Mohamed Khider University of Biskra
Faculty of Exact Sciences
Department of Computer Science

MASTER THESIS

Speciality : Master - Networks and Information and Communication Technologies - Ref : P1189/2026

TITLE

AI-Powered Ocular Image Analysis System

Prepared and defended by

Soundous FODIL

Lili FETTAH

Defense date : 08/06/2026

DEFENSE COMMITTEE

Full Name	Grade	Role
Djihane BABAHENINI	MCB	**President**
Abdelhak MERIZIG	MCA	**Supervisor**
AHLEM BENTRAH	MCB	**Examiner**
Faouzia HAGAIN	vacataire	**Incubator Representative**

Academic Year : 2025/2026

Acknowledgements

First and most importantly, all praise and gratitude belong to Allah, the Almighty, for inspiring **us** with the strength, patience, and determination required to bring this work to completion.

We wish to express **our** profound gratitude to **our** supervisors, **Mr. Merizig Abdelhak** and **Mr. Zerari Abd El Mouméne**, for their exceptional guidance, unwavering support, and insightful mentorship throughout this research journey. Their profound expertise and encouragement not only shaped this thesis but also served as a constant source of motivation. **We** are deeply indebted to them for the invaluable time and energy they dedicated to **our** academic growth.

We would also like to extend **our** sincere appreciation to the esteemed members of the jury for their time, effort, and commitment to reviewing this work. Their constructive feedback and valuable insights are highly appreciated.

We are also grateful to **Dr. Nader Shabat** of Clinic El Razzi, Biskra, and to the medical staff at **Dr. Ghamri's** clinic, for their generous guidance and valuable medical insights, which greatly enriched the practical dimension of this work.

We extend heartfelt gratitude to **our families** for their unconditional love, sacrifices, and encouragement, which have been the foundation of **our** academic journey. Their unwavering belief made this achievement possible.

We sincerely thank **our** friends and colleagues for their support, knowledge, and companionship, which made this journey rewarding. Special appreciation goes to **Hiouni Salah Eddine** for his steadfast support and to **Imane Fodil** for her constant encouragement and presence.

Fettah Lili
Fodil Soundous

Abstract

This thesis presents the design and implementation of a stack web-based clinical decision support system for automated ocular disease classification using multimodal imaging, including slit-lamp, fundus, and Optical Coherence Tomography (OCT) images. The proposed system integrates three specialized deep learning models, each trained to classify diseases within its respective imaging modality. To ensure seamless usability in real clinical environments, an automatic modality classification module is introduced to first identify the input image type and route it to the appropriate diagnostic model without requiring manual selection from the user.

The platform is deployed as a web application with a role-based access architecture that supports System Administrators, Medical Doctors, and Nursing Staff. Medical users can register patients, upload ocular images, obtain real-time AI-driven diagnostic predictions with associated confidence scores, and generate explainable attention maps that pinpoint the regions that most influence the model's decisions, thereby enhancing interpretability and fostering clinical trust.

Experimental evaluation demonstrates strong performance across all modalities, achieving validation accuracies of 98.17% for slit-lamp images, 96.17% for OCT images, and 92.24% for fundus images. These results confirm the effectiveness of multimodal deep learning combined with automated modality routing in improving diagnostic accuracy and workflow efficiency.

Overall, this work contributes a scalable and interpretable AI-assisted ophthalmology system that enhances diagnostic support through multimodal imaging

and web-based clinical integration.

Keywords : Ocular Disease Classification, Multimodal Imaging, Clinical Decision Support System, Deep Learning, Ophthalmology, Optical Coherence Tomography (OCT), Fundus Imaging, Slit-Lamp Imaging, Explainable AI, Medical Image Analysis, Web-Based Healthcare System, Automated Diagnosis.

Résumé

Cette thèse présente la conception et la mise en œuvre d'un système d'aide à la décision clinique basé sur le Web, destiné à la classification automatisée des maladies oculaires à l'aide d'une imagerie multimodale, incluant les images de lampe à fente (slit-lamp), du fond d'œil (fundus) et de la tomographie par cohérence optique (Optical Coherence Tomography, OCT). Le système proposé intègre trois modèles spécialisés d'apprentissage profond, chacun entraîné pour classifier les maladies correspondant à sa modalité d'imagerie spécifique.

Afin de garantir une utilisation fluide dans des environnements cliniques réels, un module automatique de classification des modalités est introduit pour identifier d'abord le type d'image en entrée et l'orienter vers le modèle diagnostique approprié, sans nécessiter de sélection manuelle de la part de l'utilisateur.

La plateforme est déployée sous forme d'une application Web dotée d'une architecture d'accès basée sur les rôles, prenant en charge les administrateurs système, les médecins et le personnel infirmier. Les utilisateurs médicaux peuvent enregistrer les patients, téléverser des images oculaires, obtenir des prédictions diagnostiques en temps réel générées par l'intelligence artificielle avec leurs scores de confiance associés, ainsi que produire des cartes d'attention explicables mettant en évidence les régions ayant la plus grande influence sur les décisions du modèle, améliorant ainsi l'interprétabilité et renforçant la confiance clinique.

L'évaluation expérimentale démontre de solides performances pour l'ensemble des modalités, avec des précisions de validation atteignant 98,17 % pour les images de lampe à fente, 96,17 % pour les images OCT et 92,24 % pour les images du fond

d'œil. Ces résultats confirment l'efficacité de la combinaison de l'apprentissage profond multimodal et du routage automatique des modalités pour améliorer la précision diagnostique et l'efficacité des flux de travail cliniques.

Dans l'ensemble, ce travail apporte une contribution significative à travers le développement d'un système ophtalmologique assisté par l'intelligence artificielle, évolutif et interprétable, qui améliore l'aide au diagnostic grâce à l'imagerie multimodale et à son intégration dans un environnement clinique basé sur le Web.

Mots-clés : Classification des maladies oculaires, Imagerie multimodale, Système d'aide à la décision clinique, Apprentissage profond, Ophtalmologie, Tomographie par cohérence optique (OCT), Imagerie du fond d'œil, Imagerie à la lampe à fente, Intelligence artificielle explicable, Analyse d'images médicales, Système de santé basé sur le Web, Diagnostic automatisé.

Table des matières

Acknowledgements **i**

Abstract **i**

Résumé **iii**

List of Figures **xi**

List of Tables **xiv**

List of Abbreviations **xvi**

1 Introduction **1**

1.1 General Context 1

1.2 Problem 2

1.3 Problematic 3

1.4 Objectives 4

1.5 Thesis structure 5

2 Ocular Problems overview **6**

2.1 Introduction 6

2.2 Ocular Disorders 7

2.3 Major Retinal Diseases 8

2.3.1 Glaucoma 8

2.3.2 Cataracts 9

2.3.3 Corneal Disorders 10

2.3.4 Macular Hole 10
2.3.5 Diabetic retinopathy 11
2.3.6 Age-related macular degeneration (AMD) 12
2.4 Overview of Ocular Imaging Modalities 13
2.4.1 Fundus Photography 14
2.4.2 Optical Coherence Tomography 15
2.4.3 Slit Lamp 16
2.5 Challenges in Diagnosis and Early Detection 18
2.5.1 Asymptomatic Progression : The "Silent Thief" Phenomenon 18
2.5.2 Shortage of Specialized Human Resources 18
2.5.3 Image Quality and Artifact Interference 19
2.5.4 Screening Gaps for High-Risk Populations 19
2.6 Related work 19
2.7 Conclusion 22

3 AI techniques applied for ocular problems 23

3.1 Introduction 23
3.2 Machine Learning and Deep Learning in Medical Imaging 24
3.2.1 Theoretical Foundations 24
3.3 Manual and Traditional Computer–Vision–Based Diagnostic Methods 26
3.3.1 Manual grading 26
3.3.2 Traditional Computer Vision 27
3.4 Artificial Neural Networks (ANN) 29
3.4.1 The Mathematical Neuron Model 29
3.4.2 Network Architecture 29
3.4.3 Activation Functions 30
3.4.4 Forward Propagation, Backpropagation, and Loss Functions 31
3.5 Convolutional Neural Networks (CNN) 31
3.5.1 Core CNN Architecture 31

3.5.2 Popular CNN Architectures 33
3.6 Transfer Learning 35
3.6.1 Approaches to Transfer Learning 36
3.7 Conclusion 37

System Design **38**

3.8 Introduction 38
3.9 Overview of the Proposed Multi-Modal System Architecture . . . 39
3.9.1 System block diagram 39
3.9.2 Why Separate Models per Modality 40
3.10 Dataset Description 40
3.10.1 Optical Coherence Tomography Dataset 41
3.10.2 Fundus Photography Dataset 41
3.10.3 Slit–Lamp Dataset 42
3.10.4 Modality Classification Dataset 42
3.11 Preprocessing Pipelines 43
3.11.1 Image Resizing 43
3.11.2 Contrast-Limited Adaptive Histogram Equalization (CLAHE) 43
3.11.3 Pixel Intensity Normalisation 44
3.12 Data Augmentation 44
3.13 Dataset Splitting Strategy 45
3.13.1 Held-out Test Set 45
3.13.2 Stratified K-Fold Cross-Validation 45
3.14 Convolutional Block Attention Module (CBAM) 45
3.15 Training Phase of the Proposed Deep Learning Architectures . . . 46
3.15.1 Fundus Branch MobileNetV2 with CBAM 46
3.15.2 OCT Branch DenseNet121 with CBAM 47
3.15.3 Slit-Lamp Branch MobileNetV2 48
3.15.4 MobileNetV2 for Modality Classification 49
3.16 Explainability : Gradient-weighted Class Activation Mapping . . . 50

3.17 Prediction Phase, ... 50
3.17.1 Accuracy 50
3.17.2 Precision 51
3.17.3 Recall 51
3.17.4 F1-score 51
3.18 Conclusion 52

Results and Discussion **53**

3.19 Introduction 53
3.20 Environments, Implementations, and Development Tools 53
3.20.1 Hardware Configuration 54
3.20.2 Development Tools 54
3.20.3 PostgreSQL 56
3.20.4 JavaScript 57
3.20.5 Packages and APIs 57
3.21 Experimental Configuration 60
3.22 Interface Demonstration 61
3.22.1 Login Interface 61
3.22.2 Administrator Interfaces 61
3.22.3 Nurse Interfaces 63
3.22.4 Doctor Interfaces 65
3.23 Model Performance Results 69
3.23.1 Optical Coherence Tomography Classification (OCT) Results 69
3.23.2 Fundus Photography Classification Results 76
3.23.3 Slit Lamp Classification Results 83
3.23.4 MobileNetV2 for Modality Classification Results 90
3.24 Comparison of Proposed Models with Existing Approaches 92
3.24.1 OCT Model Comparison 93
3.24.2 Fundus Photography Comparison Table 93
3.24.3 Slit Lamp Comparison Table 94

3.25 Conclusion 94

General conclusion and Future work **95**

References **97**

Table des figures

2.1 Visual system [1]. 7

2.2 Normal VS Glaucoma [2]. 9

2.3 Human eye effected by catarats [3]. 10

2.4 Macular hole disease [4]. 11

2.5 Diabetic retinopathy (DR) [5]. 12

2.6 Macular-Degeneration-AMD disease [6]. 13

2.7 Colour fundus photography (CFP) modality[7]. 15

2.8 Stereo images improve diagnosis by adding depth information to 2D images, making it easier to locate and identify eye abnormalities [7]. 15

2.9 Spectral-domain optical coherence tomography (SD-OCT) image of retinal layers. [8]. 16

2.10 High-resolution digital slit lamps [9]. 17

2.11 Slit-lamp images of the right and left eyes [10]. 17

3.1 AI vs ML vs DL and NN. 25

3.2 Machine Learning and Deep Learning. 25

3.3 The structure of ANN model [11]. 30

3.4 Convolutional layer [12]. 32

3.5 Pooling layer [12]. 32

3.6 Fully-connected layers [13]. 33

3.7 Example of Residual CNNs [12]. 34

3.8 The dense block structure and the overall DenseNet architecture [14]. 35
3.9 Technique of Transfer Learning [15]. 36

3.10 The proposed architecture of our system. 40
3.11 The proposed Fundus Photography model of our system. 47
3.12 The proposed OCT model of our system. 48
3.13 The proposed Slit L model of our system. 49

3.14 Python logo. 54
3.15 Anaconda logo. 55
3.16 Jupyter logo. 55
3.17 Google Colab interface. 55
3.18 HTML and CSS concepts. 56
3.19 PostgreSQL. 56
3.20 JavaScript. 57
3.21 Login interface of the Eye Clinic Management System. 61
3.22 Administrator interface for creating and managing clinics. 62
3.23 Administrator interface for creating user accounts (doctors and nurses). 62
3.24 Administrator interface showing the global list of registered patients. 63
3.25 Administrator dashboard summarizing system-wide metrics. . . . 63
3.26 Nurse interface for registering a new patient and recording consent. 64
3.27 Nurse interface for uploading ocular images and assigning them to patients. 64
3.28 Nurse interface displaying the list of registered patients and image counts. 64
3.29 Doctor interface displaying the list of patients with demographic information and image counts. 66

3.30 Cataract diagnosis on a slit-lamp image with AI prediction, confidence scores, and Grad-CAM visual explanation. 67
3.31 Glaucoma diagnosis on a fundus image with AI prediction, confidence scores, and Grad-CAM visual explanation. 68
3.32 Macular hole diagnosis on an OCT image with AI prediction, confidence scores, and Grad-CAM visual explanation. 68
3.33 Training and validation accuracy curves across all 5 folds. 70
3.34 Training and validation loss curves across all 5 folds. 70
3.35 Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance. 71
3.36 Averaged training and validation accuracy curves across all 5 folds. 72
3.37 DenseNet121 Held-out Test Confusion Matrix. 72
3.38 K-Fold Cross-Validation Summary. 73
3.39 DenseNet121 Classification Report. 74
3.40 EfficientNet-B3 K-Fold Cross-Validation Results. 75
3.41 Confusion matrix on the held-out test set for the EfficientNet-B3+CBAM model. 75
3.42 Classification Report for the EfficientNet-B3+CBAM model. . . . 76
3.43 Training and validation accuracy curves across all 5 folds. 77
3.44 Training and validation loss curves across all 5 folds. 77
3.45 Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance. 78
3.46 Averaged training and validation accuracy curves across all 5 folds. 79
3.47 MobileNetV2 K-fold Held-out Test Confusion Matrix. 79
3.48 MobileNetV2 K-fold Cross-Validation Summary. 80
3.49 MobileNetV2 K-fold Cross-Validation Classification Report. . . . 81
3.50 Training and validation loss and accuracy curves for the ResNet50 model over 30 epochs. 82
3.51 ResNet50 Confusion Matrix on the held-out test set. 82

3.52 ResNet50 Test Set Classification Report. 83
3.53 Training and validation accuracy curves across all 5 folds. 84
3.54 Training and validation loss curves across all 5 folds. 85
3.55 Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance. 86
3.56 Averaged training and validation accuracy curves across all 5 folds. 86
3.57 MobileNetV2 Held-out Test Confusion Matrix. 87
3.58 MobileNetV2 K-Fold Cross-Validation Summary. 87
3.59 MobileNetV2 K-Fold Cross-Validation Classification Report. . . . 88
3.60 Training and validation loss and accuracy curves for the EfficientNet-B0 model over 20 epochs. 89
3.61 EfficientNet-B0 Confusion Matrix. 89
3.62 EfficientNet-B0 Test Set Classification Report. 90
3.63 MobileNetV2 Modality Classifier Accuracy. 91
3.64 MobileNetV2 Modality Classifier Confusion Matrix. 91

Liste des tableaux

2.1 Comparison of ophthalmic imaging modalities and their characteristics. 14

3.1 Manual grading process in retinal disease studies [16]. 26
3.2 Primary descriptors used in computer-vision methods. 28

3.3 OCT Dataset Description 41
3.4 The Fundus Dataset Description 42
3.5 Slit-Lamp Dataset Description 42

3.6 Deep Learning Packages and APIs Overview. 59
3.7 Our proposed Deep Learning Configurations. 60
3.8 Comparison of OCT models based on validation accuracy 93
3.9 Comparison of Fundus Photography models based on validation accuracy 93
3.10 Comparison of Slit Lamp models based on validation accuracy . . 94

List of Abbreviations

AI Artificial Intelligence

OCT Optical Coherence Tomography

SE Squeeze-and-Excitation block

CBAM Convolutional Block Attention Module

ML Machine Learning

DL Deep Learning

CNN Convolutional Neural Network

CLAHE Contrast-Limited Adaptive Histogram Equalization

ReLU Rectified Linear Unit

API Application Programming Interface

FC Fully Connected

XAI Explainable Artificial Intelligence

Chapitre 1

Introduction

1.1 General Context

The eye is one of the five senses that distinguish living beings, through which they can see objects and distinguish between colors and other things. Vision occurs through the transmission of nerve signals from the eye to the brain's visual cortex. Any disruption in this neural pathway negatively affects vision, which is reflected in a person's daily life, as the eye is an essential organ. With the spread of aging, the incidence of eye diseases such as visual impairment, cataracts, and glaucoma is increasing. Diagnosis and treatment methods vary depending on doctors' experience and the level of medical advancement available. In addition, social and economic status play important roles in the prevention and treatment of eye diseases, especially in low-income countries and remote areas that lack health services and specialists. This leads to the spread of eye diseases and delays in their diagnosis and treatment, increasing the risk of vision loss and negatively impacting the lives of individuals and society.

Therefore, advanced ocular imaging modalities, including optical coherence tomography and fundus photography. play a major role in ophthalmology in the recognition, screening, and follow-up of eye problems by providing detailed images of structures such as the retina and optic nerve. Besides enabling early diagno-

sis and disease monitoring, their analysis is typically subjective, time-consuming, and requires expertise, and the ever-growing volume of imaging data is creating a major problem for the health sector.

Simultaneously, health care systems are facing major problems, including a shortage of ophthalmologists, unequal distribution of eye care, and a rapid increase in patient demand, especially in remote areas where specialized facilities are lacking. This hampers large-scale screening and early diagnosis through conventional methods, thereby increasing the need for AI tools that can be scaled up.

Lately, Deep Learning in general, and CNNs in particular, have shown outstanding performance in analyzing medical images, such as retinal and OCT images, enabling highly accurate feature extraction and the detection of various diseases. AI is transforming the way ophthalmologists diagnose diseases by providing accurate, fast diagnostics, thereby minimizing errors and workload, and telemedicine, along with cloud technology, is making it possible to offer remote services.

This thesis bridges AI research and clinical ophthalmology, aiming to build a practical, workflow-integrated system that builds trust and expands access to eye care. It starts by exploring ocular diseases and diagnostic hurdles.

1.2 Problem

Despite the availability of modern medical imaging technologies, early and accurate detection of eye diseases remains a major challenge. The primary reason is the large number of patients. According to [17], at least 2.2 billion people worldwide live with vision impairment or blindness, including no fewer than 1 billion people whose conditions could be prevented or treated but who lack access to appropriate care.

Population aging boosts vision impairment and loss risks, mainly from cataracts and refractive errors, the top causes of moderate or severe vision issues in those over 50. Early intervention requires substantial scaling up of eye care services to address this largely preventable burden [18]. Developing countries and rural areas face acute shortages of medical resources and limited healthcare technology, hindering timely, efficient, and accessible eye care for all.

Traditional ophthalmology relies on manual fundus and OCT image exams, but these are time-intensive, subjective, and prone to missing subtle early signs of diseases such as glaucoma or diabetic retinopathy, especially amid variations in clinician experience and image overload. Rising image volumes from hospitals overwhelm specialists, delaying diagnoses in screenings and clinics, where asymptomatic early stages risk irreversible vision loss without timely treatment.

These limitations highlight the urgent need for automated, AI–driven systems to enable scalable, accurate early detection and improve outcomes amid growing patient loads and specialist shortages.

1.3 Problematic

How can artificial intelligence and deep learning algorithms be best incorporated into an intelligent system for ocular image analysis for accurate, fast, and automatic detection of multi-eye diseases from heterogeneous ocular image datasets with a high level of reliability, scalability, and practical usability in real clinical environments ?

1.4 Objectives

This thesis presents the creation of an AI-powered system aimed at simplifying the analysis of eye images (Retinal and OCT). The main objective is to aid doctors in the recognition, diagnosis, and ongoing care of main eye disorders such as glaucoma, diabetic retinopathy, and macular degeneration.

To help bridge the worldwide scarcity of eye doctors and the rising number of patients, this research intends to develop a fullyautomatic diagnostic tool. Besides increasing the overall patient safety, the proposed system by making diagnostics faster and more accurate, will be able to prevent permanent loss of vision. It will also facilitate healthcare delivery through telemedicine.

To realize the general goal, we will focus on four measurable milestones :

— Data Collection and Preprocessing : collecting high-quality and diverse datasets and conducting techniques such as normalization and augmentation to make sure the AI performs well in real clinical scenarios throughout the world.

— Disease Classification and Detection : The evolution of deep learning models that can accurately recognize the different sides of healthy and diseased eyes in the four major ocular diseases.

— Performance Evaluation : Quantifiable measures (e. g. accuracy sensitivity, and specificity) will be utilized to verify the system for clinical reliability and false negative minimization.

— Interface Development : an ophthalmologist–friendly interface is developed which is very easy to use so that an eye doctor can upload images without any hassle and also receive quick diagnostic support along with

risk assessments.

Ultimately, this research seeks to connect AI innovation with practical clinical workflows, reducing the burden on medical professionals while significantly improving patient outcomes through early intervention.

1.5 Thesis structure

This thesis unfolds in four chapters, each building upon the last to present a comprehensive solution to the challenges outlined above.

Chapter 2 outlines the medical and technical foundations of ocular diseases, imaging techniques, and diagnostic hurdles. It also surveys prior research and highlights recent AI-driven advances in ophthalmology.

Chapter 3 delves into artificial intelligence methods applied to ocular image analysis, covering classical computer vision, deep learning paradigms, convolutional neural networks, transfer learning strategies, and current AI systems for disease identification.

Chapter 4 details the proposed framework and system architecture, including dataset curation, preprocessing pipelines, model design, and the integrated workflow of the diagnostic platform.

Chapter 5 discusses the implementation, experimental setup, evaluation metrics, and results. It further examines the systemâs limitations and suggests directions for future research.

Chapter 6 concludes the thesis by synthesizing the key contributions, objectives achieved, and overall findings. It also reflects on future opportunities and potential enhancements for AI-assisted ocular diagnosis.

Chapitre 2

Ocular Problems overview

2.1 Introduction

Eye health plays a major role in human life. The eye is the primary organ responsible for vision and is highly suited to sensing light. It directs light landing on the light-sensitive retina, where specialized retinal nerve cells turn it into electrical signals that are transmitted to the brain through the optic nerve [19]. However, ocular disorders are grow in prevalence worldwide. According to The Lancet Global Health [20], hundreds of millions of people worldwide suffer from visual impairment, and this number is expected to increase by 2050.
Diseases of the optic nerve cause detectable structural changes such as thinning of the nerve fiber layers and modifications to the optic disc. These changes act as objective biomarkers and are commonly evaluated through optical coherence tomography and fundus imaging [21].

This chapter describe the medical background needed to understand eye diseases, which is the focus of our AI-based diagnostic systems. It emphasizes the importance of combining clinical experience with data-driven techniques to enhance the reliability of the outcomes. This chapter describes the major eye disorders, and also includes other categories and discusses challenges that exist in diagnosing and detecting these conditions early. As well it introduces various types of ocular imaging modalities used to diagnose eye disease, which include

fundus photography, optical coherence tomography, and slit lamp imaging, and will compare manual grading and traditional computer vision techniques used in ophthalmic diagnosis.

2.2 Ocular Disorders

Ocular disorders are defined as all medical conditions that affect the eye and produce some degree of loss of normal vision or complete blindness. This includes conditions that damage parts of the eye, the retina, optic nerve, lens, cornea, and/or any other component of the visual system [22]. (**Figure** 2.1) illustrates the anatomy of the visual system [1].

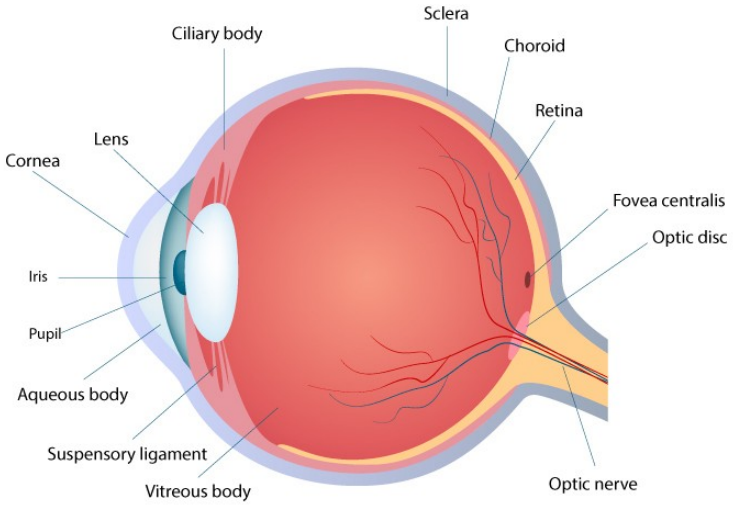


FIGURE 2.1 – Visual system [1].

It is necessary to categorize these diseases by their effects on the visual system to assess the extent of automated analysis.

— **The anterior segment :** is the anterior section of the eye play a key role for focusing light toword the retina. It consists of the cornea, iris, crystalline lens, and two fluid-filled chambers containing aqueous humour.

Consequently, Disorders of the anterior segment include dry eye, conjunctivitis, refractive errors, keratoconus, glaucoma, and cataracts [23].

— **The posterior segment :** is the back piece of the eye that involve the vitreous humour, retina, choroid, and optic nerve. It plays a significant role in vision by converting light into neural signals and transmitting them to the brain. Disorders of the posterior segment involve macular degeneration, retinal detachment, diabetic retinopathy, vitreous disorders, optic nerve diseases, and inherited retinal conditions [24].

Despite advances in diagnostic tools, the increasing number of patients and the growing volume of ocular images present significant challenges for clinicians. These challenges highlight the need for more efficient and objective diagnostic approaches, paving the way for automated and AI-assisted systems to support ophthalmologists in detecting and managing ocular disorders.

2.3 Major Retinal Diseases

Although the human eye is vulnerable to a multitude of diseases, several will be targets of automated diagnostic systems. The reason is that these diseases are very common and have characteristic visual features (biomarkers) in medical images.

2.3.1 Glaucoma

Glaucoma is a term used for various types of eye conditions that share the common feature of gradual optic nerve damage. This damage results in the loss of retinal ganglion cells, thinning of the retinal nerve fiber layer, and progressive enlargement of the optic disc cupping [25]. Affecting 76 million people worldwide, and this number is likely to exceed 111.8 million by 2040 [26]. It is a serious disease and a leading cause of blindness.

— Therefore, when evaluating using optical coherence tomography (OCT),

macular cube scans should be part of the assessments instead of only relying on optical disc imaging. Probability maps have higher sensitivity than global indices for initial detection, and a systematically analyzed, wide-field OCT scan can achieve equal or even higher diagnostic accuracy than conventional diagnostic methods [27].

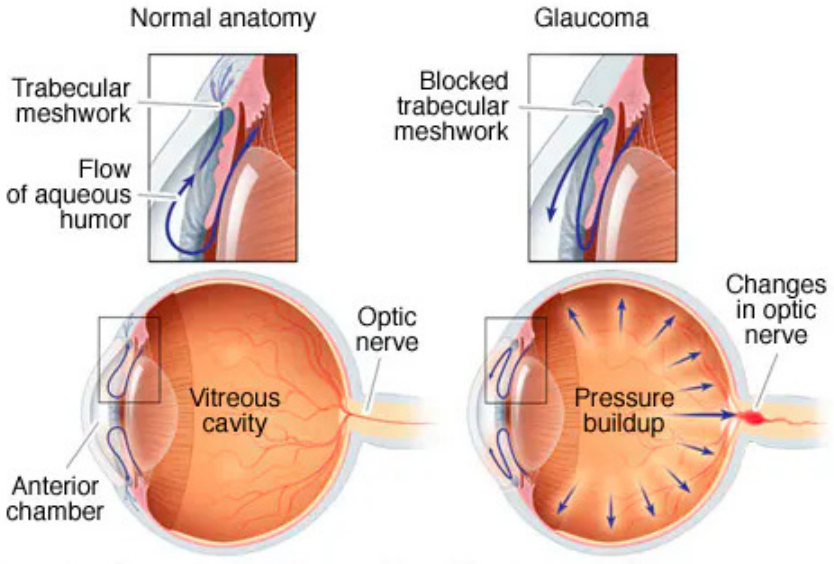


FIGURE 2.2 – Normal VS Glaucoma [2].

2.3.2 Cataracts

Cataracts are among the most common eye disorders, causing the eye's natural lens to become cloudy and leading to a significant reduction in vision. They are one of the leading case of vision loss globally, especially in the elderly, and can mostly be avoided. Some habits, like smoking and excessive exposure to UV rays, are the key danger factors for the development of cataracts. The diagnosis is confirmed by a thorough eye examination, including testing of vision, dilation of the pupil to examine the lens, and measuring the eye pressure to check the condition of the eye in general [28].

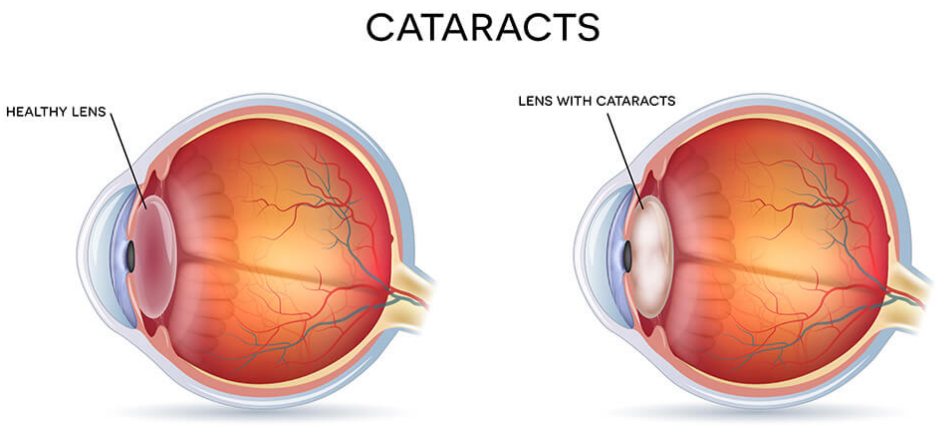


FIGURE 2.3 – Human eye effected by catarats [3].

2.3.3 Corneal Disorders

Corneal disease encompasses a variety of disorders that affect the cornea, which is the clear outermost layer of the eye. Such diseases may result in loss of vision, eye pain, and in severe cases, they may become a significant burden not only for patients' quality of life but also for the healthcare system. Typical disorders associated with the cornea are dry eye syndrome (DES), Fuchs endothelial dystrophy (FED), and keratoconus (KCN)[29].

2.3.4 Macular Hole

Macular holes are retinal defects that occur in the middle of the fovea and can lead to significant vision loss. Previously considered to have an unknown cause, current understanding suggests that macular holes result from the vitreous body pulling on the fovea. The formation of the hole is primarily due to dynamic forces caused by the movement of the posterior cortical vitreous, and not solely to static vitreous contraction. The article states that the prevalence of macular holes in the general population is approximately 3.3 per 1,000 people, and that they may resolve spontaneously, remain unchanged, or progress to a nearly full-thickness defect [30], (**Figure** 2.4).

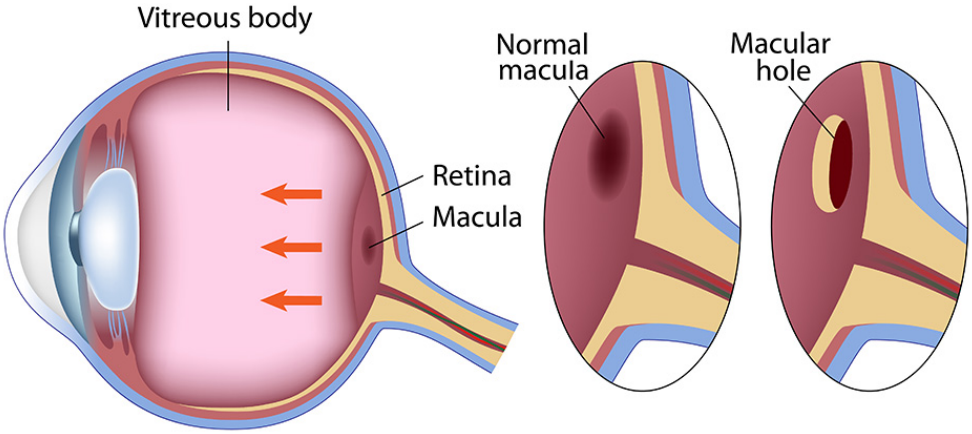


FIGURE 2.4 – Macular hole disease [4].

2.3.5 Diabetic retinopathy

Diabetic retinopathy (DR) is an eye complication caused by diabetes. It is mainly due to pathological alterations in the retinal blood vessels. These alterations include vascular swelling and increased permeability, leading to the leakage of blood and fluid into retinal tissues. Without proper treatment, the disorder may cause considerable vision loss. The mechanism by which the disease progresses is the retinal vascular leakage resulting in the accumulation of fluid in the vitreous humor. Malignant tumors (MA) and hemorrhages (HM) are visible as red spots, while soft and hard exudates (EX) are bright spots that interfere with the normal visual function. As the disease progresses, the retina is deprived of oxygen, leading to the formation of new, abnormal blood vessels (neovascularization), which ultimately results in a significant loss of vision [31].

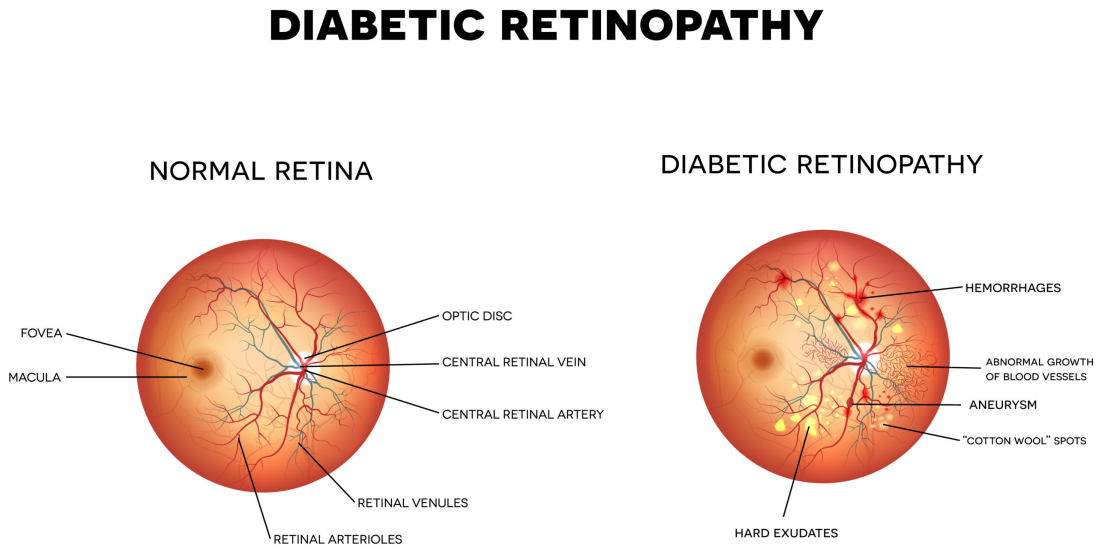


FIGURE 2.5 – Diabetic retinopathy (DR) [5].

2.3.6 Age-related macular degeneration (AMD)

Age-related macular degeneration (AMD) was defined in [32] as the presence of geographic atrophy or neovascular AMD. Geographic atrophy refers to areas of retinal degeneration within 3000 μm of the macula's center, typically associated with drusen or chromosomal abnormalities, and occurring in the absence of neovascular features. Neovascular age-related macular degeneration (AMD) is characterized by features such as retinal pigment epithelial detachment, subretinal hemorrhages, and the formation of new blood vessels visible in the choroid, or fibrous scars that mimic previous neovascularization in the choroid, provided that no other retinal disease can explain these features.

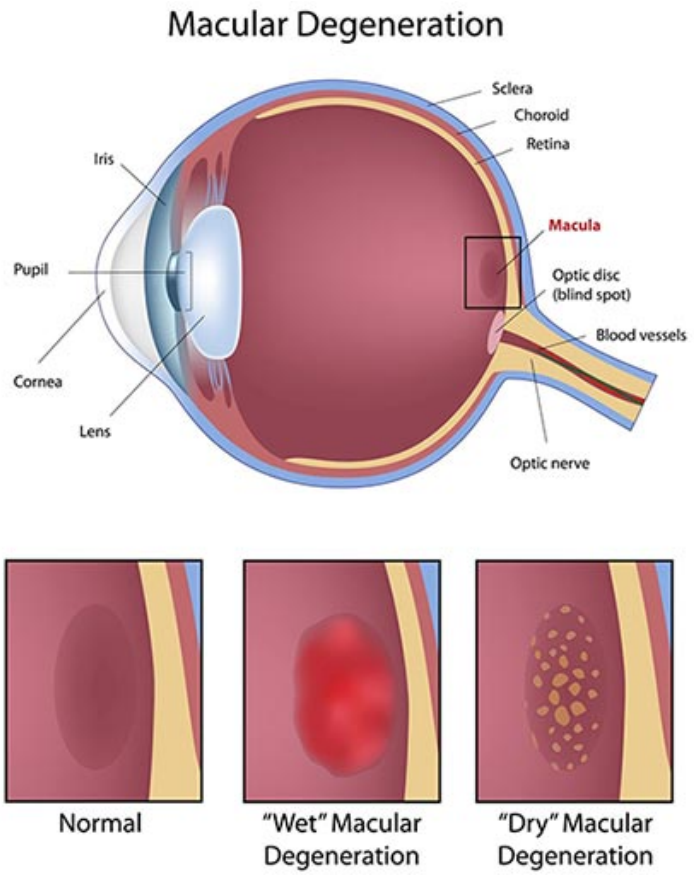


FIGURE 2.6 – Macular-Degeneration-AMD disease [6].

2.4 Overview of Ocular Imaging Modalities

Precisely identifying eye conditions requires top-notch imaging that can detect changes in the structure and function of different parts of the eye. There are several methods for capturing images for eye diagnosis, and each tool has its own pros, cons, and degree of compatibility with automated analysis. Understanding these imaging techniques is a prerequisite to knowing how AI algorithms handle and decode eye data.

Modality	Target Structure	Resolution	Cost	AI Suitability
Optical Coherence Tomography (OCT)	Retina layers, optic nerve head (cross-sectional)	∼2–10 μm (can reach ∼2–3 μm) [33].	High	Very High
Fundus Photography	The retina, macula, and optic nerve	∼10–20 μm (limited depth, 2D) [34].	Low-Medium	High
Slit Lamp	Anterior segment, cornea, lens	∼10–20 μm (depends on optics, magnification) [35].	Medium	Moderate
OCT Angiography (OCTA)	Retina layers, optic nerve head (cross-sectional)	∼5–10 μm [36].	High	Very High

TABLE 2.1 – Comparison of ophthalmic imaging modalities and their characteristics.

2.4.1 Fundus Photography

Color fundus photography is a key modality technique used to assess and document the condition of the retina. It relies on a fundus camera that captures two-dimensional (planar) images of the retina. Traditional versions provide a field of view of 30 to 55 degrees, allowing imaging of the macula and optic nerve in a single frame, while modern (wide-field and ultra-wide-angle) techniques allow imaging of up to 200 degrees of the retina in a one image [34].

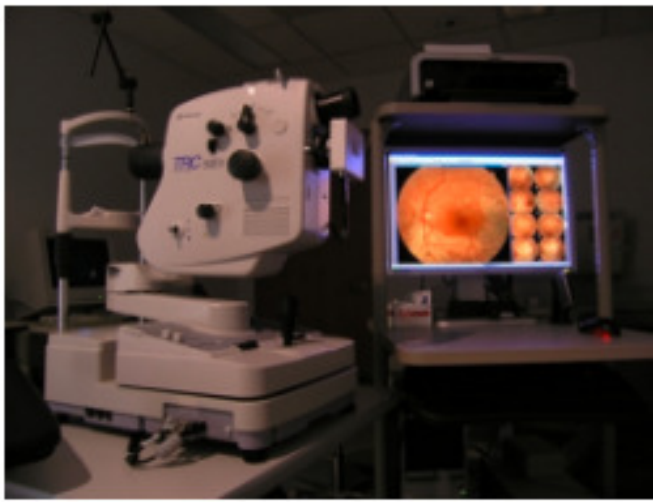


FIGURE 2.7 – Colour fundus photography (CFP) modality[7].

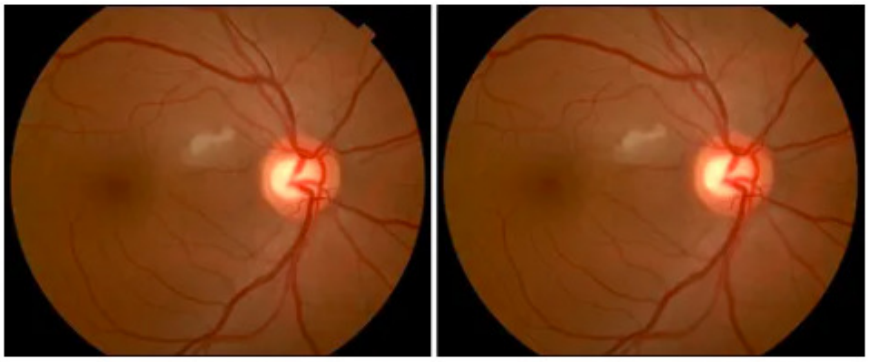


FIGURE 2.8 – Stereo images improve diagnosis by adding depth information to 2D images, making it easier to locate and identify eye abnormalities [7].

2.4.2 Optical Coherence Tomography

Optical coherence tomography (OCT) is an imaging technique used in medicine to produce high-resolution, micron-scale, three-dimensional images. This technique relies on optical reflectance measurement, a method that measures the reflection of backlight as it passes through transparent or semi-transparent materials such as biological tissues. This is achieved by analyzing both the intensity of the reflected light and the time delay of its return from the tissue. In practice, light from a broadband source is split into two paths : a reference path and a sample path, with the latter reflecting off structures at different depths within

the posterior region of the eye [7]. Figure 2.9, It displays a cross-sectional image in the spectral field (tomogram) of the retina and choroid.

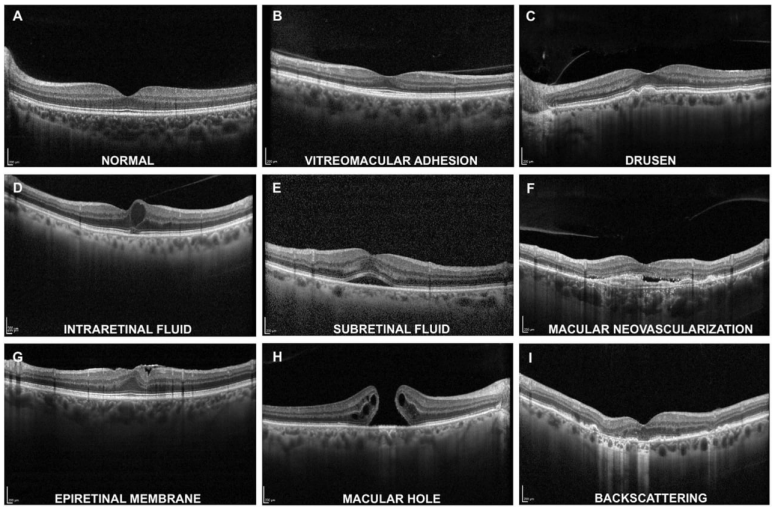


FIGURE 2.9 – Spectral-domain optical coherence tomography (SD-OCT) image of retinal layers. [8].

OCT data poses a challenge for AI due to its 3D structure and larger data size compared to fundus images. OCT deep learning models have to deal with speckle noise, shadow artifacts from blood vessels, and inter-device variability from the OCT manufacturers. Despite these challenges, seminal works such as De Fauw et al. [37] have shown that AI systems processing OCT volumes can achieve expert-level diagnostic accuracy across multiple pathologies simultaneously.

2.4.3 Slit Lamp

An optical device with a high-intensity light source that can project as a slit is called a slit lamp (**Figure** 2.10). A focused light beam with adjustable length, angle, and width is projected from this stereoscopic biomicroscope. The three-dimensional anatomical details of the anterior segment and adnexa are easier to see. A +90 D or +78 D lens can also be used to see the posterior portion. The slit lamp can also be utilized with a variety of accessories [38].



FIGURE 2.10 – High-resolution digital slit lamps [9].

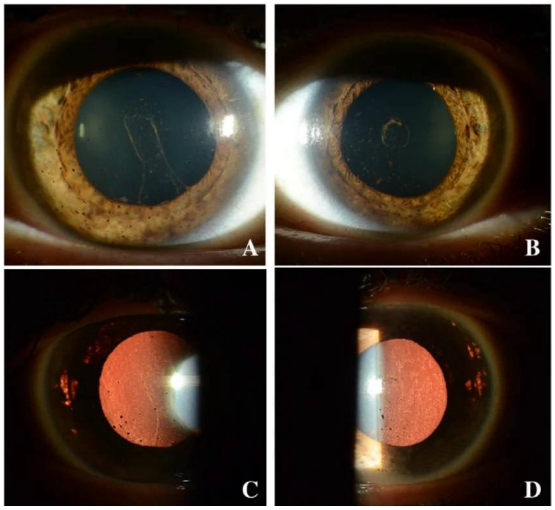


FIGURE 2.11 – Slit-lamp images of the right and left eyes [10].

In the area of AI-based diagnosis, slit-lamp images have been applied mostly to identify cataracts, specifically grading lens nuclear opacity according to defined scales such as the Lens Opacities Classification System III (LOCS III). AI systems using convolutional neural networks trained on slit-lamp images have demonstrated performance comparable to trained ophthalmologists in cataract grading tasks [39].

2.5 Challenges in Diagnosis and Early Detection

One of the greatest therapeutic challenges in treating eye conditions, especially those whose pathology progresses progressively. There are such medical barriers that it is impossible to detect eye diseases at an early stage, thus not giving the patient the possibility of timely treatment, and the global situation with people deprived of their vision is gradually worsening.

2.5.1 Asymptomatic Progression : The "Silent Thief" Phenomenon

Many sight-threatening conditions, including glaucoma and early-stage diabetic retinopathy (DR), progress insidiously without noticeable symptoms until substantial, irreversible damage has occurred [40]. This creates a critical reliance on population-based screening programs rather than symptom-driven presentation, a model that is difficult to implement universally.

2.5.2 Shortage of Specialized Human Resources

It is obvious that there is a significant worldwide gap between the number of patients who need screening and the number of available trained ophthalmologists.

— **The Specialist Bottleneck :** In most developing countries as well as rural areas, the ratio of ophthalmologists to the population is dangerously low.

— **Scalability Issues :** Retinal images manual screening takes several hours. A human expert can only review a limited number of images per day ; hence, large-scale population screening programs will be quite economically and logistically infeasible without robotic assistance.

2.5.3 Image Quality and Artifact Interference

The diagnostic performance of deep learning systems is highly sensitive to the quality of input data. Ophthalmic images frequently suffer from uneven illumination, motion artifacts, speckle noise (in OCT), and reflective glare. These artifacts obscure critical anatomical boundaries, such as the optic disc margin and retinal layers, and blur small but clinically significant lesions, such as microaneurysms and drusen [41]. As a result, models trained on clean datasets may perform poorly in real clinical settings, leading to reduced accuracy and missed diagnoses.

2.5.4 Screening Gaps for High-Risk Populations

Global statistics on blindness and visual impairment from diseases like diabetic retinopathy and glaucoma do not fully reflect the real impact of these conditions. Many people have early-stage disease that still requires monitoring and treatment. High-risk groups such as premature infants, older African Americans at risk for glaucoma, and millions of individuals with diabetic retinopathy often do not have access to proper screening programs, leading to late diagnosis and complications [20]. Improving early detection and expanding access to regular eye screening could significantly reduce preventable vision loss worldwide.

Collectively, these challenges highlight a systemic gap between the theoretical need for early detection and the practical capacity of current healthcare infrastructures to deliver it consistently and equitably.

2.6 Related work

In this section, we review and discuss several related works that have significantly advanced ocular image segmentation and disease classification. In addition, this section provides some foundation on which our proposed approach builds.

— **Hybrid CNN–Random Forest for Multi–Class Eye Disease De-**

tection

Mittal et al. (2025) [42] introduced a hybrid CNN–Random Forest model for multi-class eye disease detection. The authors, in their method, cover conditions such as cataract, glaucoma, and uveitis, and report an impressive accuracy of 98.13% on a Kaggle dataset. In addition they highlights the promise of hybrid approaches ; however, it leaves several gaps, including the absence of external clinical validation, lack of attention mechanisms (such as SE or CBAM), limited comparisons with modern architectures like EfficientNet, MobileNetV2, and ViT, no cross-validation, minimal hyperparameter reporting, and no interpretability analysis. Building on this foundation, the present thesis seeks to address these limitations by incorporating attention-integrated lightweight CNNs, validating across multiple datasets, and embedding explainable AI techniques to enhance both performance and clinical relevance.

— **Fusion–Attention Diagnosis Network (FADNet) : Joint Segmentation and Classification**

Xiao et al. (2025) [43] proposed FADNet (Fusion-Attention Diagnosis Network), an end-to-end framework that unifies optic disc segmentation and ocular disease classification, addressing the missed synergy in prior independent approaches. The model employs an attention U-Net enhanced with CBAM for segmentation, integrates outputs via dynamic weighted feature fusion, and uses a ResNet-34 backbone for classification under a multi-task loss (Dice + cross-entropy). It achieved state-of-the-art accuracy of 90.4% on a multi-class retina dataset, outperforming VGG-16, ResNet-50, DenseNet-121, Inception-V3, and EfficientNet. Ablation studies confirmed the critical role of fusion, CBAM, and joint optimization,

with accuracy drops when components were removed. The framework tackles challenges of task independence, feature overlap, and cross-dataset generalizability, though limitations remain in dataset size, domain shifts, and real-world deployment.

— **Deep Convolutional Ensemble for Retinal Disease Classification**

Pandey et al. [44] developed a deep convolutional ensemble (DCE) model to classify retinal fundus images as diabetic retinopathy, glaucoma, age-related macular degeneration, or normal, addressing the challenge of detecting multiple pathologies simultaneously. Their approach combined five InceptionV3 networks trained on bootstrap-aggregated samples with weighted cross-entropy loss, while final predictions were generated through majority voting. Trained on more than 43,000 images from 12 public datasets, the DCE achieved 79.2% accuracy and a 79.9% F1-score on a balanced test set of 100 images, significantly outperforming seven board-certified ophthalmologists, particularly in detecting mild diabetic retinopathy cases often incorrectly classified as normal. Although the model demonstrated strong calibration and reliability, several limitations remained, including a small test set, inconsistent ground-truth labels across datasets, and potential bias due to overlapping data distributions. This study highlights the potential of ensemble deep learning for multi-disease retinal screening, particularly in population-level applications within resource-limited settings, while also emphasizing the persistent challenge of label inconsistency in medical imaging AI.

2.7 Conclusion

This chapter has discussed the intricacies of eye diseases, looking in, depth at the crucial contribution of the rear region of the eye, mainly the retina and optic nerve, to the overall functionality of the human eye. We went through the major eye disorders and saw how they disclose their markers visually in different ways.

Chapitre 3

AI techniques applied for ocular problems

3.1 Introduction

With advanced technology in medical machine learning, we are seeing great strides in the interpretation and analysis of medical images. The use of detailed diagnostic imaging is especially prevalent in optometry, providing ideal conditions for intelligent diagnostic tools. Many conditions affecting the eye create distinct structural changes that can be captured by retinal imaging and cross-sectional scans, enabling algorithmic pattern recognition and increasingly accurate clinical decisions.

In this chapter, we introduce the technical foundation by discussing the shift from traditional imaging modalities to new ones that use Deep Learning in medical imaging. This chapter explains the underlying mathematics of how Artificial Neural Networks operate, the ability of CNNs to extract spatial features, and, finally, how Transfer Learning can help mitigate insufficient data for training. Complementing this architecture with an analysis of today's market gaps enables us to develop a framework that supports and justifies our proposed diagnostic solution.

3.2 Machine Learning and Deep Learning in Medical Imaging

3.2.1 Theoretical Foundations

The automated analysis of ophthalmic medical imagery represents a key area at the intersection of clinical medicine and computer science. To create the foundational of the technique employed in this research, it is compulsory to define the anatomy of relevant computational paradigms : Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

Artificial Intelligence refers to the capability of a machine or computer system to perform tasks typically associated with human intelligence. These functions involve perceiving, reasoning, learning, dealing with the environment, issue-solving, decision-making, and even proving ingenuity, Actually ML is a part of AI, which represented by algorithms that enhance their effectiveness on a given function among expertise without being specifically programmed [45]. While Deep learning is a machine learning concept based on artificial neural networks with multiple hidden layers (deep neural networks) that can automatically learn feature representations from raw, high-dimensional data, allowing them to outperform shallow machine learning models and traditional data analysis approaches in many applications [46].

Deep learning excludes the require of human-based feature engineering by automatically learning hierarchical representations from raw retinal images. Early network layers detect basic features such as edges and color gradients, while deeper layers combine them to recognize complex patterns, including textures, vascular structures, and clinically relevant pathological biomarkers for ocular disease detection [47].

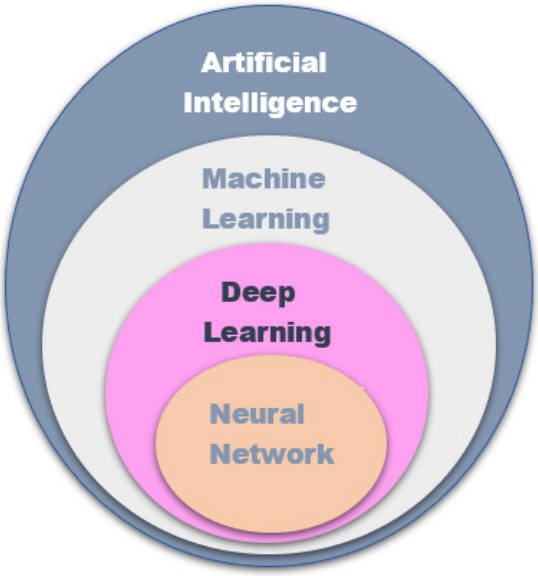


FIGURE 3.1 – AI vs ML vs DL and NN.

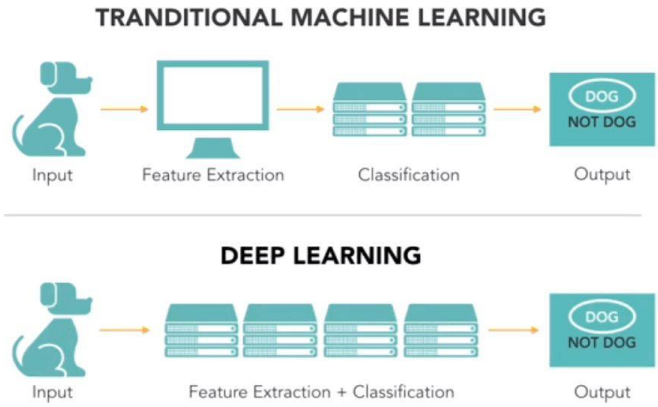


FIGURE 3.2 – Machine Learning and Deep Learning.

The ability to automatically identify complex representations required to classify or detect features from pixel values is one of the main reasons deep learning is now considered the most advanced approach for analyzing ocular images.

3.3 Manual and Traditional Computer–Vision–Based Diagnostic Methods

3.3.1 Manual grading

Before the prevalance of deep learning, there were two primary models for analyzing medical eye images : manual expert assessment and traditional computer vision techniques. The main explanations of those two approaches and their respective limitations are extremely helpful for understanding how modern AI methods have truly changed the game.

The study used manual grading by expert readers at the Wisconsin Fundus Photograph Reading Center as the gold standard to validate the AI system [16]. Independent graders analyzed fundus photographs and OCT images using standardized criteria, with final decisions based on majority voting. All graders were masked to avoid bias. This enabled evaluation of the AI system's performance against expert consensus.

Aspect	**Description**
What is manual grading ?	Human experts visually interpreting retinal images to diagnose disease severity
Who performed manual grading ?	Wisconsin Fundus Photograph Reading Center (FPRC) — a centralized, certified reading center
Number of graders per image	Three experienced readers
Decision method	Majority voting paradigm
What was graded ?	1. ETDRS severity scale from widefield stereoscopic fundus photos 2. Diabetic macular edema (DME) from OCT scans
Masking procedures	Graders were masked to AI outputs, fundus photo results when reading OCT, and OCT results when reading fundus photos
Role in the study	Served as the gold standard reference to validate the AI system
Why used ?	To determine the truth of whether a patient had mtmDR
Comparison with AI	AI's sensitivity (87.2%) and specificity (90.7%) were calculated against these manual grades

TABLE 3.1 – Manual grading process in retinal disease studies [16].

3.3.2 Traditional Computer Vision

Traditional computer-vision methods for grading ocular disease historically con- sists of two main components, initial image preprocessing followed by the extraction of engineered descriptors, and concluding with classification via conventional machine learning algorithms.

3.3.2.1 Pre-processing and Enhancement

Common techniques included :

— **Illumination Normalization :** Contrast Limited Adaptive Histogram Equalization (CLAHE) was used at most times to normalize the local contrast [48].

— **Morphological Filtering :** Transforms like top-hat were usually applied to highlight the potential lesions from the noise background.

— **Structural Localization :** Optic disc localisation was typically performed using geometric methods such as Template Matching or Hough Transform.

3.3.2.2 Feature Extraction

Researchers have relied on domain-specific descriptions to obtain quantitative representations of pathological modifications before the hierarchical feature learning process associated with deep learning.

Feature Category	**Primary Descriptors**	**Diagnostic Application**
Texture Analysis	GLCM	The Gray-Level Co-occurrence Matrix (GLCM) method quantifies texture by examining the frequency of pixel pairs with the same gray level that correspond to specific spatial characteristics [49].
Structural Patterns	LBP	The Local Binary Pattern (LBP) operator assigns a label to each pixel in an image based on thresholding its eight neighboring pixels relative to the gray scale of the central pixel [50].
Shape and Gradient	HOG	The Histogram of Oriented Gradients (HOG) descriptor measures shape and appearance based on local objects by splitting an image into a number of small spatial "cells" and creating a histogram based on the directional gradients for each cell [51].
Keypoint Detection	SIFT and SURF	Characteristics that were scale invariant interest points between images of different resolutions [52].

TABLE 3.2 – Primary descriptors used in computer-vision methods.

3.3.2.3 Classification

The previous descriptors were utilized as input variables in classical machine learning-based models, including :

— **Support-Vector Network :** Support Vector Machines (SVM) build an optimal hyperplane in feature space to increase the margin between classes. By exploiting kernel functions, they generalize efficiently to unseen data, allowing robust classification in high-dimensional spaces [53].

— **Random Forests :** Random forest is a classifier that consists of multiple tree classifiers, where each tree is trained on a random vector sampled independently from an identical distribution. Every tree votes for the most prevalent class, and the forest chooses the class with more votes [54].

— **K-nearest neighbor :** K-nearest neighbor (k-NN) is a type of supervised learning that assigns to an unknown sample, the majority class of its k closest known samples pulled from a training set. It is based on the assumption that similar samples lie in the same class. The class most com-

mon among a sample's k neighbors is assigned to the sample. The distance function and k value are critically important that matters [55].

3.4 Artificial Neural Networks (ANN)

The foundational architecture of deep learning is the ANN. According to [56], an ANN is a computational model based on biological neurons ; it consists of numerous interconnected artificial neurons that communicate with one another via weighted connections. The biological incentive concludes from the structure of neurons, which consists of dendrites (receiving inputs), a cell body (processing inputs), and an axon (transmitting the output signal to subsequent neurons).

3.4.1 The Mathematical Neuron Model

Artificial Neural Networks (ANNs) consist of artificial neurons that work together to process input by means of using weighted connections and a bias term, as follows :

$$\mathbf{x} = [x_1, x_2, \ldots, x_n]^T,$$

The output of a neuron is calculated as [57] :

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) = f\left(\mathbf{w}^T\mathbf{x} + b\right)$$

Where w_i represents the weights, b denotes the bias, and $f(\cdot)$ represents the activation function, which adds non-linearity to this overall processing.

3.4.2 Network Architecture

The structure of an ANN as described in [58] :

— **Input Layer :** It takes the raw data in the shape of a one-dimensional vector of image pixels, which is a drawback for its use in ophthalmic image processing

as it results in loss of the spatial topology of the data.

— **Hidden Layers :** Are the layers where the network extracts deep features, and the number of hidden layers and the number of neurons in each hidden layer decide the power of the network in learning complicated functions.

— **Output Layer :** Makes the final decision ; for example, it uses 1 neuron for binary classification and multiple neurons for multi-class classification, such as retinopathy grading

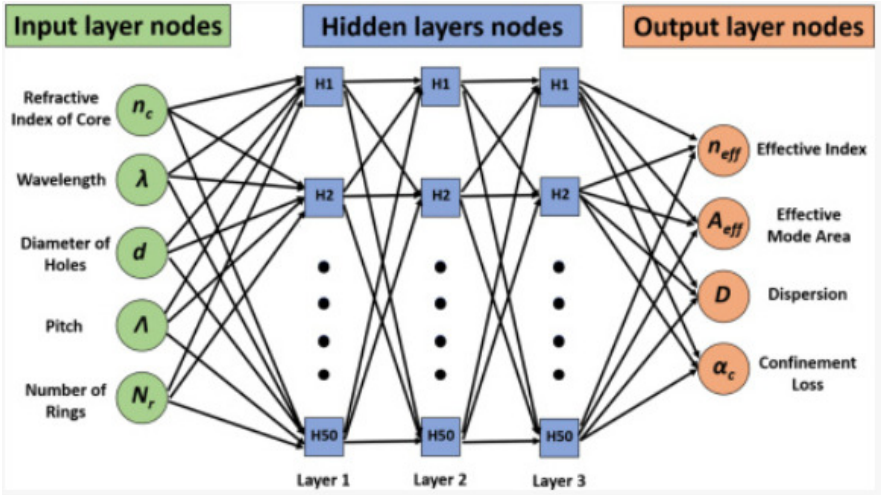


FIGURE 3.3 – The structure of ANN model [11].

3.4.3 Activation Functions

Activation functions alowing non—linearity, make it capture complex relationships. Typical functions are [58] :

— **Rectified Linear Unit (ReLU) :** Is Defined as

$$f(x) = \max(0, x)$$

RELU is popular in hidden layers for its low computational cost and helps mitigate vanishing gradient issues.

— **Sigmoid :** Converts input to the range of problems like disease detection. range, ideal for binary classification problems like disease detection.

— **Softmax :** It is mainly applied in the output layer in multi-classification. It normalizes the output to a probability distribution that uqual to 1.

3.4.4 Forward Propagation, Backpropagation, and Loss Functions

Training an ANN includes two steps : Forward Propagation, in which data is passed through the network to generate predictions, and Backpropagation, in which an error is updated using a Loss Function via a process called Gradient Descent to adjust the network's weights. The use of this cyclical technique results in lower variance between the ANN's predictions and the true labels in the clinical data [59].

3.5 Convolutional Neural Networks (CNN)

ANNs can be considered mathematically strong, but they ignore the spatial arrangement of pixels and have so many parameters to train that they are poor performers on images. CNNs were developed specifically to address these shortcomings [60]. A convolutional neural network (CNN) is a type of deep learning model whose design was inspired by how humans and animals perceive the world through their eyes, particularly how the visual cortex detects visible light within its receptive fields [61].

3.5.1 Core CNN Architecture

A CNN is consist of diverses layers, mostly including convolutional layers, pooling layers, and fully-connected layers.

— **Convolutional layer :** Apply small filters to the input image. These filters slide across the image to identify local patterns such as edges, textures, and more complex features. Using very small filters allows the network to be deeper while keeping the number of parameters relatively low [62].

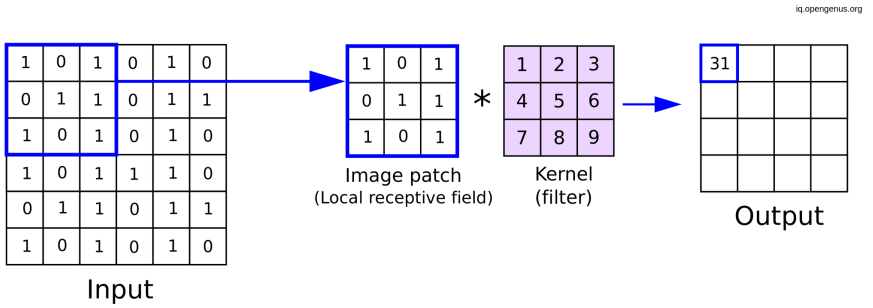


FIGURE 3.4 – Convolutional layer [12].

— **Pooling layer :** The pooling layer, also known as subsampling, is used to minimize the size of feature maps, enabling most important information. Max pooling, a widely used method, takes the maximum value within a region, while average pooling computes the average. This layer, by reducing data dimensionality, diminishes both computational complexity and the number of network parameters. Overall, it makes overfitting a less likely scenario and encourages better model generalization [63]. As described in Figure 3.5 below :

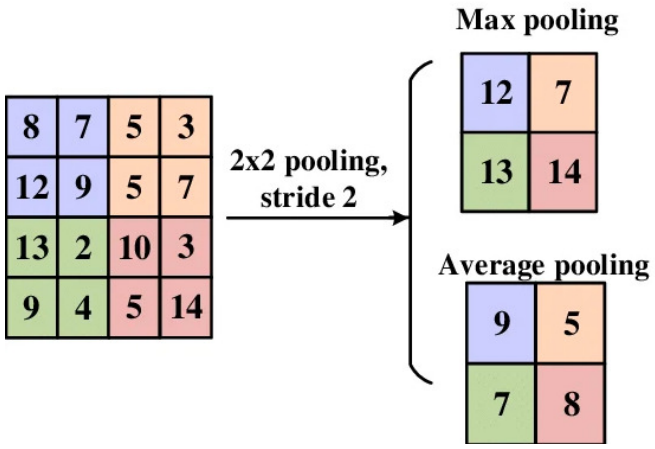


FIGURE 3.5 – Pooling layer [12].

— **Fully-connected layers :** The fully connected layer, or dense layer, has a set of characteristics, including complete connectivity between its neurons and the neurons of adjacent layers. This design allows each neuron to integrate inputs from all previous activation stages, enhancing the networkâs ability to analyze and synthesize extracted features for both classification and prediction tasks

[13].

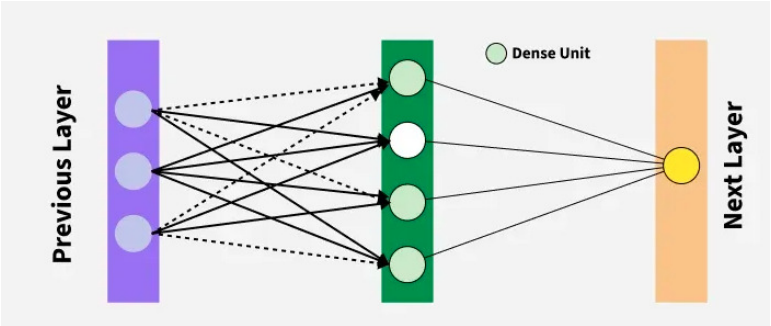


FIGURE 3.6 – Fully-connected layers [13].

3.5.2 Popular CNN Architectures

The development of convolutional neural networks has been driven by improved feature extraction, reduced computational cost, and the resolution of problems related to deep network training. Deep learning relies on several important architectures.

— **VGG Architecture :** The VGG model was designed by Karen Simonyan and Andrew Zisserman in 2014 [64], Later studies demonstrated that increasing network depth could improve image classification performance. Models such as VGG-16 and VGG-19 enabled deeper architectures by using only small convolution filters of size 3×3, which achieved similar feature extraction and classification performance to larger filters such as 5×5 and 7×7. This design reduced the number of parameters, increased the use of non-linear activation functions, and improved transfer learning capabilities. Despite their strong performance and transferability, VGG models have several drawbacks, including high computational cost, large memory requirements, and a significant number of parameters, particularly in the fully connected layers.

— **ResNet Architecture :** Is a deep neural network architecture. It relies on residual learning to facilitate deep network training. Instead of directly learning the mapping function $H(x)$, the layers learn the residual function $F(x) = H(x) - x$, and the final output is reconstructed as $F(x) + x$. This

is achieved using bypass connections that transfer inputs directly to outputs without additional parameters or computational costs. ResNet enables the creation of very deep models, such as 152 layers or even more than 1000 layers, to avoid performance degradation and achieve higher accuracy [65].

The next figure represents a block of residues, in which the identity shortcut x is applied to the residue map $\mathcal{F}(x)$ to produce $\mathcal{F}(x) + x$ before final activation by the ReLU function.

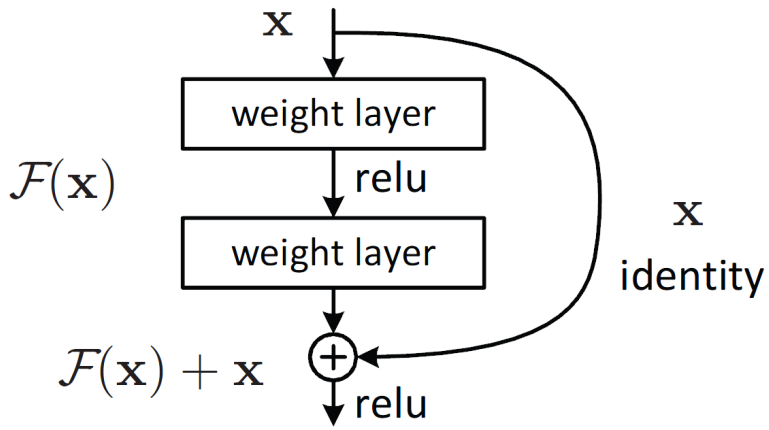


FIGURE 3.7 – Example of Residual CNNs [12].

— **DenseNet Architecture :** DenseNet is a convolutional neural network architecture in which each layer takes input from all preceding layers and passes its feature maps to all subsequent layers (Figure 3.8). This results in a network with $\frac{L(L+1)}{2}$ direct connections for an L-layer network, instead of the usual L connections found in conventional architectures. Unlike ResNets, which pool features in totality, DenseNet joins features via concatenation, enabling any layer to access the combined "knowledge" of the preceding layers, without summing and potentially removing usable information [66].

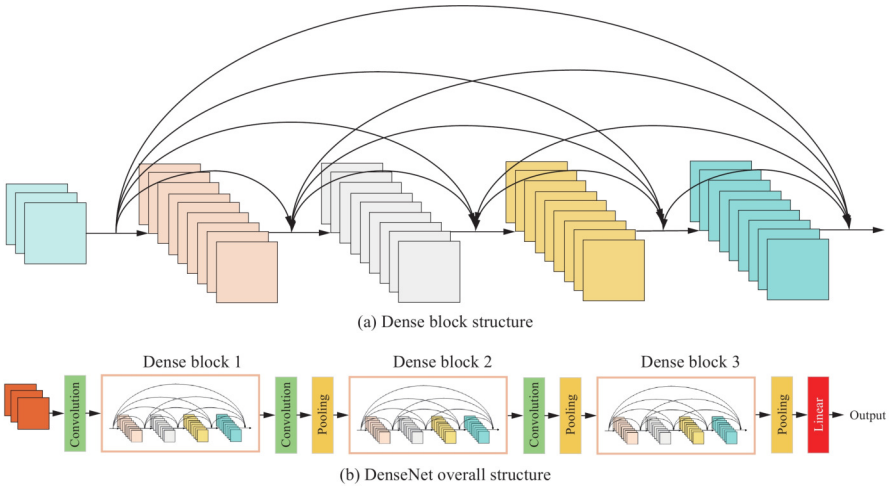


FIGURE 3.8 – The dense block structure and the overall DenseNet architecture [14].

— **MobileNet Architecture :** MobileNet is an efficient CNN model mainly designed using depthwise separable convolutions. These convolutions split the standard convolution operation into two separate parts : a depthwise layer that applies a filter to each input channel independently, and a pointwise layer that uses 1×1 convolutions to combine the outputs. As a result, computation cost and model size are significantly reduced while maintaining high accuracy [67].

3.6 Transfer Learning

Transfer learning improves the performance of a target predictive function $f_T(\cdot)$ by reusing knowledge learned from a source domain $\mathcal{D}_S$ and a source task $\mathcal{T}_S$. In general, a domain $\mathcal{D}$ is described by a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, \ldots, x_n\} \in \mathcal{X}$. For a given domain, a task $\mathcal{T}$ is characterized by a label space $\mathcal{Y}$ and a predictive function $f(\cdot)$, which can be represented probabilistically as the conditional distribution $P(Y \mid X)$. Unlike traditional machine learning, which assumes that source and target data come from the same distribution ($\mathcal{D}_S = \mathcal{D}_T$ and $\mathcal{T}_S = \mathcal{T}_T$), **transfer learning** is used when $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$. These differences usually appear in one or both of the following cases [68] :

— **domain discrepancy**, where the feature spaces differ ($\mathcal{X}_S \neq \mathcal{X}_T$) or the marginal distributions are different ($P(X_S) \neq P(X_T)$).

— **task discrepancy**, where the label spaces differ ($\mathcal{Y}_S \neq \mathcal{Y}_T$) or the conditional distributions are not the same ($P(Y_S \mid X_S) \neq P(Y_T \mid X_T)$).

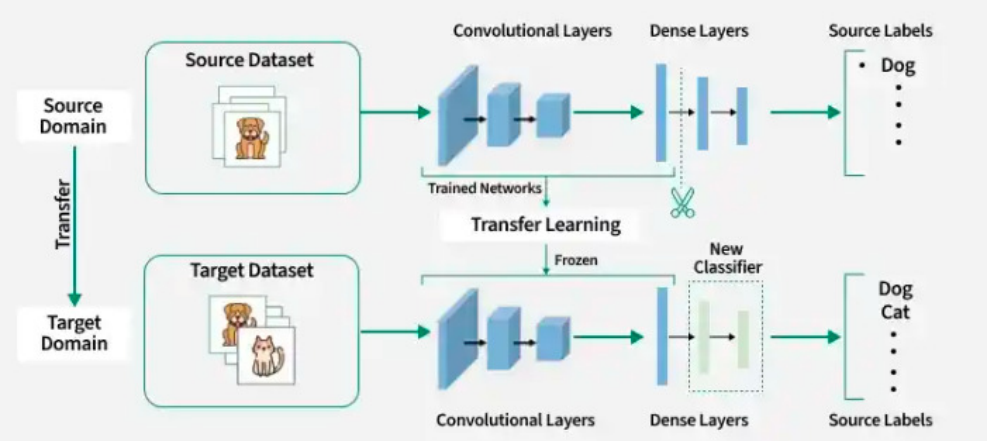


FIGURE 3.9 – Technique of Transfer Learning [15].

3.6.1 Approaches to Transfer Learning

— **Features Extraction :** Feature extraction in transfer learning involves extracting useful data features from various depth levels in a pre-trained model using fixed earlier layers for general patterns and adaptive later layers for field-specific details and then concatenating them across a projection head to improve performance on relevant but distinct tasks, like adapting a ResNet50 model trained on ImageNet for medical image classification [69].

— **Fine-tuning :** Fine-tuning involves taking a pre-trained network and adjusting its parameters for use with a target dataset and newly added classifier layers. In fine-tuning, some layers of the original pre-trained base network are selectively unfrozen and trained along with the new classifier layers. This allows us to refine both the added classification layers and parts of the original base network, enabling the model to better fit the characteristics of the target task [70].

3.7 Conclusion

This chapter serves as the technical framework for this research by reviewing the development and evolution of deep learning in medical imaging. The chapter discusses the transition from simple ANNs and the mechanics of the mathematical neuron to the more sophisticated CNNs, which are designed to extract spatial features. It examines fundamental concepts such as backpropagation and activation functions, as well as practical applications such as transfer learning, to address potential data scarcity and develop the theoretical basis and architecture required to implement and analyze the diagnostic models presented in future sections.

System Design

3.8 Introduction

This chapter describes the overall system design of the proposed multi-modal ocular disease detection framework. Specifically, the primary objective is to detail the end-to-end architecture from raw clinical image ingestion to final diagnostic output. Despite differences among various ophthalmic imaging modalities, as a result of distinctively different physical and morphological characteristics, there are distinct â or bespoke â data ingestion pipelines and image processing methodologies (pre-processing protocols) for all three ophthalmic imaging modalities (slit-lamp, fundus, Optical Coherence Tomography [OCT]) developed within the ODD framework.

In addition, the use of deep learning models designed for pathology classification associated with each imaging modality is rigorously described throughout this section. Finally, the chapter concludes with a discussion of integrating these separate, modality-specific computational modules into a cohesive, single, observable, and fault-tolerant diagnostic application.

3.9 Overview of the Proposed Multi-Modal System Architecture

3.9.1 System block diagram

The high-level operational flow of the proposed multi-modal digital diagnostic system is described in this Fugire, which brings the individual modules' functional breakdown. The architecture is conceptually divided into three distinct sequential phases : input ingestion, parallelized modality-specific inference, and result aggregation. The inference system begins at the top of the flow of ingestion module. Images acquired from a slit-lamp, fundus camera, or OCT scanner are fed into the system via this ingestion module. The image is then transferred to a modality classification module, which routes it, once identified, to its dedicated image-specific deep learning classifier. The image-specific classifier consists of this pipeline : an image-specific preprocessing block and a dedicated CNN classifier. For example, the model used for slit-lamp images is binary, whereas the models used for fundus and OCT images are multi-class CNNs. After the image propagates through the classifier, individual class probabilities, as well as class id labels, are calculated. The individual probabilities and labels are then collected by the system and transmitted to the go-to operation interface (GUI).

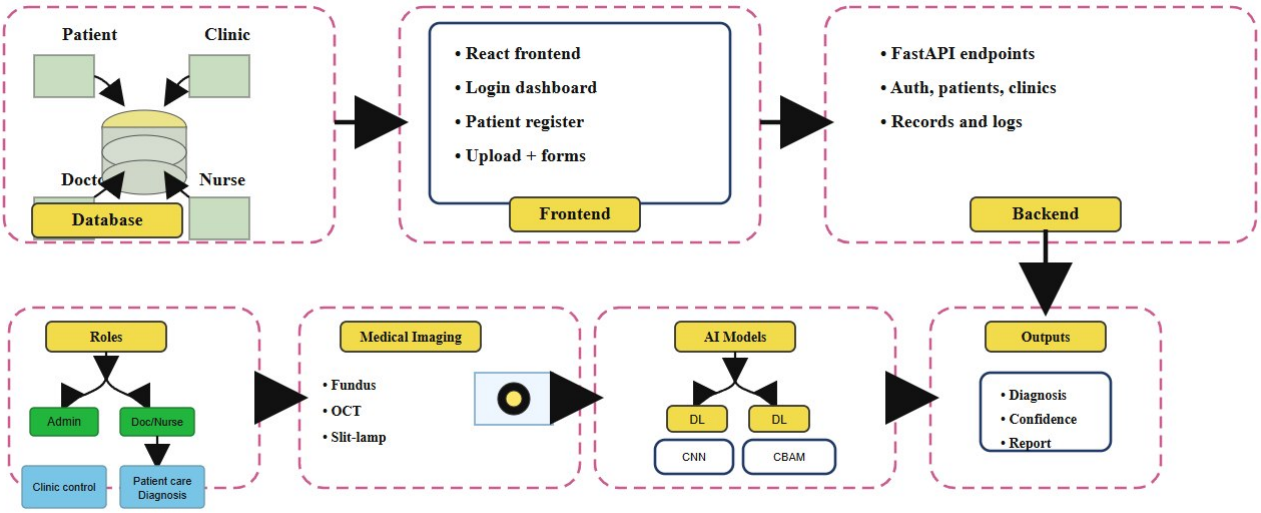


FIGURE 3.10 – The proposed architecture of our system.

3.9.2 Why Separate Models per Modality

Given the intrinsic differences among slit-lamp, fundus, and OCT imaging, this study recommends a modular, modality-specific deep learning framework rather than an end-to-end model. The three optical imaging modalities differ in how they capture anatomical structures of the eye. One network is prone to feature interference, in which modality-specific cues may confuse the model and reduce diagnostic performance. By separating the system into specialized modules, each network can focus on its own domain, improving sensitivity to unique biomarkers, while also enhancing scalability and clinical practicality since modules can be independently updated or replaced without disrupting the overall pipeline.

3.10 Dataset Description

The effectiveness of any deep learning model is highly dependent on its training set. To this end, we created three separate datasets, each representing one

imaging modality and the disease types associated with it.

3.10.1 Optical Coherence Tomography Dataset

The OCT dataset comprises cross-sectional retinal images with five pathological conditions :

Attribute	**Details**
Dataset Modality	Optical Coherence Tomography(OCT) Images
Source of Images	Public dataset collected from Kaggle [71] [34]
Number Images perclass	3000
Total Images	15000
Train / Validation / Test Split	70% / 15% / 15%
Image Format	JPG / JPEG / PNG
Annotation Procedure	Dataset provided class labels (diabetic macular edema (DME)/diabetic retinopathy (DR)/DRUSEN/macular hole (MH)/NORMAL)

TABLE 3.3 – OCT Dataset Description

3.10.2 Fundus Photography Dataset

The fundus dataset consists of retinal images captured using fundus cameras, classified into four categories :

Attribute	**Details**
Dataset Modality	Fundus Photography Images
Source of Images	Public dataset collected from Kaggle [72]
Number Images perclass	1007
Total Images	4028
Image Format	JPG / JPEG / PNG
Annotation Procedure	Dataset provided class labels (diabetic retinopathy, glaucoma, cataract, normal)

TABLE 3.4 – The Fundus Dataset Description

3.10.3 Slit–Lamp Dataset

The slit-lamp dataset is a binary classification dataset for cataract detection :

Attribute	**Details**
Dataset Modality	Slit-Lamp Images
Source of Images	Public dataset collected from Kaggle [73]
Normal Images	306
Cataract Images	306
Total Images	612
Image Format	JPG / JPEG / PNG
Annotation Procedure	Dataset-provided class labels (Normal/Cataract)

TABLE 3.5 – Slit-Lamp Dataset Description

3.10.4 Modality Classification Dataset

To construct the modality classifier dataset, 375 images were selected from each of the three imaging modalities, Slit-Lamp Photography, Fundus Photography, and Optical Coherence Tomography (OCT), resulting in a balanced dataset of 1125 images in total. This equal distribution across modalities ensures that the classifier is not biased toward any particular imaging type during training.

3.11 Preprocessing Pipelines

3.11.1 Image Resizing

All input images, regardless of their original dimensions, are resized to a fixed spatial resolution of 224×224 pixels prior to any further processing. This standardization is necessary for two reasons. First, convolutional neural networks require fixed-size input tensors ; without standardization, images captured under different acquisition conditions â which may differ substantially in resolution â cannot be batched together during training. Second, 224×224 is the canonical input resolution for ImageNet-pretrained architectures, meaning that pretrained feature representations are maximally compatible with this resolution.
Bilinear interpolation is used for downsampling, and bicubic interpolation for upsampling, to preserve fine structural detail, such as vascular patterns in fundus images and layer boundaries in OCT scans.

3.11.2 Contrast-Limited Adaptive Histogram Equalization (CLAHE)

Fundus and OCT images are inherently susceptible to non-uniform illumination, lens artefacts, and acquisition variability, all of which depress the local contrast in diagnostically critical regions. To address this, Contrast-Limited Adaptive Histogram Equalization (CLAHE) is applied as a dedicated preprocessing step for the fundus and OCT branches.

CLAHE improves upon standard global histogram equalization by the way it equalizes images. Global histogram equalization redistributes the pixel intensities of an entire image, whereas CLAHE creates a grid of non-overlapping rectangular tiles, each processed independently ; histogram equalization is performed within each tile. A clip limit for the histogram of each tile is determined before equalization to prevent overamplification of noise in low-spatial-frequency regions.

Bilinear interpolation is then used to blend adjacent tiles so that there are no visible tile boundaries in the resulting image [74].

In this system, CLAHE is applied in the LAB color space to enhance contrast without distorting diagnostically important color cues. Only the luminance (L) channel is equalized, preserving chrominance details such as the reddish tones of hemorrhages in fundus images. The parameters were carefully tuned, with clip limits of 2.5 for the fundus and 2.0 for the OCT and a fixed 8×8 tile grid, based on visual inspection to balance enhancement quality across representative samples.

3.11.3 Pixel Intensity Normalisation

Normalizations are performed after resizing and CLAHE to maintain consistent values across all inputs. Each input image is standardized to zero mean and unit standard deviation using the ImageNet statistics : mean $= [0.485, 0.456, 0.406]$ and standard deviation $= [0.229, 0.224, 0.225]$ for the Red, Green, and Blue channels, respectively. This ensures that pixel distributions remain aligned with those used during backbone network training, promoting stable convergence during fine-tuning and reducing issues related to gradient instability.

3.12 Data Augmentation

The data augmentation is solely for training purposes, to artificially increase the relative size of the training dataset and to reduce overfitting. The augmentations are applied randomly ; as such, each time, the exact same image will be transformed differently at each training epoch. During validation and testing, however, no augmentation occurs, as the evaluation needs to determine how well the model performs when evaluated on real data (without using augmentations).

3.13 Dataset Splitting Strategy

3.13.1 Held-out Test Set

Before training or cross-validation, a stratified test set is held out from each branch to preserve class balance. Fundus uses an 85/15 split, while OCT and slit-lamp use pre-partitioned test folders.

This practice prevents optimistic bias by ensuring that the test set is never used, directly or indirectly, during model selection or tuning, so that the reported metrics provide a credible estimate of real-world clinical performance.

3.13.2 Stratified K-Fold Cross-Validation

For all branches, the non-test data (85% for fundus, and the merged training-validation sets for OCT and slit-lamp) undergoes stratified 5-fold cross-validation, where each fold serves as the validation set once across five iterations. The test sets remain completely separate and are only used for final evaluation.

3.14 Convolutional Block Attention Module (CBAM)

CBAM is an end-to-end trainable module that takes an intermediate feature map as input. It first generates a 1D channel attention map to focus on what is meaningful, then generates a 2D spatial attention map to focus on âwhereâ the informative parts are. These attention maps are then multiplied element-wise with the input feature map to produce adaptively refined output features [75]. CBAM modules are integrated into MobileNetV2 and DenseNet121 to refine feature maps through channel and spatial attention. This allows the models to suppress irrelevant activations and highlight diagnostically important retinal regions, improving both classification accuracy and interpretability.

3.15 Training Phase of the Proposed Deep Learning Architectures

3.15.1 Fundus Branch MobileNetV2 with CBAM

The proposed model architecture is based on MobileNetV2 as the pretrained backbone. To enhance feature representation, CBAM (Convolutional Block Attention Module) modules are integrated after the 13th and 18th layers. The CBAM inserted after the 13th layer focuses on improving mid-level feature extraction, particularly vessel structures and optic disc representations, while the module after the 18th layer enhances high-level semantic features.

The classifier head consists of dropout-regularized fully connected layers with architecture $1280 \rightarrow 512 \rightarrow 256 \rightarrow 4$. Each linear layer is passed through *batch normalization* and ReLU activation to improve training stability and nonlinear feature learning.

A transfer learning strategy is adopted by freezing the first 13 layers of the pretrained MobileNetV2 backbone, fine-tuning layers from the 14th to the 18th, and training the classifier head from scratch.

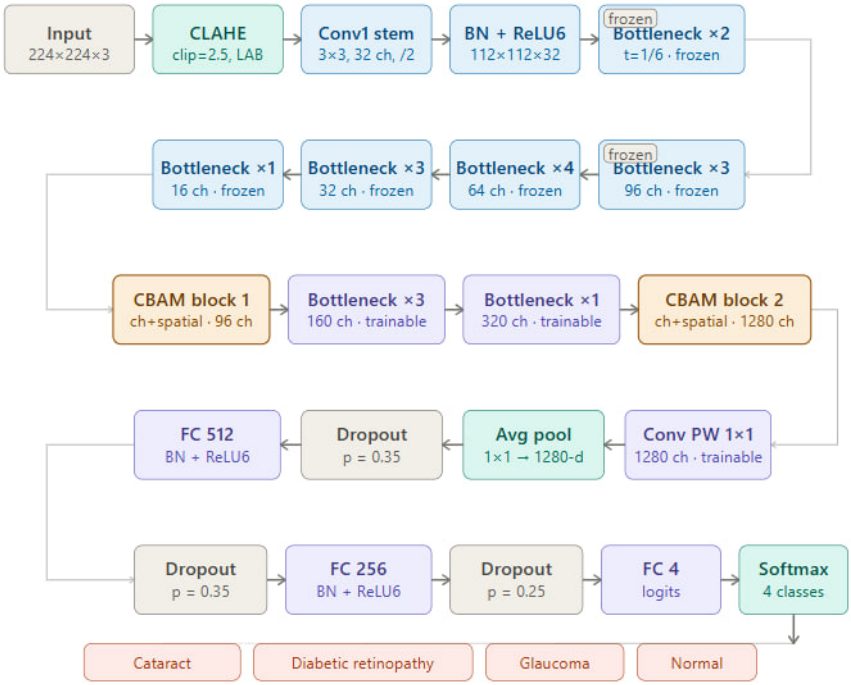


FIGURE 3.11 – The proposed Fundus Photography model of our system.

3.15.2 OCT Branch DenseNet121 with CBAM

This model is based on DenseNet121 pretrained on ImageNet-1K. To improve feature representation, a CBAM is applied after the final 1024-channel feature-extraction layer, thereby enabling enhanced attention to multi-scale retinal structures.

The classifier head consists of a three-stage dropout-regularized fully connected pipeline with architecture $1024 \to 512 \to 256 \to 5$, where each linear layer is followed by *batch normalization* and a *ReLU* activation functions.

A transfer learning strategy is adopted by freezing denseblock1–3, selectively fine-tuning denseblock4 and norm5, and training the classifier head from scratch.

The dense connectivity mechanism of DenseNet121 promotes feature reuse and efficient gradient flow, making the architecture particularly suitable for OCT image classification, where integrating retinal layer boundaries across multiple scales is essential.

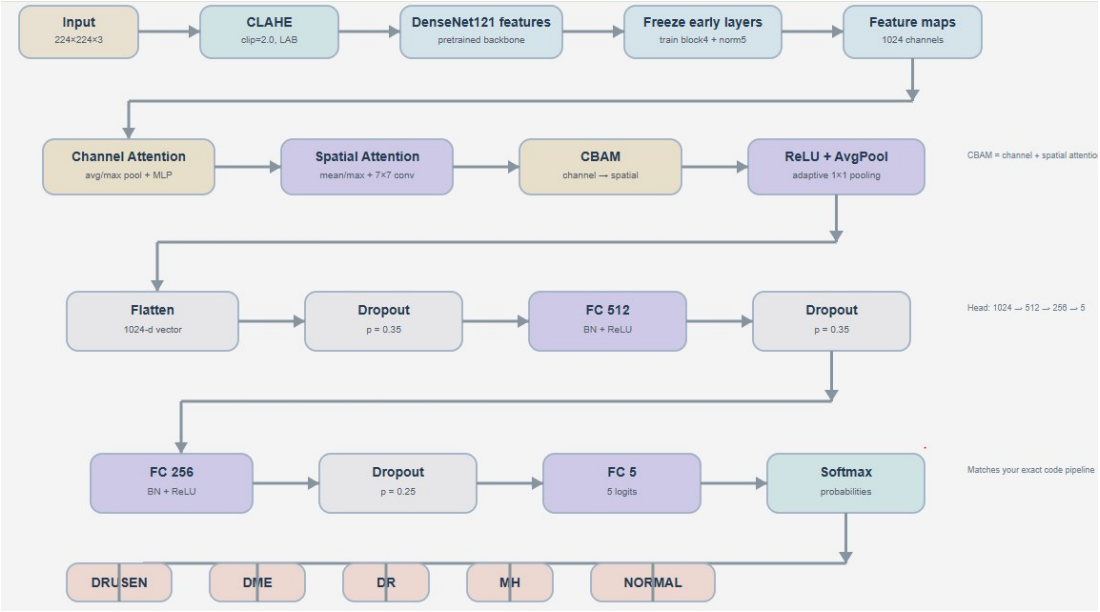


FIGURE 3.12 – The proposed OCT model of our system.

3.15.3 Slit-Lamp Branch MobileNetV2

This model uses MobileNetV2 pretrained on ImageNet-1K with a lightweight classifier head tailored for binary cataract detection. The head applies dropout regularization and two linear layers $1028 \to 100 \to 2$ with ReLU activation, balancing representational capacity and efficiency. By reducing the hidden dimension to 100 neurons, the design minimizes the risk of overfitting on the relatively small slitâlamp dataset while retaining sufficient discriminative power for the cataract-vs.-normal classification task.

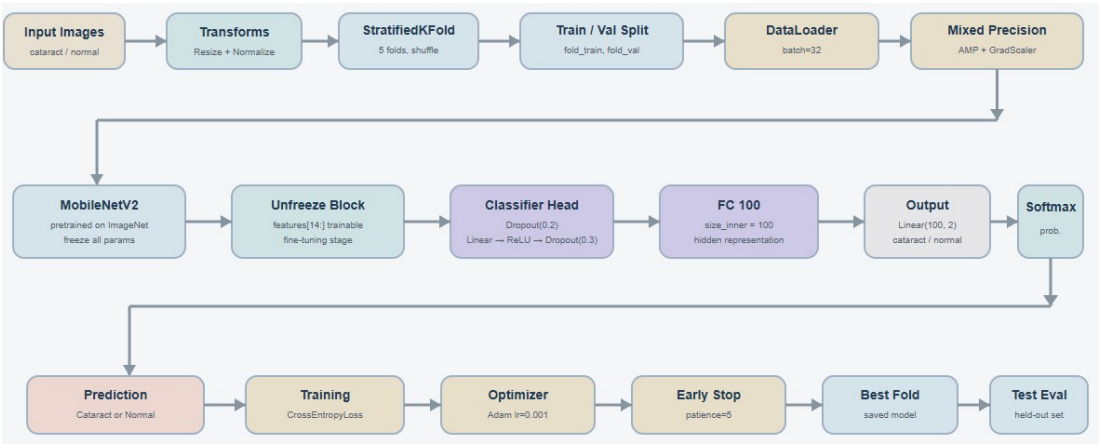


FIGURE 3.13 – The proposed Slit L model of our system.

3.15.4 MobileNetV2 for Modality Classification

A core component of the proposed system is an automated modality classifier that identifies the modality of the provided eye image before routing it through the appropriate pathway for detecting a specific disease. This avoids human intervention by allowing the system to classify among the three ocular imaging modalities :

— **Slit-Lamp Imaging**

— **Fundus Imaging**

— **Optical Coherence Tomography (OCT) Imaging**

This three-class classifier was built as a standalone DL model through training on images obtained from all three modalities. The network architecture has been designed following the same transfer learning approach applied in developing the disease classifiers, using MobileNetV2 as the feature extractor owing to its lightweight nature and high inference speed, which serves as a preprocessing pipeline.

3.16 Explainability : Gradient-weighted Class Activation Mapping

Clinical deployment of automated diagnostic systems demands that model predictions be accompanied by visual justifications that a clinician can examine and validate. . To satisfy this requirement, all three pipelines incorporate Gradient-weighted Class Activation Mapping (Grad-CAM) as a post-hoc explainability module.

Grad-CAM is a gradient-based technique for visualizing the reasoning of a CNN in image classification problems. In short, the gradients are computed between a target class logit (the score before the softmax function) and the feature maps from a convolutional layer of choice. The gradients are globally averaged to compute weights indicating how much each feature map contributes to the decision about the target class. The heatmap is produced by applying the weights as multipliers to the feature maps and then taking the result through a ReLU function so that only the feature maps that contribute positively to the network's prediction about the target class are taken into account [76].

3.17 Prediction Phase,

During the prediction phase, the trained deep learning model is deployed to infer outcomes on unseen data, translating learned representations into clinically meaningful decisions. To evaluate the performance of all classification models, several standard metrics were used.

3.17.1 Accuracy

Accuracy measures the percentage of true predictions among all predictions and brings the summary of overall prediction accuracy.

Formula :

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

Where

— **TP** = True Positives

— **TN** = True Negatives

— **FP** = False Positives

— **FN** = False Negatives

3.17.2 Precision

Precision is a measure of how many of the instances labeled as positive are relevant. It is good to use if you want to reduce false positives.

Formula :

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

3.17.3 Recall

Recall, also known as sensitivity, shows how well a model detects positive cases. It represents the fraction of true positive cases successfully detected by the model. This metric is especially useful in situations where missing a positive case is costly, making false negatives a priority.

Formula :

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN}$$

3.17.4 F1-score

The F1-score integrates Precision and Recall into a unified metric by calculating their harmonic mean. It is useful for evaluating a model when equal importance is given to false positives and false negatives, especially in datasets where one class is much more common than the other.

Formula :

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

3.18 Conclusion

The chapter has described the complete conceptual design of the suggested multi-modality ocular disease classification system. Individual design choices for CLAHE, stratified k-fold cross-validation for performance estimation, a single model branch per modality, CBAM-added backbone networks, and Grad-CAM explanations have been detailed, along with explanations for why these choices were made. The implementation, experimental results, and performance metric, along with an analysis of qualitative prediction results across three imaging modalities, will be presented in Chapter 5.

Results and Discussion

3.19 Introduction

This chapter presents the implementation and evaluation of the proposed Intelligent Ocular Diagnosis System Based on Deep Learning Techniques for Image Classification and Treatment of Medical Images. Having established the theoretical background and system design in previous chapters, this chapter focuses on how the proposal was implemented in practice and evaluated.

Furthermore, it describes how the system was implemented to assess performance, including all major aspects of system development and performance assessment, as well as an evaluation of the results. The chapter also provides information on the effectiveness of the proposed system, its limitations, and potential future modifications or improvements.

3.20 Environments, Implementations, and Development Tools

This section outlines the tools used to develop the proposed system. Some of the primary software tools, libraries, frameworks, and hardware resources are listed to illustrate what was required to build and train the models.

3.20.1 Hardware Configuration

The system used is an HP EliteBook 725 G4, equipped with an AMD PRO A12-8830B R7 processor featuring 10 compute cores (4C+6G) and a base clock speed of 2.50 GHz, along with 16 GB of RAM.

3.20.2 Development Tools

For the development of our system, various environments and tools are used to build the system, as shown in the descriptions below :

— **Python :** Is an interpreted, object-oriented, high-level language with dynamic semantics, known for its readability, modularity, and free availability across platforms. Its built-in data structures, dynamic typing, and support for modules make it well-suited for rapid application development and scripting. Developers value its productivity, as the lack of compilation enables a fast edit-test-debug cycle, with errors handled through exceptions and stack traces rather than crashes. Debugging is further supported by a source-level debugger and simple print statements, reflecting Python's introspective power and ease of use [77].



FIGURE 3.14 – Python logo.

— **Anaconda :** Anaconda provides a set of tools that help companies adopt and develop Artificial Intelligence (AI) technologies quickly and securely, including tools to manage Python packages securely and manage secure workspaces (using Anaconda Core) and tools for building and deploying artificial intelligence models using validated software (using AI Catalyst) [78].



FIGURE 3.15 – Anaconda logo.

— **Jupyter :** The Jupyter project is a community-driven, open-source project that offers software, services, and standards to support interactive computing. Users can develop and analyze code and data in more than 40 languages, including Python, R, and Julia [79].



FIGURE 3.16 – Jupyter logo.

— **Google Colab :** Colab is a free, online, cloud-based Jupyter Notebook environment by Google that does not have to be set up on your local machine. It has been designed as an online solution that allows users to write and run Python code, making it a favored tool for many applications, including Machine Learning (ML), Data Analysis (DA), and Education, to mention but a few [80].



FIGURE 3.17 – Google Colab interface.

— **HTML :** HTML, which stands for hypertext markup language (essentially,

hypertext markup), is an authoring tool for producing or redefining both the content and structure of webpages [81].

— **CSS :** The Web's presentation layer is represented by CSS, which enables the formatting of HTML documents. Therefore, when designing and implementing websites, you will use CSS alongside HTML to create the desired look.



FIGURE 3.18 – HTML and CSS concepts.

3.20.3 PostgreSQL

PostgreSQL, originating from a 1977 UC Berkeley project [82], is the world's most advanced open-source Object-Relational Database Management System (ORDBMS), offering enterprise-level features comparable to those of costly commercial products. While it is free and transparentâunlike proprietary systems such as Oracleâorganizations may still incur costs for internal support, research, or consulting. For those who need commercial backing, supported derivatives based on its open-source core, such as the Red Hat Database, are available.



FIGURE 3.19 – PostgreSQL.

3.20.4 JavaScript

JavaScript is a programming language that runs in web browsers to enable dynamic behavior on websites. It works alongside HTML and CSS to handle user interactions, update page content on the fly, and provide features like animations, form validation, and interactive maps without reloading the page [83].



FIGURE 3.20 – JavaScript.

3.20.5 Packages and APIs

This section presents several packages and APIs used in our project.

Packages and APIs	Definition
	TensorFlow is an integrated, open-source machine learning ecosystem consisting of extensive tools, dedicated ML libraries, and community resources that enable researchers and practitioners to achieve state-of-the-art results and efficiently develop and launch ML-based products [84].
	Keras is a deep learning API for Human Developers, not for Computers. It allows you to develop deep learning models quickly, and it produces high-quality code that is easy to debug. By using Keras, you create a smaller codebase that is much more likely to be improved over time [85].

	Seaborn is a library that offers statistical graphics. It is built on top of Matplotlib. Seaborn has a more intuitive interface than traditional methods for creating plots. Therefore, creating high-quality, readable representations of complex data [86].
	Matplotlib, a data visualization library for Python widely used in the scientific community, is a leading tool for producing static, animated, or interactive graphs and is easy to use and customize [87].
 	OpenCV is an open-source computer vision library that started as such, focusing on computer vision, and now provides functionality for machine learning. OpenCV aims to provide an open infrastructure that enables a wide range of vision-based applications and accelerates the integration of machine perception into commercial products [88]. Pillow is a relatively friendly fork of PIL (the Python Imaging Library), an open-source library that adds image processing capabilities to Python [89].
	FastAPI is a high-performance Python API framework that competes with Node.js and Go in terms of speed. This takes Python type hints and delivers them in an intuitive, easy-to-learn manner. It dramatically accelerates development and minimizes bugs. The framework automatically produces interactive documentation and is built on OpenAPI specifications that provide robust, production-ready code [90].

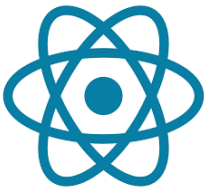
	Psycopg is the leading PostgreSQL adapter for Python, distinguished by its full compliance with the Python DB API 2.0 specification and robust thread-safe connection sharing. It is specifically engineered for demanding, highly multi-threaded environments that require rapid cursor creation and destruction, as well as high volumes of concurrent INSERTs and UPDATEs [91].
	React is a JavaScript library used to create user interfaces, particularly for web applications. It was developed by Meta, previously known as Facebook. React allows developers to build applications with components, which are reusable parts of the interface [92].
	JSX is a JavaScript syntax extension commonly used with React to define the user interface (UI). Although it resembles a classic markup language, it actually leverages the full power of JavaScript. It allows you to generate React "elements," the rendering of which in the DOM will be explained in the next section. The rest of this document will present the JSX fundamentals necessary to get started [93].

TABLE 3.6 – Deep Learning Packages and APIs Overview.

3.21 Experimental Configuration

The experiments of our three architectures were conducted under the following configuration :

Parameter	**MobileNetV2 + CBAM + CLAHE**	**DenseNet121 + CBAM + CLAHE**	**MobileNetV2**
Modality	**Fundus Photography**	**Optical Coherence Tomography Classification (OCT)**	**Slit Lamp**
Input Resolution	224×224 pixels	224×224 pixels	224×224 pixels
Hold-Out Test Split	15% stratified test set	Independent held-out test set	Independent held-out test set
Batch Size	32	32	32
Optimizer	AdamW	AdamW	Adam
Initial Learning Rate	3×10^{-4}	3×10^{-4}	1×10^{-3}
Weight Decay	1×10^{-4}	1×10^{-4}	Not Used
Learning Rate Scheduler	`ReduceLROnPlateau` (`factor=0.5`, `patience=2`)	`ReduceLROnPlateau` (`factor=0.5`, `patience=2`)	None
Loss Function	Cross-Entropy + Class Weighting + Label Smoothing (`0.1`)	Cross-Entropy + Class Weighting + Label Smoothing (`0.1`)	Cross-Entropy Loss
Dropout Rate	0.35 (FC layers), 0.25 (final layer)	0.35 (FC layers), 0.25 (final layer)	0.2 and 0.3
Maximum Epochs per Fold	30	30	30
Pooling Layer	Adaptive Average Pooling	Adaptive Average Pooling	Global Average Pooling (MobileNetV2 default)
Activation Function	ReLU	ReLU	ReLU
Explainability Method	Grad-CAM	Grad-CAM	None
Model Selection Criterion	Best Validation Macro F1-score	Best Validation Macro F1-score	Best Validation F1-score
Random Seed	42	42	42
Hardware	NVIDIA GPU (CUDA-enabled)	NVIDIA GPU (CUDA-enabled)	NVIDIA GPU (CUDA-enabled)

TABLE 3.7 – Our proposed Deep Learning Configurations.

3.22 Interface Demonstration

In this section, the various application interfaces are examined to illustrate their respective functionalities, user interactions, and contributions to the overall system workflow.

3.22.1 Login Interface

The login interface serves as the primary authentication gateway to the Eye Clinic Management System. It implements a secure, role-based access control mechanism that distinguishes between three user categories : System Administrator, Medical Doctor, and Nursing Staff, and its design prioritizes clinical professionalism, usability, and information security.

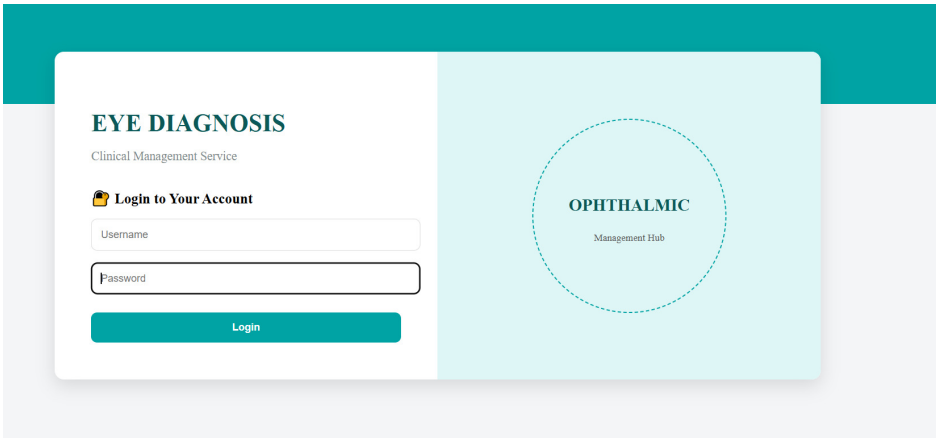


FIGURE 3.21 – Login interface of the Eye Clinic Management System.

3.22.2 Administrator Interfaces

The administrator panel enables system-wide configuration and supervision through five core modules : Clinic Management for creating and monitoring healthcare facilities (Figure 3.22), User Management for registering doctors and nurses with automatically generated credentials (Figure 3.23), Patient Management for inspecting the global patient registry across clinics (Figure 3.24), System Overview for displaying key operational indicators such as clinics, staff, pa-

tients, images, records, and pending diagnoses (Figure 3.25), and Activity Logs for comprehensive auditing of user actions. This interface grants the administrator complete supervisory control over the Eye Clinic Management System while preserving strict role-based access separation between administrative tasks and routine clinical operations.

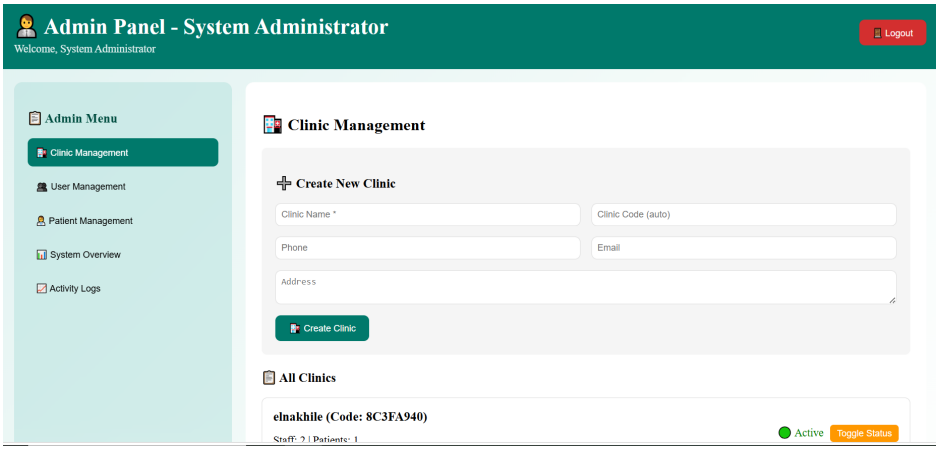


FIGURE 3.22 – Administrator interface for creating and managing clinics.

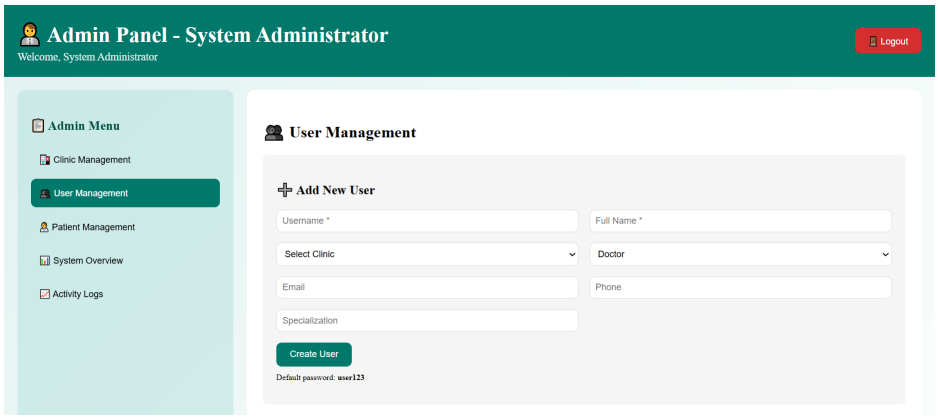


FIGURE 3.23 – Administrator interface for creating user accounts (doctors and nurses).

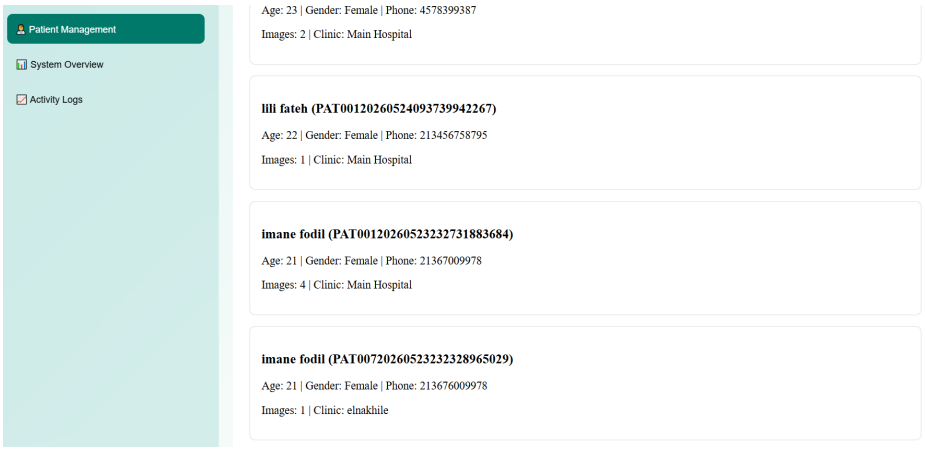


FIGURE 3.24 – Administrator interface showing the global list of registered patients.

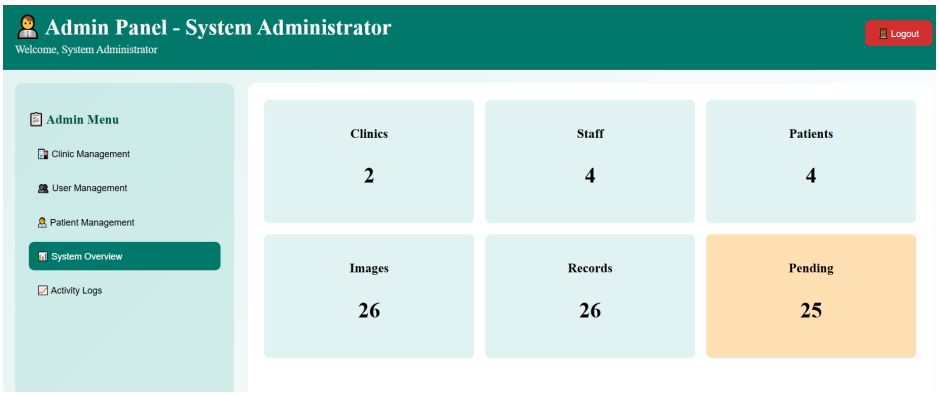


FIGURE 3.25 – Administrator dashboard summarizing system-wide metrics.

3.22.3 Nurse Interfaces

The nurse dashboard provides three core clinical workflows : patient registration through a form capturing demographic information and consent status (Figure 3.26), medical image upload that assigns images to specific patients and routes them to the responsible doctors for AI-based analysis (Figure 3.27), and patient list viewing to track all registered patients and their uploaded image counts (Figure 3.28), with all actions recorded in the activity log. This interface enables nursing staff to efficiently manage patient intake and image acquisition while maintaining a clear separation from diagnostic functions, which remain restricted to physician accounts.

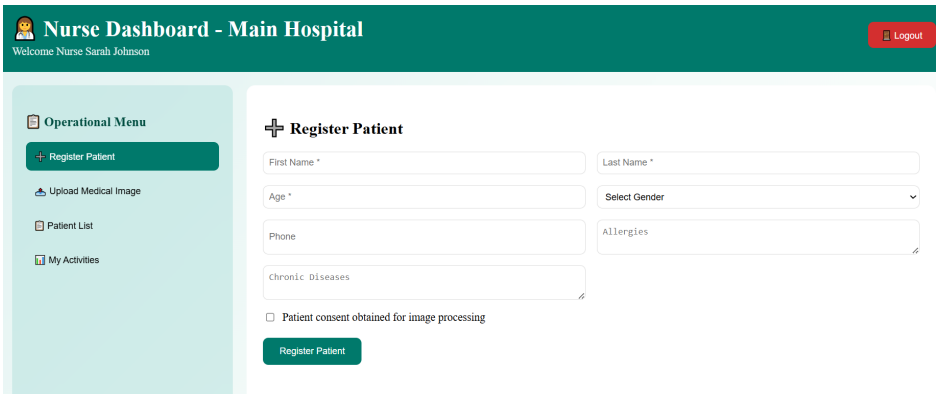


FIGURE 3.26 – Nurse interface for registering a new patient and recording consent.

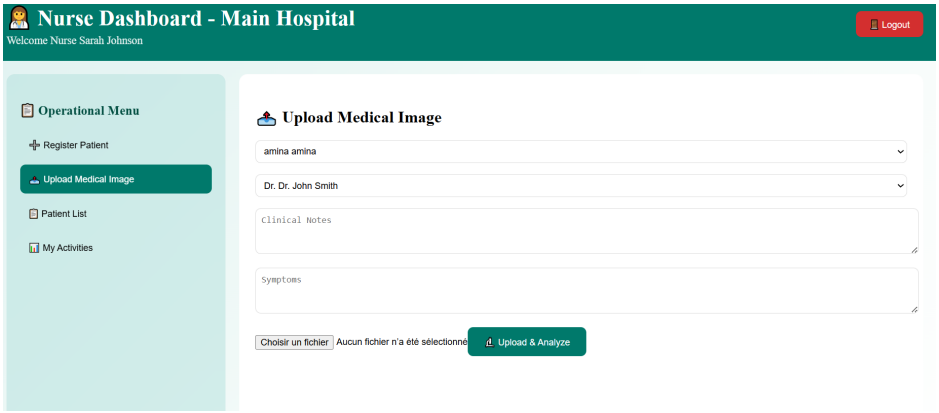


FIGURE 3.27 – Nurse interface for uploading ocular images and assigning them to patients.

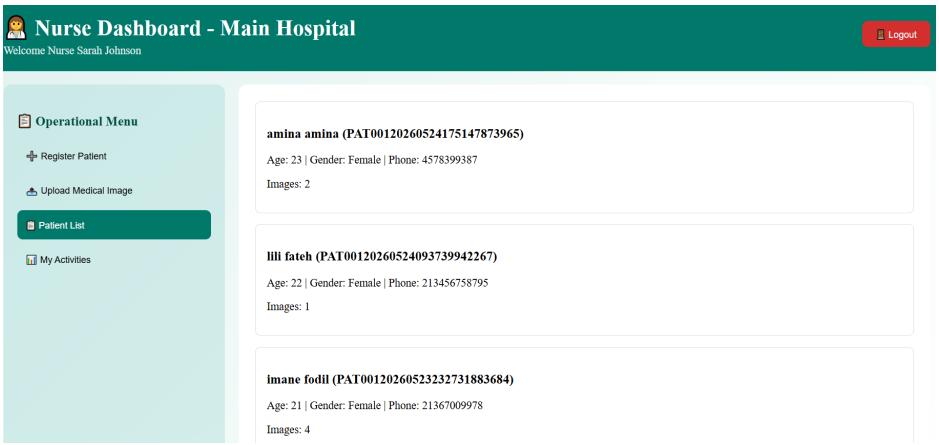


FIGURE 3.28 – Nurse interface displaying the list of registered patients and image counts.

3.22.4 Doctor Interfaces

The doctor panel provides a complete diagnostic workflow that begins with patient selection, where physicians can view all patients together with their demographic information and the number of available ocular images (Figure 3.29). From this list, the doctor can open an individual examination and analyze each medical image, for which the AI component displays the predicted modality (slit-lamp, fundus, or OCT), the disease classification with associated confidence scores, and an optional Grad-CAM-based explanation consisting of three views : the original image, a heatmap highlighting the most influential regions, and an overlay combining both representations. This interface tightly integrates artificial intelligence assistance with clinical decision-making, allowing doctors to either adopt the AI suggestion or override it based on their professional judgment, while recording the final diagnosis, prescription, and planned follow-up date in the patientâs electronic medical record.

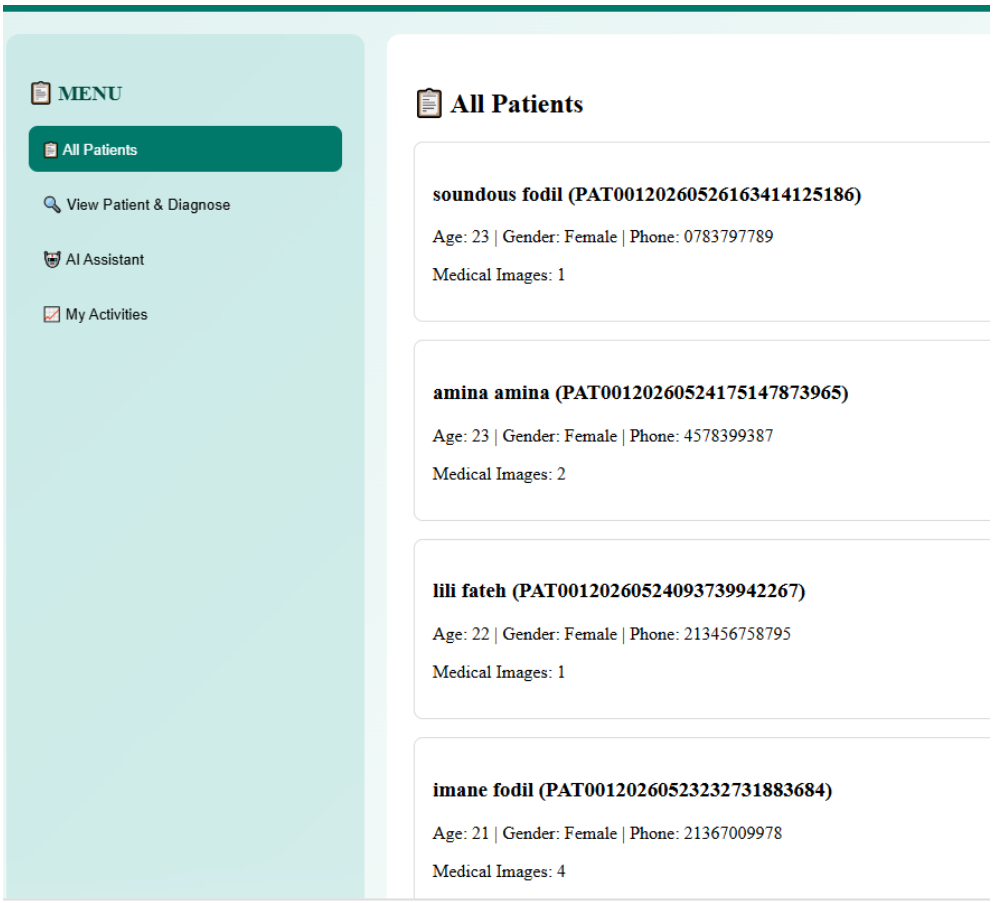


FIGURE 3.29 – Doctor interface displaying the list of patients with demographic information and image counts.

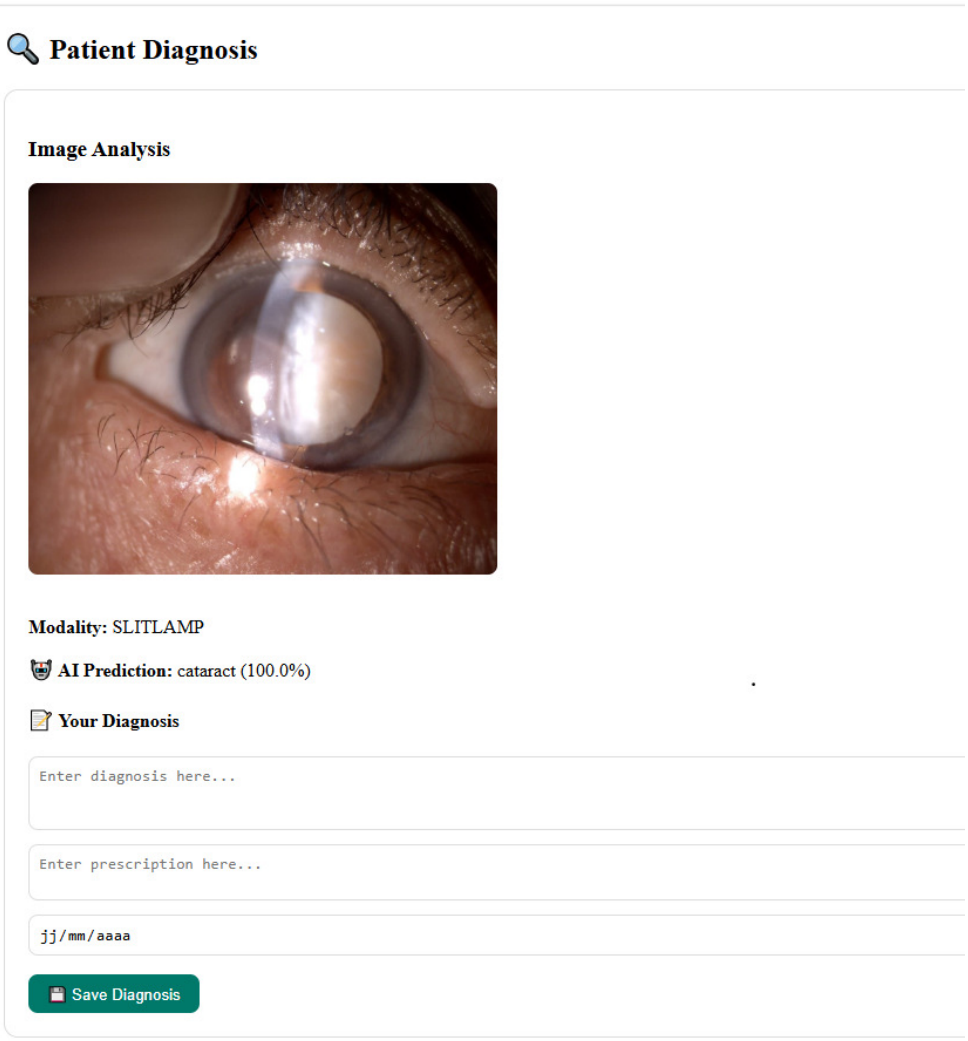


FIGURE 3.30 – Cataract diagnosis on a slit-lamp image with AI prediction, confidence scores, and Grad-CAM visual explanation.

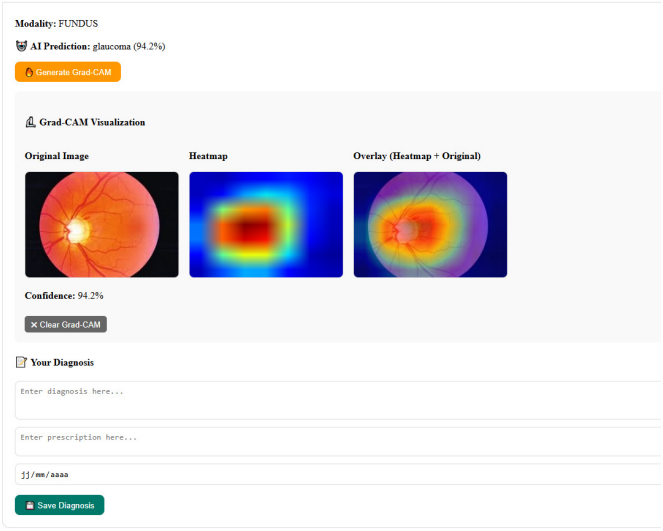


FIGURE 3.31 – Glaucoma diagnosis on a fundus image with AI prediction, confidence scores, and Grad-CAM visual explanation.

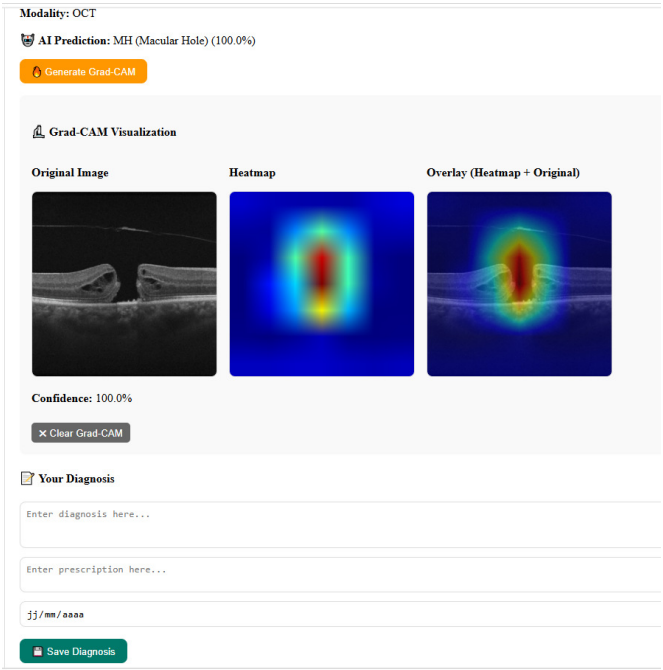


FIGURE 3.32 – Macular hole diagnosis on an OCT image with AI prediction, confidence scores, and Grad-CAM visual explanation.

3.23 Model Performance Results

In this section, we present the experimental results of the proposed multimodal ophthalmic disease classification system.

3.23.1 Optical Coherence Tomography Classification (OCT) Results

Two deep learning models were developed and evaluated for retinal Optical Coherence Tomography Classification image classification. DenseNet and EfficientNet-B3 models were trained using CLAHE preprocessing and a 5-fold cross-validation strategy. The performance of each model is analyzed using quantitative metrics and visual interpretability techniques.

3.23.1.1 DenseNet Results

— **K-Fold Cross-Validation Results**

Figures 3.33 and 3.34 present the detailed training histories (loss and accuracy) for each fold. These plots reveal that all folds follow a similar convergence trajectory. The consistency in convergence patterns across independent data splits further validates the architectural choices and hyperparameter configuration. **Accuracy :**

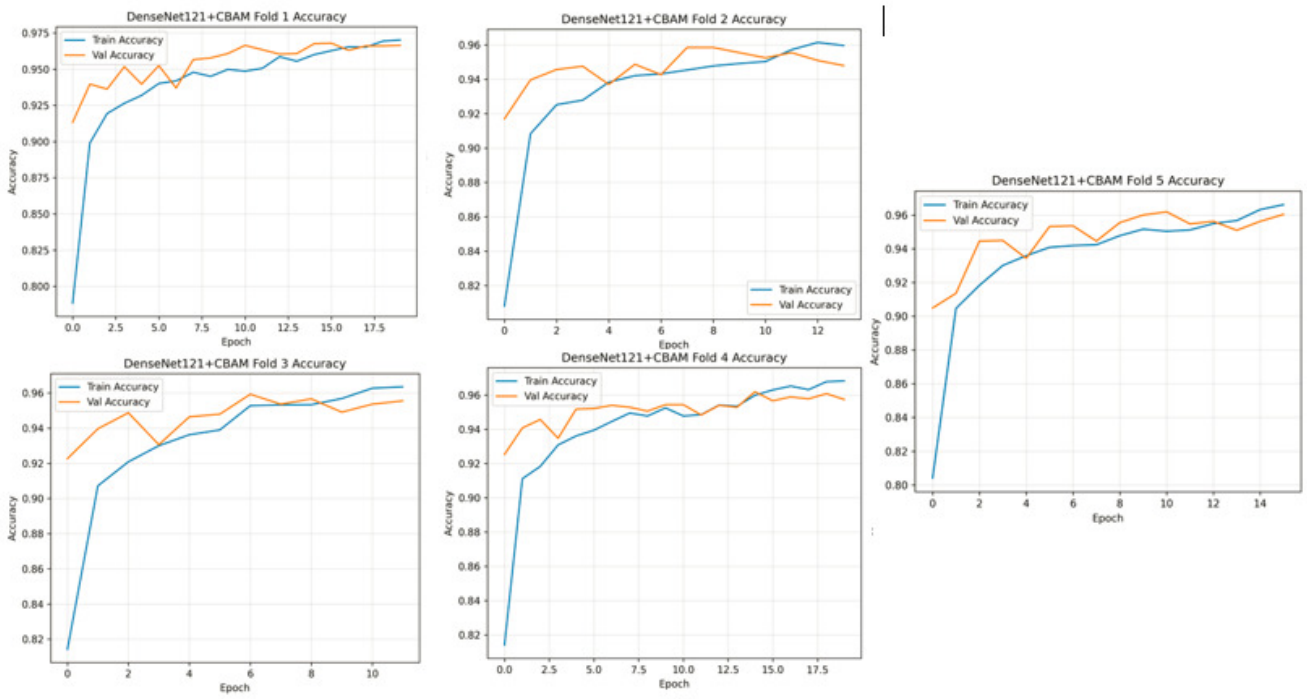


FIGURE 3.33 – Training and validation accuracy curves across all 5 folds.

Loss :

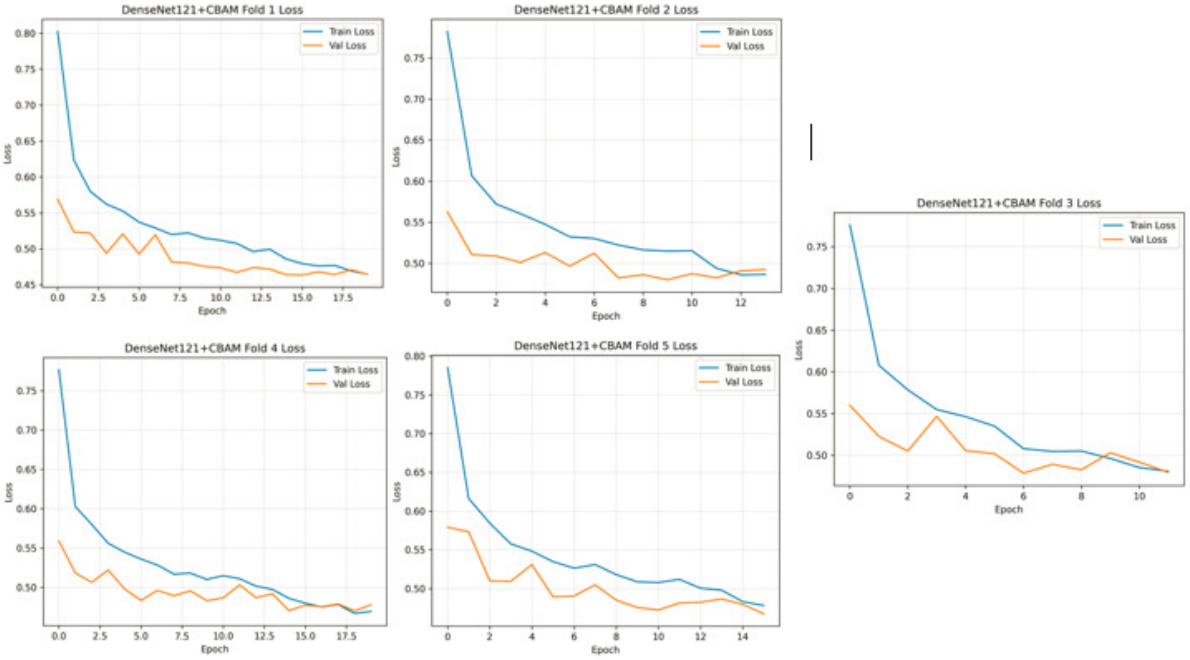


FIGURE 3.34 – Training and validation loss curves across all 5 folds.

— **K-Fold Confusion Matrix**

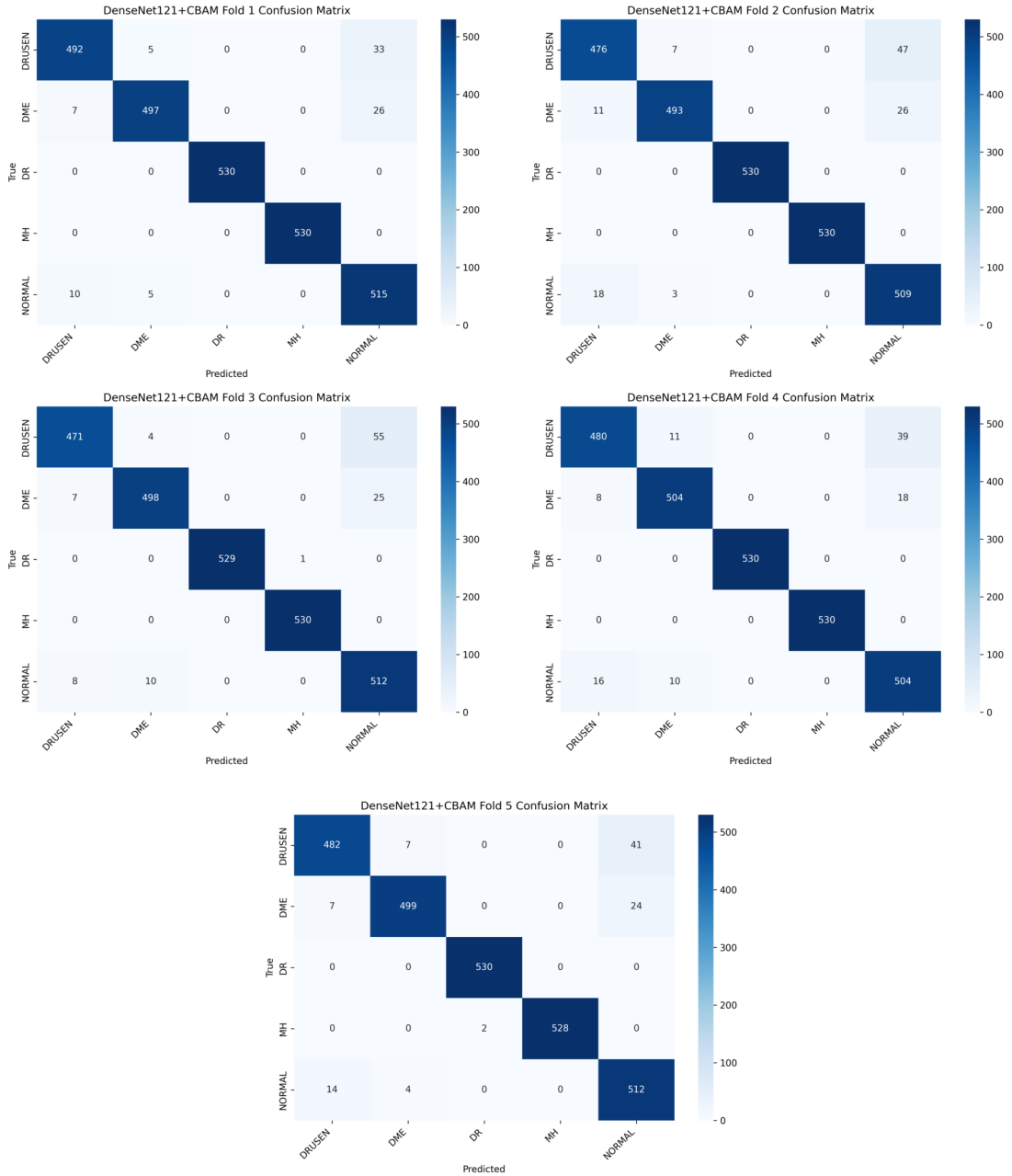


FIGURE 3.35 – Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance.

— **Final average accuracy and loss across all 5 folds**

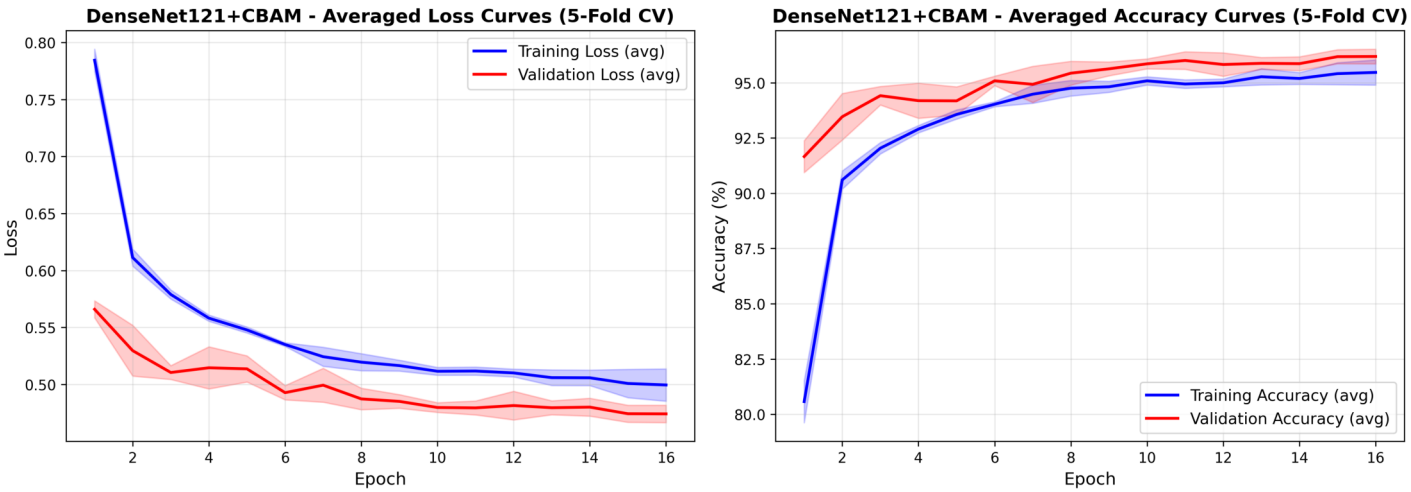


FIGURE 3.36 – Averaged training and validation accuracy curves across all 5 folds.

— **Held-Out Test Set Confusion Matrix**

The model from the best-performing fold (Fold 1), selected based on the highest validation macro F1-score of 96.77%, was evaluated on the completely held-out test set that was never exposed during any phase of training or hyper-parameter tuning.

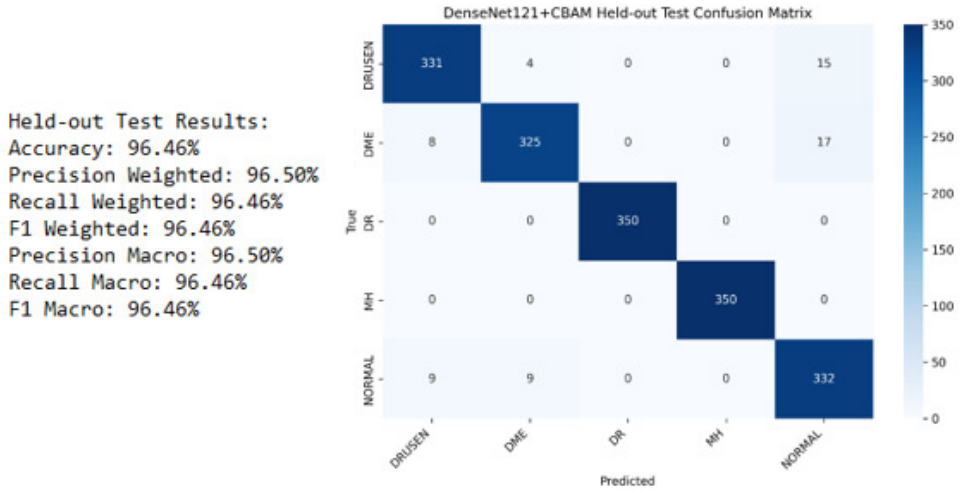


FIGURE 3.37 – DenseNet121 Held-out Test Confusion Matrix.

The confusion matrix shows that the DenseNet121 model achieved excellent classification performance, with **DR! (DR!)** and **MH! (MH!)** classes perfectly classified. The **NORMAL** class also showed very high accuracy with minimal errors. Minor confusion occurred between **DRUSEN! (DRUSEN!)**, **DME! (DME!)**, and **NORMAL** classes due to similarities in retinal features and subtle pathological signs. Overall, the model demonstrated strong reliability and effective discrimination across all OCT classes.

— **K-Fold Cross-Validation Summary**

```
============================================================
  K-FOLD CROSS-VALIDATION SUMMARY  (Mean ± Std)
============================================================
  Accuracy                  :  96.16% ± 0.39%
  Precision Macro           :  96.31% ± 0.36%
  Recall Macro              :  96.16% ± 0.39%
  F1 Macro                  :  96.18% ± 0.39%
  Precision Weighted        :  96.31% ± 0.36%
  Recall Weighted           :  96.16% ± 0.39%
  F1 Weighted               :  96.18% ± 0.39%
============================================================

Per-fold results:
 fold val_accuracy val_precision_macro val_recall_macro val_f1_macro val_precision_weighted val_recall_weighted val_f1_weighted
    1       96.75%              96.88%           96.75%       96.77%                 96.88%              96.75%          96.77%
    2       95.77%              95.95%           95.77%       95.80%                 95.95%              95.77%          95.80%
    3       95.85%              96.10%           95.85%       95.88%                 96.10%              95.85%          95.88%
    4       96.15%              96.22%           96.15%       96.16%                 96.22%              96.15%          96.16%
    5       96.26%              96.41%           96.26%       96.28%                 96.41%              96.26%          96.28%
```

FIGURE 3.38 – K-Fold Cross-Validation Summary.

Figure 3.38 presents the comprehensive cross-validation results across all five folds. The proposed model demonstrated exceptional consistency, with validation accuracies ranging from 95.77% to 96.75% and macro F1-scores ranging from 95.80% to 96.77%. The narrow performance range (approximately 1 percentage point) across folds indicates that the model is robust to data partitioning and not sensitive to specific train/validation splits.

— **Classification Report**

	precision	recall	f1-score	support
DRUSEN	0.9511	0.9457	0.9484	350
DME	0.9615	0.9286	0.9448	350
DR	1.0000	1.0000	1.0000	350
MH	1.0000	1.0000	1.0000	350
NORMAL	0.9121	0.9486	0.9300	350
accuracy			0.9646	1750
macro avg	0.9650	0.9646	0.9646	1750
weighted avg	0.9650	0.9646	0.9646	1750

FIGURE 3.39 – DenseNet121 Classification Report.

Discussion

DenseNet121 achieved a validation accuracy of 96.17% across 5-fold cross-validation on the five-class retinal OCT dataset comprising DRUSEN, DME, DR, MH, and NORMAL categories. These results reflect the model's capacity to learn discriminative representations across pathologically distinct and visually similar retinal conditions. The dense connectivity mechanism of DenseNet121, in which each layer receives feature maps from all preceding layers, enables the network to reuse low-level and mid-level features throughout the learning process. This architectural property is particularly beneficial for retinal OCT classification, where diseases such as DME and DR share overlapping morphological characteristics, including retinal fluid accumulation and layer disruption. The consistent performance observed across all five folds indicates that DenseNet121 generalizes well to unseen data without overfitting to fold-specific patterns, a critical requirement for models intended for clinical deployment, where image acquisition conditions vary significantly.

3.23.1.2 EfficientNet-B3 Results

— **K-Fold Cross-Validation Results**

Fold Results:

fold	val_accuracy	val_precision_macro	val_recall_macro	val_f1_macro
1	0.986792	0.987478	0.986792	0.986876
2	0.987547	0.987593	0.987547	0.987561
3	0.984528	0.984951	0.984528	0.984598
4	0.989811	0.989825	0.989811	0.989816
5	0.976604	0.977257	0.976604	0.976690

Cross-Validation Summary (n=5 folds)

Accuracy	98.51% ± 0.46%
Precision (Macro)	98.54% ± 0.44%
Recall (Macro)	98.51% ± 0.46%
F1 (Macro)	98.51% ± 0.45%
Precision (Weighted)	98.54% ± 0.44%
Recall (Weighted)	98.51% ± 0.46%
F1 (Weighted)	98.51% ± 0.45%

FIGURE 3.40 – EfficientNet-B3 K-Fold Cross-Validation Results.

— **Held-Out Test Set**

The model from the best-performing fold (Fold 4), selected based on the highest validation macro F1-score of 98.98%, was evaluated on the completely held-out test set that was never exposed during any phase of training or hyper-parameter tuning.

Held-out Test Results:

Accuracy:	98.17%
Precision Weighted:	98.20%
Recall Weighted:	98.17%
F1 Weighted:	98.18%
Precision Macro:	98.20%
Recall Macro:	98.17%
F1 Macro:	98.18%

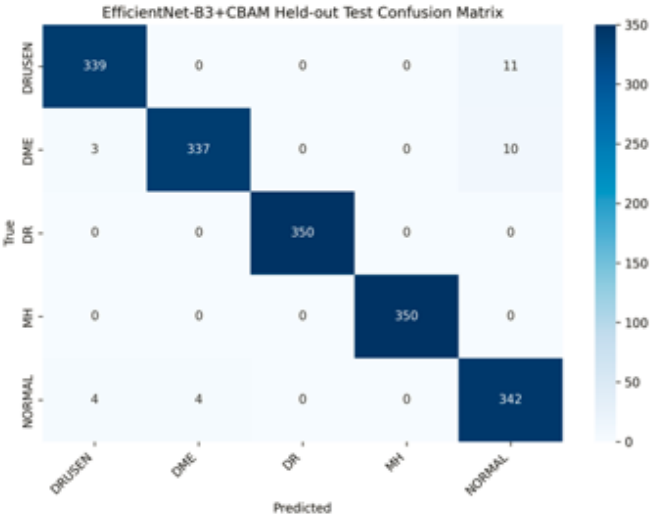


FIGURE 3.41 – Confusion matrix on the held-out test set for the EfficientNet-B3+CBAM model.

— **Classification Report**

	precision	recall	f1-score	support
DRUSEN	0.9798	0.9686	0.9741	350
DME	0.9883	0.9629	0.9754	350
DR	1.0000	1.0000	1.0000	350
MH	1.0000	1.0000	1.0000	350
NORMAL	0.9421	0.9771	0.9593	350
accuracy			0.9817	1750
macro avg	0.9820	0.9817	0.9818	1750
weighted avg	0.9820	0.9817	0.9818	1750

FIGURE 3.42 – Classification Report for the EfficientNet-B3+CBAM model.

Discussion

EfficientNet-B3 achieved the highest validation accuracy (98.17% across 5-fold CV) due to its compound scaling strategy, which effectively captures both local and global retinal OCT features. However, it was not chosen as the final model because high accuracy alone does not guarantee clinical reliability. Its architecture risks overfitting dataset-specific patterns, limiting generalization across diverse imaging conditions. DenseNet121 was ultimately preferred for its more consistent and clinically robust performance across all five retinal disease categories.

3.23.2 Fundus Photography Classification Results

In the second modality, fundus imaging, two deep learning models, **MobileNetV2** and **ResNet50**, were developed for retinal disease classification. The **MobileNetV2** model was integrated with the CBAM attention mechanism and trained using a 5-fold cross-validation strategy to evaluate robustness and generalization. In contrast, the **ResNet50** model was trained without cross-validation, using a standard data-splitting approach. The performance of both models was analyzed using quantitative evaluation metrics and interpretability techniques.

3.23.2.1 MobileNetV2 with K-Fold Cross-Validation Results

— **K-Fold Cross-Validation Results**

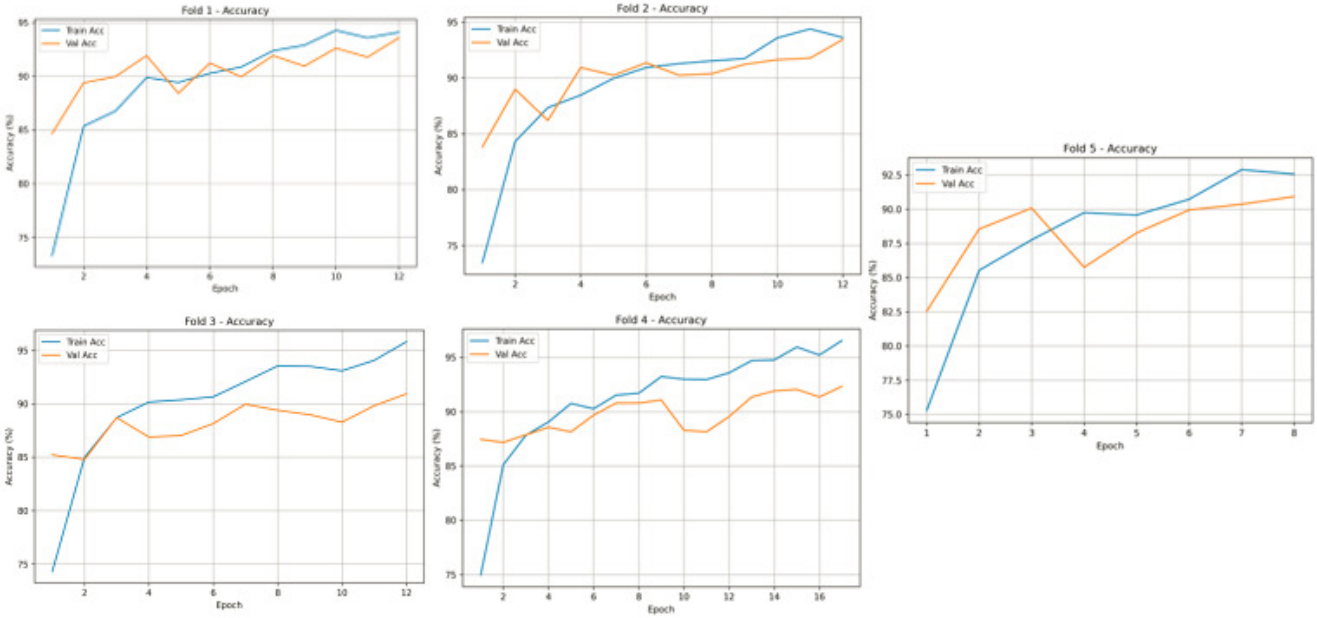


FIGURE 3.43 – Training and validation accuracy curves across all 5 folds.

Loss :

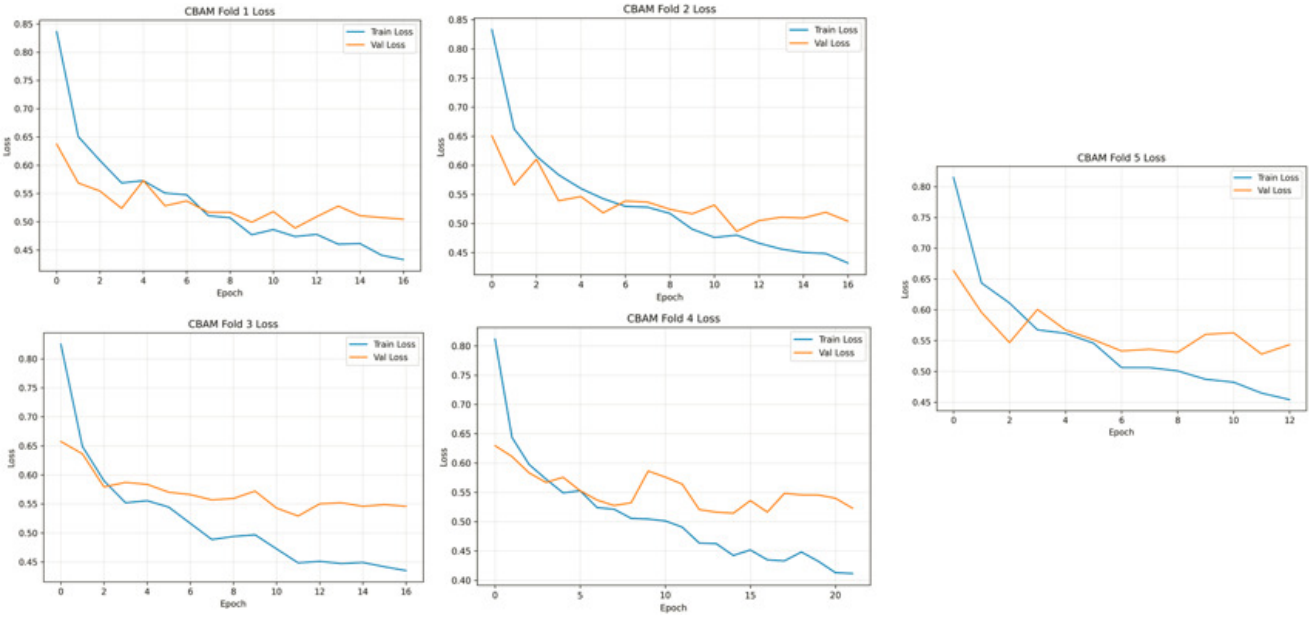


FIGURE 3.44 – Training and validation loss curves across all 5 folds.

Per-fold training dynamics reveal consistent patterns across all folds (**Figures 3.43–3.44**), with each fold showing characteristic rapid initial learning followed by progressive refinement. The early stopping mechanism successfully prevented overfitting in all folds, with training terminating between epochs 8 and 17, depending on validation performance.

— **K-Fold Confusion Matrix**

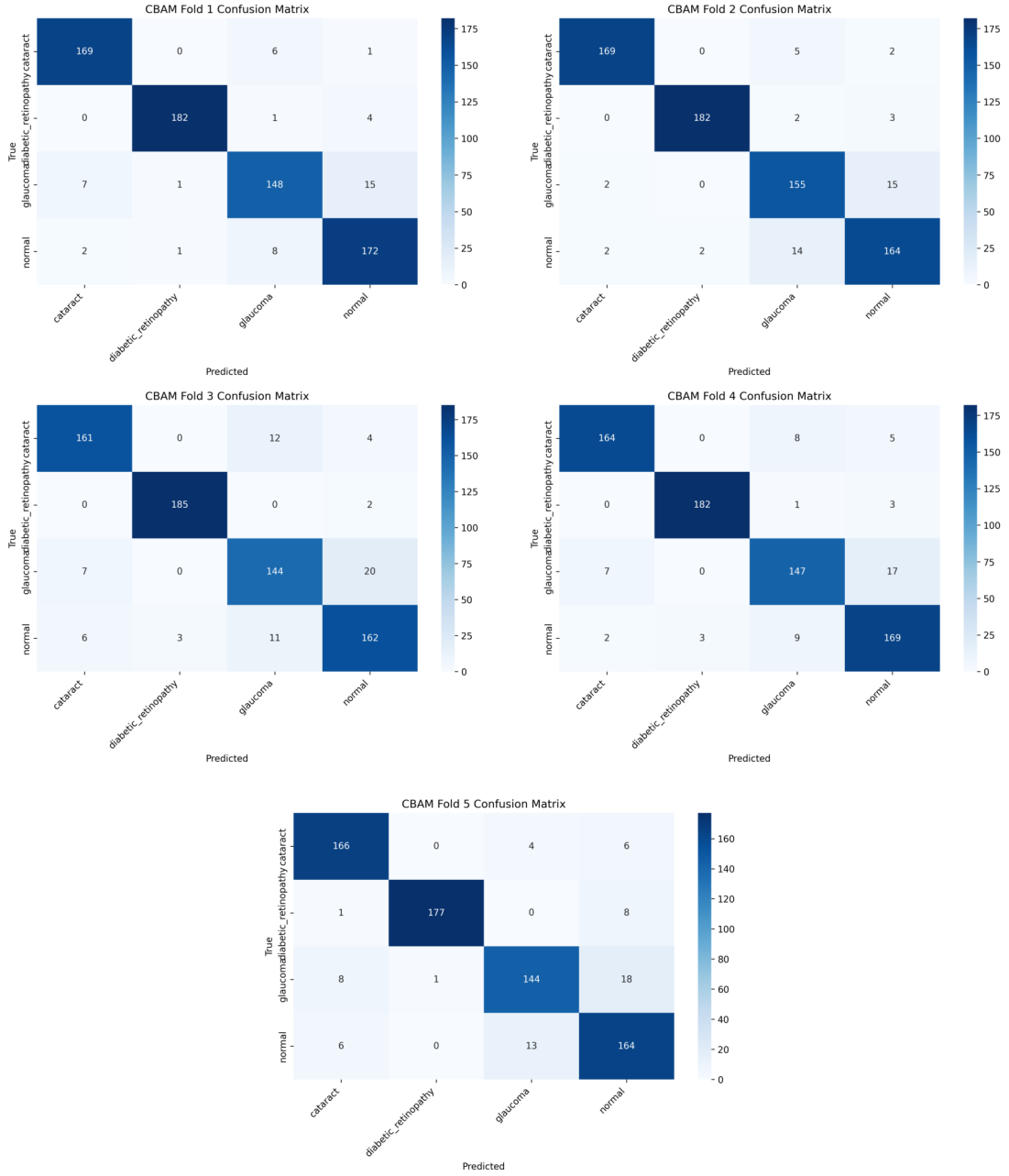


FIGURE 3.45 – Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance.

— **Final average accuracy and loss across all 5 folds**

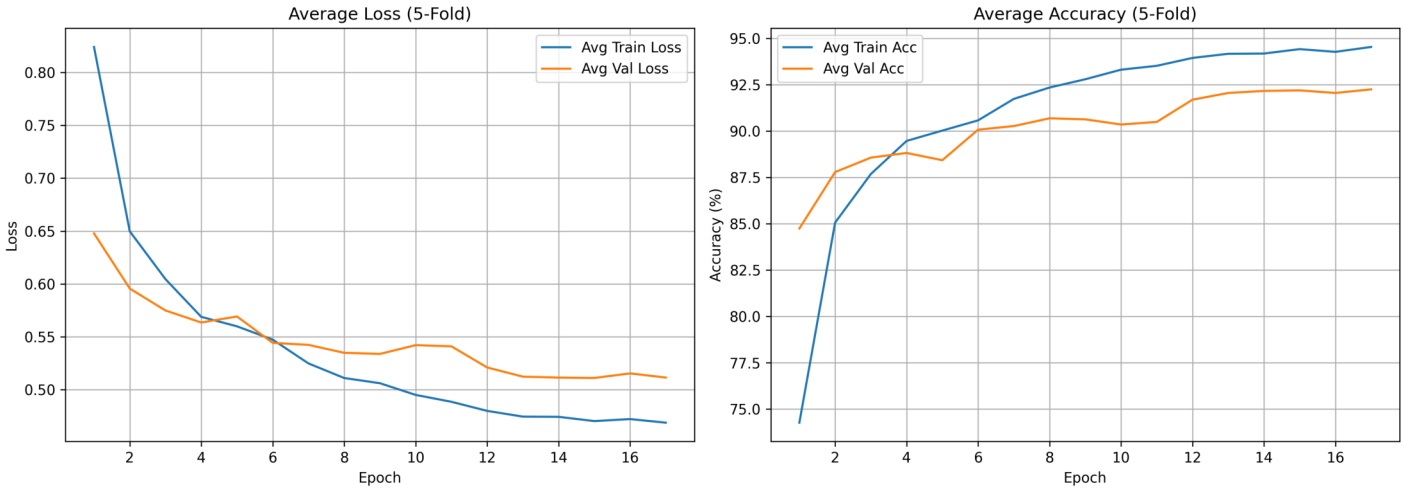


FIGURE 3.46 – Averaged training and validation accuracy curves across all 5 folds.

— **Held-Out Test Set Confusion Matrix**

The best performing model (Fold 1, validation macro F1 = 93.48%) was evaluated on the independent held-out test set comprising 633 images. Figure 3.47 summarizes the comprehensive test set performance metrics.

```
Held-out Test Results:
Accuracy: 90.68%
Precision Weighted: 90.66%
Recall Weighted: 90.68%
F1 Weighted: 90.66%
Precision Macro: 90.54%
Recall Macro: 90.51%
F1 Macro: 90.52%
```

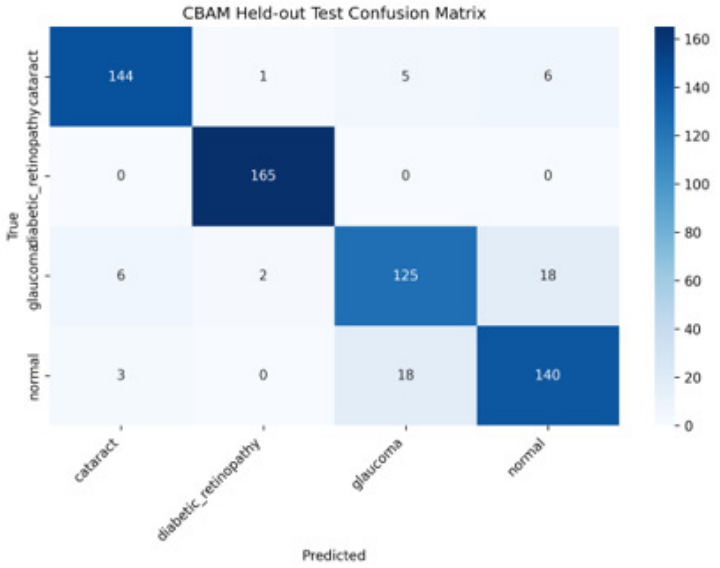


FIGURE 3.47 – MobileNetV2 K-fold Held-out Test Confusion Matrix.

The confusion matrix shows that **DR!** was classified with very high classification performance and minimal errors, while **cataract** showed occasional confusion with glaucoma and normal cases. **Glaucoma** was the most challenging class, often misclassified as **normal** or **cataract**, and some normal images were also confused with **glaucoma**, reflecting the visual similarity between these conditions in fundus images.

— **K-Fold Cross-Validation Summary**

Figure 3.48 presents the per-fold validation performance metrics for the MobileNetV2 model across all five folds of the stratified cross-validation.

```
Cross-Validation Mean ± Std:
Accuracy:           92.24% ± 1.30%
Precision Macro:    92.25% ± 1.28%
Recall Macro:       92.14% ± 1.32%
Macro F1:           92.17% ± 1.31%
Precision Weighted: 92.32% ± 1.27%
Recall Weighted:    92.24% ± 1.30%
Weighted F1:        92.26% ± 1.29%
```

Fold Results:

fold	val_accuracy	val_precision_macro	val_recall_macro	val_f1_macro	val_precision_weighted	val_recall_weighted	val_f1_weighted
1	0.935844	0.935594	0.934719	0.934874	0.936220	0.935844	0.935751
2	0.934449	0.934499	0.933938	0.934148	0.935276	0.934449	0.934793
3	0.909344	0.908327	0.907781	0.907948	0.909443	0.909344	0.909288
4	0.923291	0.923451	0.922049	0.922457	0.924043	0.923291	0.923374
5	0.909218	0.910663	0.908269	0.908851	0.911223	0.909218	0.909605

FIGURE 3.48 – MobileNetV2 K-fold Cross-Validation Summary.

The results demonstrate consistent performance across all folds, with low variation between validation scores, indicating strong model stability and good generalization across different data splits.

— **Classification Report**

Figure 3.49 presents the detailed per-class classification report for the held-out test set.

	precision	recall	f1-score	support
cataract	0.9412	0.9231	0.9320	156
diabetic_retinopathy	0.9821	1.0000	0.9910	165
glaucoma	0.8446	0.8278	0.8361	151
normal	0.8537	0.8696	0.8615	161
accuracy			0.9068	633
macro avg	0.9054	0.9051	0.9052	633
weighted avg	0.9066	0.9068	0.9066	633

FIGURE 3.49 – MobileNetV2 K-fold Cross-Validation Classification Report.

Discussion

The MobileNetV2 model, integrated with the CBAM attention mechanism and trained using 5-fold cross-validation, demonstrated robust and consistent performance, achieving a validation accuracy of 92.24%. The incorporation of CBAM likely enhanced the model's ability to focus on relevant features in fundus images, thereby contributing to superior discrimination between disease classes, particularly in identifying diabetic retinopathy. The use of k-fold cross-validation further reinforced the stability and generalizability of this approach, as evidenced by low variation in performance metrics across different folds. This combination of lightweight architecture and attention modules reflects a promising balance between computational efficiency and high classification accuracy, making it well-suited for deployment in real-world screening environments.

3.23.2.2 ResNet50 Results

The ResNet50 architecture was employed as a second model for the classification of fundus images. The model was configured with pretrained ImageNet weights and fine-tuned. The dataset was partitioned into training, validation, and test sets. The training set comprised 2,819 samples (70%), the validation set contained 604 samples (15%), and the test set consisted of 605 samples (15%).

— **Training and Loss Curves**

The training and validation curves illustrating the ResNet50 learning progression over 30 epochs are presented in Figure 3.50.

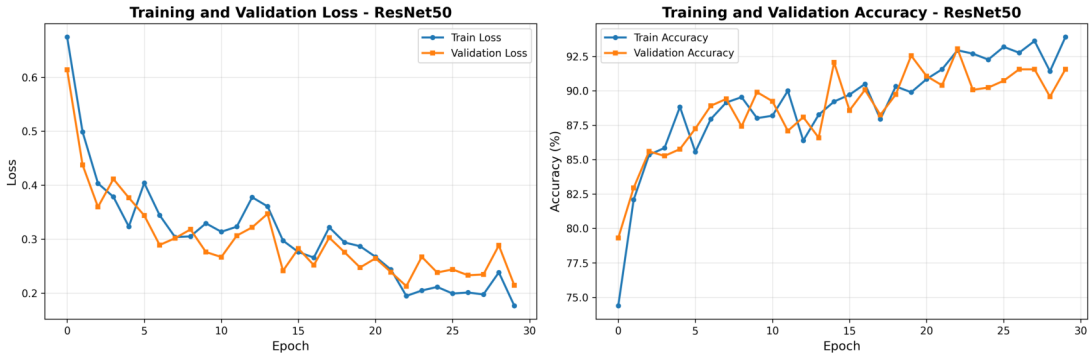


FIGURE 3.50 – Training and validation loss and accuracy curves for the ResNet50 model over 30 epochs.

— **Confusion Matrix Analysis**

The confusion matrix presented in Figure 3.51 provides a detailed breakdown of the ResNet50 model prediction patterns across all four classes.

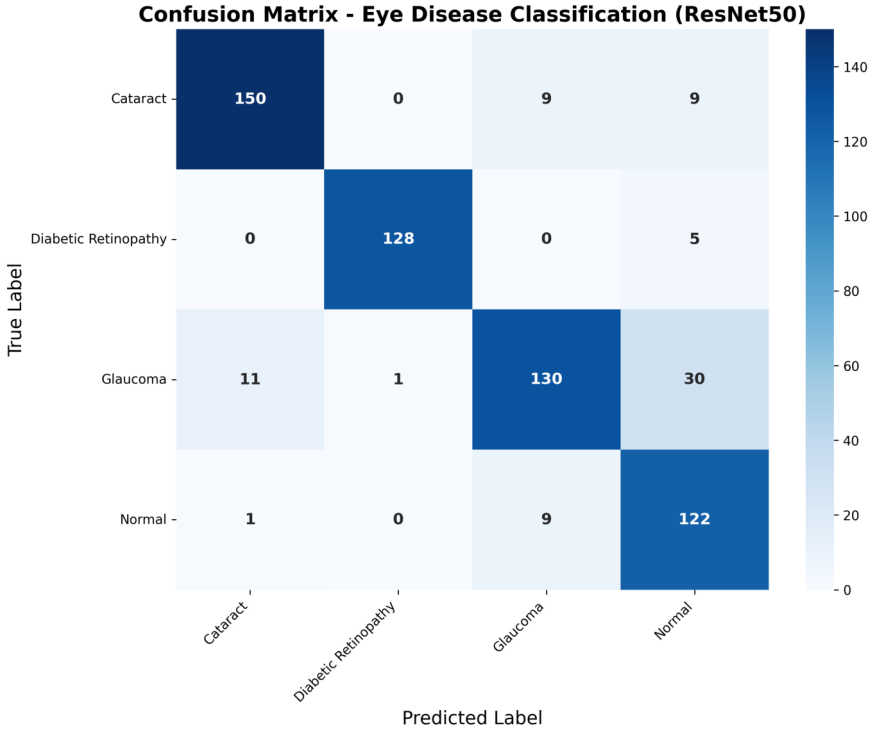


FIGURE 3.51 – ResNet50 Confusion Matrix on the held-out test set.

— **Classification Report**

The trained ResNet50 model was evaluated on the held-out test set comprising 605 samples. The overall classification performance is summarized in Figure 3.52.

Classification Report:

	precision	recall	f1-score	support
Cataract	0.93	0.89	0.91	168
Diabetic Retinopathy	0.99	0.96	0.98	133
Glaucoma	0.88	0.76	0.81	172
Normal	0.73	0.92	0.82	132
accuracy			0.88	605
macro avg	0.88	0.88	0.88	605
weighted avg	0.89	0.88	0.88	605

FIGURE 3.52 – ResNet50 Test Set Classification Report.

Discussion

The ResNet50 model, trained without k-fold cross-validation or attention mechanisms, achieved a validation accuracy of 91.56%. While ResNet50 is a deep and powerful architecture capable of capturing complex features, the absence of CBAM may have limited its ability to emphasize the most salient regions in fundus images, potentially affecting its discriminative capacity. Additionally, training without k-fold validation restricts the assessment of the model's robustness and raises concerns about overfitting to the specific data split. This comparison underscores the importance of incorporating attention mechanisms and systematic validation strategies to enhance model performance and reliability in classifying ophthalmic images.

3.23.3 Slit Lamp Classification Results

In the third modality, slit-lamp imaging, two deep learning models, MobileNetV2 and EfficientNet, were developed for ocular disease classification. Both models were evaluated using quantitative evaluation metrics to determine their predictive performance and classification capability.

3.23.3.1 MobileNetV2 with K-Fold Cross-Validation Results

— **K-Fold Cross-Validation Results**

Accuracy :

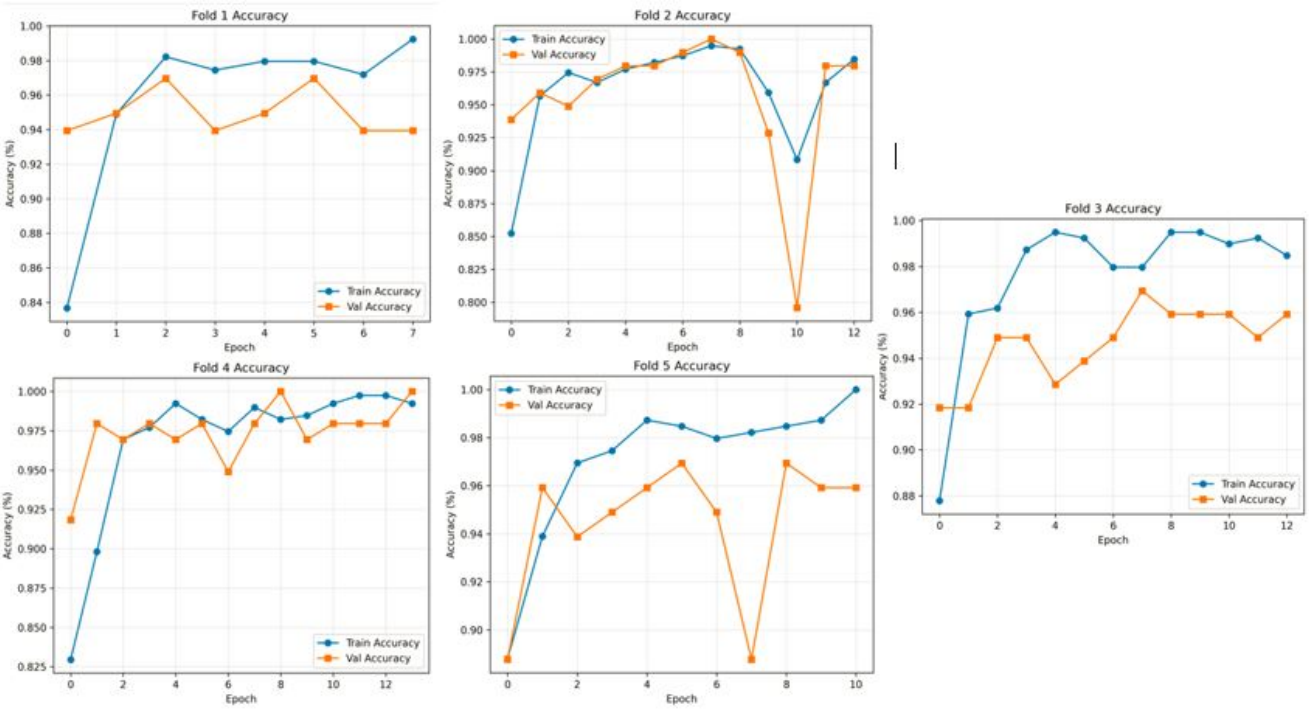


FIGURE 3.53 – Training and validation accuracy curves across all 5 folds.

Loss :

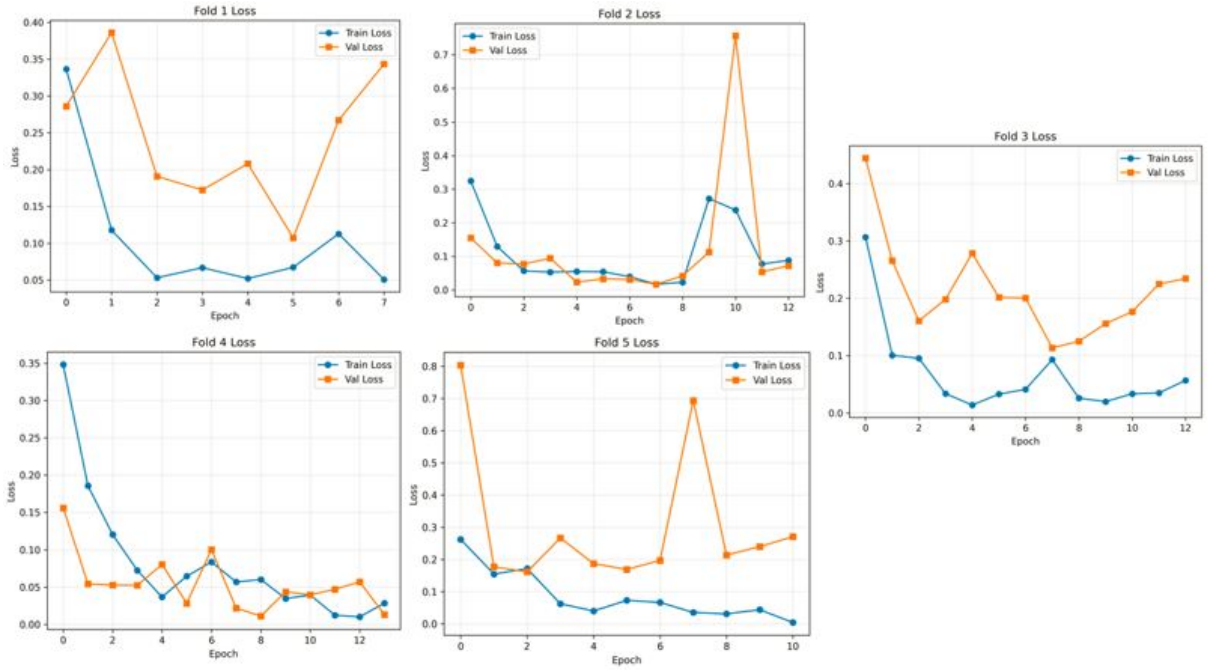


FIGURE 3.54 – Training and validation loss curves across all 5 folds.

The training dynamics across the five folds are illustrated in Figure 3.53 and Figure 3.54. Across all folds, the training and validation accuracy curves showed steady convergence, with validation accuracy stabilizing above 96% in later epochs. Similarly, the loss curves exhibited a consistent downward trend with no significant divergence between training and validation losses, indicating minimal overfitting.

— **K-Fold Confusion Matrix**

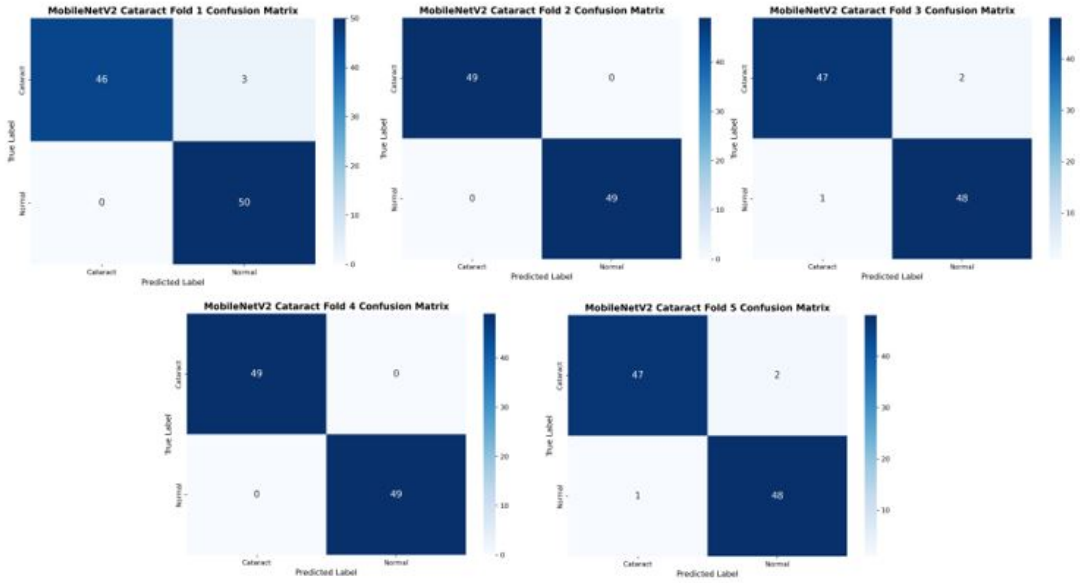


FIGURE 3.55 – Confusion matrices for each fold of the cross-validation, illustrating per-class classification performance.

— **Final average accuracy and loss across all 5 folds**

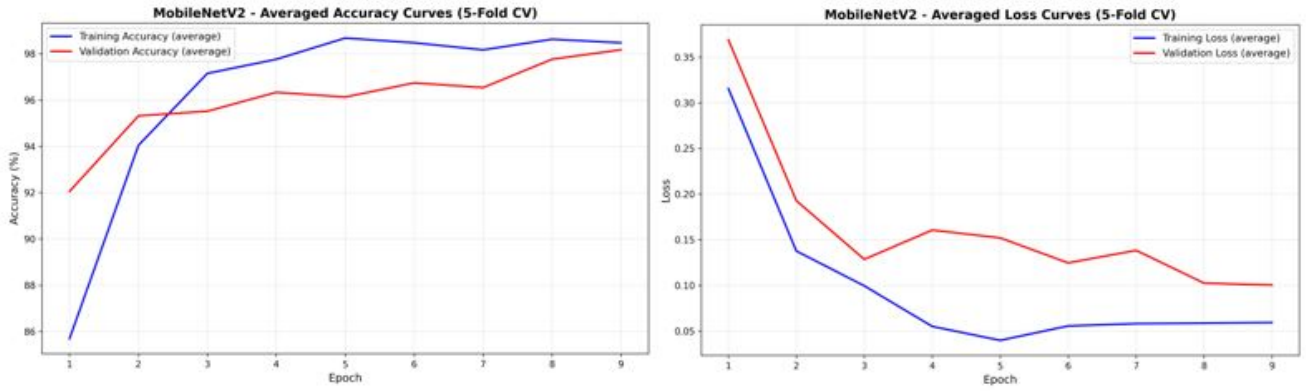


FIGURE 3.56 – Averaged training and validation accuracy curves across all 5 folds.

The averaged accuracy and loss curves across the 5-fold cross-validation are presented in Figure 3.56. The averaged validation accuracy closely tracked the training accuracy, plateauing at a high level, while the averaged validation loss stabilized near zero, further confirming the model's stable optimization and strong generalization capability.

— **Held-Out Test Set Confusion Matrix**

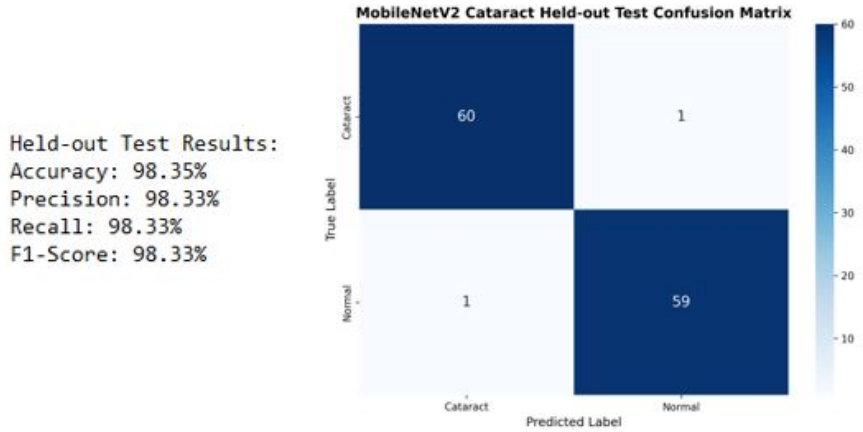


FIGURE 3.57 – MobileNetV2 Held-out Test Confusion Matrix.

The confusion matrix for the held-out test set is depicted in Figure 3.57. Out of 121 test samples, only 2 instances were misclassified : one **Cataract** sample was falsely predicted as **Normal** (False Negative), and one **Normal** sample was falsely predicted as **Cataract** (False Positive).

— **K-Fold Cross-Validation Summary**

The detailed validation metrics for each of the five folds are summarized in Figure 3.58. The model demonstrated exceptional performance, achieving perfect classification results (Accuracy, Precision, Recall, and F1-score of 100%) in Folds 2 and 4. The lowest performance was observed in Folds 3 and 5, which still achieved an F1-Score of 96.97%.

```
Fold Results:
 fold  val_accuracy  val_precision  val_recall   val_f1
    1      0.969697       0.943396    1.000000 0.970874
    2      1.000000       1.000000    1.000000 1.000000
    3      0.969388       0.960000    0.979592 0.969697
    4      1.000000       1.000000    1.000000 1.000000
    5      0.969388       0.960000    0.979592 0.969697

Cross-Validation Mean ± Std:
Accuracy: 98.17% ± 1.67%
Precision: 97.27% ± 2.58%
Recall: 99.18% ± 1.12%
F1-Score: 98.21% ± 1.64%
```

FIGURE 3.58 – MobileNetV2 K-Fold Cross-Validation Summary.

The cross-validation mean and standard deviation metrics further quantify the model's reliability.

— **Classification Report**

Classification Report:

	precision	recall	f1-score	support
Cataract	0.98	0.98	0.98	61
Normal	0.98	0.98	0.98	60
accuracy			0.98	121
macro avg	0.98	0.98	0.98	121
weighted avg	0.98	0.98	0.98	121

FIGURE 3.59 – MobileNetV2 K-Fold Cross-Validation Classification Report.

Discussion

The MobileNetV2 model, trained with 5-fold cross-validation, achieved a significantly higher validation accuracy of 98.17%. The k-fold validation process provided a more reliable and stable estimate of the modelâs performance across different data splits, reducing the likelihood of overfitting. This comprehensive validation approach enhanced the modelâs robustness and demonstrated its superior generalization capability, making it more suitable for deployment in real-world clinical settings where data variability is common.

3.23.3.2 EfficientNet-B0 Results

— **Training and Loss Curves**

The EfficientNet model was trained for 20 epochs on fundus images resized to an inner dimension of 500 pixels. Figure 3.60 presents the training and validation loss and accuracy curves.

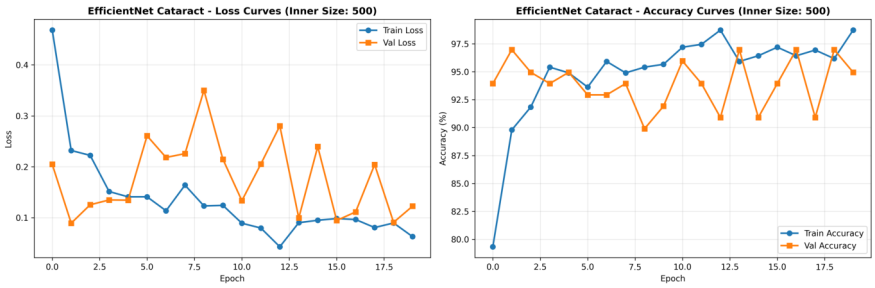


FIGURE 3.60 – Training and validation loss and accuracy curves for the EfficientNet-B0 model over 20 epochs.

— **Confusion Matrix Analysis**

The figure 3.61 represents the confusion matrix of the EfficientNet-B0 model, with a validation accuracy of 94.21%, showing low false-positive predictions and a small number of missed cataract cases.

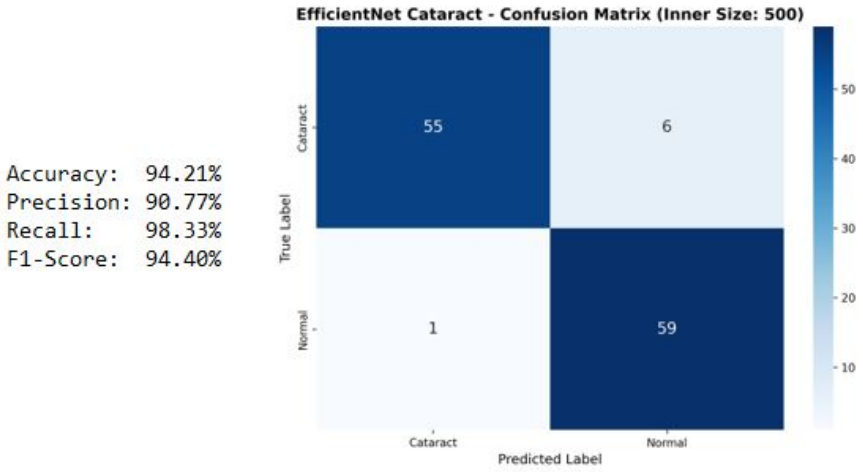


FIGURE 3.61 – EfficientNet-B0 Confusion Matrix.

— **Classification Report**

Figure 3.62 summarizes the performance metrics derived from the confusion matrix (Figure 3.61), providing a comprehensive evaluation of the modelâs classification results.

Classification Report:

	precision	recall	f1-score	support
Cataract	0.98	0.90	0.94	61
Normal	0.91	0.98	0.94	60
accuracy			0.94	121
macro avg	0.94	0.94	0.94	121
weighted avg	0.95	0.94	0.94	121

FIGURE 3.62 – EfficientNet-B0 Test Set Classification Report.

Discussion

The EfficientNet model trained without k-fold cross-validation achieved a validation accuracy of approximately 94.21%. While this accuracy is quite high, the absence of k-fold validation means the model's performance estimate may be slightly optimistic since it was evaluated on a single train-test split. Without a robust validation process, there is a risk of overfitting to the training dataset, which might limit its ability to generalize to new, unseen data in clinical practice.

3.23.4 MobileNetV2 for Modality Classification Results

The modality classifier model was evaluated using a 5-fold cross-validation strategy to ensure robust and unbiased performance estimation. Figure 3.63 illustrates the average training and validation accuracy and loss curves across all five folds, demonstrating a consistent convergence pattern with no significant signs of overfitting, as evidenced by the close alignment between training and validation accuracy throughout the training epochs, alongside a steady decrease in loss values confirming stable model optimization and generalization across all folds.

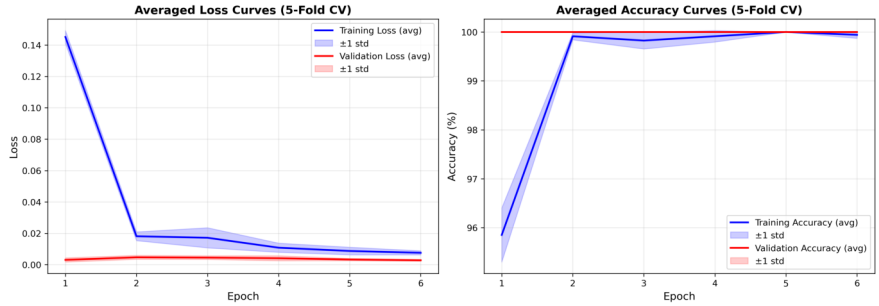


FIGURE 3.63 – MobileNetV2 Modality Classifier Accuracy.

— **Held-out Test Confusion Matrix**

The final classification performance on the test set is summarized in the confusion matrix presented in Figure 3.64. The model achieved reliable discrimination between the three imaging modalities â Slit-Lamp Photography, Fundus Photography, and OCT â with minimal inter-modality confusion, validating the effectiveness of the proposed automated routing mechanism within the overall system pipeline.

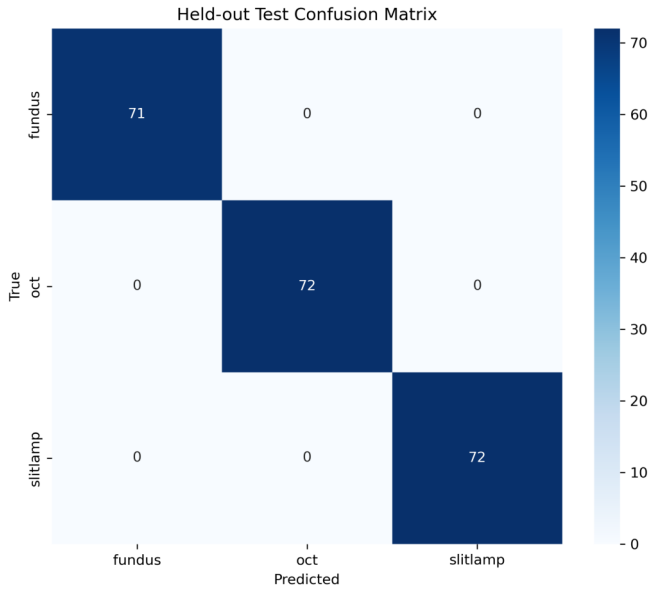


FIGURE 3.64 – MobileNetV2 Modality Classifier Confusion Matrix.

3.24 Comparison of Proposed Models with Existing Approaches

The following tables compare the proposed deep learning models with previous approaches for slit-lamp, fundus, and OCT image classification. The comparison is based on the validation accuracy achieved by each model.

3.24.1 OCT Model Comparison

Model	**Validation Accuracy (%)**
MobileNetV2 [94]	87.40%
Densenet with attention [95]	91%
ViT-Base with Multi-source training [96]	93.61%
Hybrid DL with Ant Colony Optimization [97]	93%
EfficientNet-B0 [98]	94%
InceptionV3 [99]	94.82%
Our proposed model	96.17%

TABLE 3.8 – Comparison of OCT models based on validation accuracy

3.24.2 Fundus Photography Comparison Table

Model	**Validation Accuracy (%)**
2D-CNN [100]	82.13%
MobileNet [95]	89.64%
Inceptionv3 [101]	91.76%
Our proposed model	92.24%

TABLE 3.9 – Comparison of Fundus Photography models based on validation accuracy

3.24.3 Slit Lamp Comparison Table

Model	**Validation Accuracy (%)**
ResNet101 [94]	80%
Vision Transformer ViT-B/16 [102]	82.13%
ResNet50-IBN [103]	94.24%
AutoML [104]	95.70%
Our proposed model	98.17%

TABLE 3.10 – Comparison of Slit Lamp models based on validation accuracy

3.25 Conclusion

In this chapter, we provide an overview of the tools, libraries, and technologies used in developing our system. We also presented experimental results for multimodal deep learning models for ocular disease classification, highlighting their performance across imaging modalities.

Additionally, we compared our proposed models with existing methods to evaluate their performance and effectiveness. The analysis of these results helps identify the system's capabilities and limitations, as well as its potential applications in supporting diagnostic and operational workflows in ophthalmology.

General conclusion and Future work

This thesis successfully developed a scalable, web-based, AI-assisted ophthalmology system that leverages multimodal deep learning models to automate the detection of ocular diseases from slit-lamp, fundus, and OCT images. It integrates an automatic modality classification module, facilitating seamless, user-friendly diagnostic workflows in clinical settings. The system demonstrated high accuracy across various image modalities, enhancing the efficiency and reliability of ophthalmic diagnostics, and provides explainable AI toolsâsuch as attention mapsâthat increase clinical trust and interpretability.

Despite the promising performance of the proposed web-based ocular disease detection system, several avenues for improvement remain. Future work could expand the system to support additional imaging modalities and a wider spectrum of ocular diseases, thereby enhancing its clinical relevance. Increasing dataset size and diversity through multi-institutional and multi-device image collection would strengthen model generalization. Incorporating advanced attention mechanisms and hybrid deep learning architectures could further refine feature extraction and classification, while optimizing models for faster inference would enable deployment in real-time and resource-constrained environments.

Equally important is integrating stronger explainability features by combining Grad-CAM with other XAI techniques to provide richer visual interpretations of

predictions. Enhancements to the web application, such as patient management tools, automated reporting, and cloud-based storage, would improve usability. Embedding the system within telemedicine platforms could expand access in underserved regions, while clinical validation studies with ophthalmologists would ensure reliability in clinical practice. Finally, incorporating complementary medical dataâsuch as patient history, demographics, and results from other ophthalmic examinationsâwould improve diagnostic accuracy and broaden the systemâs applicability across diverse eye diseases.

References

[1] Retina & Eye Consultants, "Eye conditions," https://retinaandeye.com.au/eye-conditions/, n.d., accessed : 2026-01-28.

[2] Mayo Clinic Staff, "Glaucoma : Symptoms and causes," https://www.mayoclinic.org/diseases-conditions/glaucoma/symptoms-causes/syc-20372839, Mayo Clinic, 2024, accessed : 2026-01-28.

[3] Delaware Ophthalmology Consultants, "Cataract," https://www.delawareeyes.com/cataract/, n.d., accessed : 2026-01-28.

[4] Cape Fear Retina, "Macular holes," https://www.capefearretina.com/macular-holes/, 2026, accessed : 2026-04-28.

[5] Lions Eye Institute, "Diabetic retinopathy," https://www.lei.org.au/services/eye-health-information/diabetic-retinopathy/, n.d., accessed : 2026-01-28.

[6] NYC Eye Doctors, "Age-related macular degeneration (amd)," https://www.eyedoctorophthalmologistnyc.com/treatment/age-related-macular-degeneration-amd/, 2026, accessed : 2026-04-28.

[7] EYE-PIX, "Fundus photography," http://eye-pix.com/fundus-photography/, 2016, accessed : 2026-04-17.

[8] I. Leandro, B. Lorenzo, M. Aleksandar *et al.*, "Oct-based deep-learning models for the identification of retinal key signs," *Scientific Reports*, vol. 13, p. 14628, 2023.

[9] Topcon Healthcare, "Slit lamps," https://topconhealthcare.ca/fr/products/slit-lamps/, 2026, accessed : 2026-04-17.

[10] ResearchGate, “Slit lamp photography of the right and left eye showing pigment deposits,” https://www.researchgate.net/figure/Slit-lamp-photography-of-the-right-1a-and-left-1b-eye-showing-pigment-deposits-on-the fig1_373997602, 2026, accessed : 2026-04-17.

[11] NUSOD, “Testing an artificial neural network (ann),” https://nusod.wordpress.com/2020/01/05/testing-an-artificial-neural-network-ann/, 2020, accessed : 2026-04-28.

[12] SuperAnnotate, “Convolutional neural networks : 1998–2023 overview,” https://www.superannotate.com/blog/guide-to-convolutional-neural-networks, 2023, accessed : 2026-05-06.

[13] GeeksforGeeks, “What is fully connected layer in deep learning ?” https://www.geeksforgeeks.org/deep-learning/what-is-fully-connected-layer-in-deep-learning/, 2025, accessed : 2026-05-06.

[14] MI Research, “Supplementary material,” https://www.mi-research.net/supplement/2fd644e5-3c8e-4c47-8342-caccbcb49dc3, 2026.

[15] GeeksforGeeks, “Ml | introduction to transfer learning,” https://www.geeksforgeeks.org/machine-learning/ml-introduction-to-transfer-learning/, 2025.

[16] M. D. AbrÃ moff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, “Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices,” *npj Digital Medicine*, vol. 1, p. 39, 2018.

[17] W. H. Organization, “World report on vision,” Geneva, 2019, iSBN : 978-92-4-151657-0. [Online]. Available : https://iris.who.int/handle/10665/328717

[18] S. R. Flaxman, R. R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen *et al.*, “Global causes of blindness and distance vision impairment 1990–2020 : a systematic

review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 12, pp. e1221–e1234, 2017.

[19] R. Armstrong and R. Cubbidge, "The eye and vision : an overview," *Handbook of Nutrition, Diet and the Eye*, pp. 3–9, 2014.

[20] M. J. Burton, J. Ramke, A. P. Marques, R. R. Bourne, N. Congdon, I. Jones, B. A. A. Tong, S. Arunga, D. Bachani, C. Bascaran *et al.*, "The lancet global health commission on global eye health : vision beyond 2020," *The Lancet Global Health*, vol. 9, no. 4, pp. e489–e551, 2021.

[21] J. Forrester, A. Dick, P. McMenamin, F. Roberts, and E. Pearlman, "The eye : Basic sciences in practice (4. baskı)," 2016.

[22] ScienceDirect, "Ophthalmic disorder," https://www.sciencedirect.com/topics/neuroscience/ophthalmic-disorder, Elsevier, n.d., accessed : 2026-01-28.

[23] Barraquer, "Functions and pathologies of the anterior segment of the eye," https://www.barraquer.com/en/news/functions-pathologies-anterior-segment-eye, 2022, accessed : 2026-01-28.

[24] ——, "Functions and pathologies of the posterior segment of the eye," https://www.barraquer.com/en/news/functions-pathologies-anterior-segment-eye, 2022, accessed : 2026-01-28.

[25] A. K. Schuster, C. Erb, E. M. Hoffmann, T. Dietlein, and N. Pfeiffer, "The diagnosis and treatment of glaucoma," *Deutsches Ärzteblatt International*, vol. 117, no. 13, p. 225, 2020.

[26] I. Inamullah, I. Razzak, and S. Jameel, "The eye as a window to systemic health : A survey of retinal imaging from classical techniques to oculomics," *The Journal of Precision Medicine : Health and Disease*, p. 100023, 2025.

[27] D. C. Hood, "Improving our understanding, and detection, of glaucomatous damage : an approach based upon optical coherence tomography (oct)," *Progress in retinal and eye research*, vol. 57, pp. 46–75, 2017.

[28] A. A. Abdullah, A. Aldhahab, and H. M. Al Abboodi, "Review of eye diseases detection and classification using deep learning techniques," in *BIO Web of Conferences*, vol. 97. EDP Sciences, 2024, p. 00012.

[29] A. Elsawy, T. Eleiwa, C. Chase, E. Ozcan, M. Tolba, W. Feuer, M. Abdel-Mottaleb, and M. Abou Shousha, "Multidisease deep learning neural network for the diagnosis of corneal diseases," *American journal of ophthalmology*, vol. 226, pp. 252–261, 2021.

[30] G. Bikbova, T. Oshitari, T. Baba, S. Yamamoto, and K. Mori, "Pathogenesis and management of macular hole : Review of current advances," *Journal of Ophthalmology*, vol. 2019, no. 1, p. 3467381, 2019. [Online]. Available : https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/3467381

[31] I. Razzak, S. Jameel *et al.*, "The eye as a window to systemic health : A survey of retinal imaging from classical techniques to oculomics," *arXiv preprint arXiv :2505.04006*, 2025.

[32] D. S. Friedman, J. Katz, N. M. Bressler, B. Rahmani, and J. M. Tielsch, "Racial differences in the prevalence of age-related macular degeneration : the baltimore eye survey," *Ophthalmology*, vol. 106, no. 6, pp. 1049–1055, 1999.

[33] G. A. Saleh, N. M. Batouty, S. Haggag, A. Elnakib, F. Khalifa, F. Taher, M. A. Mohamed, R. Farag, H. Sandhu, A. Sewelam, and A. El-Baz, "The role of medical image modalities and ai in the early detection, diagnosis and grading of retinal diseases : A survey," *Bioengineering*, vol. 9, no. 8, p. 366, Aug. 2022, basel ; Epub 2022 Aug 4.

[34] O. Nanegrungsunk, D. Patikulsila, and S. R. Sadda, "Ophthalmic imaging in diabetic retinopathy : A review," *Clinical and Experimental Ophthalmology*, vol. 50, no. 9, pp. 1082–1096, Dec. 2022, epub 2022 Sep 25.

[35] LINK Eye Group, "Sl-400 five-step magnifications slit lamp," https://linkeyegroup.com/product/sl-400-five-step-magnifications-slit-lamp/, 2024, accessed : 2026-04-17.

[36] R. Rocholz, M. M. Teussink, R. Dolz-Marco, C. Holzhey, J. F. Dechent, A. Tafreshi, and S. Schulz, "SPECTRALIS Optical Coherence Tomography Angiography (OCTA) : Principles and Clinical Applications," Heidelberg Engineering GmbH, White paper 210111-001 GL.AE18, 2018, comprehensive overview of SPECTRALIS OCTA module design principles, algorithms, slab definitions, artifact removal, and multimodal imaging. [Online]. Available : https://www.heidelbergengineering.com

[37] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tufail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A. Keane, and O. Ronneberger, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[38] K. Kaur and B. Gurnani, "Slit-lamp biomicroscope," in *StatPearls [Internet].* Treasure Island (FL) : StatPearls Publishing, 2026, updated 2023 Jun 11. [Online]. Available : https://www.ncbi.nlm.nih.gov/books/NBK587440/

[39] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on

color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, 2018.

[40] K. Mansouri, F. A. Medeiros, and R. N. Weinreb, "Global rates of glaucoma surgery," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 251, no. 11, pp. 2609–2615, 2013.

[41] Z. Lan, Y. Hu, S. Yang, M. Wu, Y. Huang, and H. Zhang, "Deep learning in ophthalmic image analysis : A task-driven review of segmentation, diagnosis, and progress prediction," *Neurocomputing*, p. 132282, 2025.

[42] S. Mittal, S. Singh, A. Gupta, P. Sahu, and M. Sharma, "Unveiling visual wellness : An advanced deep learning approach for early eye disease detection and classification," *IAENG International Journal of Computer Science*, vol. 52, no. 7, 2025.

[43] Y. Xiao, X. Ding, S. Liu, Y. Ma, T. Zhang, Z. Xiang, R. Zhang, T. Fukuyama, J. Zhao, Y. Yu *et al.*, "Fusion-attention diagnosis network (fadnet) : An end-to-end framework for optic disc segmentation and ocular disease classification," *Information Fusion*, vol. 124, p. 103333, 2025.

[44] P. U. Pandey, B. G. Ballios, P. G. Christakis, A. J. Kaplan, D. J. Mathew, S. Ong Tone, M. J. Wan, J. A. Micieli, and J. C. Wong, "Ensemble of deep convolutional neural networks is more accurate and reliable than board-certified ophthalmologists at detecting multiple diseases in retinal fundus photographs," *British Journal of Ophthalmology*, vol. 108, no. 3, pp. 417–423, 2024.

[45] N. Kühl, M. Schemmer, M. Goutier, and G. Satzger, "Artificial intelligence and machine learning," *Electronic Markets*, vol. 32, no. 4, pp. 2235–2244, 2022.

[46] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning : C. janiesch et al." *Electronic markets*, vol. 31, no. 3, pp. 685–695, 2021.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[48] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. Academic Press, 1994, pp. 474–485.

[49] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[50] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in *European conference on computer vision*. Springer, 2000, pp. 404–420.

[51] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.

[52] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[53] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[54] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[55] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, "K-nearest neighbor classification," in *Data mining in agriculture*. Springer, 2009, pp. 83–106.

[56] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717–727, 2000. [Online]. Available : https://www.sciencedirect.com/science/article/pii/S0731708599002721

[57] M. Zakaria, A. Mabrouka, and S. Sarhan, "Artificial neural network : a brief overview," *neural networks*, vol. 1, no. 2, 2014.

[58] I. Goodfellow, Y. Bengio, and A. Courville, "Deep feedforward networks," *Deep learning*, vol. 1, pp. 161–217, 2016.

[59] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Forward propagation, backward propagation, and computational graphs," https://d2l.ai/chapter_multilayer-perceptrons/backprop.html, 2023, accessed : 2026-04-28.

[60] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[61] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[62] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv :1409.1556*, 2014.

[63] DataBird, "Qu'est ce qu'un cnn ? un convolutional neural network," https://www.data-bird.co/blog/cnn, 2025, accessed : 2026-05-06.

[64] ApX Machine Learning, "Evolution of cnn architectures : Alexnet to resnet," https://apxml.com/courses/cnns-for-computer-vision/chapter-1-cnn-foundations-modern-architectures/cnn-architecture-evolution, 2026.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[66] G. Huang, Z. Liu, G. Pleiss, L. J. P. van der Maaten, and K. Q. Weinberger, "Convolutional networks with dense connectivity," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, vol. 44, pp. 8704–8716, 2019. [Online]. Available : https://api.semanticscholar.org/CorpusID:167218833

[67] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets : Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv :1704.04861*, 2017.

[68] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, p. 9, 2016.

[69] R. Jha, V. Bhattacharjee, and A. Mustafi, "Transfer learning with feature extraction modules for improved classifier performance on medical image data," *Scientific Programming*, vol. 2022, no. 1, p. 4983174, 2022.

[70] A. Bosco, "Transfer learning : Feature extraction and fine tuning," https://medium.com/data-reply-it-datatech/transfer-learning-feature-extraction-and-fine-tuning-db7d82767992, 2024, published on Medium, accessed 2026-05-08.

[71] P. Mooney, "Kermany2018 : Labeled optical coherence tomography (oct) and chest x-ray images for classification," https://www.kaggle.com/datasets/paultimothymooney/kermany2018, 2018.

[72] G. Doddi, "Eye diseases classification," https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification, 2022.

[73] N. Padia, "Cataract image dataset," https://www.kaggle.com/datasets/nandanp6/cataract-image-dataset, 2023.

[74] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11 782–11 792, 2018.

[75] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam : Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[76] K. Vinogradova, A. Dibrov, and G. Myers, "Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract)," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 10, 2020, pp. 13 943–13 944.

[77] Python Software Foundation, "Python," https://www.python.org/, 2026, accessed : 2026-05-08.

[78] Anaconda, Inc., "Anaconda," https://www.anaconda.com, 2026, accessed : 2026-05-08.

[79] Project Jupyter, "Project jupyter," https://jupyter.org/, 2026, accessed : 2026-05-08.

[80] Real Python, "Google colab," https://realpython.com/ref/tools/google-colab/, 2026, accessed : 2026-05-09.

[81] MDN Web Docs, "Html : Hypertext markup language," https://developer.mozilla.org/fr/docs/Web/HTML, 2026, accessed : 2026-05-09.

[82] J. D. Drake and J. C. Worsley, *Practical PostgreSQL.* " O'Reilly Media, Inc.", 2002.

[83] Amazon Web Services, "Qu'est-ce que javascript ?" https://aws.amazon.com/fr/what-is/javascript/, 2026.

[84] Google Open Source, "Tensorflow," https://opensource.google/projects/tensorflow, 2026, accessed : 2026-05-09.

[85] K. Developers, "Keras : Deep learning for humans," https://keras.io/, n.d., accessed : 2026-05-09.

[86] B. des Talents, "Seaborn," 2026, glossaire tech, consulted May 2026. [Online]. Available : https://www.bureaudestalents.com/glossaire-tech/seaborn

[87] W. C. School, "DÃ©finition matplotlib - qu'est-ce que matplotlib," 2026, lexique tech, accessed 2026-05-09. [Online]. Available : https://www.wildcodeschool.com/lexique/definition-matplotlib

[88] OpenCV Foundation, "About opencv," https://opencv.org/about/, n.d., accessed : 2026-05-09.

[89] Python Packaging Authority, "Pillow 2.2.1," https://pypi.org/project/pillow/2.2.1/, 2013, accessed : 2026-05-09.

[90] S. Ramírez, "Fastapi : Modern, fast web framework for building apis with python," 2018, accessed : 2026-05-25. [Online]. Available : https://fastapi.tiangolo.com/

[91] The Psycopg Team, "Psycopg documentation," https://www.psycopg.org/docs/, 2026.

[92] Meta Platforms, Inc., "React documentation," https://react.dev/, 2026.

[93] ——, "Introducing jsx," https://fr.legacy.reactjs.org/docs/introducing-jsx.html, 2026.

[94] H. Cheng, X. Ning, B. Xu, Y. Qin, C. Li, R. Ling, Y. Shen, W. Jia, and J. Zhong, "Research on deep learning-based lesion identification in optical coherence tomography," *BMC ophthalmology*, 2026.

[95] F. M. Talaat, A. A. A. Ali, R. ElGendy, and M. A. ELShafie, "Deep attention for enhanced oct image analysis in clinical retinal diagnosis," *Neural Computing and Applications*, vol. 37, no. 2, pp. 1105–1125, 2025.

[96] B. Guo, D. Wang, Z. Zhao, W. Liu, J. Hou, R. Liang, and L. Zhang, "Evaluation of classification performance for six types of fundus diseases in oct images based on multi-source training strategy," *Frontiers in Medicine*, vol. 13, p. 1775911, 2026.

[97] S. Agarwal, A. K. Dohare, P. Saxena, J. Singh, I. Singh, and U. K. Sahu, "Hdl-aco hybrid deep learning and ant colony optimization for ocular optical coherence tomography image classification," *Scientific Reports*, vol. 15, no. 1, p. 5888, 2025.

[98] I. D. A. P. P. Tentriajaya and I. G. N. L. Wijayakusuma, "Eye disease classification using efficientnet-b0 based on transfer learning," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1415–1422, 2025.

[99] M. T. Noor, S. Abrar, J. A. Mahi, M. P. Mia, A. Hridoy, and S. Ghosh, "Retinavision : Xai-driven augmented regulation for precise retinal disease classification using deep learning framework," 2026. [Online]. Available : https://arxiv.org/abs/2602.19324

[100] S. Vidivelli, P. Padmakumari, C. Parthiban, A. DharunBalaji, R. Manikandan, and A. H. Gandomi, "Optimising deep learning models for ophthalmological disorder classification," *Scientific Reports*, vol. 15, no. 1, p. 3115, 2025.

[101] Y. Pan, J. Liu, Y. Cai, X. Yang, Z. Zhang, H. Long, K. Zhao, X. Yu, C. Zeng, J. Duan *et al.*, "Fundus image classification using inception v3 and resnet-50 for the early diagnostics of fundus diseases," *Frontiers in Physiology*, vol. 14, p. 1126780, 2023.

[102] D. Z. Chen, C. Liu, J. Wu, L. Zhu, and B. C. Ooi, "Prediction of cataract severity using slit lamp images from a portable smartphone device : A pilot study," *Sensors*, vol. 26, no. 6, p. 1954, 2026.

[103] Z. Cui, Y. Cheng, S. Pan, Y. Zhu, and W. Dai, "A deep learning-driven cataract screening model derived from multicenter real-world dataset," *Frontiers in Medicine*, vol. Volume 12 - 2025, 2025. [Online]. Available : https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1691419

[104] N. Yavari, S. S. Mohammadi, J. T. Sokol, D. E. Feky, M. Rajabi, J.-H. Hung, C. Or, O. Elaraby, F. A. Anover, A.-O. Saengsirinavin *et al.*, "Utilization of an automated machine learning approach for the detection of granular corneal dystrophy via slit lamp photographs," *BMC ophthalmology*, vol. 25, no. 1, p. 657, 2025.

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي و البحث العلمي

جامعة محمد خيضر بسكرة .

عنوان المشروع:

AI-powered ocular image analysis system

مشروع لنيل شهادة مؤسسة ناشئة في اطار القرار الوزاري 1275

صورة العلامة التجارية

OphtaLS

الاسم التجاري

OphtaLS

السنة الجامعية

2023 _ 2022

بطاقة معلومات:

حول فريق الاشراف وفريق العمل

1- فريق الاشراف:

فريق الاشراف	
المشرف الرئيسي (01): مرزيق عبد الحق	التخصص: إعلام ألي
المشرف المساعد: زراري عبد المؤمن	التخصص: إعلام ألي

2- فريق العمل:

فريق المشروع	التخصص	الكلية
الطالبة:سندس فضيل	RTIC	العلوم الدقيقة
الطالبة: فتاح ليلي	RTIC	العلوم الدقيقة

فهرس المحتويات

المحور الأول:تقديم المشروع 4

فكرة المشروع (الحل المقترح) 4

القيم الطبية، التقنية، الاجتماعية والاقتصادية 8

فريق العمل والخلفية المعرفية 8

يتكون فريق العمل من طالبتين تحت إشراف أستاذ مشرف، وقد خضع الفريق لعدة دورات تكوينية تقنية وريادية ساهمت في تطوير مهاراتهم في مجال ريادة الأعمال والابتكار، من بينها: 8

أهداف المشروع 9

الجدول الزمني لتحقيق المشروع 10

المحور الثاني: الجوانب الإبتكارية 11

طبيعة الابتكارات 11

مجالات الابتكارات 12

المحور الثالث: التحليل الاستراتيجي للسوق 13

السوق المحتمل 13

الفئة المستهدفة 13

تحليل المنافسة 14

تحليل SWOT 15

المحور الرابع :خطة الإنتاج والتنظيم 16

الاستراتيجية التسويقية 16

خطوات تشغيل النظام 17

اليد العاملة 18

الشراكات والتعاونات 18

المحور الخامس: الخطة المالية PLAN FINANCIER 20

جدول الاستثمارات والتكاليف الأولية 20

مصادر التمويل المحتملة والمتاحة للمشروع 21

جدول تطور رقم الأعمال السنوي المتوقع 23

جدول السيناريوهات المالية المتوقعة 23
جدول 2.5 السيناريوهات المالية المتوقعة لمشروع منصة عيادات طب العيون 23
جدول حسابات النتائج المتوقعة للسنة الأولى 24
حتياجات رأس المال العامل(BFR) 24
جدول خطة الخزينة الشهرية للسنة الأولى 25
المحور السادس: النموذج الأولي التجريبي 26

المحور الأول:تقديم المشروع

يتمثل مشروعنا في تطوير نظام ذكي لتحليل صور العين الطبية باستخدام تقنيات الذكاء الاصطناعي والتعلم العميق، بهدف مساعدة أطباء العيون في الكشف وتشخيص الأمراض العينية بدقة وسرعة وكفاءة أكبر.

يعتمد النظام على تحليل عدة أنواع من الصور الطبية الخاصة بالعين، مثل صور قاع العين(Fundus Images) ، وصور التصوير المقطعي البصري(OCT) ، بالإضافة إلى صور المصباح الشقي(Slit Lamp) ، وذلك باستخدام نماذج ذكاء اصطناعي مدربة على قواعد بيانات طبية متخصصة.

يقوم النظام بمعالجة الصور الطبية واستخراج الخصائص المهمة منها، ثم تصنيفها للكشف عن مجموعة من أمراض العيون الشائعة والخطيرة، مثل اعتلال الشبكية السكري، الوذمة البقعية السكرية(DME) ، المياه الزرقاء (Glaucoma)، الساد(Cataract) ، إضافة إلى التمييز بين الحالات السليمة والمصابة.

كما يوفر المشروع واجهة استخدام تفاعلية وسهلة تسمح للأطباء برفع الصور الطبية وتحليلها بشكل سريع، مع عرض نتائج التشخيص بطريقة واضحة تساعد في دعم القرار الطبي. ويدعم النظام كذلك تقنيات الذكاء الاصطناعي التفسيري مثلGrad-CAM ، التي تسمح بإظهار المناطق المهمة داخل الصورة والتي اعتمد عليها النموذج أثناء عملية التنبؤ، مما يزيد من شفافية النتائج وثقة الأطباء في النظام.

ويهدف المشروع إلى المساهمة في تطوير حلول صحية ذكية حديثة تساعد على تحسين جودة التشخيص الطبي وتسريع عملية الكشف المبكر عن أمراض العيون باستخدام التكنولوجيا الحديثة، بما ينعكس إيجابًا على تحسين رعاية المرضى وتقليل نسب التشخيص المتأخر.

فكرة المشروع (الحل المقترح)

تمت دراسة واقع التشخيص الطبي لأمراض العيون، حيث يعتمد الأطباء بشكل كبير على الخبرة البشرية وتحليل الصور الطبية بشكل يدوي، مثل صور قاع العين(Fundus) ، والتصوير المقطعي البصري(OCT) ، وصور المصباح الشقي Slit) .(Lampورغم التطور الطبي، إلا أن هذا الأسلوب يعاني من عدة تحديات أهمها بطء عملية التشخيص، تفاوت الدقة بين الأطباء، وصعوبة الكشف المبكر عن بعض الأمراض المعقدة، خاصة في المناطق التي تعاني من نقص في أطباء العيون المختصين.

كما تعاني المنظومة التقليدية من محدودية استخدام الأدوات الذكية في تحليل الصور الطبية، واعتمادها على التفسير البصري فقط، مما قد يؤدي إلى تأخر في التشخيص أو عدم دقة في تحديد المرحلة المرضية، إضافة إلى عدم توفر أنظمة مساعدة قرار تعتمد على الذكاء الاصطناعي بشكل واسع.

ولمعالجة هذه الإشكاليات، تم اقتراح تطوير نظام ذكي يعتمد على تقنيات **الذكاء الاصطناعي والتعلم العميق (Deep Learning)**لتحليل صور العين الطبية بشكل تلقائي، بهدف دعم الأطباء في اتخاذ القرار وتحسين دقة وسرعة التشخيص.

يعتمد الحل المقترح على بناء نموذج ذكاء اصطناعي (CNN-based Model) مدرّب على قواعد بيانات طبية متخصصة، قادر على تحليل الصور واستخراج الخصائص المهمة منها، ثم تصنيفها للكشف عن مجموعة من أمراض العيون الشائعة والخطيرة مثل: اعتلال الشبكية السكري(Diabetic Retinopathy) ، الوذمة البقعية السكرية(DME) ، التنكس البقعي المرتبط بالعمر(AMD) ، المياه الزرقاء(Glaucoma) ، الساد(Cataract) ، إضافة إلى تحديد الحالات السليمة.

ويهدف النظام إلى توفير منصة ذكية متكاملة تمكن الأطباء من رفع الصور الطبية بسهولة، والحصول على نتائج تحليل فورية مدعومة بنسبة ثقة(Confidence Score) ، مما يساعد على تحسين جودة القرار الطبي وتقليل الأخطاء البشرية.

كما يدعم المشروع تقنيات الذكاء الاصطناعي التفسيري (Explainable AI) مثل **Grad-CAM**، والتي تسمح بتوضيح المناطق الأكثر تأثيرًا في الصورة والتي اعتمد عليها النموذج في عملية التصنيف، مما يعزز شفافية النتائج ويزيد من ثقة المستخدمين في النظام.

ويتم تجسيد الحل المقترح من خلال المراحل التالية:

1بناء نموذج الذكاء الاصطناعي

يتم تطوير نموذج تعلم عميق اعتمادًا على شبكات عصبية تلافيفية (CNN) مثلMobileNetV2 ، مع تحسين الأداء باستخدام تقنيات حديثة، بهدف تصنيف صور العين بدقة عالية.

2معالجة البيانات الطبية

تشمل هذه المرحلة:

- جمع الصور الطبية (Fundus, OCT, Slit Lamp)
- تنظيف البيانات وتحسين الجودة
- توحيد حجم الصور
- تقسيم البيانات إلى تدريب، اختبار، وتحقق

3تطوير نظام التحليل

يتم دمج النموذج داخل نظام برمجي يقوم بـ:

- استقبال الصور الطبية
- تحليلها بشكل تلقائي
- استخراج الخصائص المهمة
- إصدار نتائج التشخيص

4تطوير واجهة المستخدم(User Interface)

يتم تصميم واجهة ويب بسيطة وسهلة الاستخدام تمكّن الأطباء من:

- رفع الصور الطبية
- تحليل صورة واحدة أو عدة صور دفعة واحدة (Batch Processing)
- عرض النتائج بشكل فوري
- إظهار نسبة الثقة ونتائج التصنيف

5دمج تقنيات التفسير الطبي(Explainability)

يتم استخدام Grad-CAM لتوضيح المناطق المهمة داخل الصورة، مما يساعد الطبيب على فهم سبب قرار النموذج.

6 .الاختبار والتقييم(Testing Phase)

يتم اختبار النظام باستخدام بيانات حقيقية أو تجريبية من أجل:

- قياس دقة النموذج (Accuracy)
- تقييم الأداء باستخدام F1-score و Recall
- تحسين النتائج بناءً على الأخطاء

7النشر التجريبي(Pilot Deployment)

يتم تشغيل النظام بشكل تجريبي داخل بيئة محدودة (عيادات أو مختبرات) بهدف جمع الملاحظات وتحسين الأداء قبل التعميم.

8التوسع المستقبلي

يمكن مستقبلاً تطوير النظام ليصبح منصة طبية متكاملة تدعم:

- التكامل مع المستشفيات
- ربطه بأنظمة الملفات الطبية (EHR)
- دعم التشخيص متعدد الأمراض وبشكل أوسع

يعتمد الحل المقترح على تدريب نماذج ذكاء اصطناعي باستخدام قواعد بيانات طبية متخصصة، مما يسمح للنظام بالتعرف على مجموعة من أمراض العيون مثل اعتلال الشبكية السكري، الوذمة البقعية السكرية(DME) ، المياه الزرقاء (Glaucoma)، والساد(Cataract) ، إضافة إلى التمييز بين الحالات السليمة والمصابة.

كما يوفر النظام واجهة استخدام تفاعلية تسمح للأطباء برفع الصور الطبية وتحليلها بشكل سريع، مع عرض النتائج بطريقة واضحة وسهلة الفهم. ويدعم المشروع أيضًا تقنيات الذكاء الاصطناعي التفسيري مثلGrad-CAM ، التي تساعد على إظهار المناطق المهمة داخل الصورة والتي اعتمد عليها النموذج أثناء عملية التنبؤ، مما يعزز شفافية النظام وثقة الأطباء في النتائج.

القيم الطبية، التقنية، الاجتماعية والاقتصادية

البعد	القيم المقترحة في المشروع
القيم الطبية	-تحسين دقة التشخيص المبكر لأمراض العين باستخدام الذكاء الاصطناعي . -تقليل الوقت المستغرق في تحليل صور قاع العين و OCT مقارنة بالتحليل اليدوي . -دعم الأطباء في اتخاذ القرار الطبي من خلال تقديم نتائج مساعدة وموثوقة . -المساهمة في الكشف عن الأمراض البصرية مثلMH، DME، و.Glaucoma
القيم التقنية	-توظيف تقنيات التعلم العميق (Deep Learning) في تحليل الصور الطبية . -استخدام تقنيات الرؤية الحاسوبية (Computer Vision) لاستخراج المعلومات من الصور . -تطوير نظام ذكي قابل للتحديث والتحسين حسب البيانات الطبية الجديدة . -دمج تقنيات التفسير (Explainability) مثل Grad-CAM لزيادة الشفافية والثقة.
القيم الاجتماعية	-تحسين جودة الرعاية الصحية وتسريع عملية التشخيص . -تقليل الضغط على الأطباء في المستشفيات والمراكز الطبية . -تعزيز الوصول إلى التشخيص المتقدم في المناطق ذات الإمكانيات المحدودة . -المساهمة في تقليل حالات فقدان البصر عبر الكشف المبكر.
القيم الاقتصادية	-تقليل تكاليف الفحوصات الطبية عبر التشخيص الأولي الذكي . -تقليل التنقل بين المراكز الطبية وتحسين كفاءة النظام الصحي . -دعم تطوير حلول صحية محلية منخفضة التكلفة . -فتح المجال أمام استثمارات في مجال الذكاء الاصطناعي الطبي.

جدول1.1 القيم الطبية، التقنية، الاجتماعية والاقتصادية المستهدفة في المشروع المقترح

فريق العمل والخلفية المعرفية

يتكون فريق العمل من طالبتين تحت إشراف أستاذ مشرف، وقد خضع الفريق لعدة دورات تكوينية تقنية وريادية ساهمت في تطوير مهاراتهم في مجال ريادة الأعمال والابتكار، من بينها:

- ✓ الملكية الفكرية لحماية الأفكار الابتكارية .
- ✓ إعداد نموذج العمل التجاري .(Business Model Canvas - BMC)
- ✓ التطبيقات المالية والمحاسبية في الشركات الناشئة .

- ✓ القيم المضافة في الشركات الناشئة/ التسويق الرقمي في الشركات الناشئة.

أهداف المشروع

الأهداف قصيرة المدى (من 6 إلى 12 شهر)

- ✓ تطوير نموذج أولي (PoC) للنظام يهدف إلى عرض الوظائف الأساسية، مثل رفع وتحليل صور العين الطبية والكشف الأولي عن أمراض العيون باستخدام الذكاء الاصطناعي.
- ✓ تدريب واختبار نماذج التعلم العميق على صور Fundus و OCTبهدف تحقيق نتائج دقيقة وموثوقة في التصنيف والتشخيص.
- ✓ إنشاء واجهة استخدام تفاعلية تسمح للأطباء برفع الصور الطبية والحصول على نتائج التحليل بشكل سريع وسهل.
- ✓ اختبار النظام داخل بيئة طبية محدودة من أجل تقييم الأداء واكتشاف النقائص أو المشاكل المحتملة.
- ✓ جمع ملاحظات الأطباء والمستخدمين والعمل على تحسين دقة النموذج ووظائف المنصة وفق الاحتياجات الواقعية.

الأهداف متوسطة وطويلة المدى (من 2 إلى 4 سنوات)

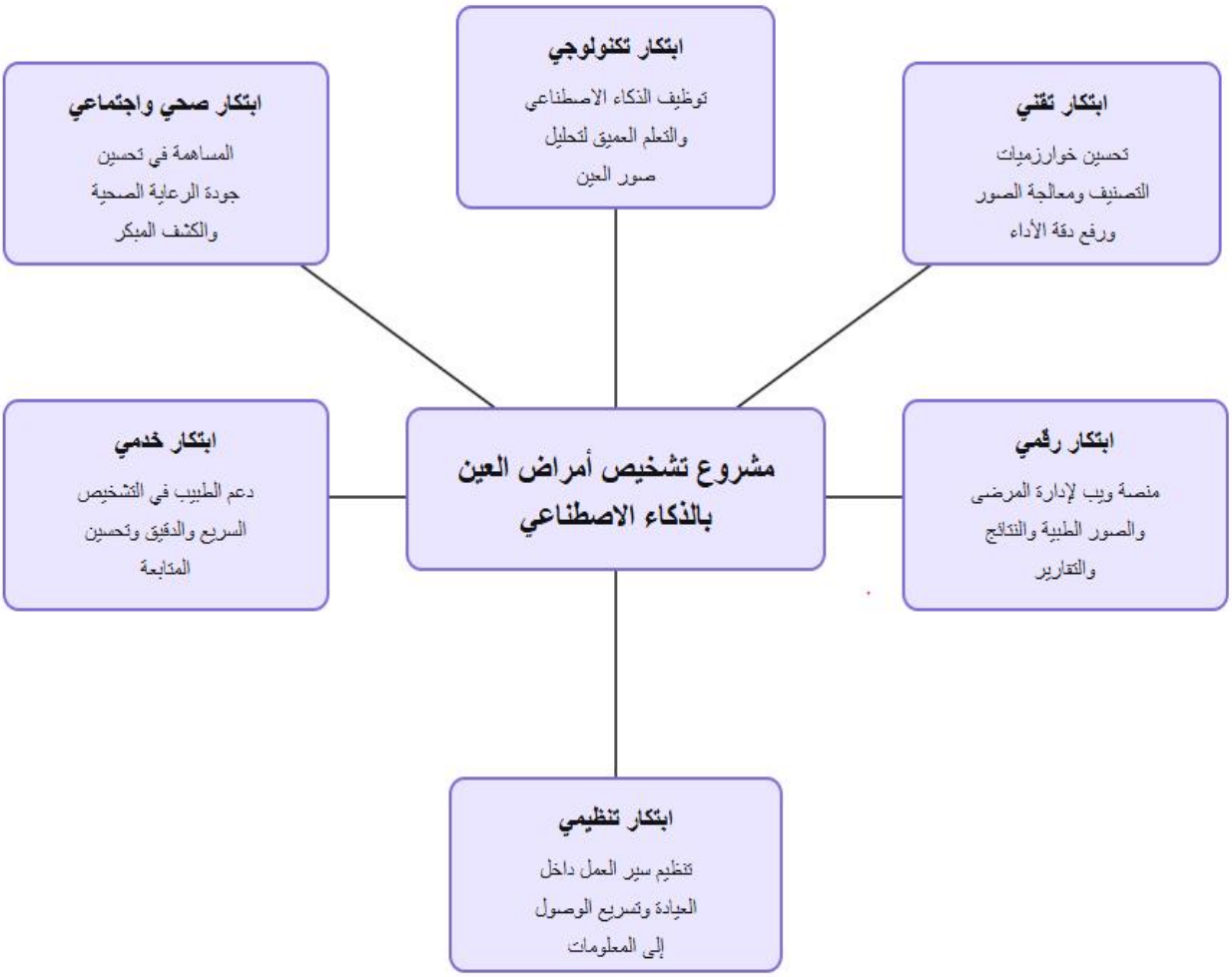
- ✓ توسيع نطاق استخدام المنصة ليشمل المستشفيات، العيادات، ومراكز طب العيون على مستوى أوسع.
- ✓ تحسين دقة النماذج الذكية بشكل مستمر من خلال تدريبها على قواعد بيانات طبية أكبر وأكثر تنوعًا.
- ✓ تحقيق التكامل مع الأنظمة الصحية الرقمية وملفات المرضى الإلكترونية لتسهيل تبادل البيانات الطبية.
- ✓ إدماج تقنيات حديثة مثل:
  1. الذكاء الاصطناعي التفسيري (Explainable AI) لتوضيح أسباب التنبؤات الطبية.
  2. الحوسبة السحابية (Cloud Computing) لتسهيل تخزين وتحليل الصور الطبية.
  3. لوحات تحكم تحليلية متقدمة لمساعدة الأطباء في متابعة الحالات والإحصائيات.
- ✓ تطوير منصة ذكية متكاملة لتحليل وتشخيص أمراض العيون، مع إمكانية توسيعها مستقبلًا لتشمل:
  1. الكشف عن أمراض طبية أخرى اعتمادًا على الصور الطبية.
  2. دعم أنواع إضافية من صور العين والفحوصات الطبية.
  3. دمج تقنيات المراقبة والمتابعة الطبية الذكية.

الجدول الزمني لتحقيق المشروع

المرحلة / الشهر	الشهر 1-2	الشهر 3-4	الشهر 5-6	الشهر 7-8	الشهر 9-10	الشهر 11-12
دراسة وتحليل المتطلبات - جمع قاعدة بيانات صور العين	✓					
معالجة وتحضير الصور الطبية - تطوير النموذج الأول		✓				
طوير النموذج الثاني (تصنيف الأمراض) - تدريب النماذج وتحسينها			✓			
تطوير الواجهة الأمامية (Frontend) - Dashboard للممرض والطبيب				✓		
تطوير الخلفية (Backend) - ربط النماذج - إنشاء قاعدة البيانات					✓	
دمج واختبار النظام - توثيق المشروع - إعداد التقرير النهائي						✓

الجدول 2.1 الجدول الزمني لتحقيق مشروع منصة تحليل صور العين بالذكاء الاصطناعي

المحور الثاني: الجوانب الإبتكارية



الشكل1.2الأبعاد الابتكارية لمشروع تشخيص أمراض العين بالذكاء الاصطناعي.

طبيعة الابتكارات

تتمثل طبيعة الابتكارات في مشروعنا في كونه يجمع بين الابتكار التكنولوجي والابتكار الرقمي والابتكار الخدمي داخل منصة واحدة متكاملة، لأن الابتكار لا يقتصر على تقديم منتج جديد فقط، بل يشمل أيضًا تطوير أساليب جديدة للحل وتحسين الخدمات والعمليات.

ابتكار تكنولوجي :يتجسد في توظيف تقنيات الذكاء الاصطناعي والتعلم العميق في تحليل صور العين والكشف المبكر عن الأمراض، وهو ما يعكس استخدام تقنيات متقدمة لتحسين الأداء والنتائج.

ابتكار رقمي :يظهر في تحويل عملية التشخيص والمتابعة إلى بيئة رقمية متكاملة تربط بين النماذج الذكية وواجهة الاستخدام وقاعدة البيانات في نظام واحد.

ابتكار خدمي :يتمثل في تقديم خدمة صحية مساعدة للطبيب من خلال تسريع التشخيص وتحسين دقته ودعم اتخاذ القرار الطبي.

ابتكار اجتماعي/صحي: يتجسد في الأثر الإيجابي للمشروع على المنظومة الصحية من خلال تحسين جودة الرعاية، وتيسير

ابتكار في العمليات :يبرز من خلال اقتراح طريقة عمل جديدة في معالجة الصور الطبية وتخزين نتائج المرضى وعرض التقارير بشكل منظم، بما يرفع الكفاءة التشغيلية ويقلل الجهد والوقت.

ابتكار تدريجي ذو قيمة تطبيقية :لأن المشروع لا يكتفي بفكرة نظرية، بل يحولها إلى تطبيق عملي قابل للاستعمال، مع إدخال تحسينات تدريجية على أساليب التشخيص والخدمة الصحية.

مجالات الابتكارات

تظهر مجالات الابتكار في المشروع في عدة جوانب مترابطة، لأن الابتكار يمكن أن يشمل المنتج والخدمة والعملية والتنظيم والنموذج الرقمي.

المجال الطبي الصحي :من خلال المساهمة في الكشف المبكر عن أمراض العيون ودعم الأطباء في التشخيص.

المجال التكنولوجي :عبر استخدام الشبكات العصبية العميقة، ومعالجة الصور الطبية، ودمج نماذج تصنيف ذكية داخل النظام.

المجال الرقمي والمعلوماتي :من خلال إنشاء تطبيق ويب يتيح إدارة المستخدمين والمرضى والنتائج والتقارير بشكل رقمي.

مجال إدارة الخدمات الصحية :لأن النظام يساعد على تنظيم البيانات الطبية، تقليل الوقت، وتحسين سير العمل داخل العيادات أو المؤسسات الصحية.

المجال البحثي والأكاديمي :لأن المشروع يربط بين الجانب النظري في الذكاء الاصطناعي والجانب التطبيقي في الصحة، مما يجعله نموذجًا بحثيًا قابلًا للتطوير لاحقًا.

المجال الاجتماعي :عبر تحسين الوصول إلى التشخيص السريع والمساهمة في الرفع من جودة الرعاية الصحية، وهو ما ينسجم مع أثر الابتكار في تحسين نوعية الحياة والخدمات.

المحور الثالث: التحليل الاستراتيجي للسوق

السوق المحتمل

يتمثل السوق المحتمل لمشروعنا في المؤسسات والجهات التي تحتاج إلى حلول رقمية ذكية للكشف المبكر عن أمراض العيون وتحسين تسيير المعلومات الطبية، خاصة مع نمو الطلب على أدوات الذكاء الاصطناعي في طب العيون وعلى خدمات الفحص عن بعد والتشخيص السريع. السوق لا يقتصر على عيادات العيون فقط، بل يمتد إلى المستشفيات، مراكز التصوير الطبي، العيادات متعددة التخصصات، وبرامج الفحص الوقائي لمرضى السكري.

ويمكن تقسيم السوق المحتمل إلى:

العيادات الخاصة لطب العيون التي تحتاج إلى تسريع التشخيص وتحسين إدارة الملفات الطبية.

المستشفيات والمراكز الطبية التي ترغب في دمج الذكاء الاصطناعي داخل خدمات الفحص والتصوير.

برامج الكشف المبكر خصوصًا لاعتلال الشبكية السكري، لأن هذا المجال يعد من أكثر تطبيقات الذكاء الاصطناعي نضجًا في طب العيون.

الأسواق الناشئة والمناطق قليلة التغطية الطبية حيث يمكن للمنصات الذكية أن تقلل أثر نقص المختصين وتدعم الوصول إلى التشخيص.

الفئة المستهدفة

تمثل الفئة المستهدفة الأساسية في **عيادات طب العيون**، حيث يهدف النظام إلى مساعدة الأطباء وطاقم العيادة في تحليل صور العين، تنظيم ملفات المرضى، وتسريع إجراءات التشخيص والمتابعة بشكل أكثر دقة وفعالية.يمكنك عرضها في التقرير بهذا الشكل:

العملاء المباشرون: عيادات طب العيون.

المستخدمون المباشرون للنظام :الأطباء، الممرضون، والتقنيون داخل العيادة.

المستفيد النهائي: المريض، لأنه يستفيد من تشخيص أسرع، متابعة أدق، وتنظيم أفضل للرعاية الصحية.

تحليل المنافسة

توجد المنافسة في هذا المجال أساسًا على المستوى الدولي من خلال منصات متخصصة في الفحص العيني المدعوم بالذكاء الاصطناعي، خاصة في مجال اعتلال الشبكية السكري. من أبرز هذه الحلول **EyeArt** من شركة Eyenuk و-**IDx DR** / **LumineticsCore**، وهما من أبرز الأنظمة التي طُوّرت لاستخدامها في نقاط الرعاية الصحية، مع قوة واضحة من حيث الاعتماد التنظيمي والخبرة السوقية.

وقد حصلت بعض هذه الحلول على اعتماد تنظيمي مهم مثل اعتمادFDA ، كما توسعت في عدة دول ومواقع طبية، وهو ما يمنحها سمعة قوية في الأسواق المتقدمة. لكن رغم هذه القوة، فإن أغلب هذه المنصات تكون موجهة لسياقات دولية محددة، وغالبًا ما تكون مرتفعة الكلفة أو غير مكيّفة بالكامل مع احتياجات المؤسسات الصحية المحلية.

السوق الجزائري

أما في **السوق الجزائري**، فالمنافسة المباشرة في مجال منصات الذكاء الاصطناعي المخصصة لطب العيون ما تزال محدودة مقارنة بالسوق العالمي، خصوصًا إذا كان الحل يجمع بين **تحليل صور العين** و**إدارة المرضى** داخل منصة واحدة. وهذا يعني أن المشروع لا يواجه بالضرورة منافسًا محليًا قويًا بنفس درجة المنصات العالمية، بل يواجه أكثر **فراغًا سوقيًا** يمكن استثماره إذا تم تقديم حل عملي ومناسب للبيئة المحلية.

العنصر	المنافسون الدوليون	مشروع
نوع الحل	حلول متخصصة غالبًا في فحص مرض محدد	منصة تجمع التشخيص وإدارة المرضى
الاعتماد والتنظيم	قوية في الأسواق المنظمة مثل FDA	فرصة للتطوير المحلي والتكييف التدريجي
الكلفة	عادة أعلى وموجهة لأسواق متقدمة	إمكانية تقديم حل أقل تكلفة
التخصيص	محدود حسب المنتج التجاري	أعلى، خاصة حسب احتياجات العيادات المحلية
اللغة والبيئة المحلية	غالبًا موجّهة لسوق دولية	ملائم أكثر للسياق المحلي

تحليلSWOT

نقاط القوة

- دمج الذكاء الاصطناعي مع منصة ويب لإدارة المرضى والنتائج في حل واحد.
- الاستجابة لاتجاه متنامٍ في الذكاء الاصطناعي الطبي وطب العيون.
- دعم الأطباء في تحليل صور العين بطريقة أسرع وأكثر تنظيمًا.
- قابلية التخصيص حسب احتياجات عيادات طب العيون.

نقاط الضعف

- الحاجة إلى بيانات طبية كافية ومتنوعة لرفع الدقة والموثوقية.
- غياب الاعتماد التنظيمي التجاري في المراحل الأولى.
- الحاجة إلى بناء ثقة المستخدمين الطبيين قبل التبني الواسع.
- محدودية الموارد مقارنة بالشركات الدولية الكبرى.

الفرص

- النمو السريع لسوق الذكاء الاصطناعي في طب العيون والتصوير الطبي.
- زيادة الحاجة إلى حلول رقمية داخل عيادات طب العيون.
- إمكانية الشراكة مع العيادات وبرامج الصحة الرقمية.
- نقص بعض الحلول المحلية الموجهة للسياق العربي أو الإقليمي.

التهديدات

- وجود منافسين عالميين لديهم اعتماد تنظيمي وانتشار سوقي.
- التحديات القانونية والأخلاقية المرتبطة بالبيانات الطبية والذكاء الاصطناعي.
- تغيرات سريعة في التكنولوجيا قد تتطلب تحديثًا مستمرًا للنظام.
- مقاومة بعض المؤسسات للتغيير أو بطء التحول الرقمي.

المحور الرابع :خطة الإنتاج والتنظيم

الاستراتيجية التسويقية

تعتمد الاستراتيجية التسويقية على **استهداف عيادات طب العيون مباشرة** من خلال إبراز قيمة النظام في **تحليل صور العين، تنظيم ملفات المرضى، وتسريع العمل داخل العيادة** .الفكرة ليست البيع للجميع، بل بناء رسالة واضحة تقول إن هذا النظام صُمم خصيصًا لاحتياجات طبيب العيون وطاقم العيادة.

1.استراتيجية التركيز

نستخدم **استراتيجية تركيز** لأن المشروع لا يستهدف السوق كله، بل شريحة محددة جدًا هي عيادات طب العيون. هذا يساعد على فهم احتياجاتهم بدقة وتقديم حل يناسبهم بدل تقديم منتج عام.

2.عرض القيمة

عرض القيمة يجب أن يكون بسيطًا ومباشرًا:

- تحليل أسرع لصور العين.
- إدارة أفضل للمرضى والنتائج.
- تقليل الوقت الضائع في التنظيم اليدوي.
- واجهة سهلة تناسب العمل اليومي في العيادة.

3.قنوات التسويق

بما أن الجمهور صغير ومحدد، فالأفضل التركيز على:

- الزيارات المباشرة للعيادات.
- العلاقات المهنية مع أطباء العيون.
- العرض التجريبي للنظام.
- الموقع الإلكتروني أو صفحة تعريفية احترافية.
- التواصل عبر البريد أو الهاتف مع العيادات المهتمة.

4.استراتيجية الترويج

الترويج يكون عمليًا أكثر من كونه إعلانيًا:

- تقديم نسخة تجريبية للنظام.
- عرض صور ولقطات من واجهة التطبيق.
- شرح الفوائد بلغة طبية بسيطة.
- إظهار كيف يوفر النظام الوقت ويحسن التنظيم داخل العيادة.

5.التسعير

اعتماد **سعر مناسب للعيادات الصغيرة والمتوسطة** أو حتى نموذج مرن حسب عدد المستخدمين أو عدد الملفات. هذا مهم لأن الحلول الدولية غالبًا تكون أعلى كلفة وأقل ملاءمة للسوق المحلي.

خطوات تشغيل النظام

يمر تشغيل النظام بعدة مراحل منظمة تبدأ من تسجيل الدخول إلى المنصة، ثم إدخال بيانات المريض ورفع صورة العين، وبعدها يقوم النظام بمعالجة الصورة وتحليلها اعتمادًا على النموذج الذكي، ليعرض النتيجة للطبيب بشكل واضح. بعد ذلك يمكن حفظ النتيجة داخل قاعدة البيانات وربطها بملف المريض لإتاحة المتابعة لاحقًا بطريقة منظمة.

وتتمثل الخطوات العملية في:

- تسجيل دخول الطبيب أو المستخدم المخول.
- إدخال بيانات المريض.
- رفع صورة العين أو ملف الفحص.
- تنفيذ التحليل بواسطة نموذج الذكاء الاصطناعي.
- عرض النتيجة الطبية المبدئية.
- حفظ التقرير داخل النظام.
- تمكين المتابعة أو الطباعة عند الحاجة.

اليد العاملة

بما أن المشروع نظام رقمي، فإن اليد العاملة المطلوبة ليست كبيرة، بل تركز على الجوانب التقنية والتنظيمية. في المرحلة الأولى يكفي فريق صغير يضم مطورًا أو أكثر، إضافة إلى شخص مسؤول عن الاختبارات أو الدعم التقني، ومعه عند الحاجة مختص طبي لمراجعة النتائج وتقييم دقة المخرجات.

ويمكن تلخيص اليد العاملة كما يلي:

- **مطور برمجي** :لتطوير الواجهة والوظائف الأساسية.
- **مختص في الذكاء الاصطناعي أو البيانات** :لتحسين نموذج التحليل.
- **مختص طبي أو طبيب عيون** :لمراجعة النتائج والتأكد من ملاءمتها.
- **مسؤول تشغيل/دعم** :لمتابعة الاستخدام والتحديثات.

الشراكات والتعاونات

شراكات طبية وعلاجية

- شراكة مع أطباء العيون والعيادات المتخصصة في طب العيون.
- دعم علمي وطبي من طرف مختصين في أمراض العين والأشعة الطبية.
- الاستفادة من خبرة الممارسين الطبيين في تقييم دقة النتائج وتحسين جودة التشخيص.

شراكات بحثية

- التعاون مع مخابر البحث الجامعية في الذكاء الاصطناعي والرؤية الحاسوبية.
- تطوير النظام مع باحثين في مجال معالجة الصور الطبية.
- اختبار النماذج وتحسينها اعتمادًا على بيانات حقيقية ومراجعة علمية.

شراكات حكومية

- التنسيق مع وزارة الصحة أو المؤسسات الصحية العمومية.
- العمل مع مراكز الصحة الرقمية لدعم التحول الرقمي في قطاع طب العيون.

- إمكانية إدماج النظام مستقبلًا في برامج الفحص والمتابعة الطبية.

شراكات تسويقية وتوزيعية

- نشر النظام والتعريف به من خلال عيادات طب العيون.
- التعاون مع مستشفيات ومراكز طبية لتجربة الحل واعتماده تدريجيًا.
- استخدام المعارض الطبية والملتقيات العلمية كوسيلة لتسويق المشروع.

المحور الخامس: الخطة المالية PLAN FINANCIER

جدول الاستثمارات والتكاليف الأولية

نوع الاستثمار	الكمية	التكلفة الوحدوية (دج)	التكلفة الإجمالية (دج)
خادم أو استضافة سحابية للمنصة	1	180,000	180,000
أجهزة حاسوب للتطوير والإدارة	2	85,000	170,000
اشتراك إنترنت عالي التدفق (12 شهرًا)	12	7,500	90,000
رواتب فريق التطوير التقني (2 أشخاص × 12 شهرًا)	24	75,000	1,800,000
تهيئة قاعدة البيانات والبيانات الطبية	1	120,000	120,000
تطوير واختبار منصة تشخيص أمراض العيون	1	220,000	220,000
خدمات قانونية واستشارية	1	100,000	100,000
التسويق والتعريف بالمشروع لدى عيادات طب العيون	1	120,000	120,000

|المجموع الكلي | **2,700,000** | | |

مصادر التمويل المحتملة والمتاحة للمشروع

يمكن تمويل مشروع **منصة ذكية لعيادات طب العيون** من خلال عدة مصادر تمويل، سواء حكومية أو خاصة، وذلك نظرًا لطبيعته الرقمية وارتباطه بقطاع الصحة والتحول الرقمي. ويهدف هذا التنوع في التمويل إلى توفير الموارد اللازمة لتطوير المنصة، تشغيلها، وتحقيق الاستقرار المالي في المراحل الأولى من المشروع.

الجهات الحكومية المباشرة

يمكن الاستفادة من تمويل أو دعم بعض الجهات الحكومية المرتبطة بقطاع الصحة أو الرقمنة، ومن بينها:

- وزارة الصحة.
- وزارة اقتصاد المعرفة والمؤسسات الناشئة والمؤسسات المصغرة.
- وزارة البريد والمواصلات السلكية واللاسلكية.
- الوكالة الوطنية لترقية وتطوير الحظائر التكنولوجية.
- برامج دعم مشاريع التحول الرقمي والخدمات الصحية الذكية.

المستثمرون الأفراد

يمكن للمشروع جذب مستثمرين أفراد مهتمين بمجالات:

- التكنولوجيا الصحية.
- الذكاء الاصطناعي في الطب.
- التحول الرقمي.
- الأنظمة الذكية الخاصة بالمؤسسات الصحية.

ويتميز هذا النوع من التمويل بإمكانية الحصول على:

- دعم مالي أولي.
- توجيه وخبرة تقنية.
- علاقات مهنية تساعد على توسيع المشروع.

الدعم الحكومي المباشر

يمكن الاستفادة من برامج الدعم الحكومي الموجهة للمؤسسات الناشئة والمشاريع الابتكارية، مثل:

- حاضنات الأعمال ومسرعات المؤسسات الناشئة.
- برامج دعم الابتكار الرقمي.
- صناديق تمويل المشاريع التكنولوجية.
- إعفاءات أو تسهيلات ضريبية للمؤسسات الناشئة.

ويساعد هذا الدعم في تقليل الأعباء المالية خلال المراحل الأولى من تطوير المشروع.

القروض البنكية

يمكن الاعتماد على القروض البنكية لتمويل جزء من التكاليف الأولية للمشروع، خاصة المتعلقة بـ:

- شراء الخوادم والمعدات.
- تطوير البنية التحتية التقنية.
- تغطية تكاليف التشغيل الأولى.

ومن بين الجهات الممكن التعامل معها:

- البنوك العمومية الجزائرية.
- بنوك أو هيئات تدعم المؤسسات الناشئة.
- مؤسسات تمويل موجهة للمشاريع الرقمية.

جدول تطور رقم الأعمال السنوي المتوقع

الملاحظات	رقم الأعمال السنوي (دج)	متوسط قيمة العقد (دج)	عدد العقود السنوية	السنة
انطلاق أولي مع عدد محدود من عيادات العيون	1,000,000	500,000	2	N+1
توسع تدريجي داخل الولاية أو الجهة	2,200,000	550,000	4	N+2
زيادة عدد العيادات المستفيدة	3,600,000	600,000	6	N+3
توسع أكبر وتحسن الانتشار	5,850,000	650,000	9	N+4
استقرار نسبي ونمو واضح	8,400,000	700,000	12	N+5

جدول1.5 تطور رقم الأعمال السنوي المتوقع لمشروع منصة عيادات طب العيون

جدول السيناريوهات المالية المتوقعة

السيناريو المتشائم (دج)	السيناريو المتوسط (دج)	السيناريو المتفائل (دج)	السنة
1,400,000	2,000,000	3,200,000	N+1
3,100,000	4,640,000	6,800,000	N+2
5,400,000	8,060,000	10,500,000	N+3
8,200,000	12,240,000	16,000,000	N+4
11,500,000	17,520,000	23,500,000	N+5

جدول 2.5 السيناريوهات المالية المتوقعة لمشروع منصة عيادات طب العيون

جدول حسابات النتائج المتوقعة للسنة الأولى

البيان	المبلغ (دج)
رقم الأعمال السنوي	2,000,000
تكاليف الخوادم والصيانة	350,000
رواتب فريق العمل	1,800,000
مصاريف تشغيلية وإدارية	290,000
إجمالي المصاريف	2,440,000
النتيجة الصافية المتوقعة	-440,000

جدول3.5 حسابات النتائج المتوقعة للسنة الأولى لمشروع منصة عيادات طب العيون

حتياجات رأس المال العامل(BFR)

المكوّن	القيمة (دج)
المخزون	**0**
مستحقات العملاء (30 يومًا)	**180,000**
ديون الموردين	**-120,000**
احتياجات رأس المال العامل	**60,000**

جدول4.5 احتياجات رأس المال العامل لمشروع منصة عيادات طب العيون

جدول خطة الخزينة الشهرية للسنة الأولى

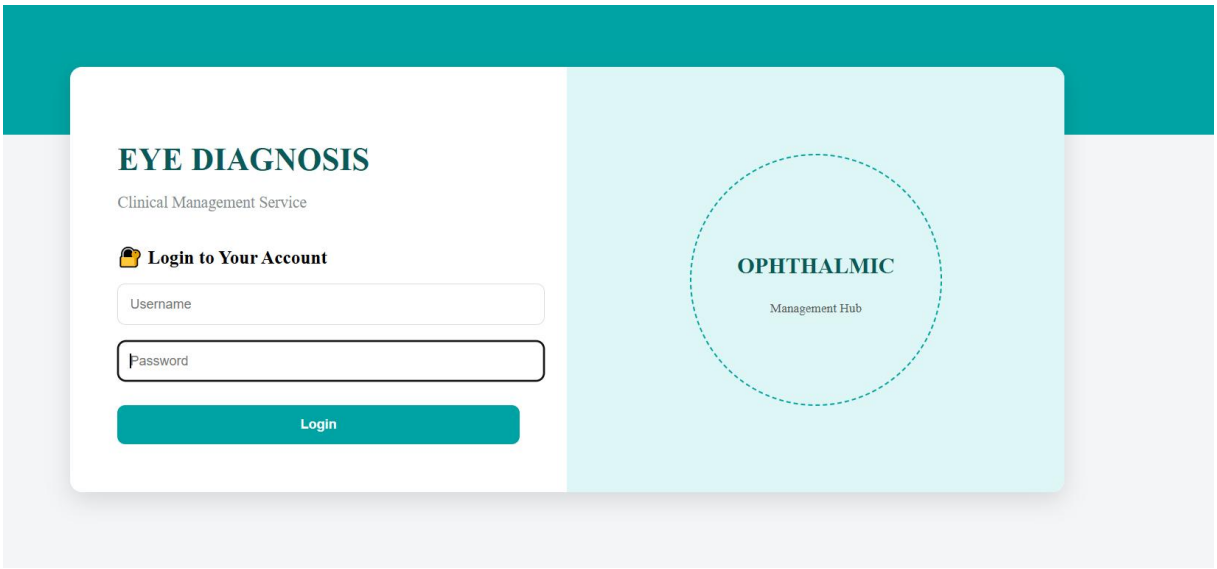
الشهر	الإيرادات(دج)	النفقات(دج)	الرصيد الشهري (دج)	الرصيد التراكمي(دج)
جانفي	0	205,000	-205,000	-205,000
فيفري	0	205,000	-205,000	-410,000
مارس	75,000	205,000	-130,000	-540,000
أفريل	75,000	205,000	-130,000	-670,000
ماي	75,000	205,000	-130,000	-800,000
جوان	75,000	205,000	-130,000	-930,000
جويلية	90,000	205,000	-115,000	-1,045,000
أوت	90,000	205,000	-115,000	-1,160,000
سبتمبر	90,000	205,000	-115,000	-1,275,000
أكتوبر	90,000	205,000	-115,000	-1,390,000
نوفمبر	120,000	205,000	-85,000	-1,475,000
ديسمبر	150,000	205,000	-55,000	-1,530,000

جدول5.5 خطةالخزينةالشهريةللسنةالأولى لمشروع منصة عيادات طب العيون

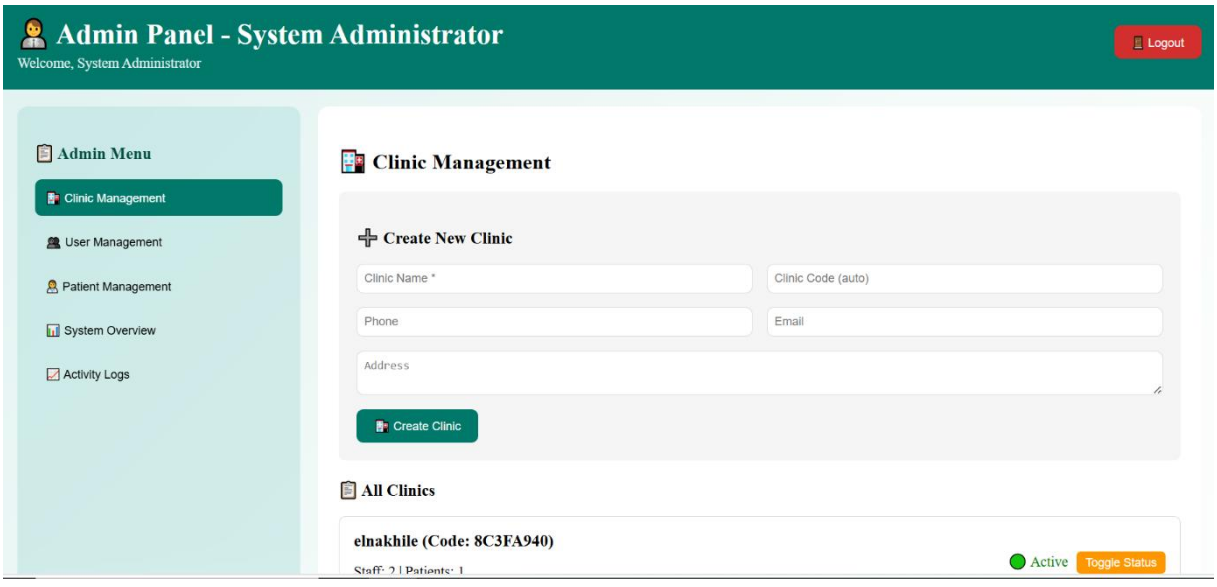
المحور السادس: النموذج الأولي التجريبي

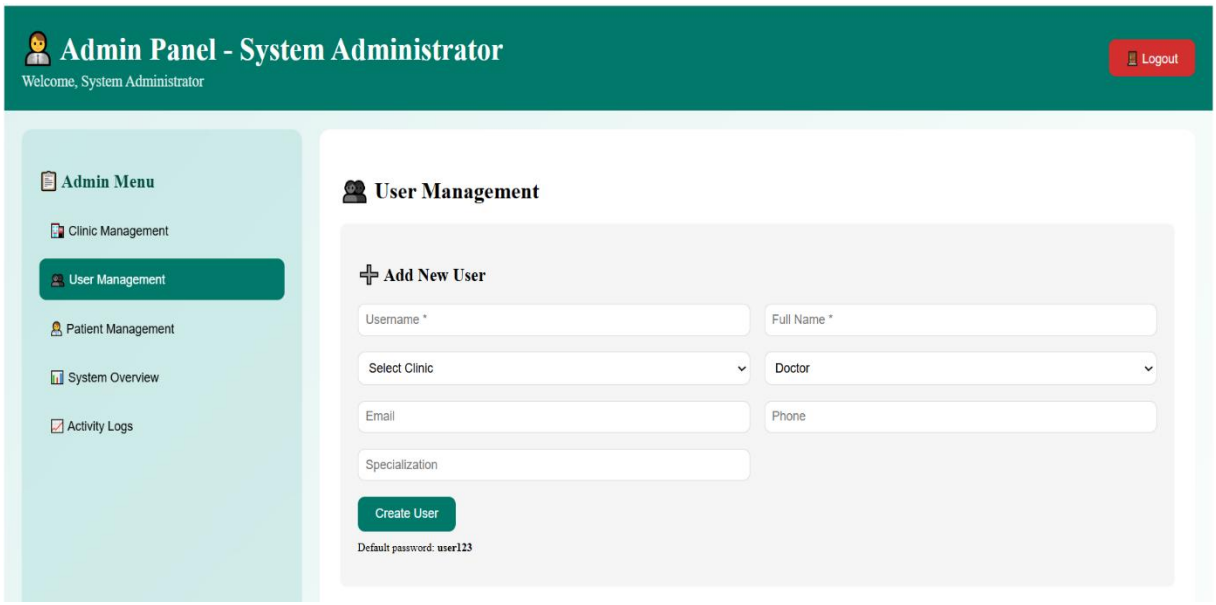
واجهة التطبيق:

واجهة تسجيل الدخول:تمثّل نقطة الدخول الأساسية إلى نظام إدارة عيادة العيون، حيث تسمح فقط للمستخدمين المصرّح لهم بالوصول. تعتمد الواجهة على آلية دخول آمنة مبنية على التحكم في الصلاحيات حسب الدور الوظيفي، بحيث تميّز بين ثلاث فئات من المستخدمين: مدير النظام، الطبيب، وممرض(ة)، مع الحرص على الحفاظ على طابع بصري مهني يلائم البيئة الطبية، وسهولة الاستخدام، وحماية معلومات المرضى.

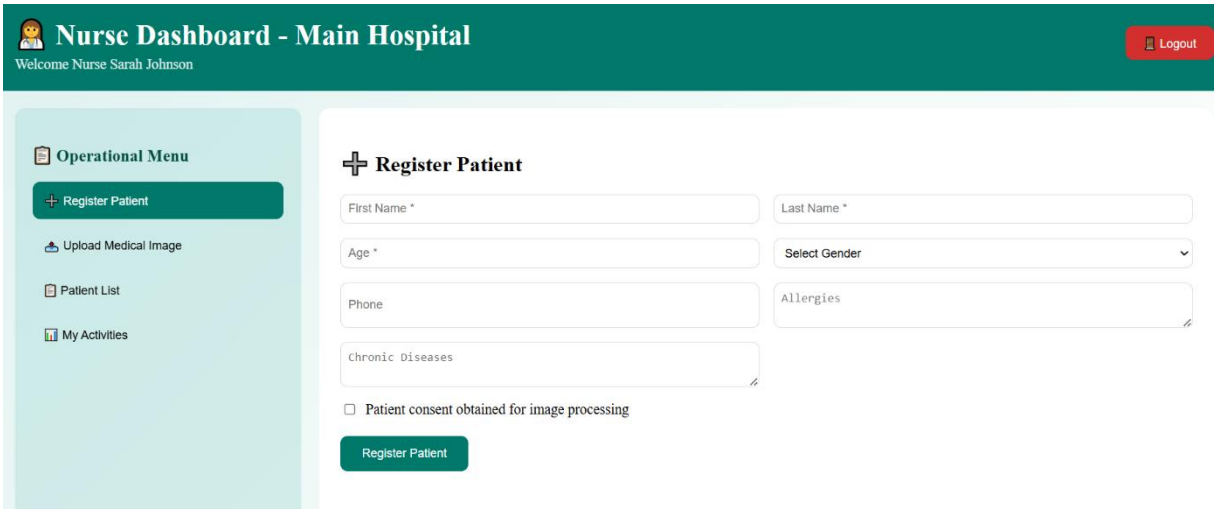


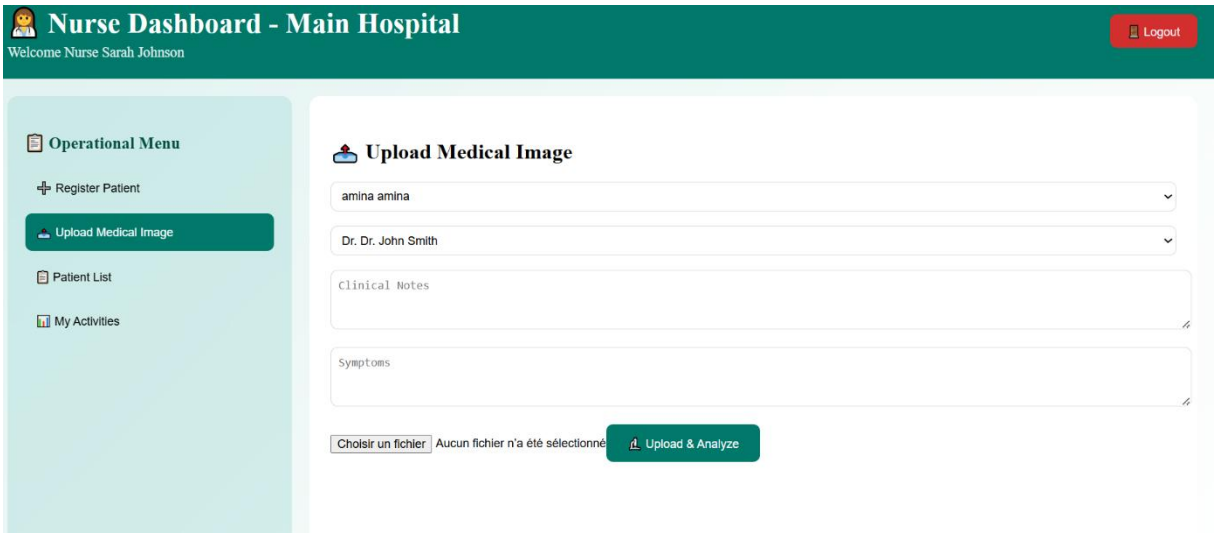
واجهات المسؤول(Administrator Interfaces) تقدّم واجهات المسؤول لوحة تحكّم مركزية تسمح بتهيئة النظام ومتابعته عبر وحدات إدارة العيادات، إدارة المستخدمين، إدارة المرضى، نظرة عامة على المؤشرات التشغيلية، وسجل الأنشطة. وتمكّن هذه الواجهة المسؤول من الإشراف الكامل على نظام إدارة عيادة العيون مع الحفاظ على فصل صارم للصلاحيات بين المهام الإدارية والعمل السريري اليومي.





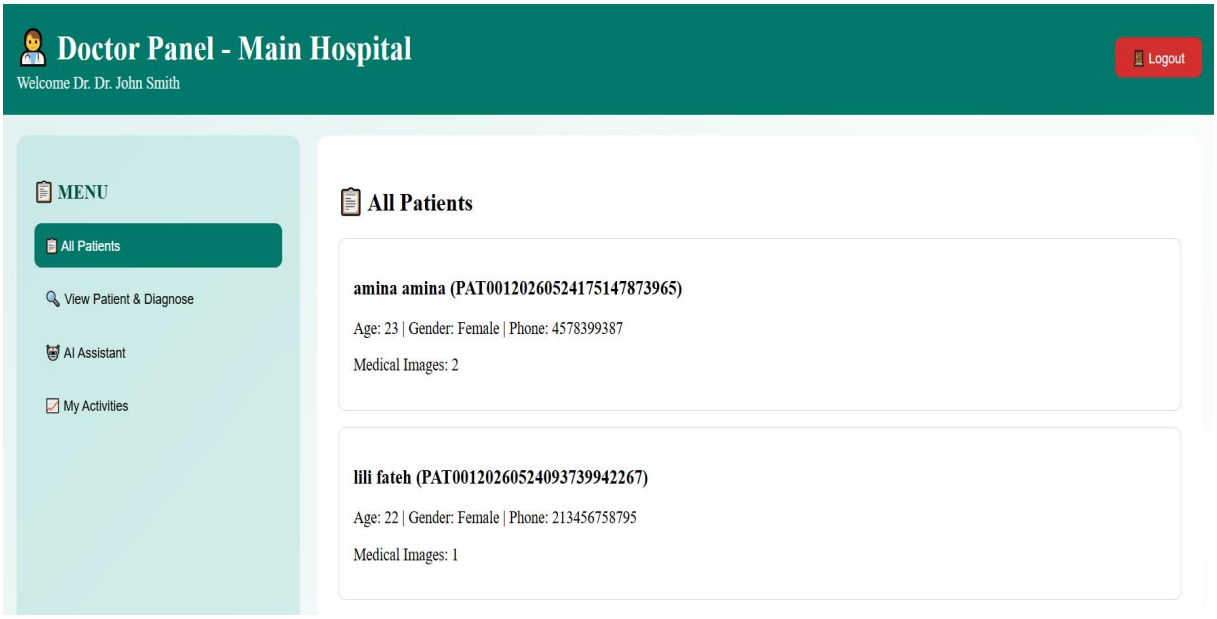
واجهة الممرّض / الممرّضةتوفّر واجهات التمريض ثلاث وظائف رئيسية: تسجيل المرضى مع إدخال بياناتهم الديموغرافية وحالة الموافقة، رفع الصور الطبية وربطها بالمرضى وتحويلها للطبيب للتحليل بالذكاء الاصطناعي، ومتابعة قائمة المرضى مع عدد الصور لكل مريض مع توثيق جميع العمليات في سجل الأنشطة. وبذلك تمكّن هذه الواجهة طاقم التمريض من إدارة الاستقبال والتصوير بكفاءة مع الإبقاء على مهام التشخيص محصورة في حسابات الأطباء فقط.

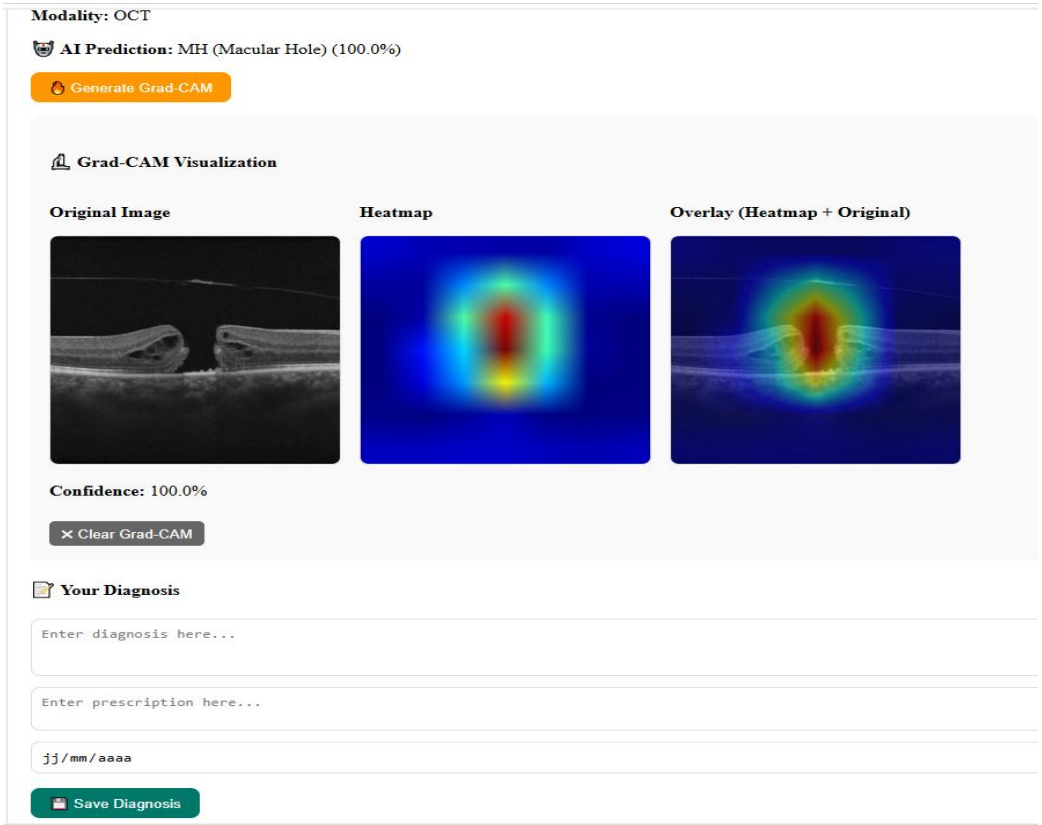


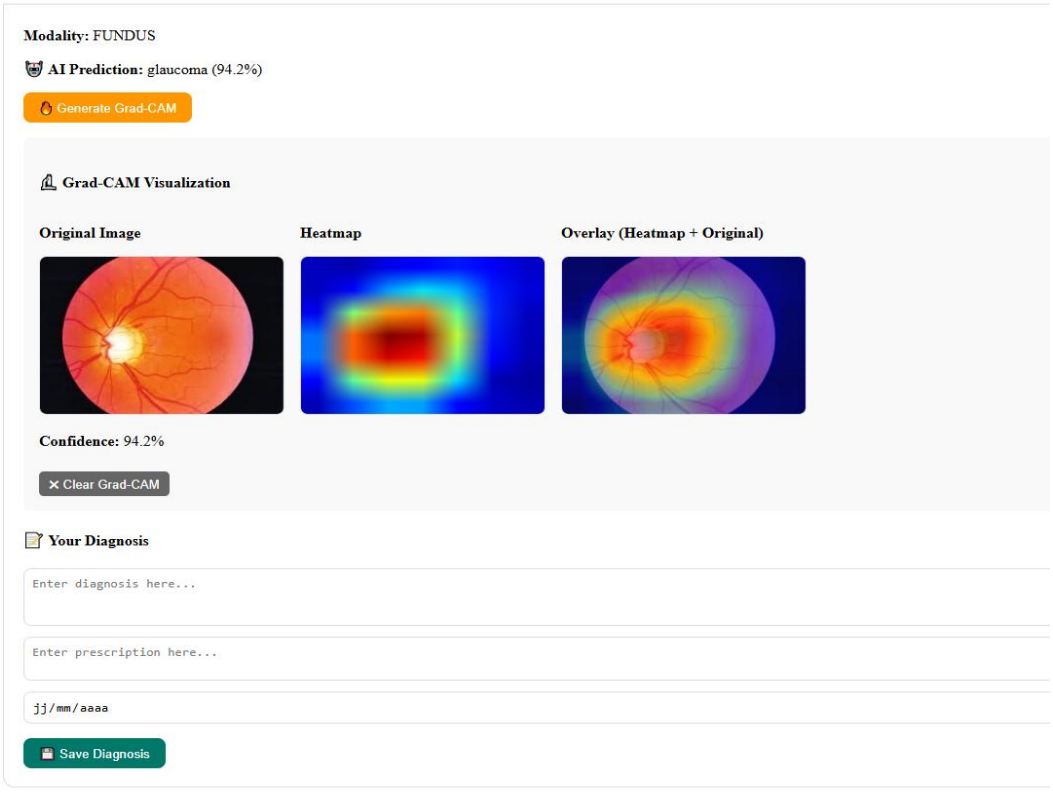


واجهة الطبيب

توفّر واجهات الأطباء مساراً تشخيصياً متكاملاً يبدأ من اختيار المريض وعرض بياناته الأساسية وعدد صوره، ثم تحليل كل صورة على حدة مع إظهار نوع الصورة، وتشخيص المرض، ودرجة الثقة، بالإضافة إلى شرح بصري باستخدام Grad-CAM يوضّح المناطق الأكثر تأثيراً في القرار. وبذلك تدمج الواجهة دعم الذكاء الاصطناعي في عملية اتخاذ القرار السريري، مع الحفاظ على سلطة الطبيب في قبول اقتراح النموذج أو تعديله وتوثيق التشخيص النهائي والعلاج والمتابعة في السجل الطبي الإلكتروني للمريض.




Modality: OCT
AI Prediction: MH (Macular Hole) (100.0%)
Generate Grad-CAM
Grad-CAM Visualization
Original Image
Heatmap
Overlay (Heatmap + Original)
Confidence: 100.0%
Clear Grad-CAM
Your Diagnosis
Enter diagnosis here...
Enter prescription here...
jj/mm/aaaa
Save Diagnosis


Modality: FUNDUS
AI Prediction: glaucoma (94.2%)
Generate Grad-CAM
Grad-CAM Visualization
Original Image
Heatmap
Overlay (Heatmap + Original)
Confidence: 94.2%
Clear Grad-CAM
Your Diagnosis
Enter diagnosis here...
Enter prescription here...
jj/mm/aaaa
Save Diagnosis

الملحق رقم 04: نموذج العمل التجاري

