softmax output layer.

Each model was compiled using the categorical cross-entropy loss function, suitable for multiclass classification, and optimized with the Adam optimizer with a learning rate of 0.001. Training was conducted with a batch size of 64 over 50 epochs. To mitigate overfitting, early stopping was employed by monitoring the validation loss, with a patience threshold set at 10 epochs. The key training hyperparameters are listed in Table 3.6.

Table 3.6: Training Hyperparameter used for all Pre-trained CNN Models

Hyperparameter	Value
Loss function	Categorical cross-entropy
Optimizer	Adam
Learning rate	0.001
Batch size	64
Epochs	50
Early Stopping	Validation loss (patience = 10 epochs)

During the process of model training, the issue of class imbalance in the CDGN dataset was carefully addressed to prevent biased learning. Specific techniques such as class weighting and training-time data augmentation were implemented to ensure balanced learning across all classes. These methods are detailed in the next subsection 3.7.1.1.

3.7.1.1 Handling Class Imbalance

In medical image classification tasks, class imbalance is a common challenge, particularly when certain disease categories are underrepresented in the dataset. The CDGN dataset used in research exhibits such imbalance, with significant differences in the number of images among the four classes: cataracts, diabetic retinopathy, glaucoma, and normal. If unaddressed, this imbalance can bias the model toward majority classes, resulting in poor

generalization for underrepresented diseases and reduced overall performance. To address this issue, two techniques were employed during model training:

a. Class Weights

Class weighting was applied to give proportionally higher importance to the underrepresented classes during training. Class weights were calculated using the compute_class_weight function from the sklearn.utils.class_weight module, based on the distribution of class frequencies within the training set. These computed weights were then incorporated into the training process. By penalizing misclassifications of minority class instances more heavily, the model was encouraged to learn more representative and discriminative features across all classes. This approach helped improve both model fairness and accuracy.

b. Training-Time Data Augmentation

To further combat class imbalance and enhance generalization, data augmentation was applied during training. Using the ImageDataGenerator function in Keras, augmentation was performed dynamically, generating modified version of training images on-the-fly. The augmentation techniques used include horizontal flipping, vertical flipping, and zooming (with a zoom range of 0.1). These augmentations introduce variability into the training data without increasing the actual dataset size, thereby helping the model generalize better to unseen images while reducing the risk of overfitting.

By incorporating these strategies, the impacts of class imbalance in the CDGN dataset were reduced, ensuring that the trained models did not disproportionately favour the majority classes and maintain fair performance across all disease categories.

3.7.2 Attention Mechanism

To further enhance model performance, particularly in focusing on most relevant regions within retinal fundus images, an attention mechanism was integrated into the CNN architectures. While CNNs inherently capture spatial hierarchies of features, they may not always emphasize the most informative areas, especially in complex medical images where key diagnostic patterns can be subtle. The attention mechanism enables the model to "focus" selectively on regions in the image that are most relevant to the task, improving both accuracy and interpretability. Originally developed within the framework of neural machine translation using sequence-to-sequence (Seq2Seq) architectures, attention mechanisms enable models to assign dynamical importance to various components of the input when generating outputs (Bahdanau et al., 2014). Inspired by human visual attention system, these mechanisms assign higher weights to more informative features during predictions, while suppressing less relevant ones. In computer vision, including medical image classification, attention modules help highlight regions of diagnostic interest, such as lesions, vascular patterns, or optic disc abnormalities in retinal fundus images.

In this research, a spatial attention mechanism was integrated into the CNN architecture to enhance ability of the model to extract and focus on disease-relevant features within retinal fundus images. This mechanism guides the model to attend to clinically significant regions—such as the blood vessels, optic disc, or lesions—that are critical for diagnosing eye diseases. By focusing on these areas, the model can better identify subtle but important patterns associated with each eye disease class, potentially improving classification performance. The spatial attention module was implemented as a custom Keras function, inspired by common attention frameworks in computer vision. It operated by first computing average pooling and max pooling across the input feature map's channel

dimension. These two pooled feature maps were then concatenated along the channel axis, resulting in a combined representation that highlights both average and maximum activation patterns. A convolution layer employing a 7×7 kernel, followed by batch normalization and a sigmoid activation function, was utilized to generate a spatial attention map that emphasizes salient image regions. This learned attention map was element-wise multiplied with the original input feature map, effectively reweighting spatial regions to emphasize diagnostically relevant features while suppressing less informative areas. The output of the spatial attention module was subsequently forwarded to the fully connected layers for classification.

The attention-based models follow a similar architecture and training configuration as the transfer learning models described in Section 3.7.1, with key difference being the integration of the spatial attention module. This module was inserted after the feature extraction by the base CNN model and before the custom classification layers. Each attention-based model consists of:

- Pre-trained CNN base model, loaded without its top layers and initialised with ImageNet weights.
- Spatial attention module integrated to refine the feature maps before classification.
- Custom classification head composed of:
    - Dense Layer with 1024 units, followed by Batch Normalization and Dropout (rate = 0.5)
    - Dense Layer with 512 units, followed by Batch Normalization and Dropout (rate = 0.5)

- Dense Layer with 256 units, followed by Batch Normalization and Dropout (rate = 0.5)

- Output Layer: Dense layer with 4 units and softmax activation.

This integration enables the model to amplify key spatial features before classification, thus improving its focus on disease-relevant regions within the retinal fundus images. A schematic overview of the modified architecture, showing the position of the attention module within the network, is illustrated in Figure 3.12.

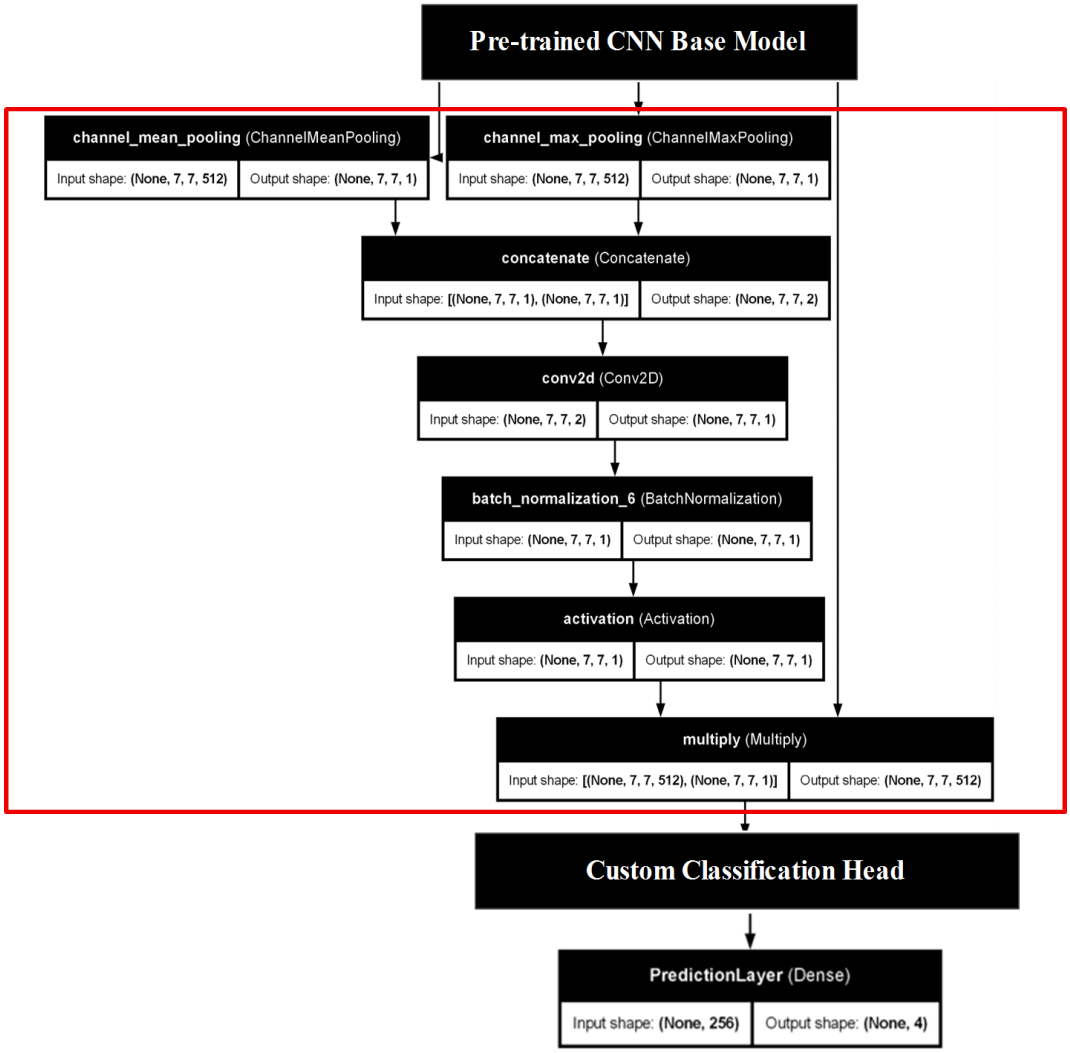


Figure 3.12: Architecture of CNN model with an integrated spatial attention module. The attention block, highlighted in red, is positioned between the base model and the custom classification head.

The training configuration for the attention-based models follows the same setup as the transfer learning models described in Section 3.7.1. All models were trained using the categorical cross-entropy loss function and optimized using the Adam algorithm (learning rate = 0.001), with a batch size of 64 over 50 epochs. To address the issue of overfitting, early stopping was implemented by monitoring the validation loss, with a patience parameter set to 10 epochs. The complete hyperparameter settings are summarized in Table 3.6. To mitigate the issue of class imbalance, the same strategies were applied during training: class weighting and training-time data augmentation (see Section 3.7.1.1).

3.7.3 Ensemble Learning

In multiclass image classification tasks, individual deep learning models often exhibit varying performance across different classes. For instance, one pre-trained model may perform better at identifying specific diseases, while another may excel in recognizing others. To capitalize on the strengths of each models while compensating for their individual limitations, ensemble learning is employed. Ensemble learning is a strategy that combines prediction from multiple models to improve the overall accuracy, robustness, and generalizability of the classification model.

There are several common ensemble learning techniques, including majority voting, averaging and weighted averaging. Majority voting, or max voting, is widely adopted for classification tasks by taking the most frequent class label predicted by multiple models as the final prediction. Averaging, on the other hand, computes the mean of predictions across models and selects the class with the highest average probability. This approach, often referred to as soft voting, is suitable when model produce probability distributions over the target classes, as is typical with CNNs using softmax output layers. Weighted averaging

extends this method by assigning different weights to each model, according to their relative contribution to the predictive outcome.

In this research, an ensemble of five attention-based CNN models—VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0—was implemented using soft voting (averaging) approach. The ensemble implementation was carried out using a custom function developed for this purpose. The function accepts a list of the trained models along with the test datasets. For each test image, the function collects prediction probabilities from all five models. These predictions are then averaged across the models to form a single, consolidated probability vector for each image. The final classification output is determined by identifying the class index corresponding to the highest probability within the averaged prediction vector. This method ensures that the ensemble benefits from the complementary strengths of each model. Figure 3.13 illustrates the architecture of the soft voting ensemble model, where predictions from five attention-based CNNs are averaged to obtain the final output.

To ensure a fair comparison, the same test set used for evaluating individual attention-based models was reused during ensemble evaluation. This consistent evaluation setting allows direct performance comparisons between the ensemble and its constituent models. The soft voting ensemble approach used here offers an effective and computationally simple strategy for multiclass classification. It capitalizes on the natural output format of deep learning, that is the probability vectors, and requires no additional model training, demonstrating suitability for enhancing performance in medical image classification applications.

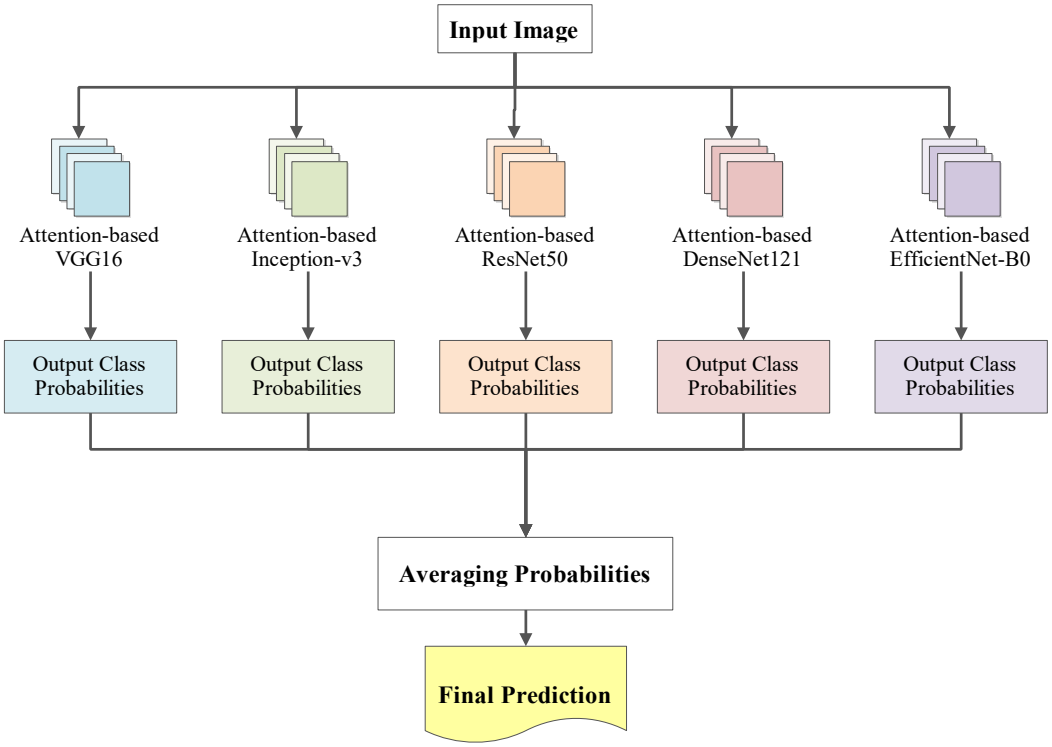


Figure 3.13: Overview of soft voting ensemble mechanism where the class probabilities predicted by five attention-based CNN models are averaged to generate the final prediction.

3.8 Modal Evaluation

This section delineates the evaluation strategies and performance metrics employed to assess the models on the test set. Two main approaches were used: confusion matrix analysis and derived performance metrics, and evaluation based on the AUC-ROC curve. These methods provide both class-specific insights and an overall understanding of the model's discriminative capability, particularly valuable in multiclass classification tasks with potential class imbalance.

3.8.1 Confusion Matrix and Derived Metrics

In this research, a confusion matrix was utilized to assess the classification performance of the deep learning models on the test set. A general structure of a 2×2 confusion matrix for binary classification is illustrated in Table 3.7.

Table 3.7: Confusion Matrix for Binary Classification

		Actual Classes	
		Actual Positive	Actual Negative
Predicted Classes	Predicted Positive	True Positive (TP)	False Positive (FP)
	Predicted Negative	False Negative (FN)	True Negative (TN)

For multiclass classification models with n target classes, the dimensions of the confusion matrix expand to $n \times n$. Table 3.8 visualizes the conceptual layout for a multiclass confusion matrix (Krüger, 2016). When evaluating performance of the model on class C_k, the term True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are interpreted accordingly based on class-wise outcomes (Krüger, 2016).

Table 3.8: Confusion Matrix for Multiclass Classification (Class C_k) (Krüger, 2016)

		Actual Classes		
		$C_o \dots C_{k-1}$	C_k	$C_{k+1} \dots C_n$
Predicted Classes	$C_o \dots C_{k-1}$	TN	FN	TN
	C_k	FP	TP	FP
	$C_{k+1} \dots C_n$	TN	FN	TN

Confusion matrix offers a comprehensive analysis of the model's predictions by comparing the actual ground truth labels with the predicted labels output by the model. Each element in the confusion matrix corresponds to a classification outcome:

- True Positive (TP) occurs when the model correctly identifies a positive instance.
- True Negative (TN) refers to the correct prediction of a negative instance by the model.

- False Positive (FP) arise when a negative instance is incorrectly classified as positive, representing a Type I Error.
- False Negative (FN) occurs when a positive instance is mistakenly predicted as negative, indicating a Type II Error.

By utilizing a confusion matrix, several evaluation metrics can be derived to quantify the performance of a classification model (Alzubaidi et al., 2021), including:

i. Accuracy quantifies the proportion of correctly classified instances relative to the total number of instances by the model, mathematically defined in Equation 3.1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Equation 3.1}$$

ii. Precision focuses on the accuracy of positive predictions by quantifying the model's ability to correctly identify positive instances among all positive predictions. Mathematically, it is defined as Equation 3.2.

$$Precision = \frac{TP}{TP + FP} \quad \text{Equation 3.2}$$

iii. Recall, also referred to as sensitivity or the true positive rate, evaluates the model's ability in identifying all positive cases. as mathematically represented by Equation 3.3.

$$Recall = \frac{TP}{TP + FN} \quad \text{Equation 3.3}$$

iv. F1-score represents the harmonic mean of precision and recall, providing a balance measure of the model performance across both positive and negative classes, which mathematically expressed in Equation 3.4.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$ Equation 3.4

These metrics were used in this research to provide a comprehensive evaluation of each model's performance, especially in handling class imbalance across multiple categories.

3.8.2 Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

In this research, the Receiver Operating Characteristic (ROC) curve was used to evaluate the classification model's performance across various thresholds values. The ROC curve is a graphical representation that illustrates the balance between the True Positive Rate (TPR) and the False Positive Rate (FPR) (Terra, 2024).

- The True Positive Rate (TPR) or recall quantifies the proportion of actual positive cases correctly identified by the model and is mathematically expressed in Equation 3.3.
- The False Positive Rate (FPR) measures the proportion of actual negatives cases that are incorrectly classified as positives, represented mathematically in Equation 3.5.

$$FPR = \frac{FP}{FP + TN}$$ Equation 3.5

In the ROC curve, the FPR is on the x-axis, and the TPR is on the y-axis. A model whose ROC curve approaches the top-left corner of the plot is indicative of better classification performance.

To summarize the performance into a single evaluative metric, the Area Under the ROC Curve (AUC) is used. It reflects the model's capacity to distinguish between the positive and negative classes. The AUC value, ranging from 0 to 1, serves as an indication

of a model's performance in term of class separability (Terra, 2024). An excellent model will have an AUC value close to 1, signifying a high separability degree between classes. Conversely, a poor model will have an AUC value approaching 0, indicating the worst separability and suggesting that the model classify the negative case as positive and vice versa. An AUC value of 0.5 implies that the model lacks any class separation ability, equivalent to random guessing.

In this research, multiclass ROC curves were generated using the One-vs-All method, where each class is treated as positive class while all others are treated as negative. In the One-vs-All approach, the model's output probabilities for each class are used directly. If two classes produce equal probabilities for a given instances, this does not affect the ROC calculation, as the metric considers the ranking of positive versus negative instances rather than the final predicted class. This metric provides a comprehensive assessment of model performance across all classification thresholds and is particularly advantageous for imbalanced datasets, where accuracy alone may not reflect true performance. For a four-class classification task (cataracts, diabetic retinopathy, glaucoma, normal), four ROC curves were plotted—one per class. This approach assesses how well the model distinguishes each class from the rest independently. The AUC scores for each class, as well as the average AUC scores, were used to evaluate and compare models.

In summary, the model performance evaluation in this research was conducted using both confusion matrix-based metrics and AUC-ROC analysis. The confusion matrix provides detailed insights into the classification outcomes for each class, facilitating the computation of essential metrics such as accuracy, recall, precision, and F1-score. In parallel, the ROC curve and the corresponding AUC values offer a threshold-independent assessment

of the model's ability to distinguish between classes. Collectively, these evaluation techniques provide a thorough and dependable assessment of the classification models' overall effectiveness.

3.9 Chapter Summary

This chapter has detailed the methodology implemented to develop and evaluate an enhanced deep learning framework for the classification of four eye conditions—cataracts, diabetic retinopathy, glaucoma, and normal—using retinal fundus images. The overall workflow comprised five main stages: data collection, data preprocessing, image enhancement, model training, and model evaluation. Publicly available datasets were combined and curated to form a larger dataset referred to as the CDGN dataset. Standardized preprocessing steps such as cropping and resizing was applied to ensure consistency and robustness across all input images. To improve image quality and emphasize clinically visual features, image enhancement techniques were applied including applying CLAHE to the L* channel, and enhancements of blood vessels and optic disc visibility through morphological operations. This process generated an enhanced version of the dataset, used in parallel with the original. Subsequently, several deep learning models based on transfer learning were developed, including attention-based variants that incorporated spatial attention mechanisms, and an ensemble strategy using soft voting to further improve classification performance. The models were trained using optimized hyperparameters and validated using standard performance metrics. Finally, model evaluation was performed through confusion matrix analysis and AUC-ROC curves to assess classification effectiveness across all classes. This structured approach lays the foundation for the experimental results and analysis presented in the following chapter.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Overview

This chapter presents the results and analysis of the performance of convolutional neural network (CNN) models in the classification of retinal fundus images. It begins with a brief description of the experimental setup, encompassing dataset characteristics, preprocessing steps, image enhancement techniques, and training configurations. This provides the necessary context for the subsequent evaluation of the five CNN architectures: VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0. Each model is evaluated on both the original and feature-enhanced CDGN dataset using standard performance metrics, including accuracy, recall, precision, F1-score, and AUC. Furthermore, confusion matrices are utilized to provide deeper insight into class-specific classification behaviour. Where applicable, visualizations such as tables, graphs and annotated images are used to enhance clarity and facilitate comparison. The analysis is extended to explore the impact of incorporating attention mechanisms and ensemble learning strategies. The chapter concludes with a comparative discussion of the findings and their implications for automated multiclass eye disease classification.

4.2 Experimental Setup

To assess the effectiveness of CNN models, a series of experiments were conducted using both original and enhanced datasets. This section provides a brief overview of the dataset, preprocessing and enhancement steps, architectural configurations, and the quantitative metrics employed to ensure a robust evaluation.

4.2.1 Dataset Description

The CDGN dataset comprises retinal fundus images categorized into four classes: Cataracts, Diabetic Retinopathy, Glaucoma, and Normal. To ensure unbiased evaluation, the dataset was partitioned into training, validation, and test subsets using a stratified split, thereby preserving the original class distribution across all subsets.

4.2.2 Image Preprocessing and Enhancement

Input images underwent cropping to remove extraneous black region and were resized to a uniform resolution compatible with the input specifications of the CNNs. For the feature-enhanced dataset, contrast-limited adaptive histogram equalization (CLAHE) and morphological operations were applied. These techniques were specifically designed to enhance the visibility of clinically relevant features such as blood vessels, optic disc, microaneurysm, cotton wool spots, and hard exudates.

4.2.3 Model Architectures

Five diverse CNN backbones were selected for this research: VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0. These architectures were initialized with weights pre-trained on the ImageNet dataset, leveraging a transfer learning approach to accelerate convergence. In addition to these baseline models, attention-based variants were implemented to refine feature localization in diagnostically significant regions. Finally, ensemble learning strategies were employed, combining and averaging predictions from multiple models to improve overall classification accuracy and enhance robustness.

4.2.4 Training Configuration

All models were trained using the Adam optimizer with a learning rate of 0.001, a batch size of 64, categorical cross-entropy as the loss function. Training was scheduled for 50 epochs, and early stopping was employed to mitigate overfitting by monitoring validation loss with a patience of 10 epochs. Model checkpointing was also implemented to ensure that only the weights corresponding to the best validation performance were retained for final testing. This combination ensures efficient and effective model training while reducing the risk of overfitting.

4.2.5 Evaluation Metrics

Model performance was assessed using standard classification metrics: accuracy, recall, precision, F1-score, and area under the receiver operating characteristic curve (AUC). To further analyse class-wise misclassifications and model sensitivity, confusion matrices were generated, ensuring an objective measure of generalization performance.

4.3 Model Performance on Original CDGN Dataset (Baseline)

This section presents the baseline performance of the five selected pre-trained CNN models: VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0 trained on the original CDGN dataset. Each model was trained and validated using consistent hyperparameters and data splits to ensure fair comparison. The performance was assessed through training and validation accuracy and loss curves, confusion matrices, classification reports, and AUC metrics. The results presented in the following subsections serve as a baseline for comparison with feature-enhanced CDGN dataset, attention-based models, and ensemble approach discussed later in the chapter.

4.3.1 VGG16

a. Training and Validation Performance

Figure 4.1 presents the training and validation accuracy and loss curves for the VGG16 model. The training accuracy exhibited a steady upward trend, while the validation accuracy followed a similar pattern, albeit with more fluctuation, and peaked at around 0.80 at epoch 24. The overall proximity between training and validation accuracy suggests that VGG16 generalizes reasonably well, with no evident signs of severe overfitting. The training loss showed a consistent decline, indicating effective learning, while the validation loss also decreased but with more variability observed. Notably, the lowest validation loss occurred at epoch 22, after which a slight divergence between the training and validation loss became apparent. This gap suggests the onset of overfitting beyond that point. Overall, the VGG16 model displays stable convergence and robust training behaviour, benefiting from effective regularization.

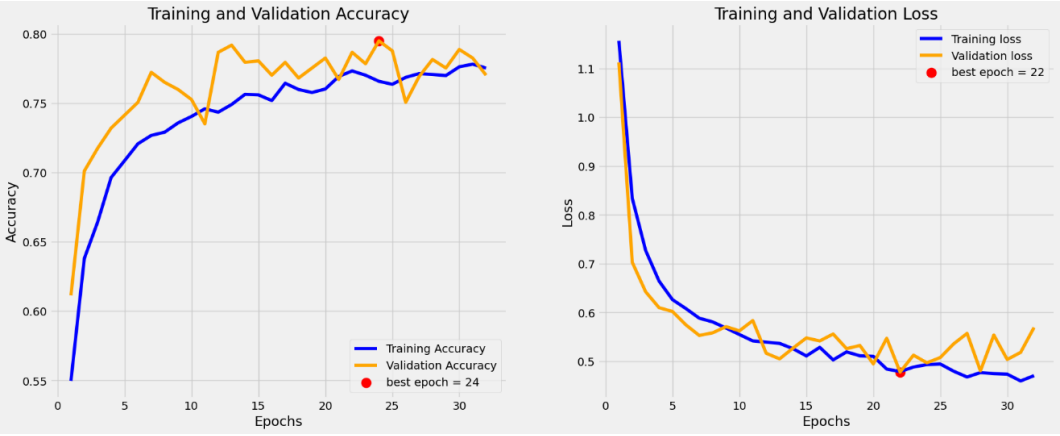


Figure 4.1: Training and validation accuracy and loss plots of VGG16

b. Confusion Matrix

The confusion matrix in Figure 4.2 provides a granular summary of VGG16's classification performance across the four classes. The diagonal elements confirm that

VGG16 is particularly effective at identifying Cataracts (n = 118) and Normal cases (n = 261), indicating that distinctive features for cataracts were effectively learned. However, a notable confusion was observed between Diabetic Retinopathy and the Normal class, with 74 diabetic retinopathy instances incorrectly classified as normal. This confusion is likely stem from subtle features of early-stage diabetic retinopathy, which often overlap with the visual characteristics of a healthy retina. Similarly, 26 Glaucoma cases were misclassified as Normal, further indicating that the model struggles disease boundaries where clinical features are less pronounced or visually similar.

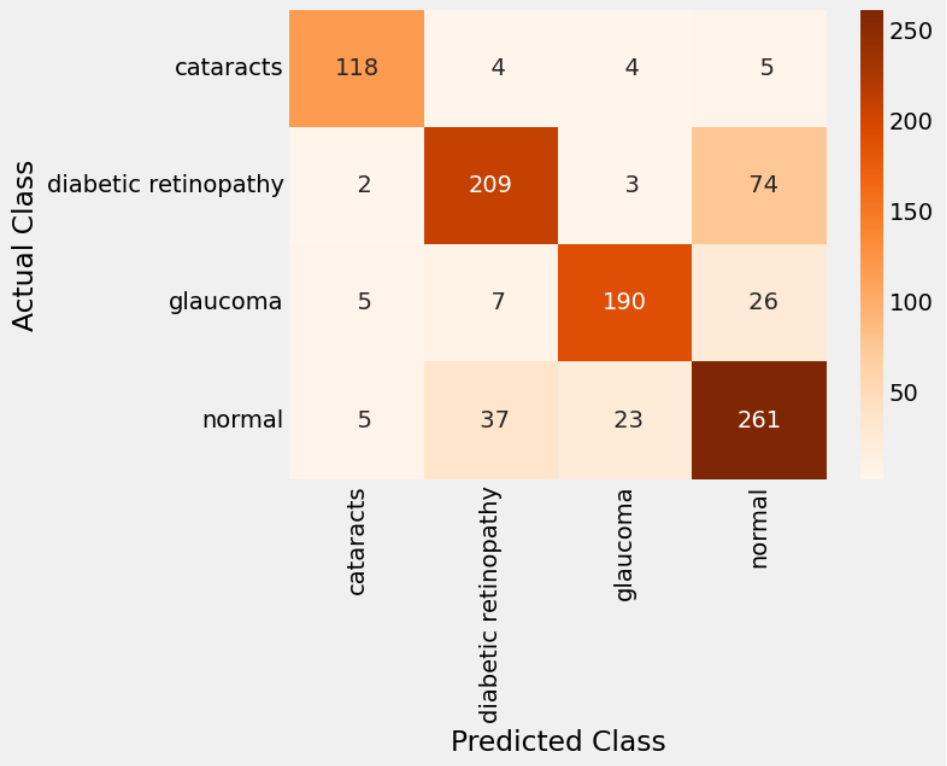


Figure 4.2: Confusion matrix of VGG16

c. Classification Report

Table 4.1 presents the classification performance metrics of the VGG16 model, including accuracy, recall, precision, and F1-score for each classes. The model achieved its highest sensitivity for Cataracts with a recall of 0.9008 and a precision of 0.9077, resulting

in a robust F1-score of 0.9042. Glaucoma detection also performed well, with yielding an F1-score of 0.8482. These results indicate that the model effectively differentiates cataracts and glaucoma from other classes. Conversely, Diabetic Retinopathy showed the lowest recall (0.7257), confirming that a significant portion of positive cases were missed—consistent with the confusion observed in Figure 4.2. However, the high precision (0.8132) indicates that when the model does predict diabetic retinopathy, it is usually correct. While the Normal class achieved a high recall (0.8006), its lower precision (0.7121) suggests a tendency to misclassify diseased images as normal. Overall, VGG16 achieved an accuracy of 79.96%. The macro-average F1-score of 0.8184 and weighted-average F1-score of 0.8003 indicate consistent performance across classes, with a slight imbalance due to lower recall for Diabetic Retinopathy.

Table 4.1: Classification Metrics of VGG16

Class	Recall	Precision	F1-score	Support
Cataracts	0.9008	0.9077	0.9042	131
Diabetic retinopathy	0.7257	0.8132	0.7670	288
Glaucoma	0.8333	0.8636	0.8482	228
Normal	0.8006	0.7131	0.7543	326
Accuracy	0.7996			973
Macro-Average	0.8151	0.8244	0.8184	
Weighted Average	0.7996	0.8042	0.8003	

d. AUC-ROC Analysis

The ROC analysis in Figure 4.3 illustrates VGG16's discriminative power. The model achieved high AUC values across all classes: Cataracts (0.99), Glaucoma (0.97), Diabetic Retinopathy (0.93), and Normal (0.90). The macro-average AUC of 0.9486 and the

weighted average AUC of 0.9388 reinforces the model's overall robustness in distinguishing between positive and negative cases. While these scores are high, the lower relative discriminative performance for Diabetic Retinopathy and Normal classes highlights the need for further refinement in reducing false negatives within those specific classes. Overall, the ROC analysis indicates that VGG16 performs well in distinguishing the four classes.

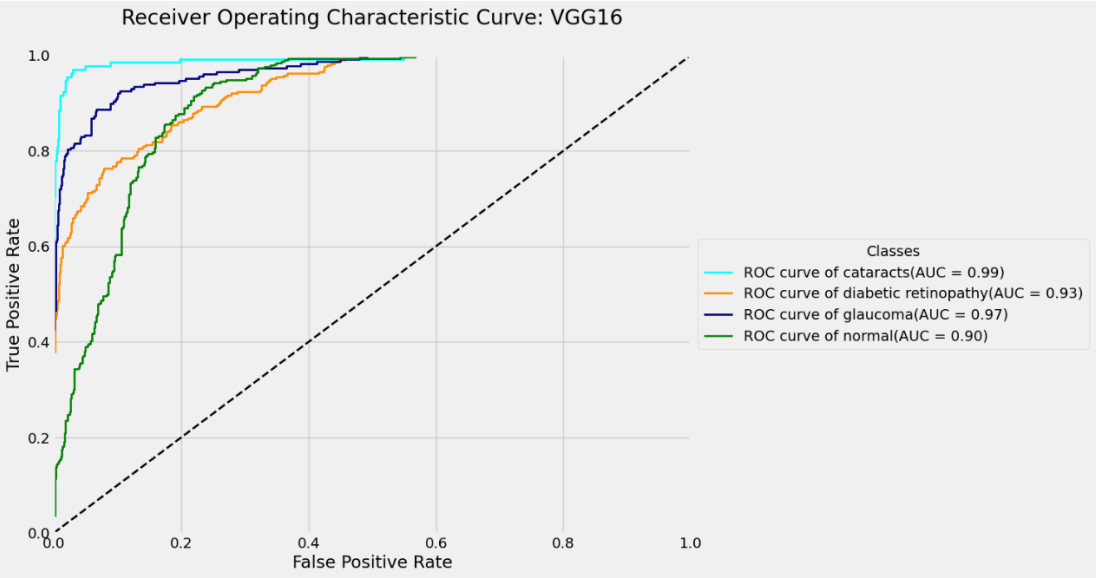


Figure 4.3: ROC curves and corresponding AUC scores for each class of VGG16

e. Discussion

The VGG16 model showed promising results in multiclass classification of eye diseases. The training and validation curves indicate stable learning behaviour with minimal signs of overfitting. Class-wise evaluation reveals high precision and recall for Cataracts and Glaucoma, indicating that the model effectively captures the distinguishing features of these conditions. The primary limitation observed is the difficulty in resolving subtle clinical features in early-stage Diabetic Retinopathy, leading to frequent misclassification as Normal. Despite being a shallower architecture that its successors, VGG16 provides a reliable benchmark for evaluating subsequent enhancements.

4.3.2 Inception-v3

a. Training and Validation Performance

Figure 4.4 illustrates the learning progression of the Inception-v3 model over 35 epochs. The training accuracy showed a gradual increase, reaching approximately 0.72 by epoch 35. Validation accuracy followed a similar trend, peaking near 0.73 at epoch 20, with minor fluctuations throughout. The close alignment between these metric indicates that the model possesses well generalization capability with minimal evidence of overfitting. Both training and validation loss curves exhibited a consistent downward trend; while the validation loss fluctuated moderately, the relatively small gap between the two loss curves confirms stable convergence and effective generalization.

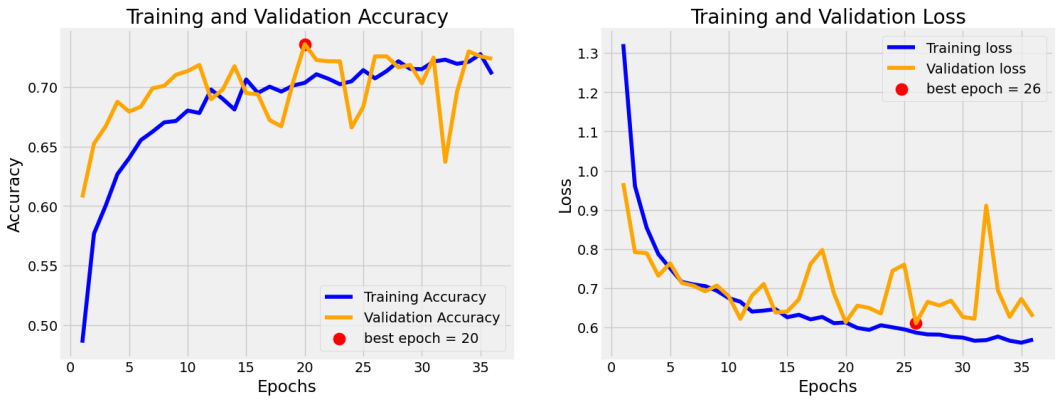


Figure 4.4: Training and validation accuracy and loss plots of Inception-v3

b. Confusion Matrix

Figure 4.5 presents the confusion matrix for Inception-v3, offering a detailed breakdown of class-wise predictive performance. The model correctly identified 112 Cataracts cases, with minor misclassifications into Diabetic Retinopathy (n = 5), Glaucoma (n = 5), and Normal (n = 9). Similarly, Glaucoma (181 correctly classified instances) and Normal (281 correctly classified instances) classes were well-represented. However, the most significant diagnostic challenge occurred with the Diabetic Retinopathy class, where

138 instances were incorrectly predicted as Normal. The substantial error rate suggests that Inception-v3 struggles to differentiate the subtle or early-stage pathological features of diabetic retinopathy from healthy retinal structures. This trend of misclassifying diseased cases as Normal—also observed to a lesser extent in Glaucoma (n = 31)—reveal a specific vulnerability in the model's ability to resolve fine-grained clinical boundaries. Overall, the model shows good performance in identifying cataracts, glaucoma, and normal fundus, while revealing substantial challenges in accurately detecting diabetic retinopathy.

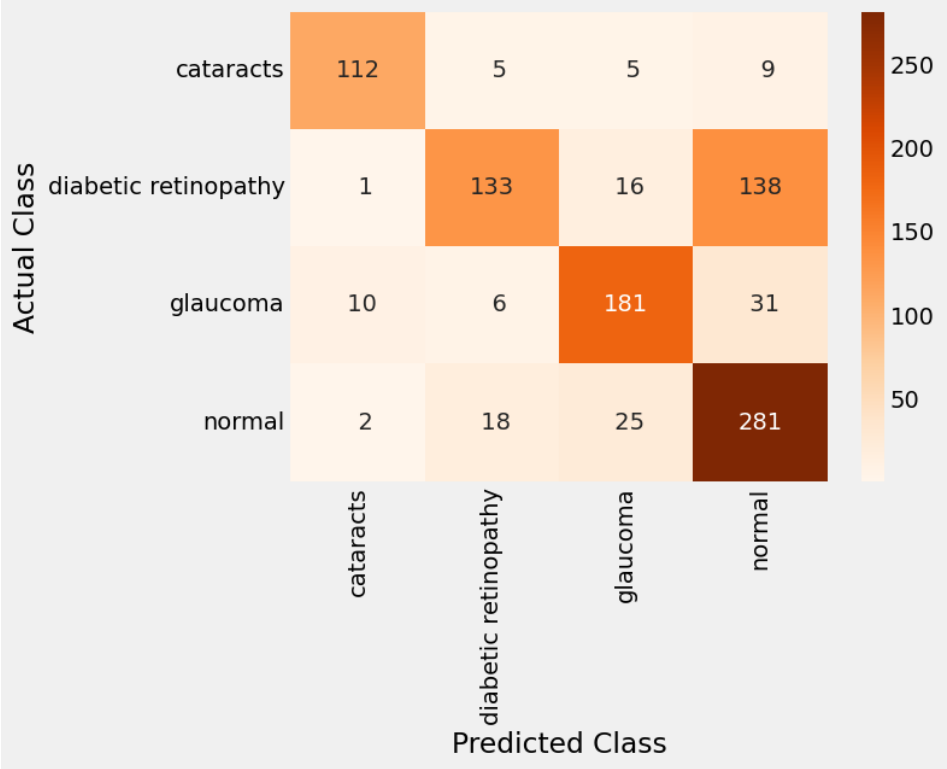


Figure 4.5: Confusion matrix of Inception-v3

c. Classification Report

Table 4.2 outlines the classification metrics for the Inception-v3 model, including accuracy, recall, precision, F1-score and support for each class. The model achieved its highest performance in Cataracts, with a precision of 0.8960 and a recall of 0.8550, resulting in a solid F1-score of 0.8750. Glaucoma detection was also consistent, with balanced

precision (0.7974) and recall (0.7939), yielding an F1-score of 0.7956. This suggests consistent performance in identifying glaucoma, with relatively low misclassification rates. In contrast, Diabetic Retinopathy exhibited poor sensitivity with a recall of only 0.4681, indicating that more than half of the cases were missed. Interestingly, the precision for Diabetic Retinopathy (0.8210) suggests that while the model is conservative in assigning this label, its predictions are generally reliable when made. The Normal class yielded the highest recall (0.8620) but lowest precision (0.6122), reinforcing the confusion matrix findings that the model frequently mislabels diseased images as healthy. Such misclassifications may have clinical implications, potentially leading to underdiagnosis. Overall, Inception-v3 reached an accuracy of 72.66%, with a macro and weighted average F1-score of 0.7444 and 0.7191, respectively. These scores suggest moderate classification performance, with relatively better results for cataracts and glaucoma than for diabetic retinopathy and normal cases.

Table 4.2: Classification Metrics of Inception-v3

Class	Recall	Precision	F1-score	Support
Cataracts	0.8550	0.8960	0.8750	131
Diabetic retinopathy	0.4681	0.8210	0.5911	288
Glaucoma	0.7939	0.7974	0.7956	228
Normal	0.8620	0.6122	0.7159	326
Accuracy	0.7266			973
Macro-Average	0.7431	0.7816	0.7444	
Weighted Average	0.7266	0.7556	0.7191	

d. AUC-ROC Analysis

The ROC curves and AUC scores for Inception-v3in Figure 4.6 reflects strong overall discriminative power, despite sensitivity issues noted above. The model achieved high AUC scores for Cataracts (0.99) and Glaucoma (0.95). However, scores for Diabetic Retinopathy (0.89) and Normal (0.87) were comparatively lower, mirroring the challenges in distinguishing these two classes. The macro-average AUC score of 0.9277 and weighted average AUC score of 0.9136 underscore a generally robust capability to separate positive and negative instances across the dataset, even if localized classification boundaries require further refinement.

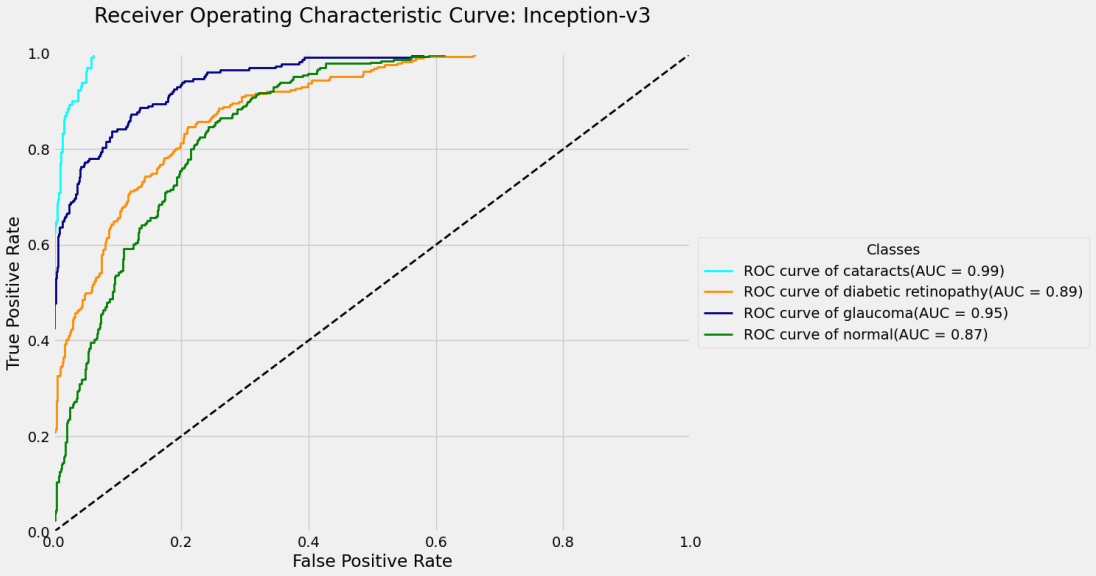


Figure 4.6: ROC curves and corresponding AUC scores for each class of Inception-v3

e. Discussion

Inception-v3 demonstrated moderate but promising performance in multiclass classification. The training and validation curves indicated stable convergence with minimal signs of overfitting. The model excelled at capturing the salient features of Cataracts and Glaucoma; however, it exhibits a distinct conservative bias toward the Normal class,

particularly when processing Diabetic Retinopathy images. This suggests that while multi-scale inception modules are effective for prominent structural changes, they may lack the sensitivity required for the micro-vascular features of early-stage diabetic retinopathy. Overall, while Inception-v3 shows strong capability in classifying certain eye diseases, its performance on diabetic retinopathy and normal cases highlights areas for further refinement. These findings emphasize the necessity of the proposed enhancement techniques and attention mechanisms to improve the detection of subtle features.

4.3.3 ResNet50

a. Training and Validation Performance

As shown in Figure 4.7, the training and validation performance of ResNet50 indicates steady learning progression.

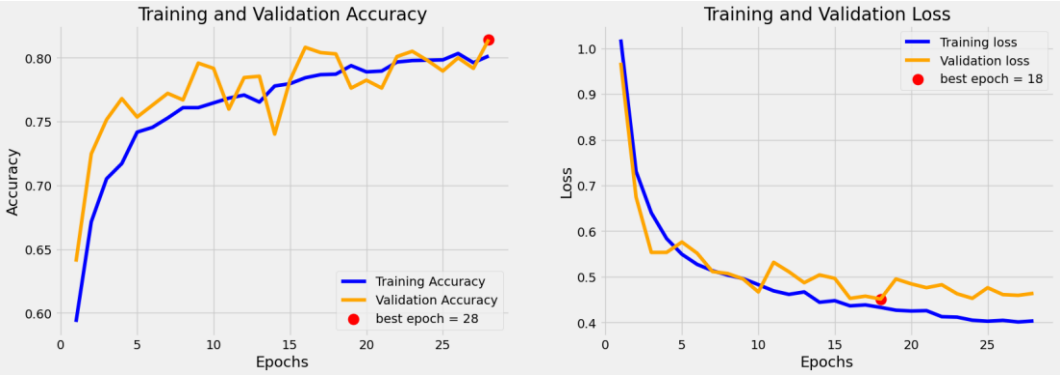


Figure 4.7: Training and validation accuracy and loss plots of ResNet50

The training accuracy increased consistently across the epochs, reaching approximately 0.80 by epoch 30. Validation accuracy closely tracked this trend, peaking at 0.82 around epoch 28, which was identified as the optimal epoch for model performance. While both training and validation loss exhibited a consistent downward trend, the validation loss reached its minimum earlier at epoch 18. This divergence suggests the potential onset of overfitting beyond this point; however, the gap between these two loss curves remains moderate,

indicating well-controlled regularization. Overall, ResNet50 demonstrated robust learning behaviour, achieving stable convergence and strong generalization capability, which indicate that the model is well-optimized for the task.

b. Confusion Matrix

The confusion matrix in Figure 4.8 details ResNet50's classification performance across the four target classes.

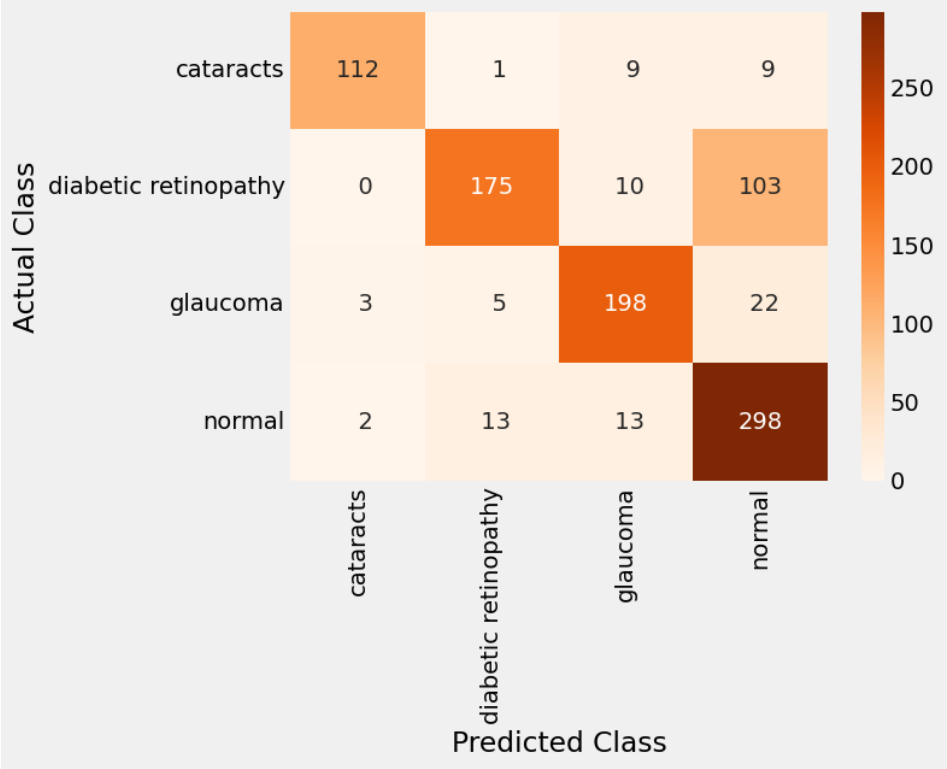


Figure 4.8: Confusion matrix of ResNet50

The model demonstrated high efficacy in identifying Cataracts (112 correctly classified cases) and Glaucoma (198 correct classified cases), with limited confusion with other classes. The Normal class achieved the highest correct classification count with 298 accurately identified instances. In contrast, Diabetic Retinopathy posed a significant challenge; although 175 cases were correctly classified, 103 instances were misclassified as Normal. This misclassification pattern—consistent with previous architectures—likely

stems from the subtle visual features of early-stage diabetic retinopathy that often mirror healthy retinal structures. Overall, ResNet50 performed particularly well in identifying Cataracts, Glaucoma, and Normal cases, while Diabetic Retinopathy remains a challenge.

c. Classification Report

Table 4.3 presents the classification performance of the ResNet50 model. The model achieved its highest precision for Cataracts (0.9573) and a strong recall of 0.8550, resulting in a high F1-score of 0.9032. Glaucoma classification also showed balanced performance, yielding a balanced F1-score of 0.8646 with minimal false positives. In contrast, Diabetic Retinopathy achieved high precision (0.9021) but a relatively low recall (0.6076), suggesting that while predictions were often correct when made, a significant number of true positive cases were missed. The Normal class attained the highest recall (0.9141), but its lower precision (0.6898) indicates that the model frequently over-predicted the Normal cases, leading to misclassification of diseased images.

Table 4.3: Classification Metrics of ResNet50

Class	Recall	Precision	F1-score	Support
Cataracts	0.8550	0.9573	0.9032	131
Diabetic retinopathy	0.6076	0.9021	0.7261	288
Glaucoma	0.8684	0.8609	0.8646	228
Normal	0.9141	0.6898	0.7863	326
Accuracy	0.8047			973
Macro Average	0.8113	0.8525	0.8201	
Weighted Average	0.8047	0.8287	0.8026	

Overall, ResNet50 achieved an accuracy of 80.47%, with a macro-average F1-score of 0.8201 and a weighted average F1-score of 0.8026. These metrics reflect good overall performance, particularly for cataracts and glaucoma, while highlighting diabetic retinopathy as the most challenging class.

d. AUC-ROC Analysis

Figure 4.9 visualizes the ROC curves for each class, highlighting ResNet50's discriminative ability. The model achieved high AUC scores: Cataracts (1.00), Diabetic Retinopathy (0.93), Glaucoma (0.97), and Normal (0.92). The perfect AUC score for Cataracts indicates near-absolute class separation. While the AUC scores for Diabetic Retinopathy and Normal remain high, their slight relative decrease aligns lower recall and higher misclassification rates noted in the confusion matrix. The macro-average AUC value of 0.9558 and weighted average AUC value of 0.9468 underscore the model's overall robust discriminative performance across diverse dataset.

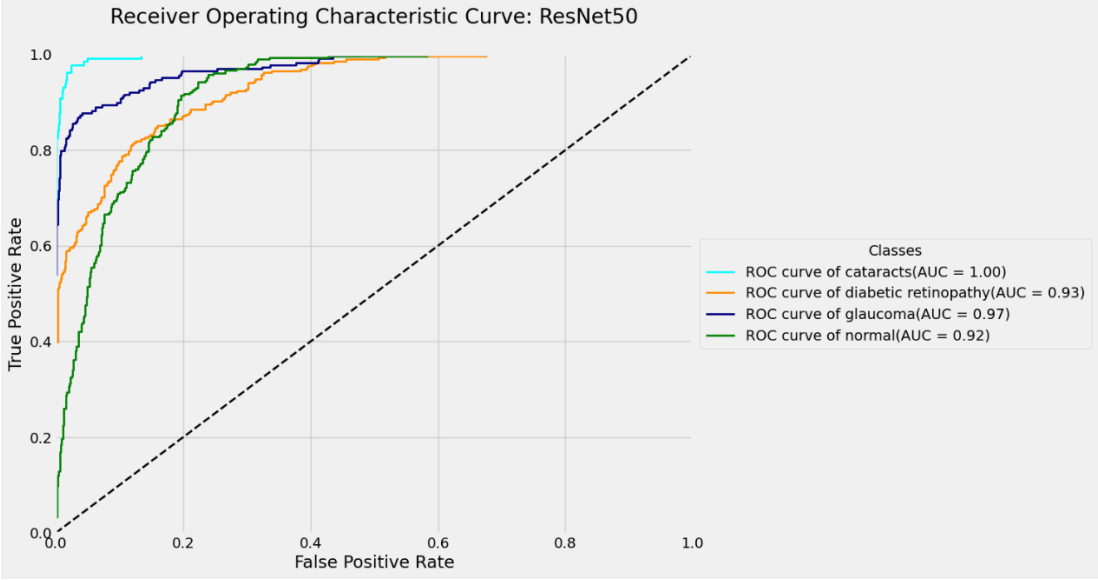


Figure 4.9: ROC curves and corresponding AUC scores for each class of ResNet50

e. Discussion

The ResNet50 model demonstrated strong performance in multiclass eye disease classification. The training and validation curves showed consistent improvement in accuracy and reduction in loss, with minimal overfitting observed. The model achieved an overall accuracy of 80.47%, indicating good generalization across all classes. Class-wise analysis reveals that the model excelled at identifying distinct structural pathologies such as Cataracts and Glaucoma. The Normal class also showed high recall but lower precision, suggesting frequent misclassification of diseased cases as normal. However, similar to other architectures, it exhibited a frequent misclassification on Diabetic Retinopathy as Normal. Such misclassification may be attributed to the subtle nature of early-stage diabetic retinopathy, which make it harder to distinguish visually. In summary, ResNet50 shows excellent discriminative capability, particularly for cataracts and glaucoma, while highlighting the need for improved detection sensitivity to boost sensitivity for subtle, early-stage disease features.

4.3.4 DenseNet121

a. Training and Validation Performance

As illustrated in Figure 4.10, the training and validation curves for DenseNet121 indicate stable and effective learning dynamics. Training accuracy increased steadily, with validation accuracy following a similar path to peak at epoch 18—marked as the best epoch. While minor stochastic fluctuations were observed in validation accuracy beyond this point, the metrics remained closely aligned with the training set, suggesting minimal overfitting. The training loss decreased steadily across epochs, reflecting continuous iterative improvement. Interestingly, validation loss reached its minimum at epoch 30; the divergence

between the best accuracy epoch and the best loss epoch suggests that while classification accuracy reached a plateau early, the model continues to gain confidence in its prediction. Overall, DenseNet121 demonstrates stable learning behaviour with good convergence and generalization capabilities across both training and validation sets.

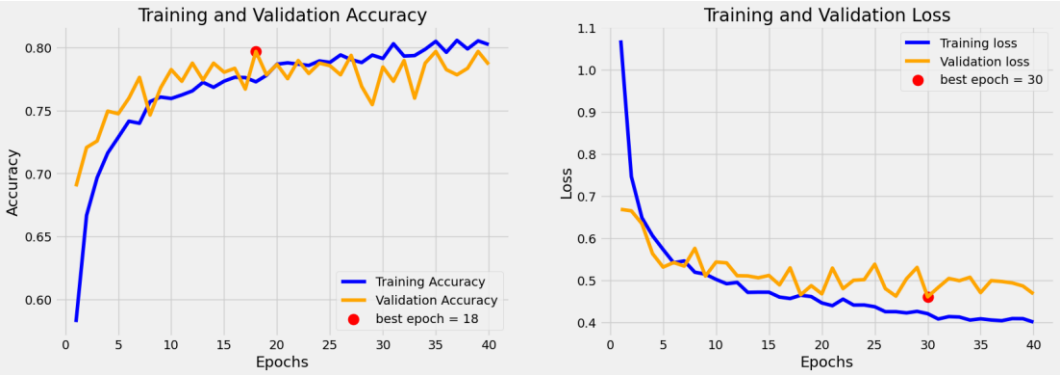


Figure 4.10: Training and validation accuracy and loss plots of DenseNet121

b. Confusion Matrix

Figure 4.11 presents the confusion matrix for DenseNet121, illustrating its classification performance across four classes. The model showed strong performance in identify Cataracts (125 correctly classified) and Glaucoma (201 correctly classified), with minimal misclassifications in both classes. However, it encountered recurring difficulties in distinguishing Diabetic Retinopathy from the Normal class. Although 198 Diabetic Retinopathy instances were correctly classified, 76 cases were misclassified as Normal. Similarly, the Normal class achieved 266 correct predictions but suffered from 39 false-positive predictions for Diabetic Retinopathy and 18 for Glaucoma. These patterns suggest that while DenseNet's feature reuse is highly effective for structural conditions, subtle overlapping features in early-stage diseases remain a point of misclassification.

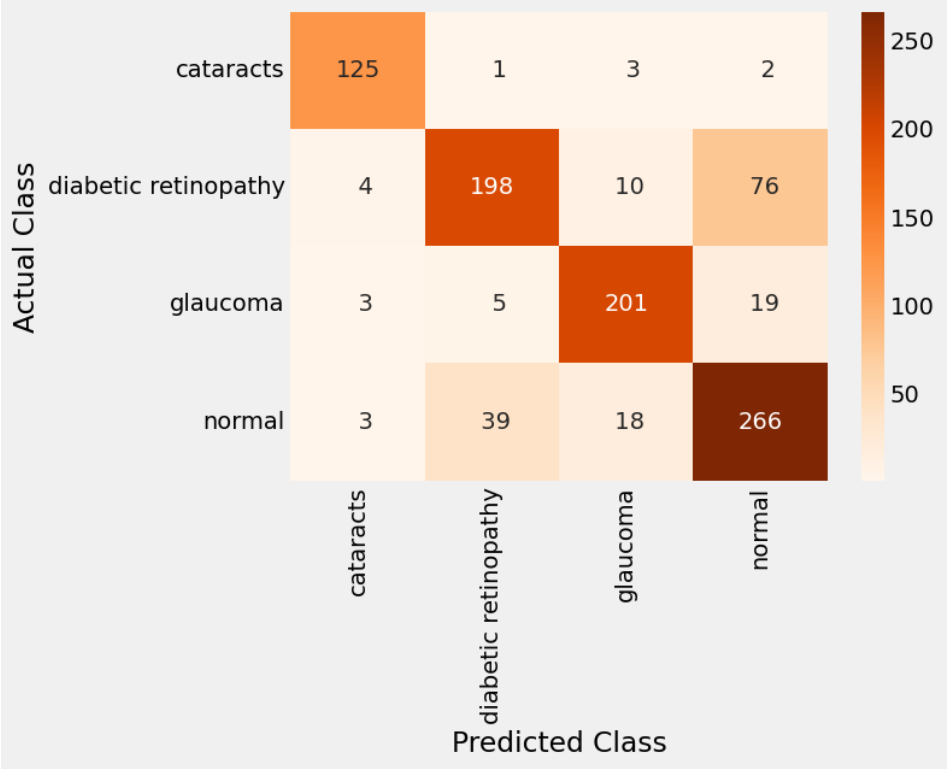


Figure 4.11: Confusion matrix of DenseNet121

c. Classification Report

As shown in Table 4.4, DenseNet121 achieved its highest sensitivity and precision in the Cataracts class, yielding a recall of 0.9542 and a leading F1-score of 0.9398. Glaucoma classification also showed robust results with a recall of 0.8816 and precision of 0.8664, yielding an F1-score of 0.8739. For Diabetic Retinopathy, the model attained a high precision (0.8148) but a lower recall (0.6875), reflecting under-detection—an observation consistent with the confusion matrix. The Normal class yielded a recall of 0.8160 and a precision of 0.7328, indicating a moderate rate of false positives from diseased classes. Overall, DenseNet121 achieved an accuracy of 81.19%, with a macro-average F1-score of 0.8329 and a weighted average F1-score of 0.8108, reflecting well-balanced multiclass performance.

Table 4.4: Classification Metrics of DenseNet121

Class	Recall	Precision	F1-score	Support
Cataracts	0.9542	0.9259	0.9398	131
Diabetic retinopathy	0.6875	0.8148	0.7458	288
Glaucoma	0.8816	0.8664	0.8739	228
Normal	0.8160	0.7328	0.7721	326
Accuracy	0.8119			973
Macro Average	0.8348	0.8350	0.8329	
Weighted Average	0.8119	0.8144	0.8108	

d. AUC-ROC Analysis

Figure 4.12 displays the AUC-ROC curves for DenseNet121, illustrating its discriminative performance across the four classes.

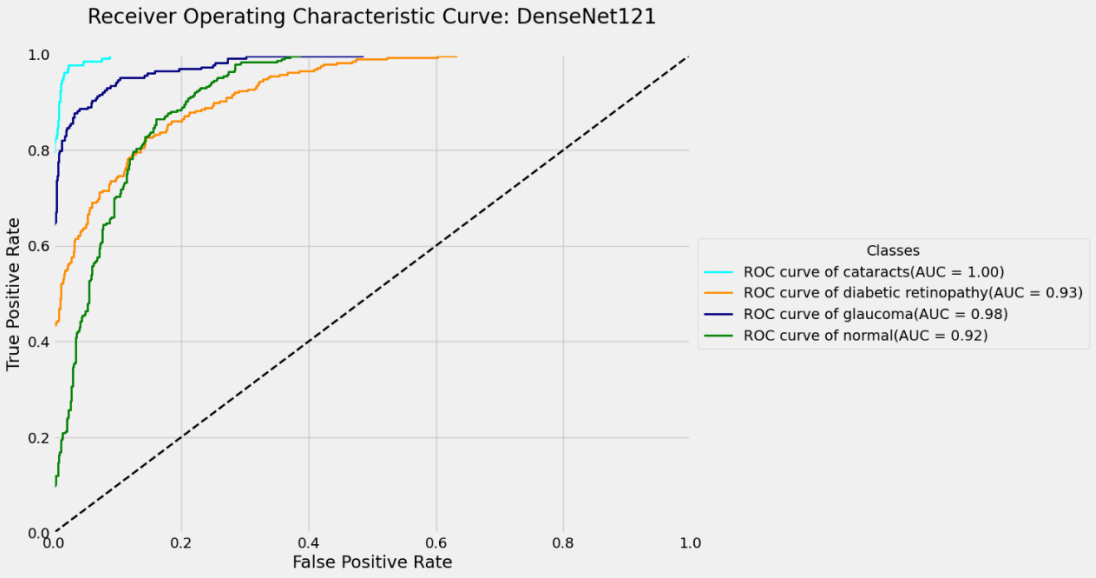


Figure 4.12: ROC curves and corresponding AUC scores for each class of DenseNet121

The model achieved consistently high AUC values, highlighting its strong ability to differentiate between eye conditions. The model achieved a perfect AUC score of 1.00 for

Cataracts and anear-perfect score of 0.98 for Glaucoma, reflecting DenseNet121's high effectiveness in identifying these diseases. Although slightly lower, the AUC scores for Diabetic Retinopathy (0.93) and Normal (0.92) remain strong, highlighting the model's ability to maintain high separability despite the classification challenges noted earlier. The macro-average AUC score of 0.9558 and weighted average AUC score of 0.9463 further solidifies DenseNet121's standing as a robust architecture for multiclass eye disease classification.

e. Discussion

DenseNet121 demonstrated exceptional and consistent performance in the multiclass classification of eye diseases, particularly for conditions with distinct pathological features such as cataracts and glaucoma. The dense connectivity of the architecture likely facilitated the extraction of complex features, resulting in stable learning dynamics and impressive generalization. While Diabetic Retinopathy remains the most challenging class due to its subtle and overlapping features with the Normal class, DenseNet121 maintained a commendable balance between sensitivity and specificity across the dataset. These findings underscore DenseNet121's potential for automated screening applications, though further optimization may be necessary to improve detection of more nuanced conditions.

4.3.5 EfficientNet-B0

a. Training and Validation Performance

Figure 4.13 delineates the training and validation accuracy and loss curves for EfficientNet-B0, illustrating strong convergence and balanced performance across training and validation sets. The accuracy curves showed a consistent upward trend, with the peak validation performance occurring at epoch 22. Concurrently, the training and validation loss

curves declined steadily, reaching their minimum values at the same epoch. The close alignment between the training and validation metrics throughout the training process suggest high generalization capability and minimal susceptibility to overfitting. The simultaneous optimization of validation accuracy and loss at epoch 22 reflects well-tuned hyperparameters and efficient training dynamic characteristic of the architecture's compound scaling method.

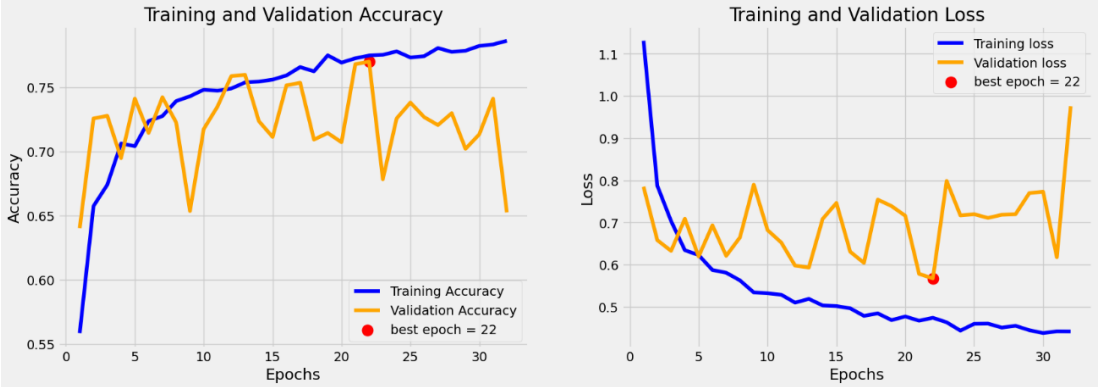


Figure 4.13: Training and validation accuracy and loss plots of EfficientNet-B0

b. Confusion Matrix

Figure 4.14 depicts the confusion matrix for EfficientNet-B0, highlighting its classification performance across four classes. The model showed high precision in identifying Cataracts (123 correct) and Glaucoma (200 correct), with minimal misclassifications. However, Diabetic Retinopathy remains a significant challenge; while 171 instances were correctly predicted, a substantial number (95 instances) were misclassified as Normal. The Normal class also experienced some confusion, with 270 correctly classifications but 33 instances misclassified as Glaucoma. These results reinforce the observation that while EfficientNet-B0 is particularly effective in classifying Cataracts and Glaucoma, it struggles to differentiate Diabetic Retinopathy from Normal class.

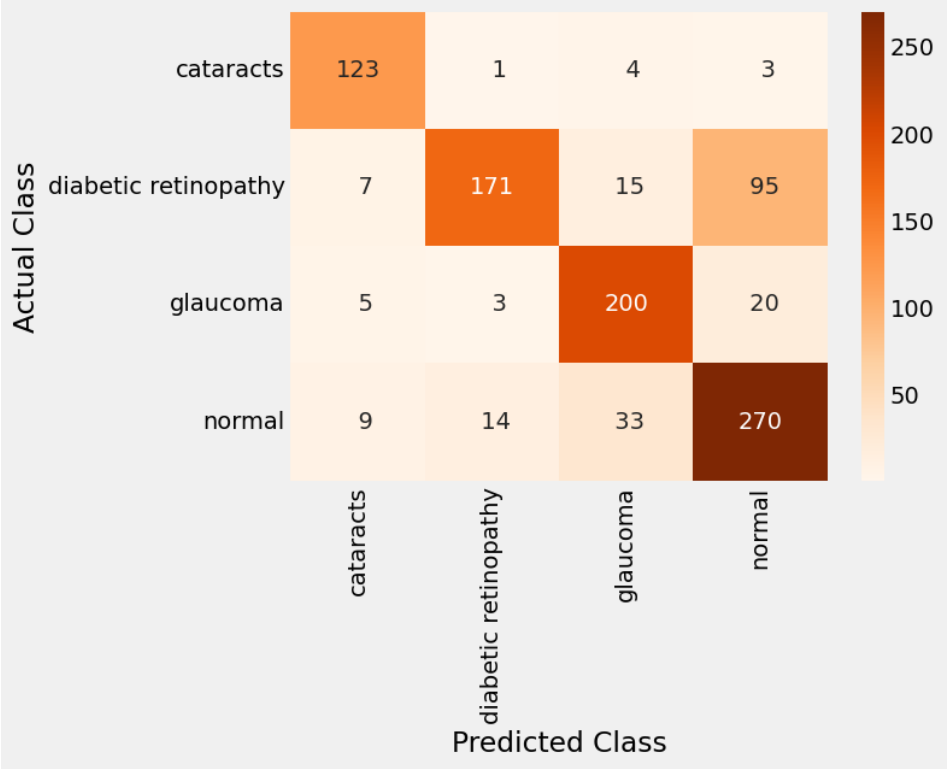


Figure 4.14: Confusion matrix of EfficientNet-B0

c. Classification Report

The classification report outlined in Table 4.5 further quantifies the performance of EfficientNet-B0. The model achieved its highest sensitivity in the Cataracts class, with a recall of 0.9389 and a robust F1-score of 0.8945. Glaucoma classification was similarly effective, yielding an F1-score of 0.8333. Conversely, Diabetic Retinopathy classification was characterized by high precision (0.9048) but notably low recall (0.5938), resulting in an F1-socre of 0.7170. This disparity indicates that while the model's positive predictions for Diabetic Retinopathy are highly reliable, it lacks the sensitivity to detect a significant number of true cases. The Normal class had a recall of 0.8282 and a precision of 0.6958, indicating frequent false positives from diseased classes. EfficientNet-B0 achieved an overall accuracy of 78.52%, with a macro-average F1-score of 0.8003 and a weighted average F1-score of 0.7813, reflecting balanced performance.

Table 4.5: Classification Metrics of EfficientNet-B0

Class	Recall	Precision	F1-score	Support
Cataracts	0.9389	0.8542	0.8945	131
Diabetic retinopathy	0.5938	0.9048	0.7170	288
Glaucoma	0.8772	0.7937	0.8333	228
Normal	0.8282	0.6958	0.7563	326
Accuracy	0.7852			973
Macro Average	0.8095	0.8121	0.8003	
Weighted Average	0.7852	0.8019	0.7813	

d. AUC-ROC Analysis

Figure 4.15 illustrates the AUC-ROC curves of each class for EfficientNet-B0, reinforcing its strong discriminative power.

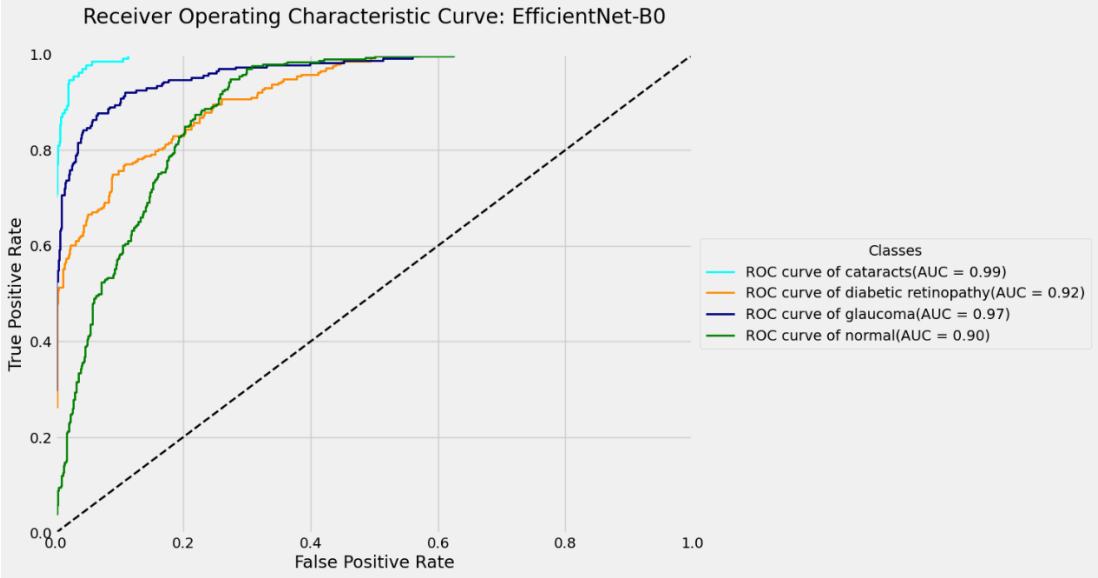


Figure 4.15: ROC curves and corresponding AUC scores for each class of EfficientNet-B0

The model achieved high AUC values for Cataracts (0.99) and Glaucoma (0.97), suggesting excellent class separability for these diseases. The comparatively lower AUC scores for

Diabetic Retinopathy (0.92) and Normal (0.90) mirror the classification challenges previously observed in the confusion matrix and performance metrics, particularly the overlap between these two classes. Nevertheless, the macro-average AUC score of 0.9446 and weighted average AUC score of 0.9333 confirm a robust overall ability to distinguish between the four classes.

e. Discussion

EfficientNet-B0 demonstrated solid performance in multiclass eye disease classification. The training and validation curves showed consistent trends, with both accuracy and loss improving steadily throughout the training process. The model effectively captures the distinct visual features associated with Cataracts and Glaucoma. However, the under-detection of Diabetic Retinopathy, with lower recall but high precision indicating a tendency to miss true cases—likely due to subtle pathological features in early-stage diabetic retinopathy. Overall, EfficientNet-B0 achieved balanced classification performance with good generalization, while highlighting areas for refinement to improve sensitivity for more nuanced diagnostically feature presentations.

4.3.6 Summary and Discussion

Table 4.6 provides a comprehensive comparison of the five baseline CNN models—VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0—evaluated on the original CDGN dataset. Among them, ResNet50, and DenseNet121 emerged as the top-performing models, each achieving accuracies exceeding 80% and a shared peak macro-AUC of 0.9558, reflecting excellent class separability. DenseNet121 offered the most balanced performance, yielding the highest macro-average F1-score of 0.8392, which suggests a superior ability to maintain consistency across different disease classes. VGG16

demonstrated competitive performance as baseline, with macro-average F1-score of 0.8184 and an AUC score of 0.9486. Despite its simpler, linear architecture, it proved capable of capturing the primary pathological features within the dataset. EfficientNet-B0 followed with moderate results, achieving an accuracy of 78.52% and a macro-AUC of 0.9446. Inception-v3 showed the weakest performance among the group, recording the lowest accuracy (72.66%) and macro-average F1-score (0.7444); however, its relatively high macro-AUC (0.9277) indicates that its probabilistic outputs remain reasonably well-calibrated for class separation.

Table 4.6: Performance Metrics of Baseline CNN Models on Original CDGN Dataset

Model	ACC	Macro-Average			Weighted-Average			Macro-AUC
		REC	PRE	F1	REC	PRE	F1	
VGG16	0.7996	0.8151	0.8244	0.8184	0.7996	0.8042	0.8003	0.9486
Inception-v3	0.7266	0.7431	0.7816	0.7444	0.7266	0.7556	0.7191	0.9277
ResNet50	0.8047	0.8113	0.8525	0.8201	0.8047	0.8287	0.8026	0.9558
DenseNet121	0.8119	0.8348	0.8350	0.8329	0.8119	0.8144	0.8108	0.9558
EfficientNet-B0	0.7852	0.8095	0.8121	0.8003	0.7852	0.8019	0.7813	0.9446

Note. ACC = Accuracy, REC = Recall, PRE = Precision, F1 = F1-score.

A consistent trend observed across all models was the high classification accuracy for Cataracts and Glaucoma, indicating that their visual features are highly discriminative and well captured by CNN models. In contrast, the Diabetic Retinopathy and Normal classes exhibited overlap, contributing to higher misclassification rates and reduced sensitivity across the board. Overall, while DenseNet121 and ResNet50 represent the most reliable baselines configurations, the persistent confusion between subtle features and normal retinal

fundus images highlights a clear need for methodological intervention. These findings establish a strong empirical foundation for subsequent stages of this research, specifically the integration of targeted image enhancement, attention mechanisms, and ensemble learning strategies to refine the classification of complex eye diseases.

4.4 Model Performance on Feature-Enhanced CDGN Dataset

This section evaluates the performance of the five selected pre-trained CNN models when trained on the feature-enhanced CDGN dataset. The enhancement pipeline—incorporating CLAHE and morphological operations—was specifically designed to amplify the visibility of the clinically retinal features such as blood vessels, optic disc boundaries, microaneurysms, and exudates. Consistent with the baseline evaluation, the models are assessed using accuracy and loss curves, confusion matrices, classification reports, and AUC scores. This analysis aims to quantify the impact of domain-specific image enhancement on classification robustness and to identify persistent diagnostic challenges.

4.4.1 VGG16

a. Training and Validation Performance

Figure 4.16 illustrates training and validation accuracy and loss curves for VGG16 on the feature-enhanced dataset. Over 40 epochs, both training and validation accuracies steadily improved, with validation accuracy peaked at approximately 81% at epoch 30. The loss curves reveal a consistent decline for both training dan validation sets with minimal overfitting observed. The optimal model state was saved at epoch 30, where validation loss was minimized. These dynamics suggest that feature enhancement process contributed to a more stable learning environment, allowing the model to converge smoothly despite its simpler architectural depth.

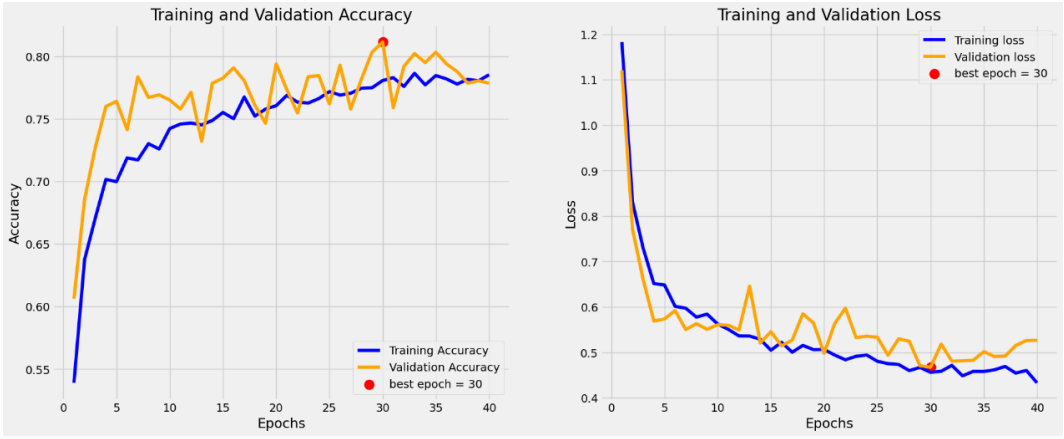


Figure 4.16: Training and validation accuracy and loss plots of VGG16 on feature-enhanced CDGN dataset

b. Confusion Matrix

The confusion matrix in Figure 4.17 details the VGG16's class-specific performance following enhancement.

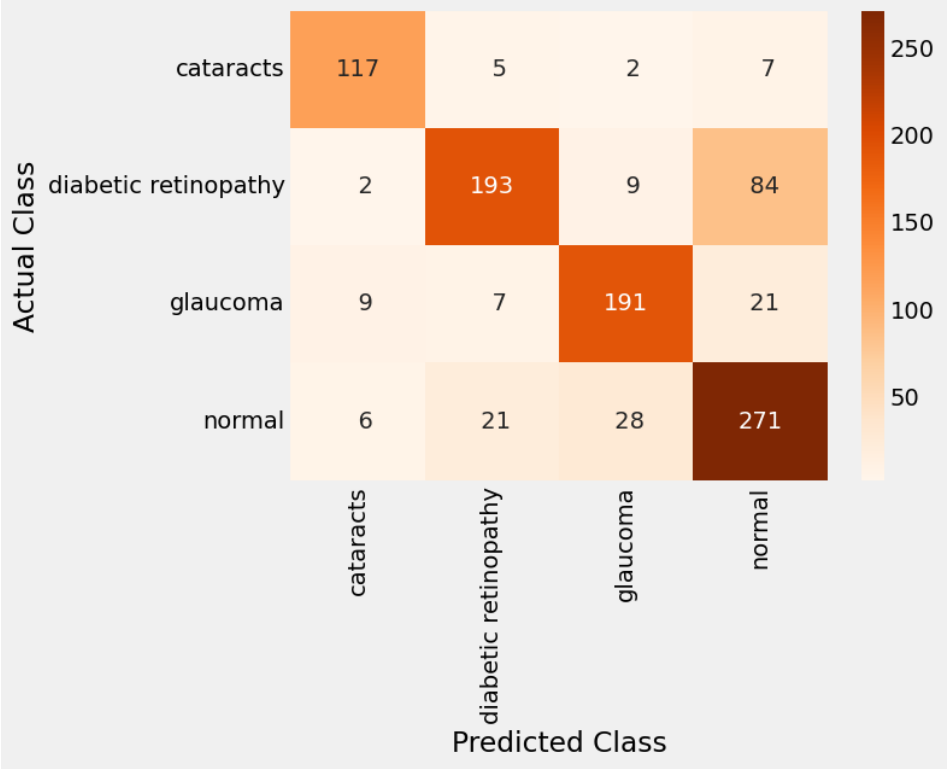


Figure 4.17: Confusion matrix of VGG16

VGG16 demonstrated strong performance in classifying Cataracts (117 correct predictions out of 131) and Glaucoma (193 correct predictions out of 228). However, Diabetic

Retinopathy and Normal classes continued to exhibit higher misclassification rates. Specifically, 84 Diabetic Retinopathy samples were misclassified as Normal, while 21 Normal samples were erroneously predicted as Diabetic Retinopathy. This persistent confusion indicates that while enhancement clarifies vascular structures, the architectural constraints of VGG16 may limit its ability to fully distinguish subtle pathological intersections between healthy and early-stage diseases retinal fundus images.

c. Classification Report

Table 4.7 provides the detailed classification metrics for VGG16 on feature-enhanced CDGN dataset.

Table 4.7: Classification Metrics of VGG16 on Feature-Enhanced CGDN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.8931	0.8731	0.8830	131
Diabetic retinopathy	0.6701	0.8540	0.7510	288
Glaucoma	0.8377	0.8304	0.8341	228
Normal	0.8313	0.7076	0.7645	326
Accuracy	0.7934			973
Macro Average	0.8081	0.8163	0.8081	
Weighted Average	0.7934	0.8020	0.7927	

The model achieved its highest recall (0.8931) and F1-score (0.8830) for Cataracts, reflecting high diagnostic sensitivity. Glaucoma also exhibited reliable performance with an F1-score of 0.8341. In contrast, Diabetic Retinopathy yielded a lower recall of 0.6701, and the Normal class exhibited the lowest precision at 0.7076. These quantitative results reinforce the confusion matrix findings regarding the difficulty of resolving boundary

between Normal and Diabetic Retinopathy. While VGG16 performed consistently across most classes, with an overall accuracy of 79.34%, macro-average F1-score of 0.8081 and weighted average F1-score of 0.7927, there remains room for improvement in Diabetic Retinopathy and Normal classification.

d. AUC-ROC Analysis

The ROC analysis in Figure 4.18 evaluates the model's discriminative ability on the enhanced images. VGG16 achieved high AUC scores across all classes: Cataracts (0.99), Diabetic Retinopathy (0.93), Glaucoma (0.97), and Normal (0.90). The macro-average AUC score of 0.9479 and weighted-average AUC score of 0.9380 confirm that the model maintains a high degree of confidence in its probabilistic classifications. These results underscore that the model is generally effective at separating classes, though the absolute thresholds for binary decision-making (as seen in recall) remain hampered by class overlap.

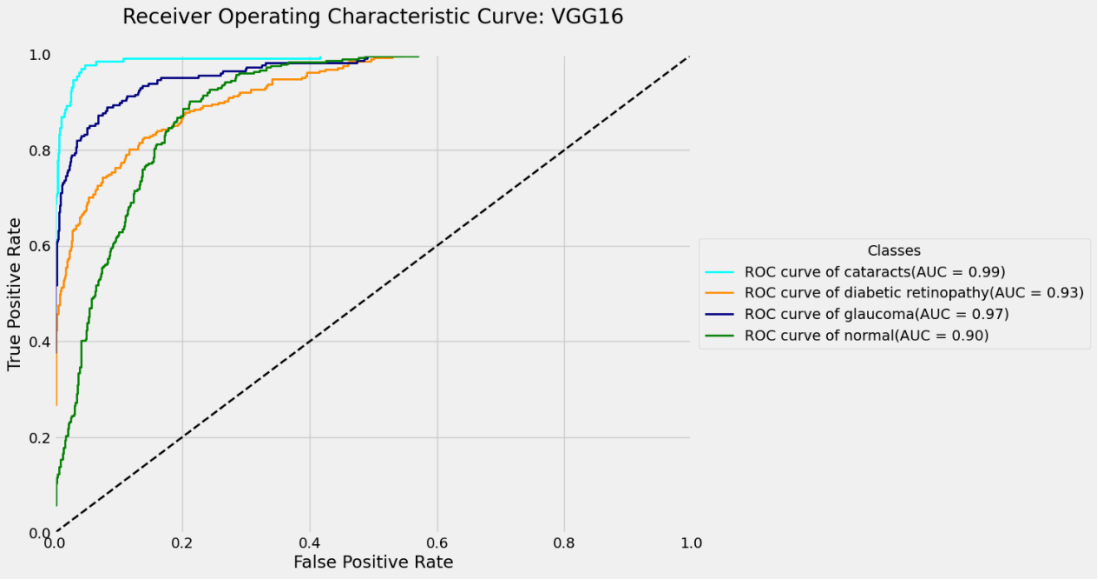


Figure 4.18: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of VGG16

e. Discussion

When compared to baseline results on original dataset, VGG16 exhibited marginal performance fluctuations, with some metrics showing a slight decline. While the overall accuracy remained stable, the gains in recall, precision, and F1-score were limited. This observation suggests that the relatively shallow architecture of VGG16 may lack the hierarchical complexity required to fully exploit the enriched features provided by the enhancement pipeline. Although the enhancement led to smother convergence and stable learning dynamics, the persistent confusion between Diabetic Retinopathy and Normal cases suggests that more advanced, deeper architectures are likely necessary to leverage the full diagnostic potential of the enhanced retinal fundus images.

4.4.2 Inception-v3

a. Training and Validation Performance

Figure 4.19 illustrates the training and validation accuracy and loss curves for Inception-v3 trained on the feature-enhanced CDGN dataset.

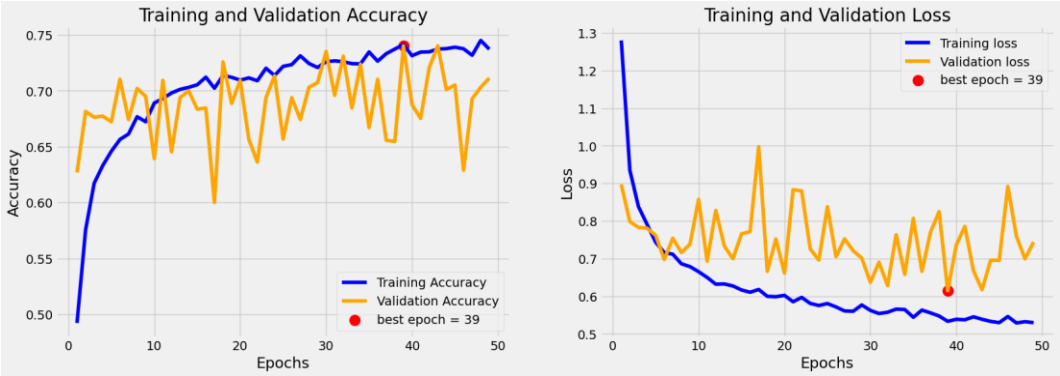


Figure 4.19: Training and validation accuracy and loss plots of Inception-v3 on feature-enhanced CGDN dataset

Both accuracy and loss metrics exhibited a general trend of improvement throughout the training process. Training accuracy stabilized at approximately 0.73, while validation

accuracy showed more pronounced stochastic fluctuations, eventually peaking at approximately 0.74 at epoch 39. Correspondingly, both loss curves followed a declining trend, with the validation loss reaching its minimum at epoch 39. However, the noticeable volatility in the validation curves suggests a degree of optimization instability. This may be attributed to Inception-v3's multi-scale architecture being highly sensitivity to the enhanced features introduced in the dataset.

b. Confusion Matrix

The confusion matrix in Figure 4.20 provides a detailed view of the class-specific performance following enhancement.

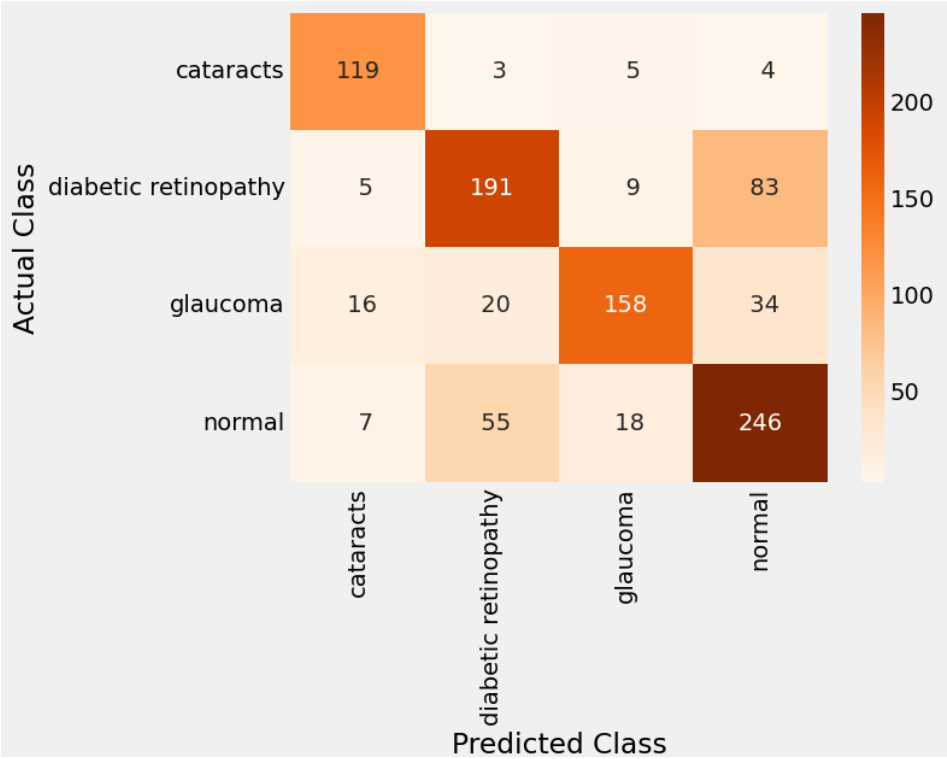


Figure 4.20: Confusion matrix of Inception-v3

The Cataracts class was predicted with high fidelity, achieving 119 true positives out of 131 samples. Diabetic Retinopathy demonstrated moderate classification accuracy with 191 correct predictions, although 83 instances remained misclassified as Normal. Glaucoma

yielded 158 correct predictions, with 34 samples misclassified as Normal. Notably, the Normal class experienced the highest rate of misclassification, with 55 samples incorrectly labelled as Diabetic Retinopathy and 18 as Glaucoma. These misclassification patterns suggest that while the enhancement techniques amplify disease-specific features, they also introduce artifacts or emphasized non-diseased features that the Inception-v3 model struggled to distinguish from true diseased states.

c. Classification Report

Table 4.8 summarizes the classification performance of Inception-v3. The model reached an overall accuracy of 73.38%, with macro-average F1-score of 0.7520 respectively. Cataracts remained the most reliably detected class, achieving a recall of 0.9084 and an F1-score of 0.8561. While Glaucoma showed relatively strong performance (F1-score = 0.7560), the Diabetic Retinopathy and Normal classes exhibited lower F1-scores of 0.6858 and 0.7100, respectively.

Table 4.8: Classification Metrics of Inception-v3 on Feature-Enhanced CDGN Dataset.

Class	Recall	Precision	F1-score	Support
Cataracts	0.9084	0.8095	0.8561	131
Diabetic retinopathy	0.6632	0.7100	0.6858	288
Glaucoma	0.6930	0.8316	0.7560	228
Normal	0.6703	0.6703	0.7100	326
Accuracy	0.7338			973
Macro Average	0.7548	0.7554	0.7520	
Weighted Average	0.7338	0.7386	0.7333	

These results reinforce the confusion matrix observations, indicating that despite the image enhancement, resolving boundary between diabetic retinopathy and healthy retina remains a significant challenge for this architecture.

d. AUC-ROC Analysis

Figure 4.21 displays the AUC-ROC curves for each class in the feature-enhanced dataset. The highest AUC score was achieved by Cataracts (0.99), followed by Glaucoma (0.94), indicating strong discriminative ability of these disease conditions. In contrast, the Diabetic Retinopathy and Normal classes yielded lower AUC scores of 0.89 and 0.88, respectively. Notably, the macro-average AUC (0.9244) and the weighted average AUC (0.9114) are slightly lower than the baseline values. This suggests a subtle decline in the model's probabilistic classification confidence, potentially due to the increase complexity of the feature space following the enhancement.

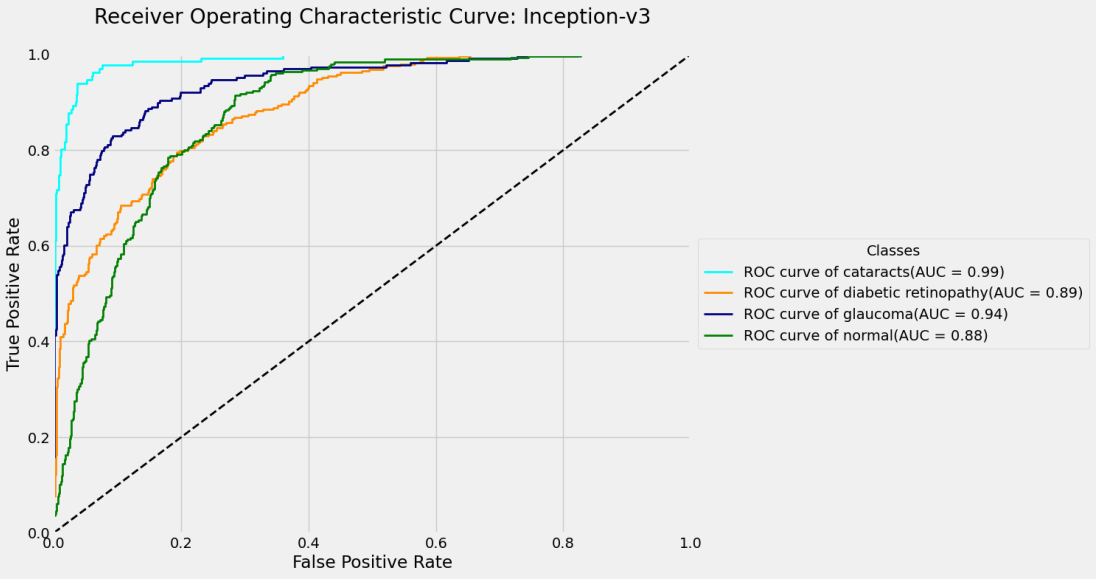


Figure 4.21: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of Inception-v3

e. Discussion

The application of enhancement techniques provided measurable gains for Inception-v3 across most primary metrics, particularly in the classification of Cataracts and Glaucoma. While Diabetic Retinopathy and Normal cases remain diagnostically difficult, their specific classification metrics showed improvement over the baseline. However, the marginal decline in AUC score suggests that while the enhancement helped the model make better discrete predictions (accuracy/F1-score), it slightly reduced its overall probabilistic confidence. This indicates that while deeper models benefit from the enhanced clinical features, careful hyperparameter tuning is essential to preserve discriminative confidence when dealing with the subtle markers of early-stage eye diseases.

4.4.3 ResNet50

a. Training and Validation Performance

Figure 4.22 illustrates the training and validation performance of ResNet50, demonstrating how the architecture adapted to the feature-enhanced CGDN dataset.

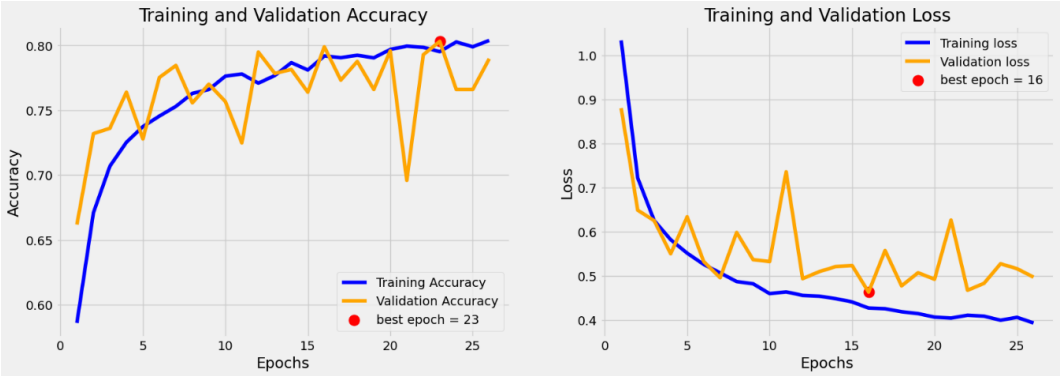


Figure 4.22: Training and validation accuracy and loss plots of ResNet50 on feature-enhanced CDGN dataset

The training accuracy showed a consistent upward trend, reaching approximately 0.80 by epochs 25. Validation accuracy closely tracked this progression, peaking at epoch 23, which

represents the model's optimal state for generalization. While the training loss decreased steadily, the validation loss showed marginal stochastic fluctuations before stabilizing, with the absolute minimum occurring at epoch 16. These trends suggest that ResNet50 effectively leveraged the amplified clinical features in the enhanced dataset, maintaining stable convergence with no significant signs of overfitting.

b. Confusion Matrix

The confusion matrix Figure 4.23 provides granular insights into the class-wise performance of ResNet50 post-enhancement.

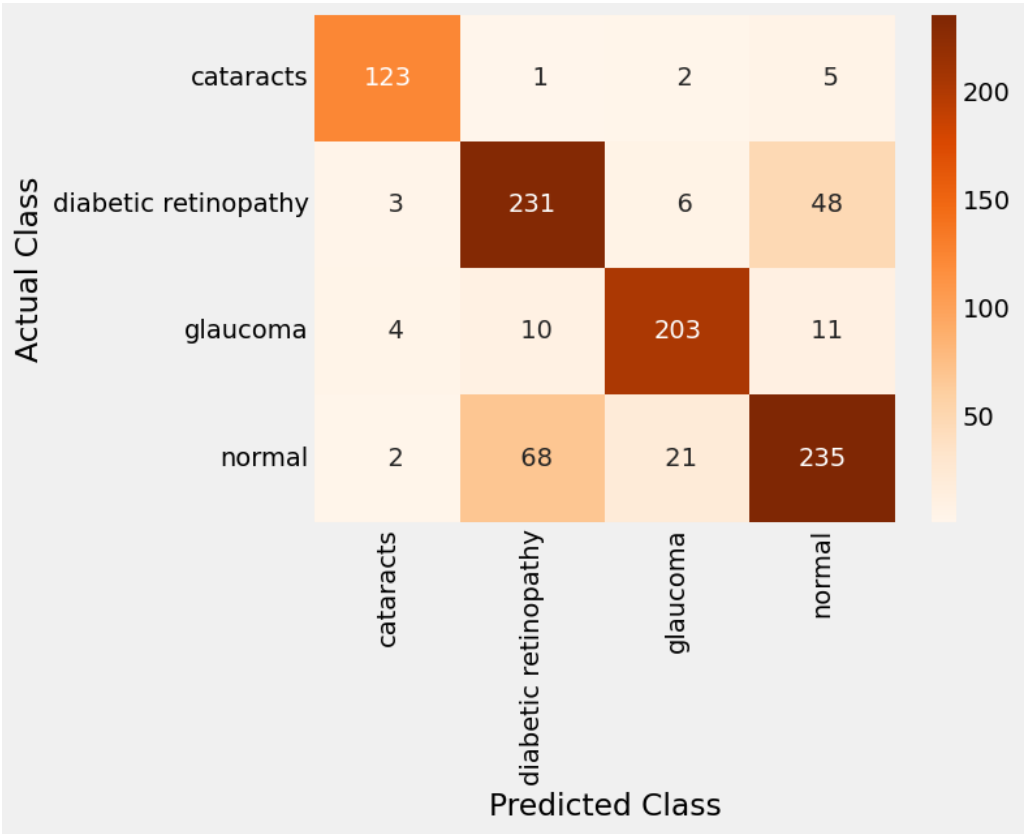


Figure 4.23: Confusion matrix of ResNet50

The model achieved high diagnostic precision for Cataracts, correctly identifying 123 out of 131 instances. Glaucoma also showed robust results with 203 correct predictions; however minor leakage was observed into the Diabetic Retinopathy (n = 10) and Normal (n = 11)

classes. Diabetic Retinopathy saw a marked improvement with 231 correct identifications, though 48 instances remained misclassified as Normal. Conversely, the Normal class recorded 235 correct predictions, with notable portion of false positives categorised as Diabetic Retinopathy ($n = 68$) and Glaucoma ($n = 21$). This suggests that while enhancement significantly aids in identifying clinical features, the visual boundary between Normal and Diabetic Retinopathy remains the primary source of classification error.

c. Classification Report

The classification performance of ResNet50 is summarized in Table 4.9.

Table 4.9: Classification Metrics of ResNet50 on Feature-Enhanced CDGN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.9389	0.9318	0.9354	131
Diabetic retinopathy	0.8021	0.7452	0.7726	288
Glaucoma	0.8904	0.8750	0.8826	228
Normal	0.7209	0.7860	0.7520	326
Accuracy	0.8140			973
Macro Average	0.8381	0.8345	0.8356	
Weighted Average	0.8140	0.8144	0.8134	

Class-wise, Cataracts were detected with high reliability, achieving the highest recall (0.9389) and a robust F1-score (0.9354). Glaucoma performance remained strong with an F1-score of 0.8826. For the Diabetic Retinopathy and Normal classes, the model achieved balanced F1-scores of 0.7726 and 0.7520, respectively. ResNet50 achieved an overall accuracy of 81.40% supported by a macro-average F1-score of 0.8356. These metrics reflect

a solid, well-rounded performance, underscoring the model's improved ability to resolve complex disease features following the image enhancement process.

d. AUC-ROC Analysis

The AUC-ROC curves for each class are depicted in Figure 4.24. ResNet50 achieved excellent class separability for Cataracts (AUC = 1.00) and Glaucoma (AUC = 0.98). The Diabetic Retinopathy and Normal classes also performed well, with AUC scores of 0.94 and 0.92, respectively. A macro-average AUC score of 0.9581 and weighted average AUC score of 0.9549 further validates the model's robustness and its high probabilistic confidence in distinguishing between the four target classes on the enhanced data.

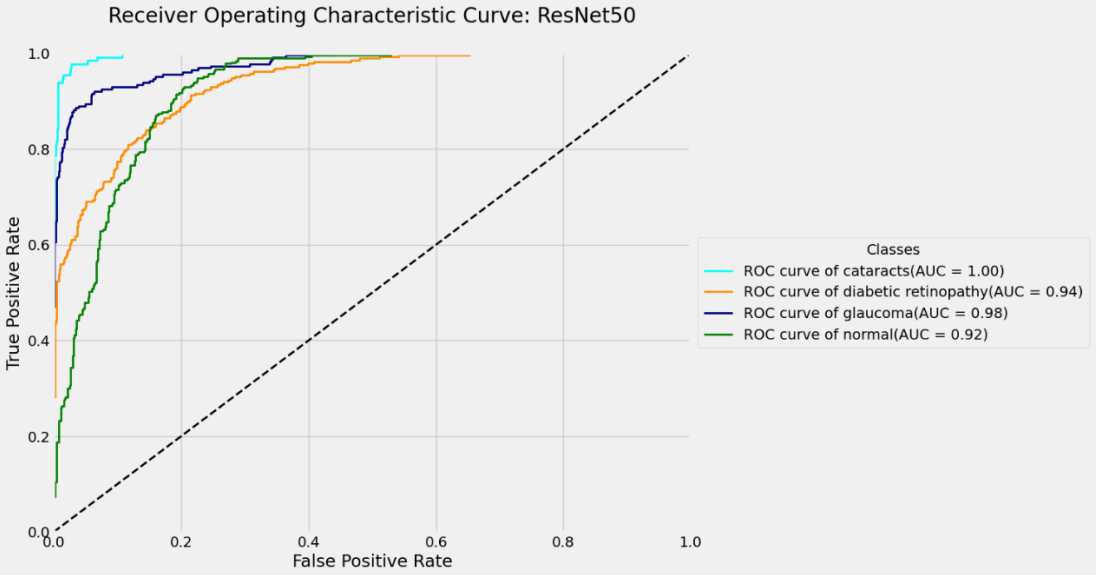


Figure 4.24: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of ResNet50

e. Discussion

Overall, ResNet50 demonstrated a substantial performance elevation compared to its baseline results on the original CDGN dataset. The model exhibited a well-balanced learning process, as reflected in the training and validation curves. The significant gains in F1-

scores—particularly for the more challenging classes—validate the effectiveness of combining domain-specific enhancement with deep residual learning. While the persistent overlap between Diabetic Retinopathy and Normal cases remains a diagnostic hurdle, the overall improvements in accuracy and AUC confirm that ResNet50 is a highly reliable architecture for automated eye disease screening when provided with optimized input data.

4.4.4 DenseNet121

a. Training and Validation Performance

Figure 4.25 illustrates the training and validation accuracy and loss curves for DenseNet121 on the feature-enhanced CDGN dataset. Both accuracy curves showed a consistent upward trend throughout the training duration, with validation accuracy reaching close to 0.80 at epoch 42. The absence of a widening gap between training and validation metrics indicates minimal overfitting. The loss curves exhibited a parallel decreasing pattern, with the validation loss reaching its minimum of approximately 0.43 at epoch 34. This relatively narrow generalization gap suggests that the model's dense connectivity effectively utilized the enhanced features that was able to generalize effectively to unseen data.

Figure 4.25: Training and validation accuracy and loss plots of DenseNet121 on feature-enhanced CDGN dataset

Table 4.15: Classification Metrics of Attention-Based ResNet50 on Feature-Enhanced CDGN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.9389	0.9462	0.9425	131
Diabetic retinopathy	0.6319	0.8708	0.7324	288
Glaucoma	0.8816	0.8933	0.8874	228
Normal	0.8834	0.7042	0.7837	326
Accuracy	0.8160			973
Macro Average	0.8340	0.8536	0.8365	
Weighted Average	0.8160	0.8304	0.8142	

d. AUC-ROC Analysis

The AUC-ROC curves illustrated in Figure 4.39 provide insight into the attention-based ResNet50 model's ability to distinguish between classes.

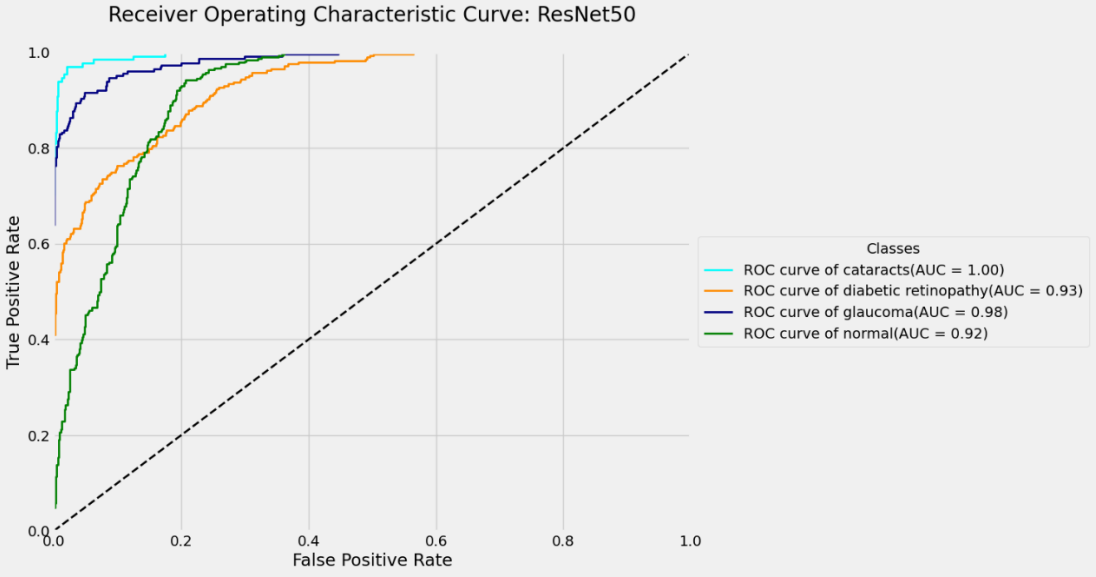


Figure 4.39: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based ResNet50

Cataracts achieved a perfect AUC score of 1.00, indicating ideal class separability, followed closely by Glaucoma at 0.98. Diabetic Retinopathy and Normal classes yielded AUC values of 0.93 and 0.92, respectively, suggesting some overlap in feature representation even after image enhancement. The macro-average AUC score of 0.9564, and the weighted-average AUC score of 0.9468 confirm that the attention-based ResNet50 maintains a high level of probabilistic confidence, effective utilizing spatial salient features to distinguish between the four target classes.

e. Discussion

Overall, the attention-based ResNet50 model demonstrated a high degree of efficacy in multiclass eye disease classification, particularly for Cataracts and Glaucoma. However, the main limitation lies in distinguishing Diabetic Retinopathy from the Normal class. This could stem from subtle visual similarities between these classes, although with image enhancement and attention mechanism integration. Compared to the non-attention baseline, the attention-based variant showed improvements across several key metrics, most notably in class-wised F1-scores. These improvement demonstrate that the spatial attention enables the model to make more informed predictions by emphasizing relevant regions within the enhanced retinal fundus images. The well-managed training curves and high AUC scores further validate ResNet50 with spatial attention as a robust architecture for automated diagnostic assistance.

4.5.4 Attention-Based DenseNet121

a. Training and Validation Performance

Figure 4.40 presents the training and validation curves for accuracy and loss for the attention-based DenseNet121 model. Both accuracy curves exhibited a consistent and

synchronised upward trend, with the validation metrics closely tracking the training curve before peaking at epoch 15. The narrow gap between training and validation accuracy indicates a highly stable generalization capability and minimal overfitting. Correspondingly, the loss curves showed a steady decline, with the validation loss reaching its minimum at epoch 15. While minor stochastic fluctuations are observed in the validation loss during later stages, they remained within a controlled range and do not diverge significantly from the training loss. These patterns indicate that the model maintains a robust learning process and generalizes well on the enhanced dataset.

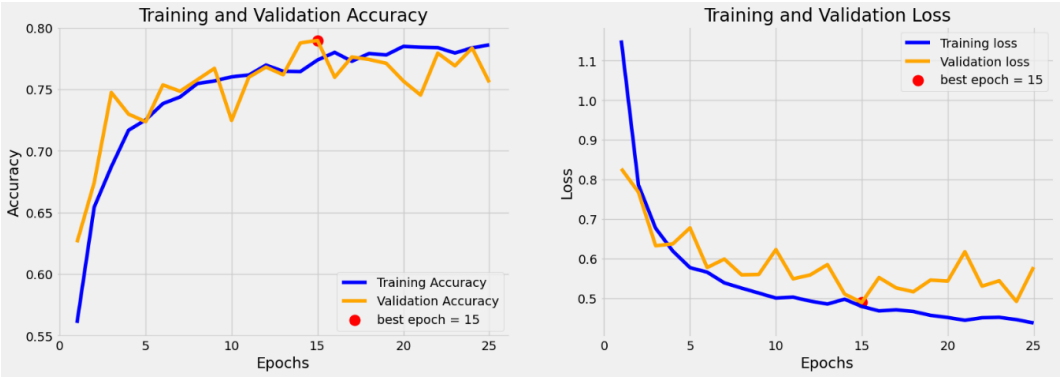


Figure 4.40: Training and validation accuracy and loss plots of attention-based DenseNet121 on feature-enhanced CDGN dataset

b. Confusion Matrix

The confusion matrix in Figure 4.41 reveals a sophisticated classification profile across the four diagnostic classes of attention-based DenseNet121 model. The model demonstrated high fidelity in predicting Normal (286 correct predictions) and Glaucoma (195 correct predictions). Cataracts were also exceptionally well-distinguished, with 123 out of 131 instances correctly identified and minimal misclassification into other classes. However, Diabetic Retinopathy continues to present a significant classification hurdle, with 89 cases misclassified as Normal. This persistent bidirectional confusion highlights the

challenge of isolating micro-vascular pathological markers from healthy retinal background, even with the localized focus provided by the spatial attention mechanism.

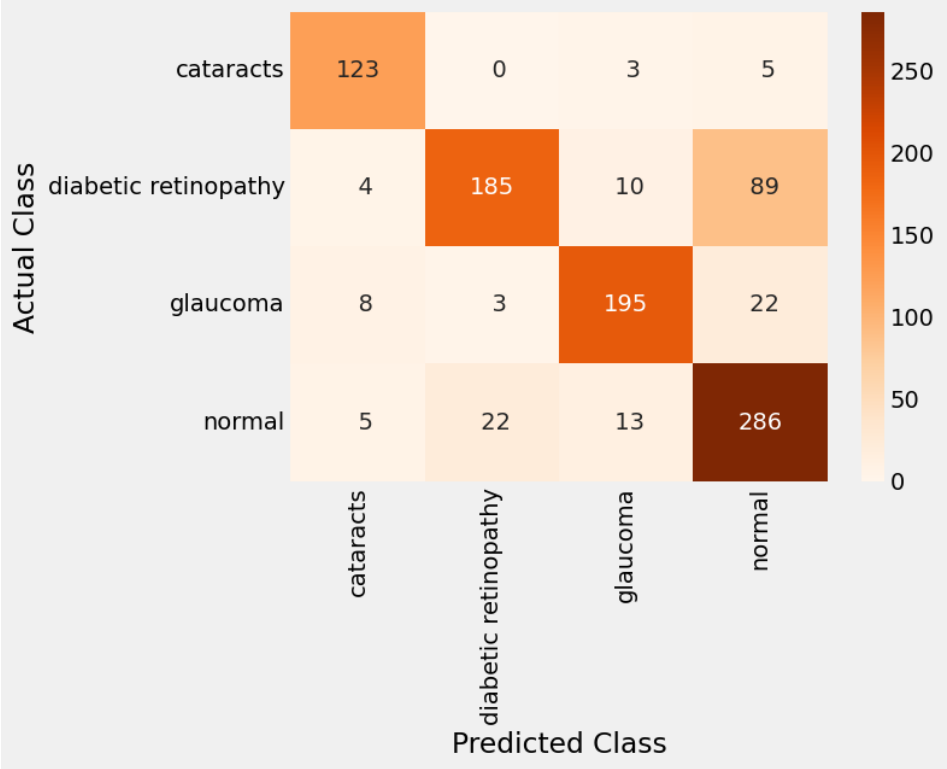


Figure 4.41: Confusion matrix of attention-based DenseNet121

c. Classification Report

Table 4.16 presents the classification metrics that quantifies the performance of the attention-based DenseNet121 model on feature-enhanced CDGN dataset. The model achieved its highest F1-scores for Cataracts (0.9077), supported by a high recall of 0.9389 and precision of 0.8786, indicating excellent sensitivity. Glaucoma also performed robustly with an F1-score of 0.8686. In contrast, Diabetic Retinopathy exhibited low recall (0.6424) paired with high precision (0.8810); this conservative behaviour implies that while the model is highly accurate when it flags DR, it misses a notable portion of subtle cases. The Normal class attained a strong recall (0.8773) but recorded the lowest precision (0.7114), as it absorbed a notable share of false positives from other classes, particularly Diabetic

Retinopathy. Attention-based DenseNet121 achieved an overall accuracy of 81.09%, with a macro-average F1-score of 0.8285 and a weighted average F1-score of 0.8089, reflecting balanced performance across all classes, albeit with variability in sensitivity and precision.

Table 4.16: Classification Metrics of Attention-Based DenseNet121 on Feature-Enhanced CDGN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.9389	0.8786	0.9077	131
Diabetic retinopathy	0.6424	0.8810	0.7430	288
Glaucoma	0.8553	0.8824	0.8686	228
Normal	0.8773	0.7114	0.7857	326
Accuracy	0.8109			973
Macro Average	0.8285	0.8383	0.8263	
Weighted Average	0.8109	0.8242	0.8089	

d. AUC-ROC Analysis

The attention-based DenseNet121 model's strong discriminative power is further affirmed by the ROC curves illustrated in Figure 4.42. The macro-average AUC score of 0.9560 and the weighted average AUC score of 0.9474 reflect robust overall diagnostic reliability. Class-wise AUC values were consistently high: 0.99 for Cataracts, 0.98 for Glaucoma, 0.93 for Diabetic Retinopathy, and 0.92 for Normal. These scores indicate near-perfect separability for Cataracts and Glaucoma, and strong discriminative potential for Diabetic Retinopathy, even where discrete classification boundaries (recall) remain difficult to establish.

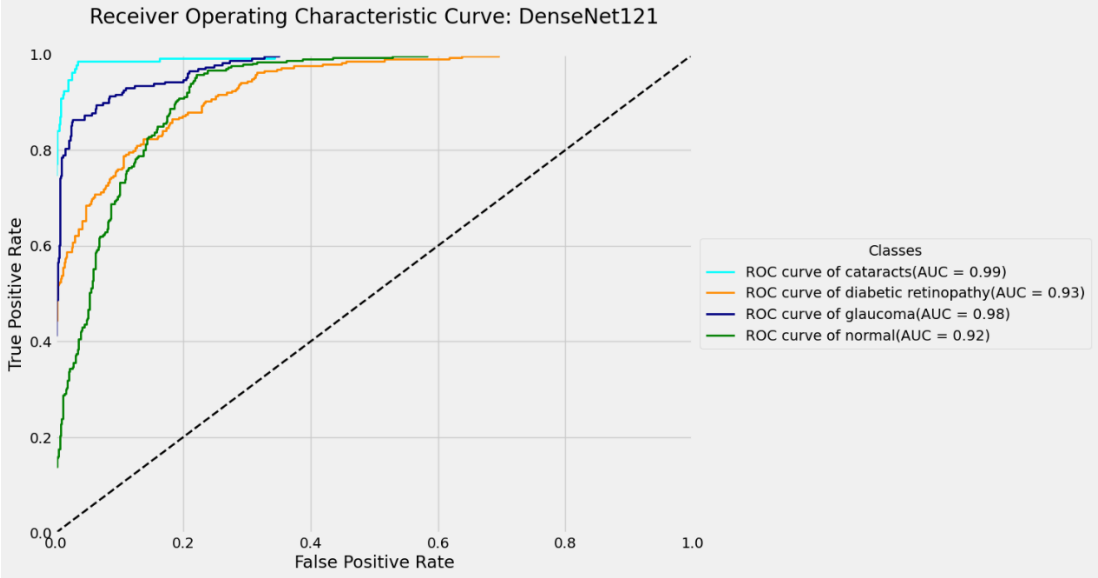


Figure 4.42: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based DenseNet121

e. Discussion

Overall, the attention-based DenseNet121 model exhibited well-rounded performance, characterised by high precision and exceptionally stable convergence. Its ability to maintain a tight convergence between training and validation performance reflects a stable and effective training process, effectively mitigating overfitting. While the model excels in precision across all classes, the primary limitation remains the sensitivity for Diabetic Retinopathy. This suggests that while attention mechanisms enhanced feature localization, the dense nature of the feature maps might require even more granular attention to fully capture the subtle disease features. Notably, while this variant performed slightly below the non-attention variant in absolute accuracy, its superior precision and AUC scores suggest it is a more reliable classifier for clinical environments where false positives must be minimized. These results indicate that that while spatial attention contributes positively to class separability and precision, its impact on overall accuracy may vary depending on dataset characteristics.

4.5.5 Attention-Based EfficientNet-B0

a. Training and Validation Performance

Figure 4.43 presents the training and validation accuracy and loss curves for the attention-based EfficientNet-B0 model. The training accuracy showed a steady upward trend, reaching approximately 0.81. Validation accuracy, however, showed more marked stochastic fluctuations, peaking at 0.79 at epoch 32, which was marked as the best-performing state. While the training loss consistently decreased, indicating effective optimization, the validation loss exhibited several oscillations. These fluctuations suggest a moderate degree of overfitting, where the model began to specialize in the training data's enhanced noise patterns. Despite this volatility, the overall alignment between the curves remains reasonable, indicating that the model benefits from the attention mechanism, even if the architecture requires more aggressive regularization to achieve complete stability.

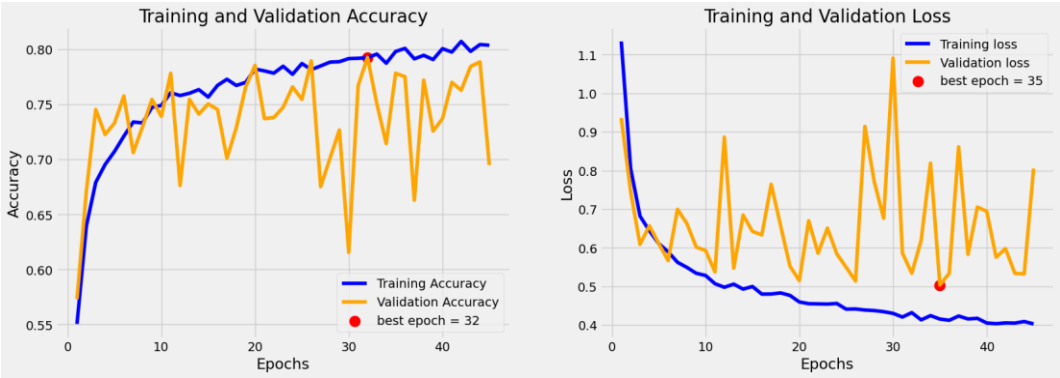


Figure 4.43: Training and validation accuracy and loss plots of attention-based EfficientNet-B0 on feature-enhanced CDGN dataset

b. Confusion Matrix

Figure 4.44 present the confusion matrix for the attention-based EfficientNet-B0 model, illustrating its class-wise prediction performance. The model demonstrated strong classification capability for Cataracts and Glaucoma, with 124 and 208 correct predictions,

respectively. Diabetic Retinopathy also showed reasonable performance with 196 correctly classified instances; however, 75 cases were misclassified as Normal, representing the primary source of diagnostic error. Conversely, Normal images were frequently misclassified as Diabetic Retinopathy (n = 33) or Glaucoma (n = 32). Glaucoma showed minimal misclassifications, with only a few instances incorrectly predicted as Cataracts (n = 4), Diabetic Retinopathy (n = 5), or Normal (n = 11). These misclassifications highlight while the attention mechanism improves focus on structural conditions, the subtle or overlapping visual features between Normal and Diabetic Retinopathy remains a challenging hurdle for this lightweight architecture.

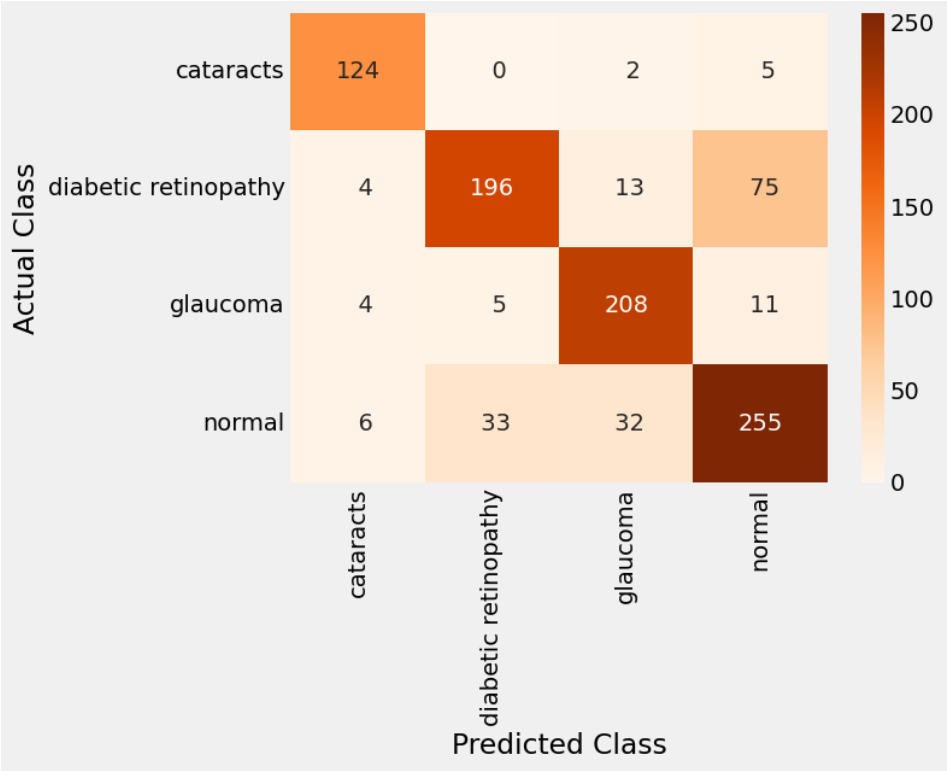


Figure 4.44: Confusion matrix of attention-based EfficientNet-B0

c. Classification Report

Table 4.17 presents the classification metrics of the attention-based EfficientNet-B0 model, revealing variable performance across the classes. Cataracts achieved the highest

recall (0.9466) and a robust F1-score of 0.9219, indicating excellent sensitivity. Glaucoma followed with a relatively good recall (0.9123) and precision (0.8157). Diabetic Retinopathy showed moderate performance with an F1-score of 0.7510, hindered by a lower recall (0.6806) despite a respectable precision (0.8376). The Normal class exhibited the lowest precision (0.7370), suggesting that a considerable number of Normal cases were misclassified as diseased. The attention-based EfficientNet-B0 reached an overall accuracy of 80.47%, with a macro-average F1-score of 0.8233 and weighted average F1-score of 0.8025. These results suggest that the model may be more effective at learning features associated with diseased classes than non-diseased ones.

Table 4.17: Classification Metrics of Attention-Based EfficientNet-B0 on Feature-Enhanced CDGN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.9466	0.8986	0.9219	131
Diabetic retinopathy	0.6806	0.8376	0.7510	288
Glaucoma	0.9123	0.8157	0.8613	228
Normal	0.7822	0.7370	0.7589	326
Accuracy	0.8047			973
Macro Average	0.8304	0.8222	0.8233	
Weighted Average	0.8047	0.8070	0.8025	

d. AUC-ROC Analysis

The ROC curves for the attention-based EfficientNet-B0 model, as depicted in Figure 4.45, further validate its class-wise prediction capability. The model achieved a near-perfect AUC score of 0.99 for Cataracts, followed by Glaucoma (0.96), Diabetic Retinopathy (0.93), and Normal (0.91). These high AUC scores indicate strong separability across the board.

The steep rise in the ROC curves, especially for Cataracts and Glaucoma, confirms that the attention mechanism effectively isolates these conditions with minimal false positive rates. The macro-average AUC score of 0.9508 and weighted average AUC score of 0.9409 further solidates EfficientNet-B0's robust overall discriminative ability. These results suggest that the attention mechanism in EfficientNet-B0 effectively enhances feature discrimination, especially for visually distinct disease classes.

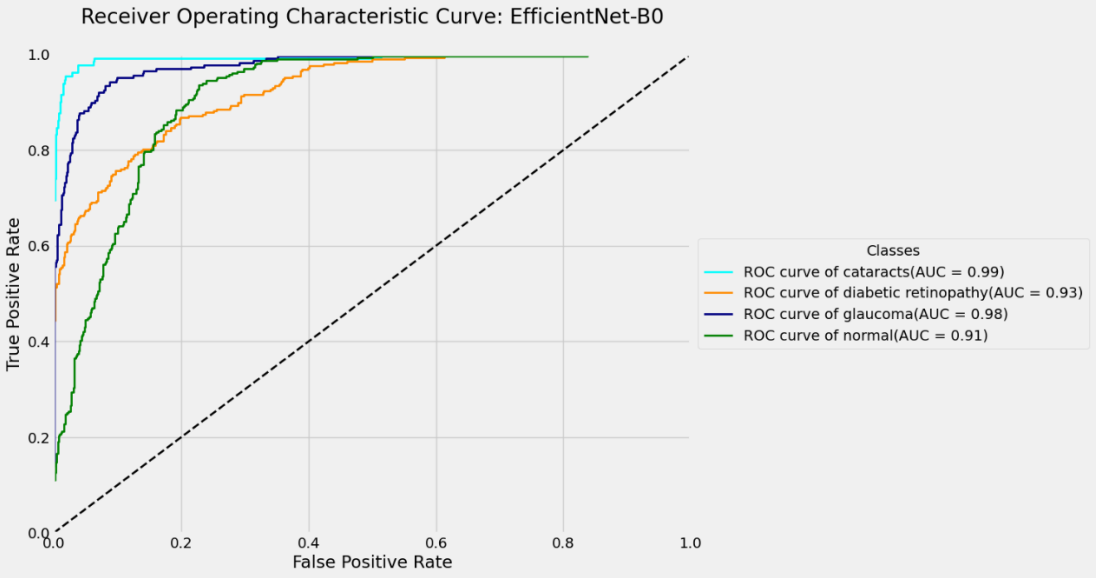


Figure 4.45: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of attention-based EfficientNet-B0

e. Discussion

The attention-based EfficientNet-B0 model demonstrated commendable classification performance in multiclass eye disease classification, showing that even a parameter-efficient model can achieve high diagnostic accuracy when augmented with spatial focus. The model's exceptional performance in detecting Cataracts and Glaucoma suggests that the spatial attention gates are well-aligned with the structural manifestations of these diseases. While the model struggled with the finer visual similarities of early-stage Diabetic Retinopathy, it represents a meaningful improvement over its non-attention and

baseline counterparts. The integration of spatial attention, combined with enhanced image features, appears to prioritize disease-specific features, which is highly beneficial for clinical screening. When compared across all configurations, the gains in F1-scores and AUC values affirm that combining feature-enhanced images with attention-weighted learning is a superior strategy for automating multiclass eye disease diagnosis.

4.5.6 Summary and Discussion

This subsection presents a comprehensive summary and comparative analysis of the five attention-based CNN models evaluated on the feature-enhanced CDGN dataset. The performance metrics, including accuracy, macro-average and weighted average recall, precision, F1-score, and macro-average AUC, are detailed in Table 4.18. This analysis examines the efficacy of integrating spatial attention into various CNN backbones for multiclass eye disease classification.

Table 4.18: Performance Metrics of Attention-Based CNN Models on Feature-Enhanced CDGN Dataset

Attention-Based Model	ACC	Macro-Average			Weighted Average			Macro-AUC
		REC	PRE	F1	REC	PRE	F1	
VGG16	0.8016	0.8144	0.8362	0.8175	0.8016	0.8188	0.8005	0.9454
Inception-v3	0.7544	0.7720	0.7856	0.7721	0.7544	0.7676	0.7528	0.9333
ResNet50	0.8160	0.8340	0.8536	0.8365	0.8160	0.8304	0.8142	0.9564
DenseNet121	0.8109	0.8285	0.8383	0.8263	0.8109	0.8242	0.8089	0.9560
EfficientNet-B0	0.8047	0.8304	0.8222	0.8233	0.8047	0.8070	0.8025	0.9508

Note. ACC = Accuracy, REC = Recall, PRE = Precision, F1 = F1-score.

Among the five attention-based models, ResNet50 and DenseNet121 emerged as the top performers. ResNet50 achieved the highest classification accuracy at 81.60%, as well as leading macro-average F1-score (0.8365) and macro-average AUC (0.9564). This suggests a superior ability to maintain a balance between sensitivity and specificity across all four classes. DenseNet121 followed closely maintaining a competitive macro-average precision of 0.8536 and macro-average AUC score of 0.9560, reflecting its consistent and confident predictions. Attention-based EfficientNet-B0 also performed commendably, achieving accuracy of 80.47% and a macro-average AUC score of 0.9508. Its macro-average F1-score (0.8233) suggest a well-balanced performance in term of sensitivity and specificity.

Attention-based VGG16 showed stable performance with a slightly lower accuracy of 80.16% while maintaining a reasonable macro-average AUC score of 0.9454. Its macro-average F1-score (0.8175) was also comparatively lower, suggesting slightly less consistent class-wise performance. Attention-based Inception-v3 underperformed relative to the other models, with the lowest accuracy of 75.44% and the macro-average AUC score of 0.9333. Despite moderate precision (0.7856) and recall (0.7720), its overall results suggest that its multi-scale architecture may struggle to align with a singular spatial attention gate, leading to limitations in capturing finer inter-class distinctions compared to more linear residual or dense frameworks.

Comparing the attention-based models to their non-attention counterparts on feature-enhanced dataset, performance gains were observed in several cases. For VGG16, the inclusion of attention resulted in a slight increase in overall accuracy (from 79.34% to 80.16%) and macro-average F1-score (from 0.8081 to 0.8362), although macro-average AUC score decreased marginally (from 0.9479 to 0.945). This suggests that the integration

of attention mechanisms contributed to more precise and balanced predictions, although the overall gains were modest. Inception-v3 showed the most notable improvement among the models. Accuracy improved from 73.38% to 75.44%, macro-average F1-score increased from 0.7520 to 0.7721, and macro-average AUC score rose from 0.9244 to 0.9333. These results suggest that the attention mechanism effectively compensated for Inception-v3's relatively shallow architecture by improving its ability to focus on relevant spatial features. For EfficientNet-B0, its accuracy increased from 79.34% to 80.47%, and the macro-average F1-score rose from 0.8115 to 0.8233. The macro-average AUC score also improved from 0.9453 to 0.9508. These results suggest that the lightweight and highly optimized structure of EfficientNet may synergize well with additional attentions modules.

In contrast, ResNet50 experienced a slight performance decline with the integration of attention. Despite slight increase in accuracy from 81.40% to 81.60%, ResNet50's macro-average F1-score and macro-average AUC score dropped from 0.8475 to 0.8365, and from 0.9581 to 0.9564 respectively. Despite these reductions, the attention-based ResNet50 still retained competitive metrics, suggesting that its residual connections may already provide sufficient feature propagation, thereby limiting the added value of attention modules. Interestingly, DenseNet121 experience decreasing accuracy from 81.60% to 81.09% and macro-average F1-score from 0.8301 to 0.8263. The macro-average AUC score also declined slightly from 0.9579 to 0.9560. These results suggest that DenseNet121's dense connectivity structure may inherently capture spatial dependencies effectively, reducing the incremental benefit of attention mechanisms.

Overall, the integration of spatial attention mechanisms provided a more focused feature extraction process, which is particularly evident in the improvement of macro-level

metrics like F1-score and AUC. These results demonstrate that attention contributes not only to better global accuracy but also to improved classification performance of minority classes. These metrics are critical in medical imaging because they reflect a model's ability to identify minority classes without being biased toward majority class. The mixed results across the five architectures highlight that the efficacy of spatial attention mechanisms is architecture-dependent. While models like VGG16, Inception-v3, and EfficientNet-B0 benefited from attention in terms of accuracy and F1-score, others such as ResNet50 and DenseNet121 which already possess high efficiency feature paths showed slight performance declines on the enhanced dataset. These findings underscore the importance of tuning attention mechanisms to complement the inherent representational characteristics of the CNN backbone.

4.6 Ensemble Model Performance on Feature-Enhanced CDGN Dataset

This section presents the performance of an ensemble model that combines the predictions of multiple attention-based CNN architectures evaluated on enhanced CDGN dataset. By combining models with varying structural philosophies—ranging from the deep feature reuse of DenseNet to the compound scaling of EfficientNet—the ensemble aims to leverage the strength of individual models to achieve more robust and accurate classification. Evaluation includes overall performance metrics, and a discussion of how the ensemble compares to individual models across previous result sets.

a. Confusion Matrix

The confusion matrix of the ensemble model, as shown in Figure 4.46, indicates strong classification performance across all four classes. A notable highlight is the near-perfect identification of Cataracts, with 126 out of 131 cataracts cases were correctly

identified, reflecting a very high true positive rate. The model also shows exceptional discriminative power in Glaucoma and Normal classes, significantly reducing the leakage observed in earlier experiments. While the transition between Diabetic Retinopathy and Normal remains the most complex boundary (75 instances of confusion), the ensemble reduces this error rate compared to individual models. This collective second opinion mechanism effectively filters out the stochastic errors of individual backbones, leading to more stable clinical predictions.

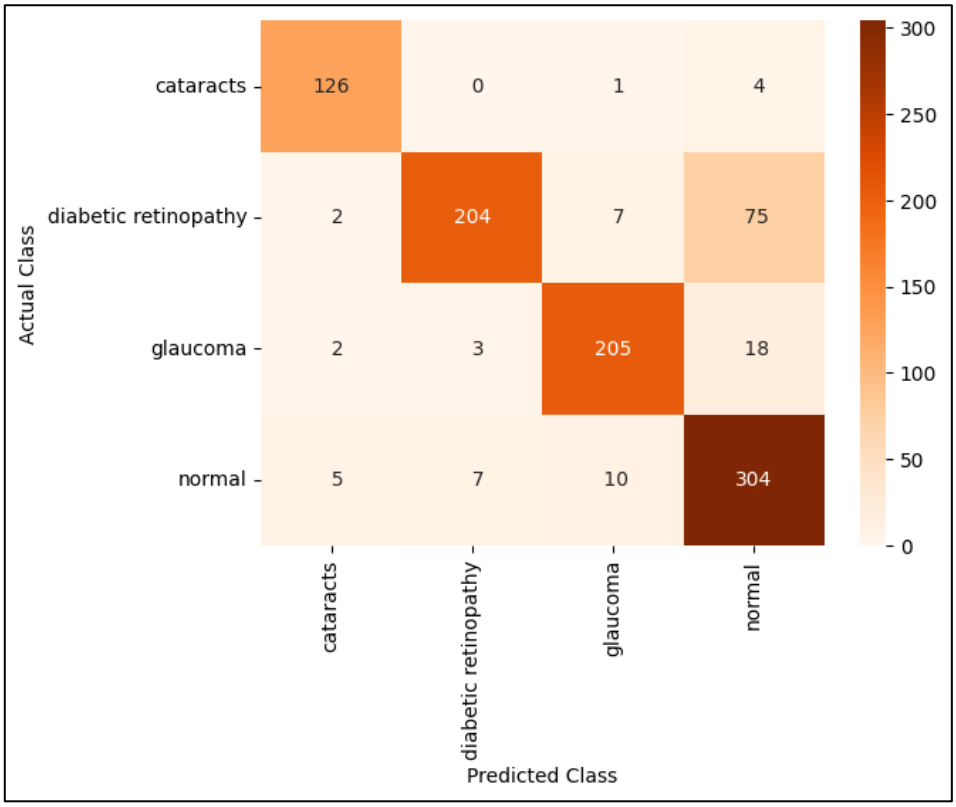


Figure 4.46: Confusion matrix of ensemble model

b. Classification Report

Table 4.19 outlines the classification metrics of the ensemble model. The ensemble model achieved an overall accuracy of 86.23%, outperforming all individual attention-based models. Class-wise metrics further underscore the effectiveness of the ensemble strategy. The ensemble's strength in its balanced sensitivity. Cataracts achieved a recall of 0.9618 and

an F1-score of 0.9474, indicating that nearly all cataract cases were successfully detected. Glaucoma also showed excellent performance, with a recall of 0.8991 and an F1-score of 0.9091, suggesting consistent detection across samples. Most importantly, the ensemble significantly bolstered the precision of Diabetic Retinopathy to 0.9533. This suggests that when the ensemble identifies a cases as Diabetic Retinopathy, the prediction is highly reliable, which is vital for reducing false alarms in clinical screening. The Normal class was also well-handled with an F1-score of 0.8363. Overall, the ensemble approach enhances class-wise generalization and help mitigate bias toward majority classes, which indicated by the macro-average F1-score of 0.8764, underscoring the model's ability to handle class imbalance effectively. The improved recall indicates better sensitivity compared to individual models.

Table 4.19: Classification Metrics of Ensemble Model on Feature-Enhanced CDGN Dataset

Class	Recall	Precision	F1-score	Support
Cataracts	0.9618	0.9333	0.9474	131
Diabetic retinopathy	0.7083	0.9533	0.8127	288
Glaucoma	0.8991	0.9193	0.9091	228
Normal	0.9325	0.7581	0.8363	326
Accuracy	0.8623			973
Macro Average	0.8755	0.8910	0.8764	
Weighted Average	0.8772	0.8772	0.8613	

c. AUC-ROC Analysis

The AUC-ROC curves for ensemble model are illustrated in Figure 4.47. The model attained a macro-average AUC score of 0.9689 and a weighted average AUC score of

0.9631, both of which represent the highest scores across all experimental phases. Class-wise, the AUC scores were consistently high: Cataracts (0.99), Glaucoma (0.98), Diabetic Retinopathy (0.94), and Normal (0.94). These AUC values reflect the model's excellent discriminatory ability across all four classes. The steepness of the curves across all categories indicates that the ensemble maintains high sensitivity while keeping the false positive rate at a minimum, a critical requirement for automated eye disease screening systems.

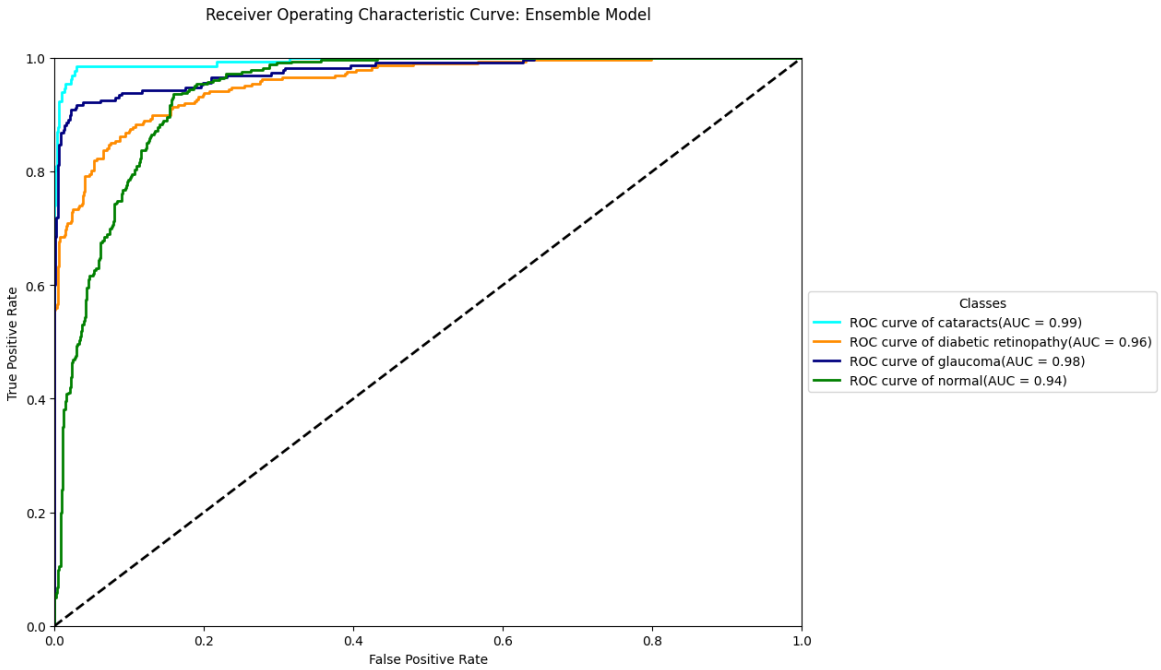


Figure 4.47: ROC curves and corresponding AUC scores for each class in feature-enhanced CDGN dataset of ensemble model

d. Discussion

The results demonstrate that combine the five attention-based models (VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0) through ensemble significantly outperforms individual configurations. While individual models displayed strengths in specifics areas—such as recall, precision, or AUC—the ensemble effectively merges these strengths and mitigates individual weaknesses such as Diabetic Retinopathy/Normal confusion in standalone tests. The performance gain is attributed to the

diversity of the base learners. While EfficientNet-B0 might be sensitive to small-scale vascular changes, ResNet50 provides a more stable global context; the ensemble leverages these varying perspectives to arrive at a more accurate final diagnosis. In the medical imaging domain, where the cost of a false negative can be vision-threatening, the ensemble's 86.23% accuracy and high AUC scores provide a more dependable and clinically viable framework for large-scale eye disease screening.

While it is theoretically expected that an ensemble would perform better, the specific use of soft voting on a heterogenous set of architectures provides the following technical insights into why these improvement occurred:

i. Leveraging Predictive Confidence via Soft Voting

The ensemble utilizes a soft voting mechanism. This is technically superior to hard voting because it accounts for the degree of certainty in each model's prediction. For instance, in complex Diabetic Retinopathy cases where a standalone ResNet50 might be only 51% certain of a "Normal" classification, high-confidence probability scores from the other four models can correct this "weak" prediction. This probabilistic averaging smooths the decision boundary, making the model more resilient to outliers and subtle feature overlaps.

ii. Architectural Diversity and Error Decorrelation

Following the review outlined by Ganaie et al. (2022), an ensemble is most effective when its members are diverse. The five models utilized—VGG16, Inception-v3, ResNet50, DensNet121, and EfficientNet-B0—represent fundamentally different approaches to feature extraction, such as DenseNet's feature reuse and EfficientNet's compound scaling. As a result, their errors are decorrelated. For example, while the attention-based EfficientNet-B0 might overfit to certain enhancement artifacts in Normal images, ResNet50's more

conservative feature propagation helps correct those false positives. The ensemble functions as a "consensus mechanism," where high-confidence correct predictions of the majority outweigh the stochastic misclassification of an individual model, effectively reducing the overall variance and generalization error.

iii. Synergy between Spatial Attention and Ensemble Learning

Each individual model in the ensemble is already optimized via a spatial attention mechanism to focus on disease relevant regions. As noted by Muller et al. (2022), ensembling diverse models allows for a pooling of strengths. Here, the ensemble acts as a meta-analyser that aggregates these localized insights. While individual models might struggle to isolate tiny lesion from healthy background noise, the ensemble can more reliably distinguish between conditions with overlapping features, such as early-stage Diabetic Retinopathy and Normal through their averaged probabilities before confirming a diagnosis.

4.7 Comparative Evaluation of Model Configurations

This section presents a comprehensive analysis of the experimental results, structured to evaluate the proposed deep learning development pipeline from two perspectives. First, a macro-level comparison is conducted across the four experimental phases—baseline using the original CDGN dataset, the feature-enhanced CDGN dataset, the attention-based models and the final attention-based ensemble—to assess the cumulative impact of each stage. Second, a granular class-specific analysis and an architectural ablation study are presented to examine how the spatial attention mechanism specifically influences the discriminative capability of each CNN backbone. This dual-layered evaluation validates the effectiveness of the image enhancement, attention mechanisms, and ensemble learning in enhancing multiclass eye disease classification.

4.7.1 Macro-Performance Summary

Table 4.20 summarizes the performance of the leading model from each experimental phase.

Table 4.20: Comparative Performance Metrics of Best-Performing Models from Each Experimental Configuration

Configuration	Best performing Model	ACC	REC	PRE	F1	AUC
Original CDGN Dataset	DenseNet121	0.8119	0.8348	0.8350	0.8329	0.9558
Enhanced CDGN Dataset	DenseNet121	0.8160	0.8269	0.8466	0.8301	0.9579
Attention + Enhanced CDGN Dataset	ResNet50	0.8160	0.8340	0.8536	0.8365	0.9564
Ensemble	Ensemble	0.8623	0.8755	0.8910	0.8764	0.9689

Note. ACC = Accuracy, REC = Recall, PRE = Precision, F1 = F1-score.

The results demonstrate a consistent upward trend in performance as the pipeline advances. The transition from the original dataset to the feature-enhanced dataset yielded modest but steady improvements, particularly in macro-average precision and AUC scores. This suggests that enhancement methods effectively improved model specificity and generalization by highlighting critical diagnostic features. The integration of spatial attention mechanisms further improved performance, indicating that the attention modules allowed the models to focus more effectively on disease-relevant regions within retinal images, thereby enhancing class-wise configuration. The improvement over the already enhanced dataset configuration highlights the complementary role of attention in refining feature

learning. Finally, the ensemble model—combining predictions from five attention-based models—achieved the highest overall performance across every metric, reaching an accuracy of 86.23% and an AUC score of 0.9689. This underscores the strength of model ensemble in leveraging complementary architectural strengths to improve diagnostic robustness.

In summary, the experimental findings affirm that each stage of the proposed research methodology pipeline—image enhancement, model development with attention mechanism and ensemble learning—contributes positively to classification performance. The progression from the original dataset to the final ensemble model represents a systematic and effective approach to enhancing deep learning models for multiclass eye disease classification using retinal fundus images.

4.7.2 The Impact of Attention Mechanisms (Ablation Study)

To isolate the impact of the attention module, an ablation study was conducted on the feature-enhanced CDGN dataset. Figure 4.48 illustrates the accuracy shifts when moving from non-attention to attention-based configurations. The side-by-side comparison reveals that while the enhancement pipeline provides high class separability, spatial attention serves as a vital refinement tool for localized feature extraction. For higher-performing models like ResNet50, the improvement was marginal (0.2%); however, for models with greater architectural variance like Inception-v3, attention provided a significant performance boost (over 2%). This confirms that attention mechanisms are particularly beneficial for allowing the models to focus on the subtle and diagnostically clinical features.

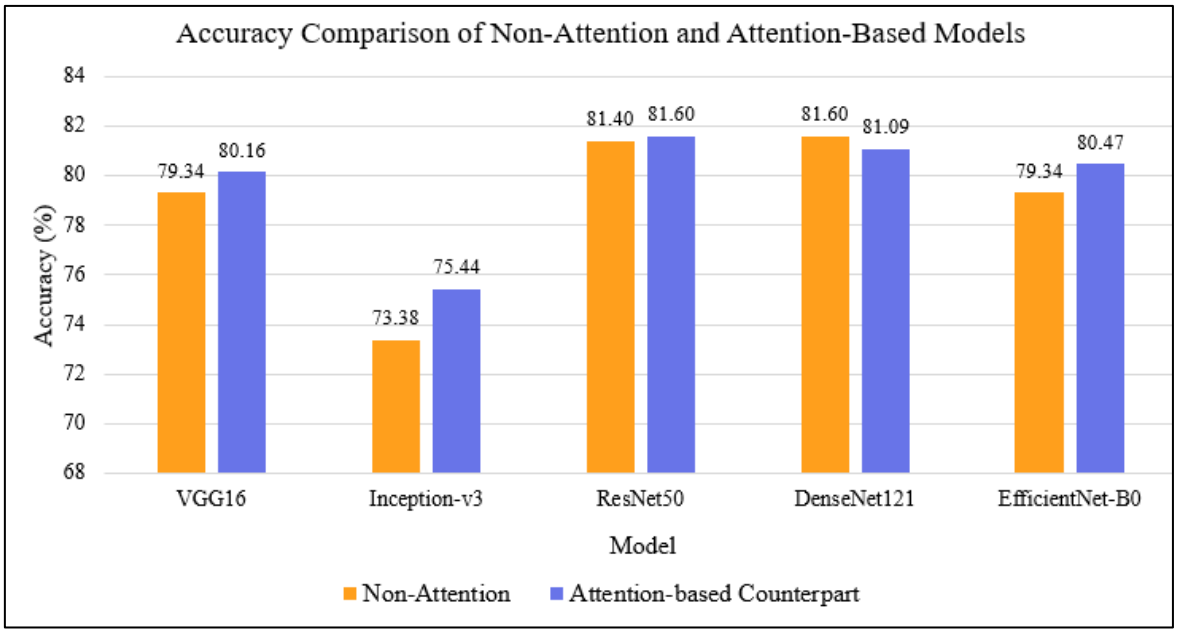


Figure 4.48: Comparison of Performance (Accuracy) of Non-Attention Model versus Its Attention Counterpart on Feature-Enhanced Dataset

4.7.3 Class-Specific Discriminative Analysis

Table 4.21 summarize the class-wise AUC scores to evaluate how different classes respond to attention-weighted feature extraction. Analysis by disease class reveals distinct levels of classification difficulty:

- Cataracts (AUC: 0.99 – 1.00)

This class consistently yielded the highest AUC scores across all architectures. The distinct visual opacity of cataracts, further amplified by image enhancement pipeline, likely provided a clear pathological signature that all models could easily distinguish.

- Glaucoma (AUC: 0.94 – 0.98)

Robustly identified by ResNet50 and DenseNet121, suggesting that structural changes in the optic disc are well captured by deeper connectivity paths.

- Diabetic Retinopathy (AUC: 0.89 – 0.94) and Normal (AUC: 0.88 – 0.92)

These classes are generally the lowest performing class, as models occasionally struggle to distinguish subtle, early-stage disease features, a common challenge in multiclass disease classification.

Table 4.21: Class-Specific AUC Performance Comparison (Non-Attention vs. Attention-based)

Model Architecture		Cataracts	DR	Glaucoma	Normal
VGG16	Non-Attention	0.99	0.93	0.97	0.90
	Attention-Based	0.99	0.92	0.97	0.90
Inception-v3	Non-Attention	0.99	0.89	0.94	0.88
	Attention-Based	0.99	0.90	0.96	0.89
ResNet50	Non-Attention	1.00	0.94	0.98	0.92
	Attention-Based	1.00	0.93	0.98	0.92
DenseNet121	Non-Attention	0.99	0.93	0.98	0.92
	Attention-Based	0.99	0.93	0.98	0.92
EfficientNet-B0	Non-Attention	0.99	0.92	0.97	0.89
	Attention-Based	0.99	0.93	0.98	0.91

A consistent observation throughout this study is that Diabetic Retinopathy consistently exhibit highest rate of misclassification, particularly into the Normal class. Unlike structural changes of Glaucoma or the clouding of Cataracts, early-stage Diabetic Retinopathy is characterized by minute lesions such as microaneurysms and small haemorrhages. These features often overlap with standard retinal textures or physiological artifacts. Even with image enhancement, the contrast between microaneurysm and the background vasculature can be insufficient for a CNN to register a diseased state without

significantly higher resolution or specific channel-wise attention. In addition, some diseased features occupy a very small percentage of the total pixel area. If the attention mechanism fails to localize these tiny features, the model defaults to the dominant features of a healthy retina.

The results demonstrate that the enhanced dataset provides a high degree of separability between classes. Most models maintain an AUC above 0.90 across all conditions validates the effectiveness of the image enhancement pipeline in highlighting diagnostically relevant features. While the attention mechanisms did not drastically shift the scores for already high-performing models like ResNet50, its ability to boost the performance of Inception-v3 and EfficientNetB0 confirms its value as a refinement tool for enhancing localized feature extraction in medical imaging. By focusing the network's computational resources on diagnostically relevant regions, the attention mechanisms capable of mitigating some of the confusion between healthy and diseased samples, resulting in a more robust and reliable multiclass diagnostic system.

4.8 Chapter Summary

This chapter presented a comprehensive evaluation and discussion of five CNN models—VGG16, Inception-v3, ResNet50, DenseNet121, and EfficientNet-B0—for multiclass eye disease classification using retinal fundus images. The evaluation was structured across four experimental configurations: baseline performance on the original CDGN dataset, performance on an enhanced version of the dataset, models integrated with attention mechanisms, and an ensemble of attention-based models. The results demonstrated a clear progression in performance from the baseline to the ensemble model. Experiments using the enhanced dataset demonstrated consistent improvements, indicating the

effectiveness of image enhancement techniques in improving model generalization and disease feature sensitivity. The integration of spatial attention mechanisms further improved model performance, particularly in precision and F1-score, by enabling models to better focus on disease-relevant regions in the retinal fundus images. Finally, the ensemble of attention-based models yielded the highest overall results, significantly outperforming individual models in all key performance metrics. These findings highlight the cumulative benefits of the proposed research pipeline in improving classification performance. The next chapter will present the overall conclusions drawn from this research, along with its contributions, limitations, and future directions.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Overview

This chapter presents the key findings and contributions of the research, summarising how the study has addressed the research objectives. It highlights the theoretical and practical contributions made to the field, particularly in the application of deep learning for eye disease classification. The limitations encountered during the research are also discussed, acknowledging constrains related to data, model performance, or methodological scope. Based on the findings and limitations, this chapter proposes recommendations for both practical implementation and future research directions. These insights aim to guide further advancement in automated eye disease detection and support the continued development of more robust, accurate, and clinically application deep learning models.

5.2 Contribution

This research makes several contributions to the field of deep learning for ophthalmic image analysis, particularly in multiclass eye disease classification using retinal fundus images. The contributions span data curation, model architecture enhancements, and experimental design, each addressing challenges identified in existing literature. The following points summarize the primary contributions of this research:

i. Synthesis of a high-diversity multiclass dataset

A primary contribution of this research is the construction of the **CDGN dataset**, a curated multiclass corpus engineered to address the generalizability gap found in existing ophthalmic datasets. While previous studies often rely on single-source repositories or small-

scale merges, the novelty of the CDGN dataset lies in its strategic synthesis of eight distinct publicly available datasets. This multi-source integration introduces a higher level of hardware and demographic variability, effectively simulating the real-world inconsistencies—such as varying imaging devices, image resolution, and illumination conditions—encountered in actual clinical practice. These inconsistencies are valuable in promoting the development of more robust and generalizable models that can perform effectively across different imaging conditions and populations. Unlike existing datasets that often suffer from limited sample sizes, CDGN dataset provides a higher number of samples per class of four major classes: cataracts, diabetic retinopathy, glaucoma, and normal fundus.

The suitability and objectivity of the CDGN dataset were validated through two primary mechanism: first, by utilizing source repositories already peer-validated and widely cited in existing literature, and second, through a rigorous manual filtration process to ensure high-fidelity disease feature representation and remove label noise. This careful selection was empirically confirmed by benchmarking the dataset across five diverse CNN architectures, where consistently high-performance metrics demonstrated that the dataset provides a robust and unbiased foundation for multiclass classification. By standardizing image through preprocessing steps—including extraneous black background removal and resolution resizing, CDGN dataset represents a scalable, publicly valuable resource for benchmarking model robustness in diverse clinical scenarios.

ii. Integrated deep learning pipeline for enhanced eye disease classification

This research proposes a comprehensive multi-stage integration of feature-specific enhancement and attention-based ensembles, forming a pipeline that aims to improve the robustness and generalizability of multiclass eye disease classification. The image

enhancement procedures utilize CLAHE on L* channel and dual-track morphological enhancement to improve quality of input retinal fundus images and enhance the visibility of vascular structures and optic disc boundaries. The pipeline moves beyond global feature extraction by incorporating spatial attention modules within pre-trained CNN backbones. This enables the model to mathematically prioritize the diagnostically relevant regions generated by the enhancement stages, creating a unique synergy between preprocessing and model architecture. Finally, this pipeline employs an averaged ensemble of multiple attention-based models. This allows the framework to leverage the diverse inductive biases of different architectures to resolve inter-class confusion. By specifically synchronizing the enhancement of disease-specific features with the spatial attention of the neural network, the pipeline achieves a level of diagnostic sensitivity significantly outperformed the baseline configuration. This demonstrates the contribution of each component in the pipeline and highlights the effectiveness of a systematic, multi-stage enhancement strategy for enhancing deep learning approaches in medical image classification.

5.3 Limitations

While this research has demonstrated promising results in the multiclass classification of eye diseases using deep learning, several limitations should be acknowledged.

i. Class imbalance

Despite employing data augmentation and class weighting to mitigate the effects of class imbalance, the issue remains a challenge. Certain disease classes were underrepresented which may have biased the models toward majority classes, affecting generalization ability of model in detecting minority classes.

ii. Coarse-grained disease categorization

Each of the four major classes—cataracts, diabetic retinopathy, glaucoma, and normal—was treated as a single class without further subdivision based on severity levels or clinical staging. For example, all diabetic retinopathy stages were grouped under one class, rather than being classified into mild, moderate, severe, proliferative forms. This coarse-grained approach may limit the clinical applicability of the model, as it does not distinguish between varying levels of disease progression.

iii. Uniform Model Configuration

To ensure a fair and consistent comparison across all five CNN architectures, the same classification head design, hyperparameter settings, and attention mechanism were applied uniformly. While this standardization supports experimental validity, a one-size-fits-all approach may have constrained each model's ability to perform optimally according to its unique structural characteristics.

iv. External validation

Model performance was evaluated using standard performance metrics and internal validation procedures. However, no external validation was conducted on independent datasets or through real-world clinical trials. This limits the generalizability of the findings to unseen populations, different imaging devices, or real-world clinical trials environments.

5.4 Recommendations and Future Works

Building upon the findings and limitations of this research, several directions are recommended for future work to further improve the performance, clinical applicability, and generalizability of deep learning models in eye disease classification:

i. Addressing class imbalance through advanced techniques

Although data augmentation and class weighting were applied in this study, future research could explore more advanced approaches such as synthetic data generations, oversampling strategies like Synthetic Minority Oversampling Technique (SMOTE), or cost-sensitive learning. Additionally, acquiring local clinical data to increase the representation of minority classes can help improve model generalizability.

ii. Fine-grained classification with disease staging

Future studies should consider classifying diseases based on severity levels or clinical stages. For instance, diabetic retinopathy can be subdivided into mild, moderate, severe, and proliferative stages. A more granular classification framework would provide better clinical relevance and support more informed decision-making.

iii. Model-specific optimization

Future work could explore model-specific optimization strategies to further enhance performance, including customizing classification heads, fine-tuning hyperparameters, and adapting attention mechanisms to better align with the model architecture. Tailoring these components could allow each model to leverage its strengths more effectively.

iv. External validation and real-world testing

To ensure model generalizability and real-world applicability, future models should be validated using independent datasets from diverse populations and imaging devices. Additionally, collaboration with healthcare professionals in clinical environments to conduct real-world trials or prospective studies would provide valuable insights into model performance in practical settings.

v. Model interpretability and explainability

Developing models with built-in interpretability mechanisms—such as saliency map, Gradient Class Activation Map (Grad-CAM), or attention heatmaps—can help clinicians understand the rationale of the predictions. This enhancement of model transparency is crucial for trust and adoption in clinical environments.

5.5 Conclusion

This research explored the application and enhancement of deep learning approach for multiclass eye disease classification using retinal fundus images in classifying major eye diseases—cataracts, diabetic retinopathy, glaucoma, along with normal fundus. Experimental configurations were conducted: baseline performance on the original CDGN dataset, improved performance a feature-enhanced dataset, further refinement through attention mechanisms, and a final ensemble of attention-based models. The results demonstrated that image enhancement, attention mechanisms and ensemble learning each contributed positively to model performance. Notably, the ensemble model achieved the highest overall metrics, supporting the effectiveness of the proposed pipeline in addressing the challenges of multiclass eye disease classification. Beyond model development, this research highlighted critical considerations such as class imbalance, the need for finer disease categorization, model-specific configuration, and external validation. These insights form the basis for future improvements and underline the importance of aligning AI-driven diagnostics tools with real-world clinical needs. In conclusion, this research demonstrates the potential of deep learning to support automated, accurate, and scalable eye disease screening. By enhancing both data and model design, this research contributes to the growing body of study aimed at integrating AI into ophthalmology and sets the stage for future development toward clinically deployable solutions.

REFERENCES

Acevedo, E., Orantes, D., Acevedo, M., & Carreño, R. (2025). Identification of eye diseases through deep learning. *Diagnostics*, *15*(7), 916. https://doi.org/10.3390/diagnostics15070916

Akter, N., Fletcher, J., Perry, S., Simunovic, M. P., Briggs, N., & Roy, M. (2022). Glaucoma diagnosis using multi-feature analysis and a deep learning technique. *Scientific Reports*, *12*(1), 8064. https://doi.org/10.1038/s41598-022-12147-y

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00444-8

Al-Zubaidy, M. (2020). How to perform fundoscopy with a direct ophthalmoscope. *Journal of the Foundations of Ophthalmology*. https://doi.org/10.48089/jfo7867006

Arif, Z., Fuadah, R. Y. N., Rizal, S., & Ilhamdi, D. (2023). Classification of eye diseases in fundus images using Convolutional Neural Network (CNN) method with EfficientNet architecture. *JRTI (Jurnal Riset Tindakan Indonesia)*, *8*(1), 125–131. https://doi.org/10.29210/30032835000

Arslan, G., & Erdaş, Ç. B. (2023). Detection of cataract, diabetic retinopathy and glaucoma eye diseases with deep learning approach. *Intelligent Methods in Engineering Sciences*, *2*(2), 042–047. https://doi.org/10.58190/imiens.2023.11

Aslam, J., Arshed, M. A., Iqbal, S., & Hasnain, H. M. (2024). Deep learning based multi-class eye disease classification: Enhancing vision health diagnosis. *Technical*

Journal, *29*(1), 7–12. https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1810/249

Ayeni, J. A. (2022). Convolutional Neural Network (CNN): The architecture and applications. *Applied Journal of Physical Science*, *4*(4), 42–50. https://doi.org/10.31248/AJPS2022.085

Babaqi, T., Jaradat, M., Yildirim, A. E., Al-Nimer, S. H., & Won, D. (2023). *Eye Disease Classification Using Deep Learning Techniques* (arXiv:2307.10501). arXiv. http://arxiv.org/abs/2307.10501

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1409.0473

Bajwa, M. N., Singh, G. a. P., Neumeier, W., Malik, M. I., Dengel, A., & Ahmed, S. (2020). G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/ijcnn48605.2020.9207664

Bakır, H., & Yılmaz, Ş. (2022). Using transfer learning technique as a feature extraction phase for diagnosis of cataract disease in the eye. *Uluslararası Sivas Bilim Ve Teknoloji Üniversitesi Dergisi*, *1*(1), 17–33. https://dergipark.org.tr/en/download/article-file/2604502

Balakrishnan, V., Shi, Z., Law, C. L., Lim, R., Teh, L. L., & Fan, Y. (2022). A deep learning approach in predicting products' sentiment ratings: A comparative analysis. *The Journal of Supercomputing*, *78*(5), 7206–7226. https://doi.org/10.1007/s11227-021-04169-6

Belden, S. (2023, January 11). *What is the optic disc? - medical definition*. All About Vision. Retrieved February 12, 2025, from https://www.allaboutvision.com/eye-care/eye-anatomy/eye-structure/optic-disc/

Bernabe, O., Acevedo, E., Acevedo, A., Carreno, R., & Gomez, S. (2021). Classification of Eye Diseases in Fundus Images. *IEEE Access, 9,* 101267–101276. https://doi.org/10.1109/ACCESS.2021.3094649

Bragança, C. P., Torres, J. M., Soares, C. P. D. A., & Macedo, L. O. (2022). Detection of glaucoma on Fundus images using deep learning on a new image set obtained with a smartphone and handheld ophthalmoscope. *Healthcare*, *10*(12), 2345. https://doi.org/10.3390/healthcare10122345

Bulut, B., Kalin, V., Bektaş Güneş, B., & Khazhin, R. (2022). Classification of Eye Disease from Fundus Images Using EfficientNet. *Artificial Intelligence Theory and Applications*, *2*(1), 1–7. https://dergipark.org.tr/en/download/article-file/2500548

Cen, L., Ji, J., Lin, J., Ju, S., Lin, H., Li, T., Wang, Y., Yang, J., Liu, Y., Tan, S., Tan, L., Li, D., Wang, Y., Zheng, D., Xiong, Y., Wu, H., Jiang, J., Wu, Z., Huang, D., . . . Zhang, M. (2021). Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25138-w

Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-supervised learning*. MIT Press (MA). https://www.molgen.mpg.de/3659531/MITPress--SemiSupervised-Learning.pdf

Chaudhari, A., Shelke, P., Thombare, P., & Sandbhor, S. (2024). Cost-Effective Real-Time Eye Disease Detection and Classification Using Deep Learning Techniques. *2024 15th International Conference on Computing Communication and Networking*

Technologies (ICCCNT), 1–8. https://doi.org/10.1109/ICCCNT61001.2024.10726189

Chavan, R., & Pete, D. (2024). Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network. *Journal of Engineering and Applied Science*, *71*(1), 26. https://doi.org/10.1186/s44147-023-00335-0

Chea, N., & Nam, Y. (2021). Classification of Fundus Images Based on Deep Learning for Detecting Eye Diseases. *Computers, Materials & Continua*, *67*(1), 411–426. https://doi.org/10.32604/cmc.2021.013390

Choudhary, S., & Kesswani, N. (2020). Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. *Procedia Computer Science*, *167*, 1561–1573. https://doi.org/10.1016/j.procs.2020.03.367

Cicinelli, M. V., Buchán, J., Nicholson, M., Varadaraj, V., & Khanna, R. C. (2023). Cataracts. *The Lancet, 401*(10374), 377–389. https://doi.org/10.1016/s0140-6736(22)01839-6

Cui, Y., Sun, R., Liu, B., Liu, Z., & Toe, T. T. (2023). Eye Diseases Classification Using Transfer Learning of Residual Neural Network. *2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS)*, 244–248. https://doi.org/10.1109/ISCTIS58954.2023.10213138

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y., Long, X., Wen, Y., Lu, L., Shen, Y., Chen, Y., Shen, D., Yang, X., Zou, H., Sheng, B., & Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, *12*(1), 3242. https://doi.org/10.1038/s41467-021-23458-5

Deepak, G. D., & Bhat, S. K. (2024). Deep learning-based CNN for multiclassification of ocular diseases using transfer learning. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *12*(1), 2335959. https://doi.org/10.1080/21681163.2024.2335959

Doddi, G. V. (2020). *Eye Disease Retinal Images* [Dataset; Kaggle]. Kaggle. https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data

Ejaz, S., Baig, R., Ashraf, Z., Alnfiai, M. M., Alnahari, M. M., & Alotaibi, R. M. (2024). A deep learning framework for the early detection of multi-retinal diseases. *PLOS ONE*, *19*(7), e0307317. https://doi.org/10.1371/journal.pone.0307317

Ejaz, S., Zia, H. U., Majeed, F., Shafique, U., Altamiranda, S. C., Lipari, V., & Ashraf, I. (2025). Fundus image classification using feature concatenation for early diagnosis of retinal disease. *Digital Health*, *11*, 20552076251328120. https://doi.org/10.1177/20552076251328120

Emir, B., & Colak, E. (2023). Performance analysis of pretrained convolutional neural network models for ophthalmological disease classification. *Arquivos Brasileiros de Oftalmologia*, *87*(5). https://doi.org/10.5935/0004-2749.2022-0124

Erdaş, Ç. B., & Arslan, G. (2024). Efficient detection of multiclass eye diseases using deep learning models: A comparative study. *EnSci Dubai 2024 – International Conference on Engineering & Sciences*, 06–16. https://doi.org/10.20319/icstr.2024.0616

Galloway, N. R., & Amoaku, W. M. K. (1999). Basic Anatomy and Physiology of the Eye. In N. R. Galloway & W. M. K. Amoaku, *Common Eye Diseases and their Management* (pp. 5–12). Springer London. https://doi.org/10.1007/978-1-4471-3625-5_2

Ganaie, M., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151

GBD 2019 Blindness and Vision Impairment Collaborators & Vision Loss Expert Group of the Global Burden of Disease Study. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *The Lancet Global Health*, *9*(2), e144–e160. https://doi.org/10.1016/s2214-109x(20)30489-7

Gheisari, S., Shariflou, S., Phu, J., Kennedy, P. J., Agar, A., Kalloniatis, M., & Golzan, S. M. (2021). A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Scientific Reports*, *11*(1), 1945. https://doi.org/10.1038/s41598-021-81554-4

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. In *MIT Press eBooks*. MIT Press. http://www.deeplearningbook.org

Guergueb, T., & Akhloufi, M. A. (2021). Ocular Diseases Detection using Recent Deep Learning Techniques. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3336–3339. https://doi.org/10.1109/EMBC46164.2021.9629763

Gupta, I. K., Patil, S., Mahadevkar, S., Kotecha, K., Mishra, A. K., & Rodrigues, J. J. P. C. (2025). Retinal fundus imaging-based diabetic retinopathy classification using transfer learning and fennec fox optimization. *MethodsX*, *14*, 103232. https://doi.org/10.1016/j.mex.2025.103232

Han, X., Zhong, Y., Cao, L., & Zhang, L. (2017). Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sensing*, *9*(8), 848. https://doi.org/10.3390/rs9080848

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Helen, D., & Gokila, S. (2023). EYENET: An Eye Disease Detection System using Convolutional Neural Network. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 839–842. https://doi.org/10.1109/ICECAA58104.2023.10212139

Hemalakshmi, G. R., Santhi, D., Mani, V. R. S., Geetha, A., & Prakash, N. B. (2021). Classification of retinal fundus image using MS-DRLBP features and CNN-RBF classifier. *Journal of Ambient Intelligence and Humanized Computing*, *12*(9), 8747–8762. https://doi.org/10.1007/s12652-020-02647-y

Hemelings, R., Elen, B., Barbosa-Breda, J., Blaschko, M. B., De Boever, P., & Stalmans, I. (2021). Deep learning on fundus images detects glaucoma beyond the optic disc. *Scientific Reports*, *11*(1), 20313. https://doi.org/10.1038/s41598-021-99605-1

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Imaduddin, H., Utomo, I. C., & Anggoro, D. A. (2024). Fine-tuning ResNet-50 for the classification of visual impairments from retinal fundus images. *International*

Journal of Electrical and Computer Engineering (IJECE), *14*(4), 4175. https://doi.org/10.11591/ijece.v14i4.pp4175-4182

Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science*, *132*, 679–688. https://doi.org/10.1016/j.procs.2018.05.069

International Business Machines Corporation [IBM]. (n.d.). *What are Neural Networks? | IBM*. IBM. Retrieved February 15, 2024, from https://www.ibm.com/topics/neural-networks

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (arXiv:1502.03167). arXiv. https://doi.org/10.48550/arXiv.1502.03167

Jain, A. (2022). *Glaucoma Fundus Imaging datasets* [Dataset; Kaggle]. https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets

Jain, P., & Patidar, S. (2023). Eyes Disease Detection Using Deep Learning Methodologies. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 174–177. https://doi.org/10.1109/UPCON59197.2023.10434618

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jhaveri, R. H., Revathi, A., Ramana, K., Raut, R., & Dhanaraj, R. K. (2022). A Review on Machine Learning Strategies for Real-World Engineering Applications. *Mobile Information Systems*, *2022*, 1–26. https://doi.org/10.1155/2022/1833507

Joint Shantou International Eye Centre. (2021). *1000 Fundus images with 39 categories* [Dataset; Kaggle]. https://www.kaggle.com/datasets/linchundan/fundusimage1000

Kallel, F., & Echtioui, A. (2024). Retinal fundus image classification for diabetic retinopathy using transfer learning technique. *Signal, Image and Video Processing*, *18*(2), 1143–1153. https://doi.org/10.1007/s11760-023-02820-8

Kaur, G., Sharma, N., Chauhan, R., Kukreti, S., & Gupta, R. (2023). Eye Disease Classification Using ResNet-18 Deep Learning Architecture. *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, 1–5. https://doi.org/10.1109/INCOFT60753.2023.10425690

Khalid, M., Sajid, M. Z., Youssef, A., Khan, N. A., Hamid, M. F., & Abbas, F. (2024). CAD-EYE: An Automated System for Multi-Eye Disease Classification Using Feature Fusion with Deep Learning Models and Fluorescence Imaging for Enhanced Interpretability. *Diagnostics*, *14*(23), 2679. https://doi.org/10.3390/diagnostics14232679

Khazaeni, L. M. (2023a, December 6). *The eye examination*. MSD Manual Consumer Version. Retrieved December 15, 2023, from https://www.msdmanuals.com/home/eye-disorders/diagnosis-of-eye-disorders/the-eye-examination

Khazaeni, L. M. (2023b, December 8). *Cataract*. Merck Manuals Professional Edition. Retrieved December 29, 2023, from https://www.merckmanuals.com/professional/eye-disorders/cataract/cataract?query=cataract#

Kovalyk, O., Morales-Sánchez, J., Verdú-Monedero, R., Sellés-Navarro, I., Palazón-Cabanes, A., & Sancho-Gómez, J. (2022a). PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01388-1

Kovalyk, O., Morales-Sánchez, J., Verdú-Monedero, R., Sellés-Navarro, I., Palazón-Cabanes, A., & Sancho-Gómez, J.-L. (2022b). *PAPILA* [Dataset]. https://figshare.com/articles/dataset/PAPILA/14798004/1?file=28454352

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Krüger, F. (2016). *Activity, context, and plan recognition with computational causal behaviour models* [PhD Dissertation, University of Rostock]. https://www.researchgate.net/publication/314116591_Activity_Context_and_Plan_Recognition_with_Computational_Causal_Behaviour_Models

Kumar, P., Bhandari, S., & Dutt, V. (2023). Pre-Trained Deep Learning-Based Approaches for Eye Disease Detection. *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, 1286–1290. https://doi.org/10.1109/ICCPCT58313.2023.10245175

Larxel. (2020). *Ocular disease recognition* [Dataset; Kaggle]. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Mannepalli, P. K., Baghela, V. D. S., Agrawal, A., Johri, P., Dubey, S. S., & Parmar, K. (2024). Transfer Learning for Automated Classification of Eye Disease in Fundus Images from Pretrained Model. *Traitement Du Signal*, *41*(5), 2459–2470. https://doi.org/10.18280/ts.410520

Marouf, A. A., Mottalib, M. M., Alhajj, R., Rokne, J., & Jafarullah, O. (2022). An Efficient Approach to Predict Eye Diseases from Symptoms Using Machine Learning and Ranker-Based Feature Selection Methods. *Bioengineering, 10*(1), 25. https://doi.org/10.3390/bioengineering10010025

McCaa, C. S. (1982). The eye and visual nervous system: anatomy, physiology and toxicology. *Environmental Health Perspectives*, *44*, 1–8. https://doi.org/10.1289/ehp.82441

Meedeniya, D., Shyamalee, T., Lim, G., & Yogarajah, P. (2025). Glaucoma identification with retinal fundus images using deep learning: Systematic review. *Informatics in Medicine Unlocked*, *56*, 101644. https://doi.org/10.1016/j.imu.2025.101644

Muchuchuti, S., & Viriri, S. (2023). Retinal Disease Detection Using Deep Learning Techniques: A Comprehensive review. *Journal of Imaging*, *9*(4), 84. https://doi.org/10.3390/jimaging9040084

Muller, D., Soto-Rey, I., & Kramer, F. (2022). An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. *IEEE Access*, *10*, 66467–66480. https://doi.org/10.1109/access.2022.3182399

Narkhede, S. (2022, March 5). *Understanding AUC - ROC Curve - towards data science*. Medium. Retrieved March 1, 2024, from https://medium.com/towards-data-science/understanding-auc-roc-curve-68b2303cc9c5

National Eye Institute. (2024, November 8). *Diabetic Retinopathy | National Eye Institute*. Retrieved February 14, 2025, from https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy

National survey shows fewer cases of blindness in those over 50. (2023, November 17). *The Star*. https://www.thestar.com.my/news/nation/2023/11/18/national-survey-shows-fewer-cases-of-blindness-in-those-over-50

Nichols, J. A., Chan, H. W. H., & Baker, M. A. B. (2018). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, *11*(1), 111–118. https://doi.org/10.1007/s12551-018-0449-9

Orlando, J. I., Fu, H., Breda, J. B., Van Keer, K., Bathula, D. R., Diaz-Pinto, A., Fang, R., Heng, P., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S. S. R., Shankaranarayana, S. M., Sikka, A., . . . Bogunović, H. (2020). REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, *59*, 101570. https://doi.org/10.1016/j.media.2019.101570

Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Giancardo, L., Quellec, G., & Mériaudeau, F. (2021a). *Retinal Fundus Multi-Disease Image Dataset (RFMID)* [Dataset]. https://doi.org/10.21227/s3g7-st65

Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Giancardo, L., Quellec, G., & Mériaudeau, F. (2021b). Retinal Fundus Multi-Disease Image Dataset (RFMID): a dataset for multi-disease detection research. *Data*, *6*(2), 14. https://doi.org/10.3390/data6020014

Pan, Y., Liu, J., Cai, Y., Yang, X., Zhang, Z., Long, H., Zhao, K., Yu, X., Zeng, C., Duan, J., Xiao, P., Li, J., Cai, F., Yang, X., & Tan, Z. (2023). Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases. *Frontiers in Physiology*, *14*. https://doi.org/10.3389/fphys.2023.1126780

Peking University. (2019). *Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019)*. grand-challenge.org. Retrieved January 31, 2023, from https://odir2019.grand-challenge.org/introduction/

Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P., & Pizer, S. M. (1998). Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, *11*(4), 193–200. https://doi.org/10.1007/bf03178082

Porter, D. (2018, April 23). *What is a Slit Lamp?* American Academy of Ophthalmology. Retrieved December 15, 2023, from https://www.aao.org/eye-health/treatments/what-is-slit-lamp

Priyadharsini, C., & Jagadeesh, R. (2023). Retinal image enhancement based on color dominance of image. *Scientific Reports*, *13*(1), 7172. https://doi.org/10.1038/s41598-023-34212-w

Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & Haq, Q. M. U. (2023). Understanding of Convolutional Neural Network (CNN): A Review. International Journal of Robotics and Control Systems, 2(4), 739–748. https://doi.org/10.31763/ijrcs.v2i4.888

Qummar, S., Khan, F. G., Shah, S. A., Khan, A., Shamshirband, S., Rehman, Z. U., & Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access*, *7*, 150530–150539. https://doi.org/10.1109/access.2019.2947484

Raj, A., Tiwari, A. K., & Martini, M. G. (2019). Fundus image quality assessment: survey, challenges, and future scope. *IET Image Processing*, *13*(8), 1211–1224. https://doi.org/10.1049/iet-ipr.2018.6212

Raza, A., Khan, M. U., Saeed, Z., Samer, S., Mobeen, A., & Samer, A. (2021). Classification of Eye Diseases and Detection of Cataract using Digital Fundus Imaging (DFI) and Inception-V4 Deep Learning Model. *2021 International Conference on Frontiers of Information Technology (FIT)*, 137–142. https://doi.org/10.1109/FIT53504.2021.00034

Saini, A., Guleria, K., & Sharma, S. (2023). An Efficient Deep Learning Model for Eye Disease Classification. *2023 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 1–6. https://doi.org/10.1109/SCSE59836.2023.10215000

Sandoval-Cuellar, H. J., Alfonso-Francia, G., Vázquez-Membrillo, M. A., Ramos-Arreguín, J. M., & Tovar-Arriaga, S. (2021). Image-based glaucoma classification using fundus images and deep learning. *Revista Mexicana De Ingenieria Biomedica*, *42*(3), 28–41. https://doi.org/10.17488/rmib.42.3.2

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., & Wang, K. (2021). Image Preprocessing in Classification and Identification of Diabetic Eye Diseases. *Data Science and Engineering*, *6*(4), 455–471. https://doi.org/10.1007/s41019-021-00167-z

Sayal, A., Jha, J., N, C., Gupta, V., Gupta, A., Gupta, O., & Memoria, M. (2023). Neural Networks And Machine Learning. *2023 IEEE 5th International Conference on*

Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), 58–63. https://doi.org/10.1109/ICCCMLA58983.2023.10346612

Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T. S., & Pfeiffer, N. (2020). The diagnosis and treatment of glaucoma. *Deutsches Arzteblatt International*. https://doi.org/10.3238/arztebl.2020.0225

Shamia, D., Prince, S., & Bini, D. (2022). An Online Platform for Early Eye Disease Detection using Deep Convolutional Neural Networks. 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), 388–392. https://doi.org/10.1109/ICDCS54290.2022.9780765

Shamrat, F. M. J. M., Shakil, R., Sharmin, Hoque Ovy, N., Akter, B., Ahmed, M. Z., Ahmed, K., Bui, F. M., & Moni, M. A. (2024). An advanced deep neural network for fundus image analysis and enhancing diabetic retinopathy detection. *Healthcare Analytics*, *5*, 100303. https://doi.org/10.1016/j.health.2024.100303

Shamsan, A., Senan, E. M., & Shatnawi, H. S. A. (2023). Automatic Classification of Colour Fundus Images for Prediction Eye Disease Types Based on Hybrid Features. *Diagnostics*, *13*(10), 1706. https://doi.org/10.3390/diagnostics13101706

Sharma, R., Gangrade, J., Gangrade, S., Mishra, A., Kumar, G., & Kumar Gunjan, V. (2023). Modified EfficientNetB3 Deep Learning Model to Classify Colour Fundus Images of Eye Diseases. *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 632–638. https://doi.org/10.1109/ICCCMLA58983.2023.10346769

Shoukat, A., Akbar, S., Hassan, S. A. E., Rehman, A., & Ayesha, N. (2021). An Automated Deep Learning Approach to Diagnose Glaucoma using Retinal Fundus Images. *2021*

International Conference on Frontiers of Information Technology (FIT), 120–125. https://doi.org/10.1109/FIT53504.2021.00031

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for Large-Scale image recognition*. https://doi.org/10.48550/arXiv.1409.1556

Singh, G., Guleria, K., & Sharma, S. (2023). A Pre-Trained VGG16 Model for Cataract Eye Disease Prediction. *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1–6. https://doi.org/10.1109/SMARTGENCON60755.2023.10442647

Sivaswamy, J., Krishnadas, S. R., Chakravarty, A., Joshi, G. D., Ujjwal, & Syed, T. A. (2015a). A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, *2*(1). https://api.semanticscholar.org/CorpusID:29553360

Sivaswamy, J., Krishnadas, S. R., Chakravarty, A., Joshi, G. D., Ujjwal, & Syed, T. A. (2015b). *Drishti-GS Dataset* [Dataset]. http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php

Sivaswamy, J., Krishnadas, S. R., Joshi, G. D., Jain, M., Ujjwal, & Abbas, S. T. (2014). Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI 2014)*. https://doi.org/10.1109/isbi.2014.6867807

Srivastava, O., Tennant, M., Grewal, P. S., Rubin, U., & Seamone, M. E. (2023). Artificial intelligence and machine learning in ophthalmology: A review. *Indian Journal of Ophthalmology*, *71*(1), 11-17. https://doi.org/10.4103/ijo.ijo_1569_22

Sushith, M., Sathiya, A., Kalaipoonguzhali, V., & Sathya, V. (2025). A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. *Scientific Reports*, *15*(1), 15166. https://doi.org/10.1038/s41598-025-99309-w

Sutton, R., & Barto, A. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054. https://doi.org/10.1109/tnn.1998.712192

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning* (arXiv:1602.07261). arXiv. https://doi.org/10.48550/arXiv.1602.07261

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). *Rethinking the Inception Architecture for Computer Vision* (arXiv:1512.00567). arXiv. https://doi.org/10.48550/arXiv.1512.00567

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.48550/arxiv.1512.00567

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015*

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *36th International Conference on Machine Learning*, *abs/1905.11946*, 6105–6114. https://doi.org/10.48550/arxiv.1905.11946

Tashkandi, A. (2025). Eye Care: Predicting Eye Diseases Using Deep Learning Based on Retinal Images. *Computation*, *13*(4), 91. https://doi.org/10.3390/computation13040091

Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*, *12*(5), 91. https://doi.org/10.3390/computers12050091

Terra, J. (2024, August 13). *What is a ROC curve, and how do you use it in performance modeling?* Simplilearn.com. Retrieved January 15, 2025, from https://www.simplilearn.com/what-is-a-roc-curve-and-how-to-use-it-in-performance-modeling-article

Thanki, R. (2023). A deep neural network and machine learning approach for retinal fundus image classification. *Healthcare Analytics*, *3*, 100140. https://doi.org/10.1016/j.health.2023.100140

Toki, S. A., Rahman, S., Billah Fahim, S. M., Al Mostakim, A., & Rhaman, Md. K. (2022). RetinalNet-500: A newly developed CNN Model for Eye Disease Detection. *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 459–463. https://doi.org/10.1109/MIUCC55081.2022.9781785

Topaloglu, I. (2023). Deep Learning Based Convolutional Neural Network Structured New Image Classification Approach for Eye Disease Identification. *Scientia Iranica*, *30*(5), 1731–1742. https://doi.org/10.24200/sci.2022.58049.5537

Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D. I., & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, *135*, 104599. https://doi.org/10.1016/j.compbiomed.2021.104599

Turbert, D. (2022, March 10). *Visual Field test*. American Academy of Ophthalmology. Retrieved December 15, 2023, from https://www.aao.org/eye-health/tips-prevention/visual-field-testing

Vardhan, K. B., Nidhish, M., C, S. K., Shameem, D. N., Charan, V. S., & Bhavadharini, R. M. (2024). Eye disease detection using deep learning models with transfer learning techniques. *ICST Transactions on Scalable Information Systems, 11*. https://doi.org/10.4108/eetsis.5971

Varghese, R. E., & Pandian, I. A. (2023). Inception-Resnet V2 Based Eye Disease Classification Using Retinal Images. *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 1–5. https://doi.org/10.1109/ICMNWC60182.2023.10435893

Wahab Sait, A. R. (2023). Artificial Intelligence-Driven Eye Disease Classification Model. *Applied Sciences*, *13*(20), 11437. https://doi.org/10.3390/app132011437

Willoughby, C. E., Ponzin, D., Ferrari, S., Lobo, A., Landau, K., & Omidi, Y. (2010). Anatomy and physiology of the human eye: effects of mucopolysaccharidoses

disease on structure and function – a review. *Clinical and Experimental Ophthalmology*, *38*(s1), 2–11. https://doi.org/10.1111/j.1442-9071.2010.02363.x

World Health Organization. (2019). *World Report on Vision*. World Health Organization. https://www.who.int/publications/i/item/world-report-on-vision

World Health Organization. (2023, August 10). *Blindness and vision impairment*. Retrieved January 25, 2025, from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

Xia, X., Zhan, K., Li, Y., Xiao, G., Yan, J., Huang, Z., Huang, G., & Fang, Y. (2022a). *Eye Disease Diagnosis and Fundus Synthesis* [Dataset]. https://github.com/xia-xx-cv/EDDFS_dataset

Xia, X., Zhan, K., Li, Y., Xiao, G., Yan, J., Huang, Z., Huang, G., & Fang, Y. (2022b). Eye Disease Diagnosis and Fundus Synthesis: A Large-Scale Dataset and Benchmark. *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, 1–6. https://doi.org/10.1109/mmsp55362.2022.9949547

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Zhang, Z., Yin, F. S., Liu, J., Wong, W. K., Tan, N. M., Lee, B. H., Cheng, J., & Wong, T. Y. (2010). ORIGA-light: An online retinal fundus image database for glaucoma analysis and research. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 3065–3068. https://doi.org/10.1109/iembs.2010.5626137