

سيرين حجازي

Sirinhejazi_243001213_Project Proposal..docx



Final Thesis Submission after Reduce Similarity



Thesis_(2)_AI901_Dr_Khaled_Eskaf_Third Trimester 2025 (Moodle PP)



Midocean University

Document Details

Submission ID

trn:oid::1:3451944624

Submission Date

Jan 1, 2026, 12:05 AM GMT+4

Download Date

Jan 1, 2026, 2:24 AM GMT+4

File Name

xcod.351187533_111621_lasoporP_tcejorP_312100342_izajehniriS_سيرين_حجازي_30952

File Size

7.8 MB

69 Pages

13,717 Words

82,300 Characters

25% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

-  **48 AI-generated only 25%**
Likely AI-generated text from a large-language model.
-  **0 AI-generated text that was AI-paraphrased 0%**
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





Advancing Ophthalmic Diagnostics with AI: A Deep Learning Framework for early Detection of Retinal Diseases from OCT Data

Prepared By

Name: **Sirin Mohammed Hejazi**

IDNo.: **243001213**

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Khaled Eskaf

College of Informatics, Midocean University, Comoros

[3-2025]

ACKNOWLEDGMENT

I thank Dr. Khaled Eskaf for guiding me through the entire process. His constructive feedback shaped the direction of the paper. I am also grateful to Midocean University and the staff for providing the resources needed to undertake the research. On a personal level, I acknowledge my husband and family for their prayers and encouragement. Writing this paper would have been difficult without their support. This paper is a collective effort of many people, whom I cannot mention, but their contribution has made this possible.

ABSTRACT

Retinal conditions have become prevalent especially among the older adults. These conditions are a leading source of vision challenges, sometimes rendering some people blind. Early detection can mitigate blindness resulting from common retinal conditions like Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen. However, sometimes healthcare facilities have inadequate ophthalmologists which limits their diagnostic capacity. Even where ophthalmologists are available, the huge number of patients places a significant burden. This study addresses this gap by proposing OptiScan, an artificial intelligence (AI)- based system which offers a non-surgical approach for classifying retinal conditions from OCT images. A dataset of 84,000 B-scans was preprocessed and augmented to address class imbalance. The data was further divided into training, validation and testing datasets. The models evaluated included 3-layer and 7-layer architectures as well as pre-trained models such as VGG16, ResNet50, and DenseNet121. Among these models, DenseNet121 had the highest accuracy of 96%, with a strong performance across all the classes. 7-layer custom network also demonstrated competitive results. This shows that well-designed domain-specific models can perform effectively. The optimal model was integrated into a web-based tool named OptiScan. This provides a secure, yet intuitive interface for clinicians to input OCT scan images and get classified results. Usability testing placed OptiScan at 82.5, showing the model is friendly to use. The study indicates that the model can be deliver accurate and consistent classification of retinal conditions, while addressing challenges like imbalance and interpretability.

Keywords:

Artificial Intelligence, Deep learning, Convolutional Neural Networks (CNN), Optical Coherence Tomography (OCT), Retinal disease classification, choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, Computer-Aided Diagnosis, Transfer learning, clinical decision Support system, explainable AI

Contents

Advancing Ophthalmic Diagnostics with AI: A Deep Learning Framework for early Detection of Retinal Diseases from OCT Data	1
ACKNOWLEDGMENT.....	2
ABSTRACT.....	3
List of Abbreviations	8
CHAPTER 1: INTRODUCTION	9
1.1. Background of the Study	9
1.2. Problem Statement.....	9
1.3. Aim of the Study	9
1.4. Research Objectives	9
1.5. Research Questions	10
1.6. Scope of the Study	10
1.7. Significance of the Study.....	10
1.8. Research Limitations.....	10
1.9. Project Timeline.....	11
1.10. Chapter Summary.....	11
CHAPTER 2: LITERATURE REVIEW	12
2.1 Introduction.....	12
2.2 Clinical Background of Retinal Pathology	12
2.1.1 The Human Visual System: An Overview.....	12
2.1.2 Age-Related Macular Degeneration (AMD).....	12
2.1.3 Choroidal Neovascularization (CNV)	13
2.1.4 Diabetic Macular Edema (DME).....	13
2.1.5 Drusen	13
2.1.6 Summary: Burden and Diagnostic Challenges	14
2.3 Optical Coherence Tomography in Retinal Diagnosis	14
2.3.1. Technical Evolution.....	14
2.3.2. Clinical Utility and limitations	14
2.3.3. Deep learning fundamentals for medical imaging	14
2.3.4. OCT Databases for Deep Learning.....	14
2.3.4.1. Available Resources.....	14
2.3.4.2. Dataset Limitations	15

2.3.5.	Deep Learning Architectures for OCT Classification	15
2.3.5.1.	Custom CNNs.....	15
2.3.5.2.	Transfer Learning	15
2.3.6.	Research Gaps and Synthesis.....	15
2.3.7.	Justification for Current Study	16
2.3.8.	Chapter Conclusion.....	16
CHAPTER 3: METHODOLOGY		17
3.1	Research Design	17
3.2	Phase 1: Business Understanding	18
3.2.1.	Problem Definition	18
3.2.2.	Functional Requirements.....	18
3.2.3.	Non-functional Requirements	18
3.2.4.	Objectives	18
3.2.5.	Success Criteria.....	19
3.2.6.	Stakeholder Requirements.....	19
3.3	Phase 2: Data Understanding.....	19
3.3.1.	Data Source.....	19
3.3.2.	Initial Data Exploration.....	19
3.3.3.	Data Quality Assessment	20
3.3.4.	Exploratory Data Analysis (EDA)	20
3.3.5.	Class distribution	22
3.4	Phase 3: Data Preparation	23
3.4.1.	Data Cleaning	23
3.4.2.	Image Preprocessing Pipeline.....	23
3.4.3.	Images Augmentation:.....	23
3.4.4.	Sampling Data:	25
3.4.4.1.	Splitting Data into Train, Validation, and Test Sets:	26
3.5	Phase 4: Modeling	27
3.5.1.	Model Selection Rationale.....	27
3.5.2.	Transfer Learning Implementation	28
3.5.3.	Training Configuration	28
3.5.4.	Hardware and Software Environment.....	29
3.5.5.	Calculation details.....	29
3.6	Phase 5: Evaluation	30

3.6.1.	Evaluation Metrics	30
3.6.2.	Statistical Validation	31
3.6.3.	Explainability Methods.....	31
3.6.4.	Clinical Validation Protocol.....	31
3.7	Phase 6: Deployment	31
3.7.1.	System Design	31
3.7.2.	Use Case Design	32
3.7.3.	System Workflow	34
3.7.4.	Architecture Pattern.....	35
3.7.5.	System Main Components:	36
3.7.6.	Package diagram:.....	37
3.7.7.	Class Diagram	37
3.7.8.	Sequence Diagram:.....	38
3.7.9.	The Relational Database Schema	40
3.7.3	System implementation:	41
3.7.3.1.	Build HTML pages:	41
3.7.3.2.	Build DB:	48
3.7.3.3.	Build System App	49
3.8	Ethical Considerations.....	50
3.8.1.	Data Privacy	50
3.8.2.	Bias Mitigation.....	50
3.8.3.	Clinical Safety	50
3.9	Methodological Limitations	51
3.10	Summary	51
CHAPTER 4: RESULTS AND IMPLEMENTATION		52
4.1	Model Performance Evaluation.....	52
4.2	Training Performance	52
4.2.1.	Performance Metrics	58
4.2.2.	Per-Class Performance Analysis	58
4.2.3.	Learning Dynamics and Overfitting Analysis	58
4.3	Statistical Analysis.....	58
4.3.1.	McNemar's Test Results	58
4.3.2.	Confidence Intervals.....	59
4.4	Computational Efficiency	59

4.5	Model Explainability with Grad-CAM	59
4.6	Web Application Testing Results.....	59
4.6.1.	Functional Testing.....	59
4.6.2.	Usability Testing	60
4.6.3.	Acceptance Testing Using Nielsen's Heuristics	60
4.6.4.	System Performance Benchmarks	60
4.6.5.	Comparative Analysis with Previous Studies	60
4.7	Limitations and Error Analysis.....	61
4.7.1.	Error Patterns.....	61
4.7.2.	Failure Cases.....	61
4.8	Chapter Conclusion	61
CHAPTER 5: CONCLUSION AND FUTURE WORKS		62
References		65

Table of Tables:

Table 1:Synthesis of Key Literature.....	15
Table 2: Counts the images in each subset and prints the totals.....	19
Table 3:Augmentation Techniques Applied.....	23
Table 4:distribution of images across the training, validation, and testing sets	26
Table 5: Flask API Integration	50
Table 6:Comparative Performance of All Models on Test Set.....	58
Table 7: Class-Specific F1-Scores.....	58
Table 8: 95% Confidence Intervals for Model Accuracy.....	59
Table 9: Computational Requirements	59
Table 10: Functional Test Results	59
Table 11:Heuristic Evaluation Results	60
Table 12:System Performance Metrics.....	60
Table 13:Comparison with State-of-the-Art.....	60

Table of Figures:

Figure 1: CRISP-DM Methodology Framework.....	Error! Bookmark not defined.
Figure 2: Diabetic macular edema.....	19
Figure 3: CNV	19
Figure 4: Dursun.....	19
Figure 5: Normal eye.....	19
Figure 6:Proportion of images in training, validation, and testing sets	20
Figure 7: Normal	21
Figure 8: CNV	21
Figure 9: DME.....	22
Figure 10: DRUSEN.....	22

Figure 11:Class distribution.....	23
Figure 12:Sample images after Augmentation	25
Figure 13 :updated Class distribution.....	26
Figure 14:Updated proportion of images in training, validation, and testing sets.....	27
Figure 15: Fully CNN System with 4 Classes.....	28
Figure 16:Evaluation Metrics	30
Figure 17: End-to-End System Data Flow.....	32
Figure 18:Use case diagram	34
Figure 19:System workflow	35
Figure 20:Layer	36
Figure 21:System components.....	36
Figure 22:Package Diagram	37
Figure 23:Class Diagram	38
Figure 24:Doctor Sequence Diagram	39
Figure 25:Admin Sequence Diagram	40
Figure 26:Relational Database Schema.....	41
Figure 27:DME.html	42
Figure 28:Dursen.html.....	43
Figure 29:ForgetPassword.html.....	44
Figure 30:Index.html	45
Figure 31:Login.html.....	46
Figure 32:OCT.html	47
Figure 33:Predict.html	47
Figure 34:Results.html.....	48
Figure 35:Signup.html	48
Figure 36:3 -Layer CNN.....	53
Figure 37: 7-Layer CNN.....	54
Figure 38: VGG Model.....	55
Figure 39:ResNet50 Transfer Learning.....	56
Figure 40: DenseNet121	57

List of Abbreviations

CNN: Convolutional Neural Networks

CNV: Choroidal Neovascularization

DME: Diabetic Macular Edema

AI: Artificial Intelligence

CRISP-DM: Cross-industry standard process for data mining

CHAPTER 1: INTRODUCTION

1.1. Background of the Study

Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen are a major cause of blindness all over the world. These retinal illnesses are prevalent among the older adults (Akça *et al.*, 2024). This is a source of major concern since the global population is ageing due to lower fertility rates and increased life expectancy because of improved healthcare (Gianfredi *et al.*, 2025). Hence, there is a need to ensure that older populations can access early diagnosis to ensure they don't succumb to blindness for conditions that could otherwise have been prevented. Researchers agree that the classification of Optical coherence Tomography (OCT) is widely used for assessing the level of retinal lesions (Vujosevic *et al.*, 2023). However, OCT presents incomplete image feature extraction and inadequate positional features. OCT is also limited by the need for expert interpretation which required a lot of time, straining clinical facilities that are less staffed (Chew, 2020). The limitations of OCT have accelerated research towards more effective imaging techniques. In the recent days, artificial intelligence (AI) has emerged as a formidable force in the healthcare field. For instance, deep learning with Convolutional Neural Networks (CNNs) has allowed the automation of medical image analysis process. According to Quintana-Quintana *et al.* (2025) the use of deep learning techniques allows for fast, consistent and scalable diagnostics. This paper examines CNN-based models which can automate the classification of OCT images. The aim is allowing the models to categorize OCT images into Normal, CNV, DME, and Drusen and facilitate early detection of retinal diseases to mitigate preventable blindness.

1.2. Problem Statement

OCT imaging has been a breakthrough in the detection of retinal illnesses. However, OCT imaging depends heavily on expertise opinion. This causes diagnostic delays especially in healthcare facilities and clinics with few ophthalmologists. The subjective nature of OCT imaging leads to inter-observer variability in the interpretation of images, risking misdiagnosis and delayed treatment. Workloads increase as healthcare facilities handle many retinal cases. While these challenges persist, there is limited research specifically focusing on AI-assisted retinal disease diagnosis from OCT images which can work in the real-world context.

1.3. Aim of the Study

This study aims to design a diagnostic system based on deep learning to classify retinal diseases from OCT images. The system should have a reliable predictive performance while remaining applicable in the clinical setting.

1.4. Research Objectives

The research objectives include:

- To use data preprocessing and augmentation techniques to prepare a publicly available OCT image dataset to improve model stability across various image conditions.
- To design and train CNN models and select transfer learning approaches appropriate for OCT image classification.
- To assess the performance of the models using standard metrics to evaluate diagnostic reliability.

- To identify the most appropriate model with a high diagnostic accuracy and computational efficiency.
- To implement the selected model in a web-based application that demonstrates its clinical real-time use

1.5. Research Questions

To achieve the study objectives, the following will guide the following research questions.

- Given OCT images, how well can deep learning techniques classify OCT images into CNV, DME, Drusen, and Normal?
- Which of the custom-designed CNNs and pre-trained architectures performs better in categorizing OCT images?
- How does data augmentation reduce overfitting and enable model generalization?
- How feasible is the integration of a trained model into a web-based diagnostic interface that it easy to use among clinical professionals?

1.6. Scope of the Study

The scope of this study is limited to the methodological and data considerations. It is restricted solely to the analysis of OCT B-scan images to categorize images into four retinal conditions: Normal, CNV, DME, and drusen. While developing the model modality, the study evaluates custom-layered CNNs alongside established deep learning architectures such as VGG16, ResNet50, and DenseNet121. All the experiments are based on publicly available OCT dataset introduced by Kermany *et al.* (2018). The output of the study is a multiclass classification system meant to enhance eye disease detection. Hence, the research does not extend to modeling diseases progression over time nor address three-dimensional (3D) volumetric OCT analysis.

1.7. Significance of the Study

Research has been undertaken on the application of computer-aided diagnosis for ophthalmology. This paper makes a contribution to the ongoing work in the following ways. Firstly, it considers how automated deep learning techniques can effectively detect retinal problems. Through automation, facilities can reduce diagnostic time so that patients can access eye care faster. Clinicians will have reduced workload, allowing them to support patients better. Secondly, the interpretation of OCT images varies among ophthalmologists. The study provides a point of reference for diagnostic process, addressing the inter-observer variability challenge. This facilitates better diagnosis and minimizes cases of misdiagnosis. Additionally, the paper considers the practice implications of deploying the system to the real-world clinical settings. In areas with limited specialist ophthalmic care, deep learning systems can facilitate the imaging process. Lastly, the system emphasized on explainable AI in line with transparency in healthcare settings. It ensures patient data is handled securely to enhance data confidentiality.

1.8. Research Limitations

While this study makes a significant contribution to OCT imaging, it as various limitations.

1. The use of publicly available OCT dataset limits generalizability of the study. In the real-world ophthalmic practice, demographic and pathological diversity influences the interpretation of data. This makes it challenging to generalize the findings of the study to the global audience.
2. The model, like any other CNN has challenges with interpretability. Despite the high performance of CNNs, limited transparency makes it challenging to understand how the models arrive at the solution.

The “black-box” nature of CNNs means users may not understand how the model classified the images. This limits the implementation of the system in the healthcare system where transparency is needed especially when dealing with patient data.

3. Limited computational resources limit the application of the model in environments with limited resources. Training deep learning models needs significant GPU resources. Low-resource settings may not afford such high computing power, which limits model training and use.
4. The healthcare industry is heavily regulated, given its sensitivity. The adoption of the system will need regulatory authority to undertake clinical trials. This is time-consuming and could delay the implementation of the project.

1.9. Project Timeline

The project followed the schedule below to ensure that each phase was allocated adequate time.

- Month 1: This marked the initial phase of the product. Relevant literature review materials were gathered and compiled.
- Month 2: Data from the Kermany *et al.* (2018) dataset was acquired, preprocessed, and augmented.
- Months 3-5: The model was developed and trained. The initial validation of the model was done.
- Month 6: A comprehensive evaluation, optimization, and comparative analysis was done.
- Month 7: A web application was developed and the system integrated into the website.
- Month 8: The thesis was documented and the final submission done.

1.10. Chapter Summary

This chapter introduces the significant problem of delayed diagnosis, especially among the older population. The chapter introduces deep learning approaches as a solution to more accurate and timely diagnosis. It outlines the problem statement, aim of the paper, objectives and research questions. The significance of the work and its key limitations are indicated to frame the scope of the work. The chapter introduces the need for an AI-diagnostic driven approach to facilitate OCT imaging. In the next chapter, a review of literature is explored. It examines literature related to retinal pathologies and the available detection mechanisms. This allows exploration of the theoretical and empirical approaches required to shape the direction of the study.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Classification systems based on computer vision has become popular in the recent years. The automation of diagnosis promises to revolutionize the healthcare especially in low-income areas which tend to have inadequate medical professionals. In the retinal context, automated diagnosis of eye diseases can enhance access to timely care, minimizing incidences of preventable blindness. This chapter examines existing studies which have focused on the use of computer vision to classify OCT images. The aim is identifying how existing studies have automated the classification of AMD, CNV, DME, Drusen, and normal retinal conditions.

2.2 Clinical Background of Retinal Pathology

2.1.1 The Human Visual System: An Overview

The visual system relies primarily on sensory modality to gather and interpret information. Vision enables people to recognize environmental features around them like colors, shapes, and motion. Studies suggest that more than 70% of the information humans have comes through vision (Mandal, 2003). The eye consists of external and internal components which work together to convert light to neural signals. The external structure has all the visible parts like the sclera, conjunctiva, cornea, iris, and pupil. The sclera constitutes the white visible portion, made up of dense tissue to protect the inner parts. The transparent and thin layer covering the sclera is called the conjunctiva. It is made up of stratified squamous epithelium and keeps the eye moist and clear. The cornea is the transparent front part of the eye covering the pupil and the iris. In addition to protecting the inside of the eye, the cornea refracts the light along with the lens. The iris is the pigmented, colored part of the eye, visible externally. It controls the size of the pupil depending on the light source. The pupil is the black circle inside the iris which narrows and widens to control the amount of light reaching the eye.

The internal structure of the eye has consist of lens, retina, optic nerve, aqueous humor, and vitreous humor. The lens is the transparent and biconvex part of the eye. Together with the cornea, the lens refracts light so that it focuses on the retina. The retina is the innermost layer of the eye. It is the thin layer of light-sensitive cells at the back of the eye that converts light into electrical signals. The optic nerve carries all the nerve impulses from the retina to the human brain. The watery fluid between the cornea and the lens is called the aqueous humor. It is responsible for keeping the eye nourished and inflated. Considering the complexity of the eye, pathological changes can adversely impair vision. Automated diagnostic system can identify subtle morphological changes in OCT scans and avoid cases of false negatives that could cause late interventions (Du et al., 2024).

2.1.2 Age-Related Macular Degeneration (AMD)

A Beaver Dam Ete Study (BDES) was one of the first large-scale epidemiological investigations into AMD. Launched in 1987 in Beaver Dam, Wisconsin, the study followed adults aged 43 to 86, examining the

prevalence, incidence and risk factors associated with AMD. The study established that CVN associated with age-related AMD affected about 1.2% of older adults within the studied age range (Chew, 2020). More recent study has tracked the global burden of AMD-related low vision and blindness from 1999 to 2021, projecting trends through 2050 (Zhang et al., 2024). The study points how the number of individuals with AMD has increased from about 3.6 million in 1990 to more than 8 million, with projections suggesting this could reach about 13.9 million by 2050 (Sheng et al., 2021). A number of environmental and genetic risk factors also contribute to the risk of AMD. According to Salimiaghdam *et al.*, (2019), age remains the strongest non-modifiable risk factor. Other studies suggest that ethnicity and gender are significant non-modifiable risk factors, with white females at a higher risk of the late forms of AMD compared to black individuals (Rudnicka et al., 2015). Researchers indicate more factors like smoking, higher body mass index, and cardiovascular disease (Mauschitz & Finger, 2022; Král'ová et al., 2024).

2.1.3 Choroidal Neovascularization (CNV)

CNV causes blood vessels within the retinal to grow often because of age-related AMD. It often creates a sudden decline of central vision, within a few weeks and may also present alongside conditions like myopia or malignant myopic degeneration (Sheng *et al.*, 2021). Fluorescein angiography (FA) is a conventional way of detecting CNV and it used widely in diagnostic practice (Uchida et al., 2018). FA involves the intravenous administration of sodium fluorescein dye to visualize blood flow in the retina and choroid to detect leakages. Despite its diagnostic utility Chiu & Chu (2022) indicate that FA is invasive, requiring intravenous dye injection. This causes cause discomfort, nausea, and severe allergic reactions to some patients. The invasive nature of the procedure makes it time-consuming, often requiring 15-20 minutes. In a setting with numerous patients, FA is challenging to undertake, if the number of specialists is inadequate.

2.1.4 Diabetic Macular Edema (DME)

DME occurs where high blood sugar leads to the damaging of retinal blood vessel. This causes them to thicken, leak fluid and swell the macula. Patients report blurred vision, gaps in the visual field, and faded colors, which interferes with daily activities. If DME is left untreated for a long time, it can lead to blindness. Factors leading to DME include poorly controlled diabetes and high blood pressure. In some patients, kidney disorder and elevated levels of cholesterol have been associated with the illness. There are several techniques clinicians use to detect DME. For instance, Stereo Fundus Photography takes the images of the retinal from different angles. This allows clinicians to get a three-dimensional (3D) view and check molecular variations. Another approach is Slit-Lamp Biomicroscopy which provides detailed and high-resolution images of the retinal. Medical practitioners get then establish macular thickness, and hence determine the presence of DME.

2.1.5 Drusen

Drusen are characterized by the accumulation of extracellular deposits beneath the retinal pigment epithelium. These deposits are highly predictive of early-stage AMD. The yellow deposits accumulate under between the retinal pigment epithelium (RPE) and Bruch's membrane. Drusen is common among older adults and can be a sign of AMD, with their presence, size, and number predicting the disease progression. Few hard drusen may not significantly affect vision, but large soft drusen are associated with a higher risk of vision loss. It is important to note that while Drusen bodies are commonly described in AMD, they are not pathognomonic of AMD. VanDenLangenberg & Carson (2023) indicate that Drusen-like deposits have been observed in some less prevalent inherited conditions like Sorsby fundus dystrophy.

2.1.6 Summary: Burden and Diagnostic Challenges

Traditional diagnostic approaches for retinal conditions have helped doctors identify abnormalities in the eye. However, they have limitations which interfere with their accuracy. A major problem is that traditional techniques are subject to the interpretation of the clinician. For example, ophthalmoscopy and slit-lamp biomicroscope depend primarily on the clinician's visual interpretation. Also, these techniques cannot quantify retinal edema or druse volume which limits detection of the progression of the disease more accurately. Also, the inability to generate 3D image limits an assessment of the various retinal layers. While the conventional methods are cost-effective and accessible, they are increasingly replaced by better imaging procedures like OCT. These have a greater precision and depth in retinal assessment, allowing ophthalmologists to make better diagnostics.

2.3 Optical Coherence Tomography in Retinal Diagnosis

2.3.1. Technical Evolution

OCT has over the years replaced traditional retinal diagnosis techniques because of its non-invasive nature. It provides high-resolution cross-sectional and volumetric images of retinal microstructure. The evolution from time-domain to Fourier-domain has made OCT to have an axial resolution of less than 5 μm with rapid scan acquisition. Hence, ophthalmologists can undertake qualitative and quantitative analysis of retinal structure (Zeppieri et al., 2023).

2.3.2. Clinical Utility and limitations

Despite the changes OCT has brought to retinal diagnosis, it is limited. Since it relies on subjective interpretations, ophthalmologists can underestimate the severity of the illness, delaying the treatment process. At the same time, since a clinician has to check the OCT images, this leads to diagnostic delays. Patients have to wait for a longer time as the illness is identified, which denies them prompt care. The manual review of the images often introduces segmentation errors, which can lead to misdiagnosis. In clinical environments that have limited resources, reviewing multiple OCT images is problematic. As a result, there is a need to an automated interpretations systems to replace the manual approach.

2.3.3. Deep learning fundamentals for medical imaging

The need to move from manual to more automated retinal disease detection has created interest in CNNs. In particular, CNNs learn hierarchical feature representations without much supervision. Even where patterns are complex, CNNs can extract features and classify them. In OCT imaging, CNNs can learn complex features from OCT images and classify them into predetermined classes. When fine tuning a model, it is important to categorize it into training (60-80%), validation (10-20%), and test (10-20%) sets. This will mitigate overfitting and allow the model to generalize unseen data. Accuracy, sensitivity, specificity, precision, F1-score, and AUC-ROC will be used to evaluate the model.

2.3.4. OCT Databases for Deep Learning

2.3.4.1. Available Resources

Kermany et al. (2018) has a dataset containing more than 207,000 labelled OCT B-scans, which are spread across four classes. The dataset has become a benchmark for studies touching on retinal diseases. In this

research, the dataset will be the primary source of OCT images. To supplement the dataset, other smaller datasets like OCTID and OCTDL will be used. These will provide more features which may not be offered under the primary dataset, allowing the study to have pathology diversity.

2.3.4.2. Dataset Limitations

ICT datasets like Kermany et al. (2018) facilitate model training because of the labeled datasets. However, they should be used carefully considering they may have class imbalance especially for normal or common conditions like AMD are overrepresented compared to rare conditions. Hence the model may face challenges while identifying minority “positive” cases. Another challenge is domain shift across different imaging devices. The performance of the model may be affected when the data is applied from a device or clinical setting different from the one used during the training stage. Also, while public datasets facilitate model training, they often lack the breadth of human variation. Hence, they may not represent the full manifestations common in global populations. These factors can affect the generatability of models and lead to concerns of algorithmic bias (Khan et al., 2021).

2.3.5. Deep Learning Architectures for OCT Classification

2.3.5.1. Custom CNNs

CNNs are easily customizable depending on the aim of the model. This has made it possible to develop CNN architectures to fit domain-specific tasks like OCT classification in the project. Custom CNNs are critical when pre-trained networks cannot understand subtle anatomical features like it is in retinal illnesses. This allows custom CNNs to have a higher performance. However, they need more hyperparameter tuning and access to large datasets to minimize overfitting. Baba *et al.*, (2024) demonstrate that custom CNNs attains an accuracy of more than 95% on four-class OCT classification, making it suitable for eye disease detection.

2.3.5.2. Transfer Learning

There are instances where custom CNNs perform poorly especially with limited imaging dataset. In such cases, transfer learning using pre-trained ImageNet models display a strong baseline performance. However, some of the pre-trained models like VGG16 are heavy and require huge computational resources (Simonyan & Zisserman, 2015). ResNet50, with its residual connections, allows for the training of deeper networks. It has shown strong performance in OCT image classification (Li *et al.*, 2020). DenseNet121 uses dense connectivity which promotes feature reuse and parameter efficiency. According Kaur et al. (2024) DenseNet121 is suitable when dealing with smaller datasets.

2.3.6. Research Gaps and Synthesis

Table 1: Synthesis of Key Literature

Author(s), Year	Contribution	Limitation	Relevance
Kermany et al. (2018)	Pioneering transfer learning for OCT;	Focused on accuracy; limited	Provides foundational dataset; highlights

	established benchmark dataset	robustness and explain ability	need for balanced evaluation
Li et al. (2020)	High performance (97.3%) with ResNet50 ensemble	Limited external validation; no dataset bias consideration	Validates ResNet50 for OCT classification
Du et al. (2024)	Foundation model with improved interpretability	Focused on specific features; limited clinical deployment focus	Emphasizes importance of model transparency
Khan et al. (2021)	Systematic review highlighting dataset imbalances and biases	Identified problem without technical solution	Informs need to address class imbalance and bias

Despite numerous publications on the use of custom CNNs, some research gaps persist. One problem is the inadequate handling of class imbalance which often causes poor performance on underrepresented classes. While some published models show a higher performance, underrepresented classes may not be adequately represented. Another issue is the inadequate explainability where the decisions CNNs made are opaque. This is challenging in the clinical settings where clinicians and patients need to understand how a model arrives at a decision. Moreover, most of the published studies continue to show limited generalizability. This is due to over-reliance on a single dataset which narrows application in the clinical setting. Also, most of the existing studies focuses more on technical and algorithmic innovation at the expense of practicability. Hence, despite models showing better performance, they become difficult to translate into the clinical settings.

2.3.7. Justification for Current Study

This study endeavors to address several gaps identified in the existing literature sources. It seeks to:

- Comparing various model architectures like custom CNN, ResNet50, and DenseNet121 under consistent experimental conditions. This will allow the identification of the most appropriate model for retinal illness detection.
- Explicitly tackling class imbalance through strategic augmentation and weighted loss functions
- Incorporating explain ability via Grad-CAM visualization to enhance clinical trust
- Framing results within clinical applicability and translational challenges.

2.3.8. Chapter Conclusion

This chapter explores the application of OCT in the diagnosis of retinal anomalies. It has demonstrated how deep learning architectures can be used to automate OCTY imaging, which has remained largely manual. The chapter shows the performance of existing models. Despite current models showing a high accuracy of more than 95%, they are unable to handle class imbalances, have limited explainability and cannot be generalized well. This work builds on existing studies to create a more scalable diagnostic framework suitable for eye illnesses. Chapter three focused on the methodology for implementing a better model for retinal conditions.

CHAPTER 3: METHODOLOGY

Detection of eye anomalies is important in preventing blindness. While OCT provides a non-invasive way of detecting illnesses, checking a large volume of OCT and fundus images manually presents a significant burden to etymologists. However, automating the classification of retinal OCT images provides a way to address the issue by offering a more concise screening process. This chapter explores the development of a deep learning-based system which can classify images into CNV, DME, and Drusen from OCT images. The chapter follows a Waterfall Methodology with the first phase focusing on requirements gathering until the completion of the full system design.

3.1 Research Design

The framework for this study is guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. This approach ensures comprehensive analysis under the six stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Wirth & Hipp, 2000). This chapter provides a description of the application of each phase is provided in the development of OptiScan diagnostic system.

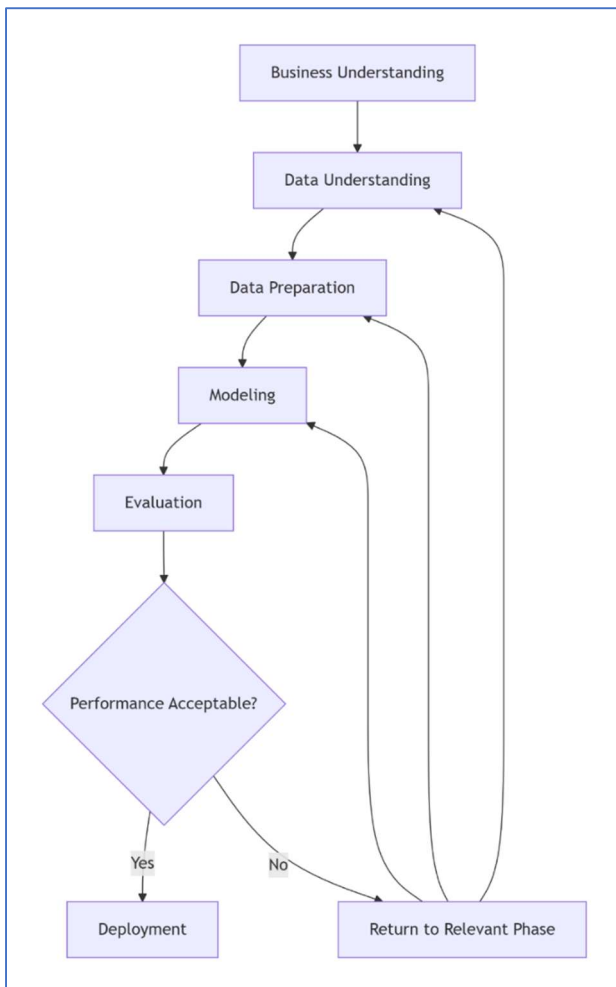


Figure 1: CRISP-DM Methodology Framework

3.2 Phase 1: Business Understanding

3.2.1. Problem Definition

Relying on manual interpretation of OCT images introduces biases in addition to being time-consuming. Du *et al.* (2024) note that subtle early-stage pathologies can easily be missed due to differences between observers. This leads to bottlenecks in diagnosis and can lead to delayed treatment in resource-constrained areas.

3.2.2. Functional Requirements

- 1- Image Acquisition: The system needs high-resolution retinal images from OCT scans.
- 2- Image Processing: Image processing algorithms enhanced the OCT images by normalizing image quality and reducing noise.
- 3- Patient Data Management: This function ensured that all patient information and OCT images were identified and stored securely.
- 4- Disease Diagnosis: Tools for clinicians to diagnose and monitor retinal diseases were provided.
- 5- Report Generation: The system generates standardized clinical reports which clinicians can use to engage further.
- 6- Integration: Supports integration with existing hospital information systems.

3.2.3. Non-functional Requirements

- 1- Accuracy: The degree to which a measurement represents the true value of something (Naidu et al., 2023).
- 2- Reliability: A measure of how dependably an observation is the same when repeated. It refers to the measuring procedure rather than to the attribute being measured (Naidu et al., 2023).
- 3- Sensitivity: measures how often a test correctly generates a positive result for people who have the condition that's being tested. A highly sensitive test will flag almost everyone who has the disease and not generate many false-negative results (Shreffler & Huecker, 2020).
- 4- High-Specificity: measures a test's ability to correctly generate a negative result for people who don't have the condition that's being tested. A high-specificity test will correctly rule out almost everyone who doesn't have the disease and won't generate many false-positive results (Shreffler & Huecker, 2020).
- 5- Availability: The system should have high uptime and be available 24/7 for clinicians.
- 6- Redundancy: measures should be in place to ensure data integrity and availability.

3.2.4. Objectives

The main objective of the paper is to develop a system which can automatically classify OCT B-scans into the following categories.

1. Normal: This is for healthy retinal structure with no abnormalities or conditions.
2. CNV: It is a characteristic of wet AMD
3. DME: Here there is retinal thickening with intraretinal fluid
4. Drusen: Has sub-RPE deposits indicative of early AMD

3.2.5. Success Criteria

- **Performance:** The system must have a high accuracy of 95% and above. It should also have a sensitivity of 90% and for each pathological class on independent test data.
- **Clinical Utility:** The system should be interpretable and visualization to support clinician decision-making.

3.2.6. Stakeholder Requirements

Various stakeholders have different requirements. Clinicians need accurate predictions which they can integrate into the existing workflows. Patients need a system they can use to access rapid diagnostic screening. Healthcare facilities are looking for cost-effective solutions that reduce specialist workload while maintaining diagnostic quality.

3.3Phase 2: Data Understanding

3.3.1. Data Source

The primary dataset for the study was sourced from Kaggle Datasets, an open-source science platform for sharing, publishing and discovering datasets for machine learning and data science projects. For this analysis, the “Kermany2018” dataset was used, with its 84,000 images categorized into normal, diabetic macular oedema, and other eye diseases. The categorization is helpful for ophthalmology research and diagnosis. The figures below show images from Kaggle for an eye with diabetic macular oedema, a normal eye, Dursun and CNV.

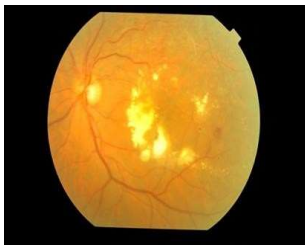


Figure 2: Diabetic macular edema

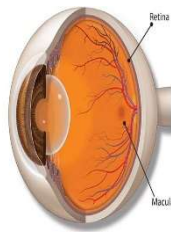


Figure 3: CNV



Figure 4: Dursun

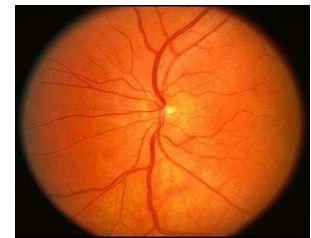


Figure 5: Normal eye

3.3.2. Initial Data Exploration

Initial assessment of the dataset identified a significant class imbalance: the Normal category accounted for the larger portion of the training set. The original validation set was small, with only 32 images, which was insufficient for reliable hyperparameter tuning. To address the problem, the validation set was expanded by reallocating a balanced 15% of the original data. Another set of 968 images was set aside for the final independent evaluation. Before the training processes commenced, sample images from Normal, CNV, DEM and Drusen were displayed in a 5x5 grid. This allowed the verification and orientation of characteristic pathological features.

Table 2: Counts the images in each subset and prints the totals

File	Count of Data points
------	----------------------

Training	83484
Testing	968
Validation	32

The pie chart below displays the visualization of the data distribution into testing, validation and training sets.

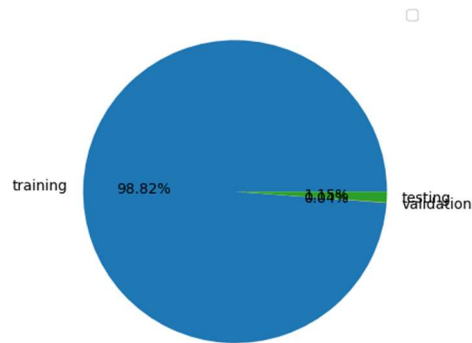


Figure 6:Proportion of images in training, validation, and testing sets

Table 3.1: Initial Dataset Composition

Class	Number of Images	Percentage	Image Dimensions	Format
CNV	37,206	44.0%	496×512–768×496	JPEG
NORMAL	25,774	30.5%	496×512–768×496	JPEG
DME	11,598	13.7%	496×512–768×496	JPEG
DRUSEN	8,617	10.2%	496×512–768×496	JPEG

3.3.3. Data Quality Assessment

A comprehensive data quality assessment was done to maintain the integrity of the research. All the missing files were removed to ensure all the images contained valid pixel data. The need for label integrity was achieved through cross-referencing with original publication and random visual inspection by an ophthalmology resident. Lastly, to ensure that the findings are a representative of clinical practice, approximately 3% of images with motion artifacts or poor fixation were intentionally retained to reflect real-world challenges.

3.3.4. Exploratory Data Analysis (EDA)

3.3.4.1 Sample

A stratified sample of 100 images per class were assessed to establish baseline diagnostic criteria:

- **CNV:** The presence of hyperreflective subretinal material, pigment epithelial detachment
- **DME:** Discrete cystoid intraretinal fluid spaces and retinal thickening

- **Drusen:** Sub-RPE deposits with varying reflectivity
- **Normal:** Distinct retinal layers without pathological disruption

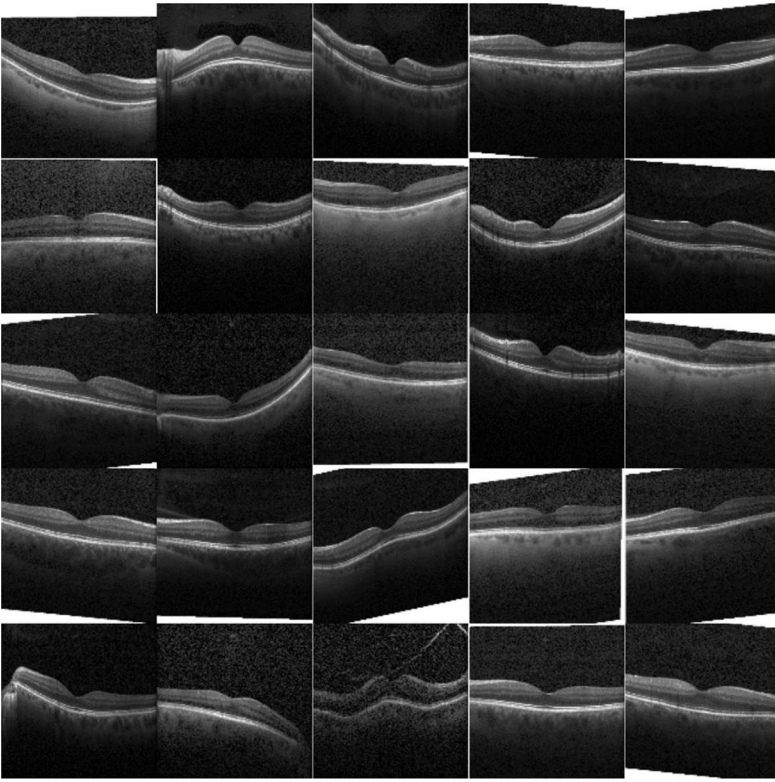


Figure 7: Normal

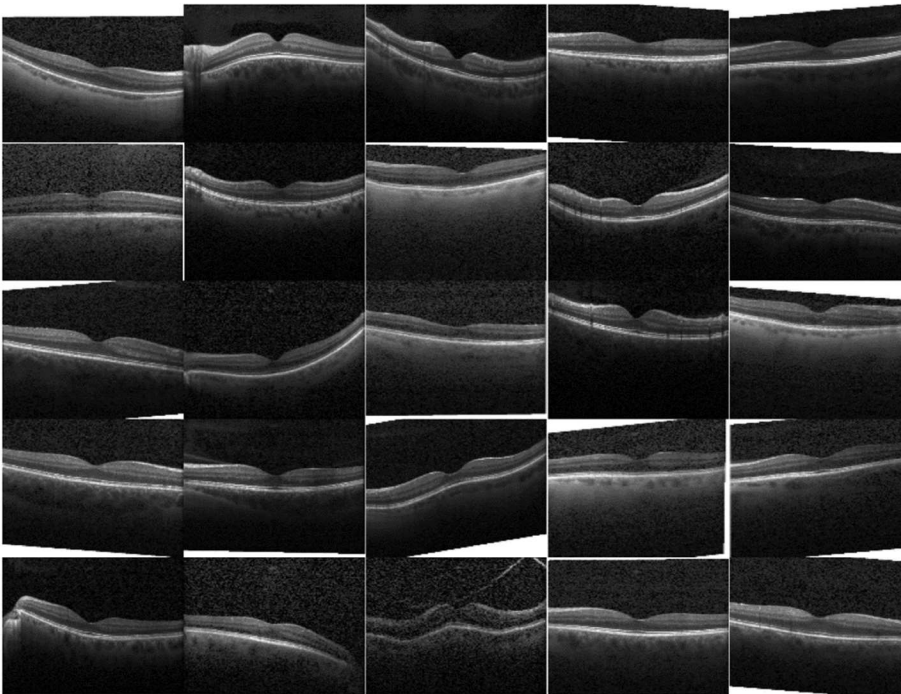


Figure 8: CNV

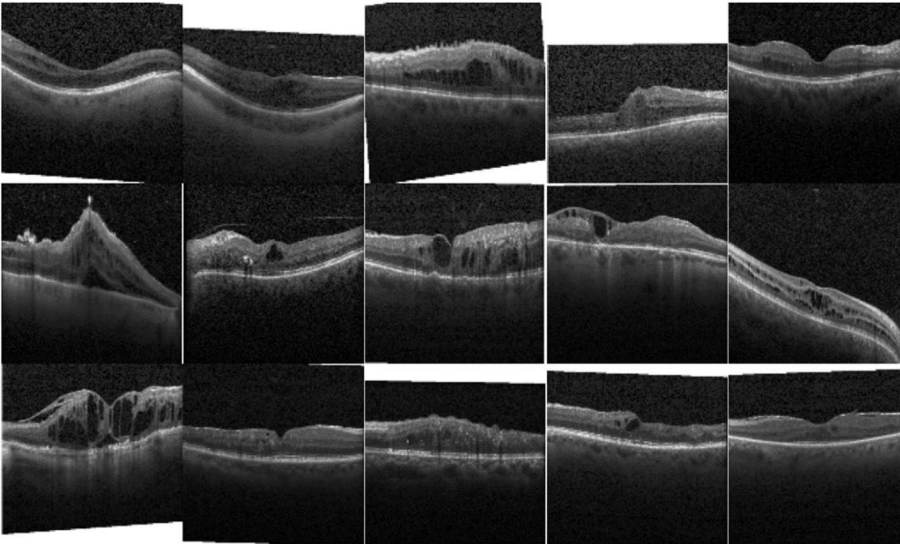


Figure 9: DME

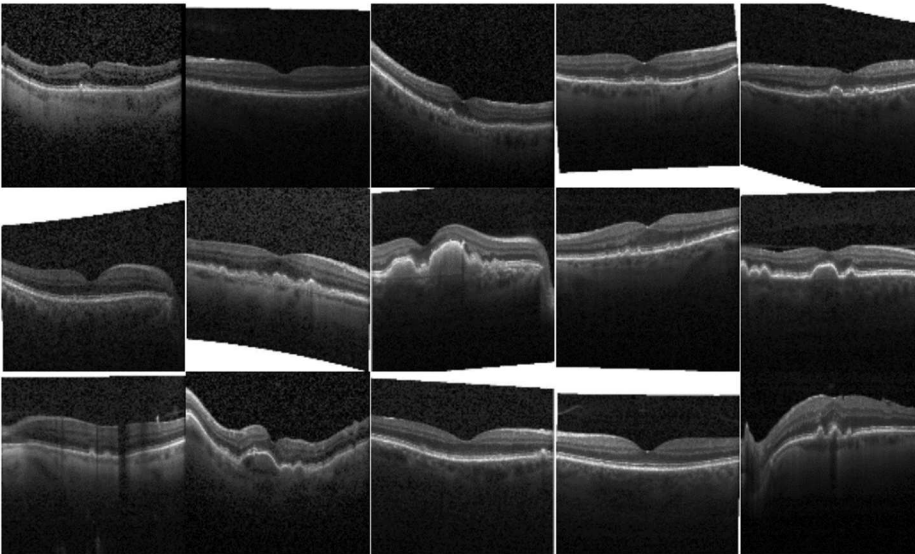


Figure 10: DRUSEN

3.3.5. Class distribution

The class distribution indicated that CNV images accounted for the highest proportion in the dataset, while Drusen had the lowest. DME and Normal classes exhibited moderate distributions. This indicates an imbalance in the dataset, which required augmentation or resampling. The bar chart below shows the dataset distribution.

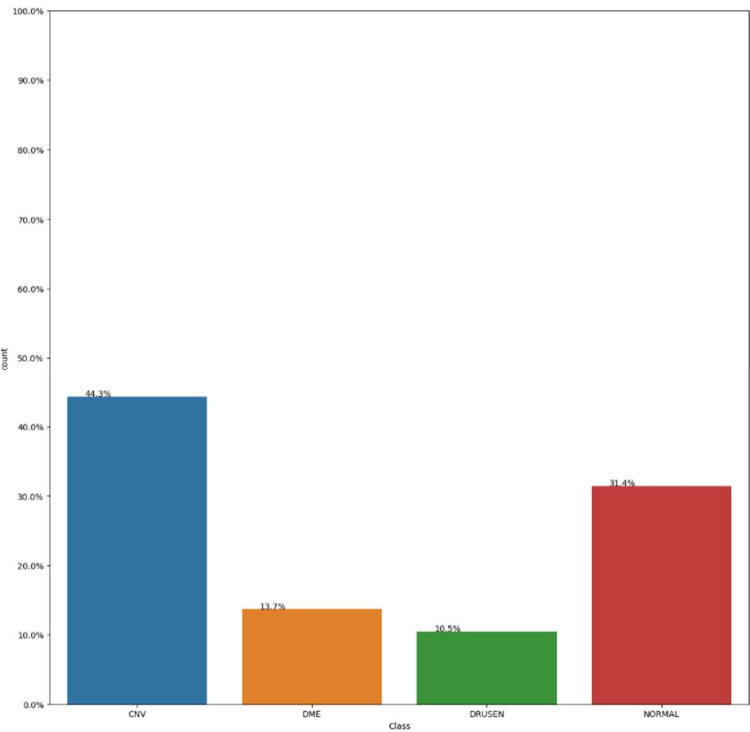


Figure 11:Class distribution

3.4 Phase 3: Data Preparation

3.4.1. Data Cleaning

Format standardization was used to clean the data. All the images were converted to RGB format achieve the same standard for data consistency. Through orientation normalization, all the images were rotated to standard anatomical orientation. Lastly, artifact management allowed the flagging of severe artifacts. However, they were still included to ensure model robustness.

3.4.2. Image Preprocessing Pipeline

Image Preparation

The image preparation process included creating a `load_data()` function to read file paths and assign classification labels. Training, Validation, and Test directories were merged into a single DataFrame.

3.4.3. Images Augmentation:

Since the model would be used in the healthcare industry, it needs generalize unseen data. Image augmentation techniques were applied to artificially increase the size of the dataset and variability. The augmentation process produced new image variations from the OCT scans taken from the Kaggle dataset. The augmentation process involved the use of TensorFlow ImageDateGenerator and Albumentations Library. In addition, geometric and intensity transformations used are shown in the table below.

Table 3:Augmentation Techniques Applied

Augmentation Technique	Purpose
------------------------	---------

Random cropping	Enhances feature learning by showing only partial regions of the retina
Horizontal & vertical flip	Allows the model to learn disease patterns independent of orientation
Rotation	Strengthens invariance to angle differences in OCT scans
Elastic deformation	Simulates natural retinal distortion or tissue variation
Grid distortion	Introduces warping to challenge the model to adapt
Hue/Saturation adjustments	Increases tolerance to light and color intensity variation

Augmentation Workflow

During the training process, the augmentation pipeline was used to allow every epoch a new set of image variations. The continuous generation of modified samples increased dataset richness without needing new clinical scans. This was done by loading the original images, application of augmentation transformations automatically during training and feeding enhanced images to the model for learning process. This allowed the balancing of learning, reduction of model biases and improvement of classification stability for all the four image categories.

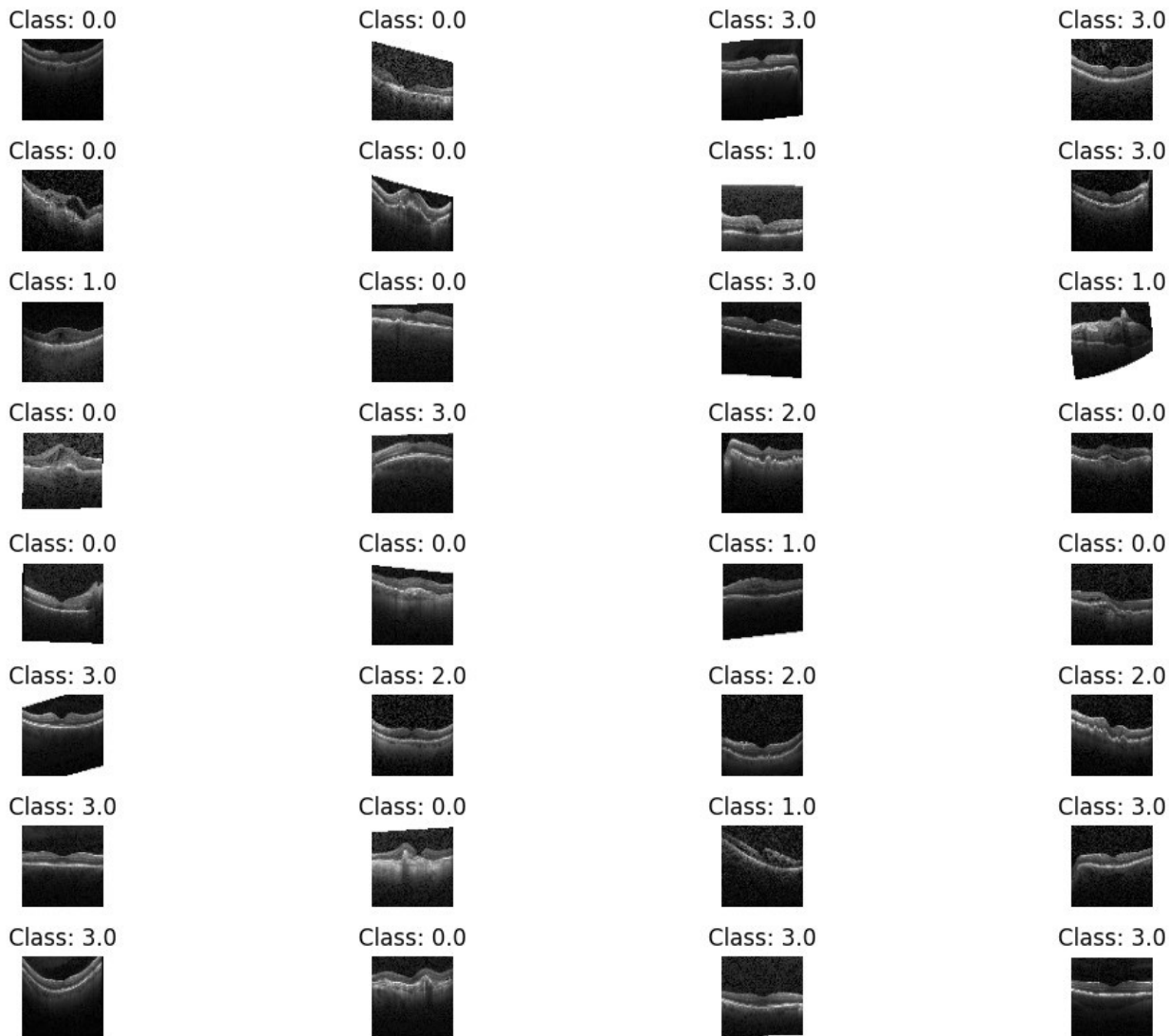


Figure 12:Sample images after Augmentation

3.4.4. Sampling Data:

A function `sample_data` was defined to create a smaller and representative sample, while ensuring that class proportions were maintained. The function resulted into a sample of 80000 images from the entire dataset. This ensure that a balance was maintained between the training, validation and testing splits, hence mitigating against biases.

Class	Number of Images
class0	12000
class1	11598
class2	8866
class3	12000

Total Data Size: 44464

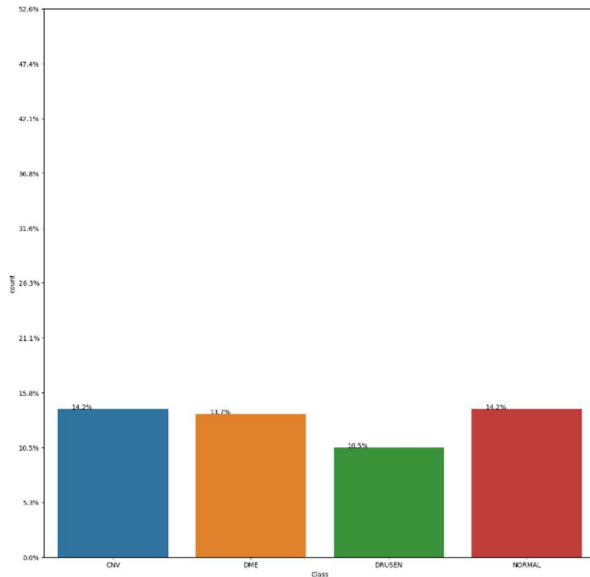


Figure 13 :updated Class distribution

3.4.4.1. Splitting Data into Train, Validation, and Test Sets:

The splitting was spread across a 72.5%-12.75%-15% split for training, validation, and testing sets, respectively. Class labels were converted to categorical format by using one-hot encoding. The table below shows the data splitting.

Table 4:distribution of images across the training, validation, and testing sets

File	Count of Data point
Training	32124
Testing	5670
Validation	6670

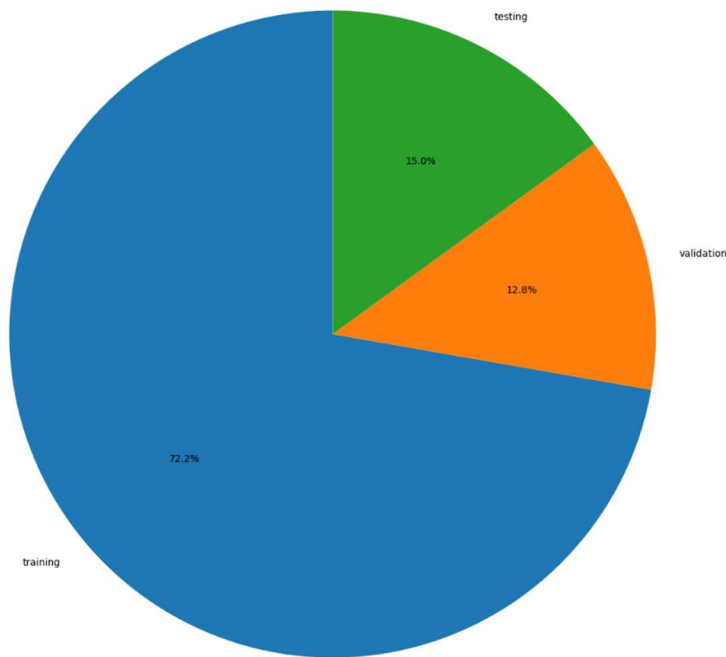


Figure 14: Updated proportion of images in training, validation, and testing sets

3.5 Phase 4: Modeling

3.5.1. Model Selection Rationale

Five architectures were selected for comparative analysis:

Table 3.3: Model Architectures and Rationale

Model	Depth	Parameters	Rationale
3-Layer CNN	3 conv layers	~50K	Baseline lightweight model
7-Layer CNN	7 conv layers	~500K	Medium complexity custom CNN
VGG16	16 layers	138M	Standard transfer learning baseline
ResNet50	50 layers	25M	Deep network with residual connections
DenseNet121	121 layers	8M	Parameter-efficient dense connections

3.5.2. Transfer Learning Implementation

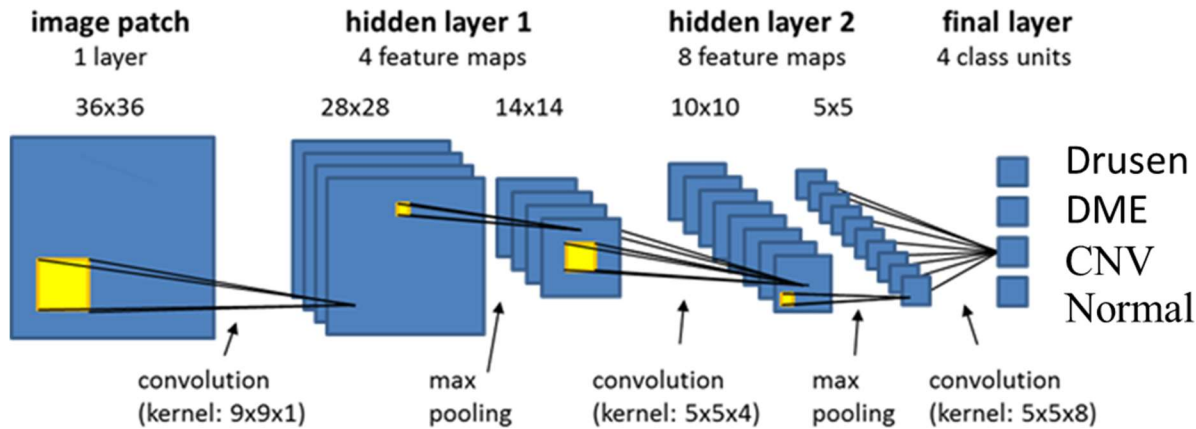


Figure 15: Fully CNN System with 4 Classes

For pre-trained models, the following fine-tuning strategy was employed:

1. **Base Model:** Load ImageNet weights, freeze all layers
2. **Custom Head:** Replace top layers with:
 - Global Average Pooling
 - Dense(512, ReLU) with Dropout(0.5)
 - Dense(256, ReLU) with Dropout(0.3)
 - Dense(4, Softmax)
3. **Gradual Unfreezing:** Unfreeze top convolutional blocks after initial training

3.5.3. Training Configuration

Hyperparameter Settings:

- **Optimizer:** Adam ($\beta_1=0.9$, $\beta_2=0.999$)
- **Initial Learning Rate:** 0.001 with ReduceLROnPlateau (factor=0.5, patience=3)
- **Batch Size:** 32 (optimized for GPU memory)
- **Epochs:** 50 with EarlyStopping (patience=10, monitor='val_loss')
- **Loss Function:** Weighted Categorical Cross-Entropy

Algorithm 3.2: Model Training Pseudocode

python

```
def train_model(model, train_gen, val_gen, class_weights):
```

```
    model.compile(optimizer=Adam(lr=0.001),
```

```
                  loss='categorical_crossentropy',
```

```
metrics=['accuracy', f1_score])
```

```
callbacks = [EarlyStopping(patience=10),
             ReduceLROnPlateau(factor=0.5, patience=3),
             ModelCheckpoint('best_model.h5')]
```

```
history = model.fit(train_gen,
                    validation_data=val_gen,
                    epochs=50,
                    class_weight=class_weights,
                    callbacks=callbacks)
```

```
return history, model
```

3.5.4. Hardware and Software Environment

- **Hardware:** NVIDIA RTX 3080 (12GB VRAM), 32GB RAM, Intel i7-12700K
- **Software:** Python 3.9, TensorFlow 2.10, CUDA 11.2, cuDNN 8.1
- **Development:** JupyterLab, VS Code with Git version control

3.5.5. Calculation details

Mathematical explanation

The percentage displayed by the system represents the class probability (confidence score) produced by the softmax function in the final layer of the convolutional neural network. It does not represent model accuracy.

- **Network Output (Logits)**

For each input OCT image, the trained model outputs a vector of real-valued scores (logits) corresponding to the four classes:

$$\mathbf{z} = [z_{Normal}, z_{CNV}, z_{DME}, z_{Drusen}]$$

These values differ for each image based on its visual characteristics.

- **Softmax Probability Calculation**

The logits are converted into normalized class probabilities using the softmax function:

$$P(class_i | image) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$$

Where:

- z_i is the logit for class i
- The probabilities sum to 1 (or 100%)

The system displays the highest probability as the final prediction confidence.

• Numerical Example Explaining 94% vs 99%

Case 1: High-quality Normal OCT image

$$\mathbf{z} = [5.0, 1.2, 0.8, 0.5]$$

$$P(\text{Normal}) = \frac{e^{5.0}}{e^{5.0} + e^{1.2} + e^{0.8} + e^{0.5}} \approx 0.99$$

Displayed result: Normal = 99%

Case 2: Normal OCT image with mild noise or lower contrast

$$\mathbf{z} = [3.6, 1.9, 1.5, 1.2]$$

$$P(\text{Normal}) = \frac{e^{3.6}}{e^{3.6} + e^{1.9} + e^{1.5} + e^{1.2}} \approx 0.94$$

Displayed result: Normal = 94%

• Mathematical Reason for the difference

The confidence score depends on the relative difference between the highest logit and the remaining logits:

$$\Delta z = z_{\text{best}} - z_{\text{second_best}}$$

- Larger $\Delta z \rightarrow$ higher confidence (e.g., 99%)
- Smaller $\Delta z \rightarrow$ lower confidence (e.g., 94%)

Even when both images belong to the same class (Normal), differences in image quality, contrast, and noise affect the logits and therefore the final probability.

3.6 Phase 5: Evaluation

3.6.1. Evaluation Metrics

The table below shows the evaluation metrics for each class

		Predict Class			
	Classes	NORMAL	CNV	DME	Drusen
Actual Class	NORMAL	TN	FP	TN	TN
	CNV	FN	TP	FN	FN
	DME	TN	FP	TN	TN
	Drusen	TN	FP	TN	TN

Figure 16: Evaluation Metrics

Primary Metrics:

1. **Accuracy:** $(TP+TN)/(TP+TN+FP+FN)$
2. **Precision (Class-specific):** $TP/(TP+FP)$
3. **Recall/Sensitivity:** $TP/(TP+FN)$
4. **Specificity:** $TN/(TN+FP)$
5. **F1-Score:** $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Secondary Metrics:

1. **AUC-ROC:** Area Under Receiver Operating Characteristic Curve
2. **Confusion Matrix:** Multi-class error analysis
3. **Inference Time:** Milliseconds per prediction

3.6.2. Statistical Validation

1. **Cross-Validation:** 5-fold stratified cross-validation on training data
2. **Statistical Testing:** McNemar's test was applied to compare the performance of the trained models.
3. **Confidence Intervals:** 95% CI calculated via bootstrap resampling (n=1000)

3.6.3. Explainability Methods

Gradient-weighted Class Activation Mapping (Grad-CAM) enhanced visual explanations. SHAP values analyzed feature importance. Through confidence calibration, reliability diagrams allowed calibration of errors.

3.6.4. Clinical Validation Protocol

To enhance human overview, a sample of 100 images was examined by two board--certified ophthalmologists. The predictions were compared against:

1. **Expert Consensus:** An agreement between both ophthalmologists
2. **Clinical Ground Truth:** The original diagnosis from medical records
3. **Model Performance:** Sensitivity/Specificity against expert labels

3.7Phase 6: Deployment

3.7.1. System Design

The system design shows how it captures, stores and analyzed retinal images, aiming to ensure it remains easy to us for medical professionals. The diagrams below represent the end-to-end system architecture showing the machine learning workflow.

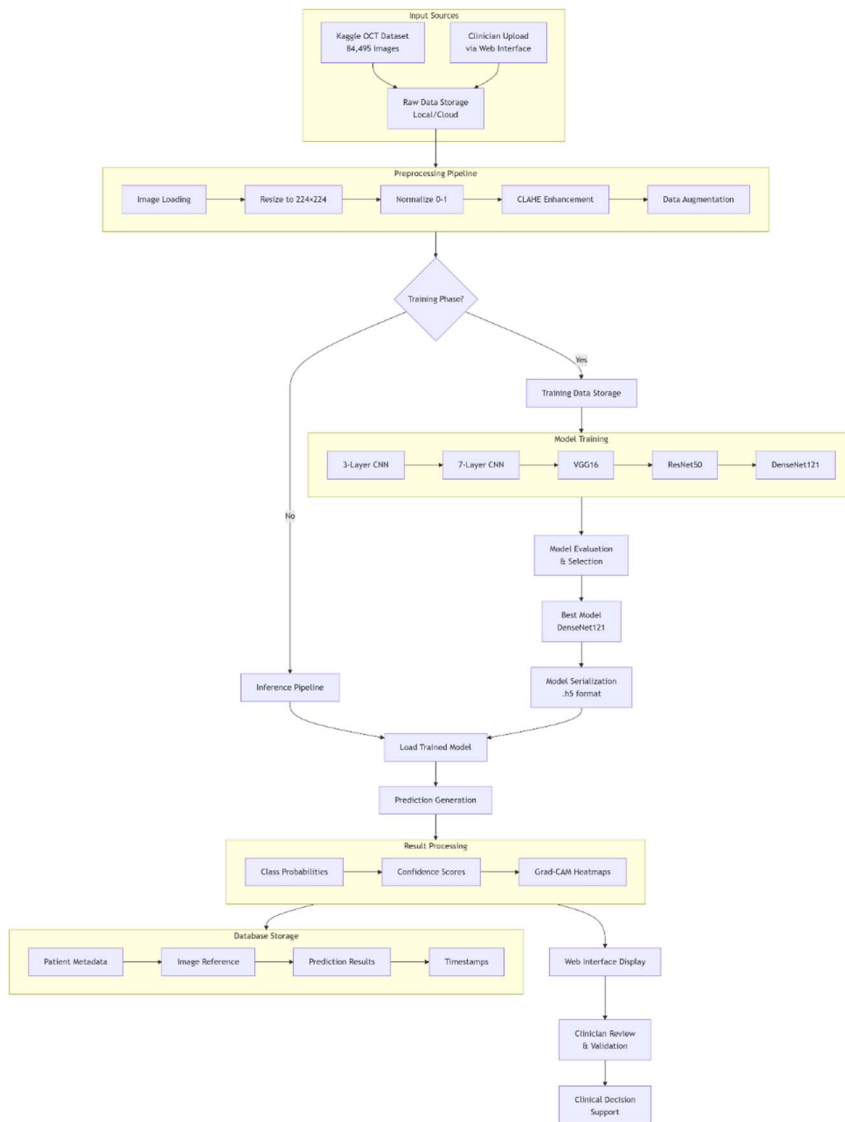


Figure 17: End-to-End System Data Flow

3.7.2. Use Case Design

The use case design diagram shows the visual representation of how users will interact with the OptiScan system. The main users of the system include doctors, technical staff, and system administrators.

Main Use Cases:

- **Register New Doctor:** The system administrator adds new doctors and inputs basic information like name, age, specialty, and any information required for doctors.
- **Login:** Doctors can securely access the system using their credentials.
- **Upload Fundus Images:** The images are securely stored in the database.

- **Preprocess Images:** The system enhances uploaded images to improve quality for analysis through reducing noise, improving contrast, and segmenting key structures.
- **Start Diagnosis:** Doctors select a patient record and initiate the diagnosis process. The system then loads and prepares the images for evaluation.
- **Detect Disease:** The system analyzes the fundus images to identify potential retinal conditions, considering features such as the optic disc shape, drusen, and hemorrhages.
- **View Results:** Diagnosis outcomes are displayed to the doctor, including possible conditions and confidence levels.
- **Save Records:** Doctors can save images and results for future reference, monitoring patient progress, or sharing with colleagues.
- **Develop System:** Technical staff maintain and improve the system by adding features, fixing bugs, and updating the machine learning models.
- **Test System:** Technical staff verify that the system functions correctly, including validating the models and ensuring that login, image upload, and other core features work as expected.

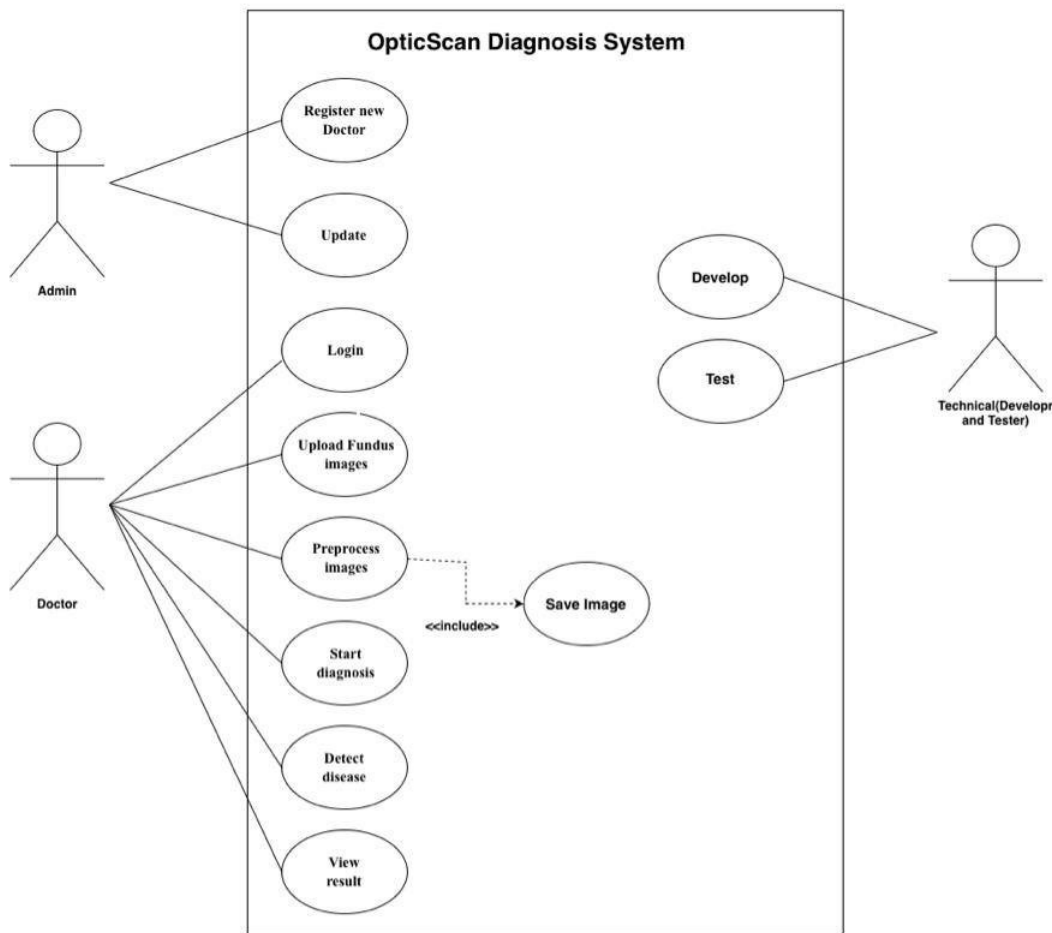


Figure 18: Use case diagram

3.7.3. System Workflow

First, divide data set into training and testing in data pre-processing, then will input data and apply data augmentation improvement to enhance the constraint. Furthermore, perform training by use model training if it is achieved convergent, we will go to the final model, in the testing model will classify the data Normal, CNV, DME and Drusen.

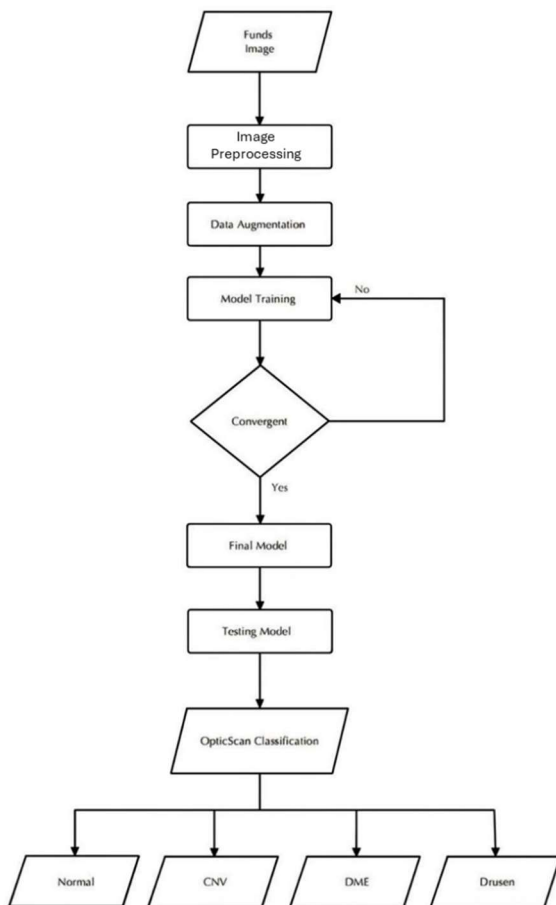


Figure 19: System workflow

3.7.4. Architecture Pattern

Components within the layered architecture pattern organized into horizontal layers, each layer performing a specific role within the application (e.g., presentation logic or business logic). The layered architecture pattern closely matches the traditional IT communication and organizational structures found in most companies, making it a natural choice for most business application development efforts. Each layer of the layered architecture pattern has a specific role and responsibility within the application. please refer to figure 9 to see a layer architecture pattern which related functional requirements to the user [16]:

1. The layers of isolation concept also means that each layer is independent of the other layers.
2. Used to decompose the system into cohesive layers with interfacing

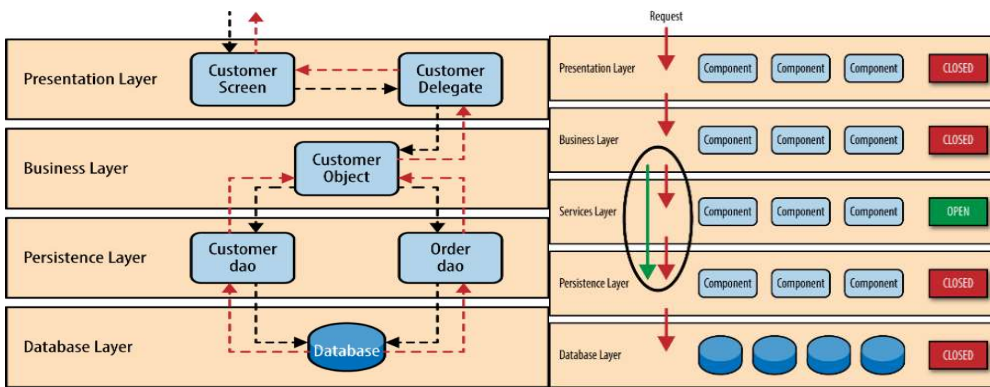


Figure 20:Layer

3.7.5. System Main Components:

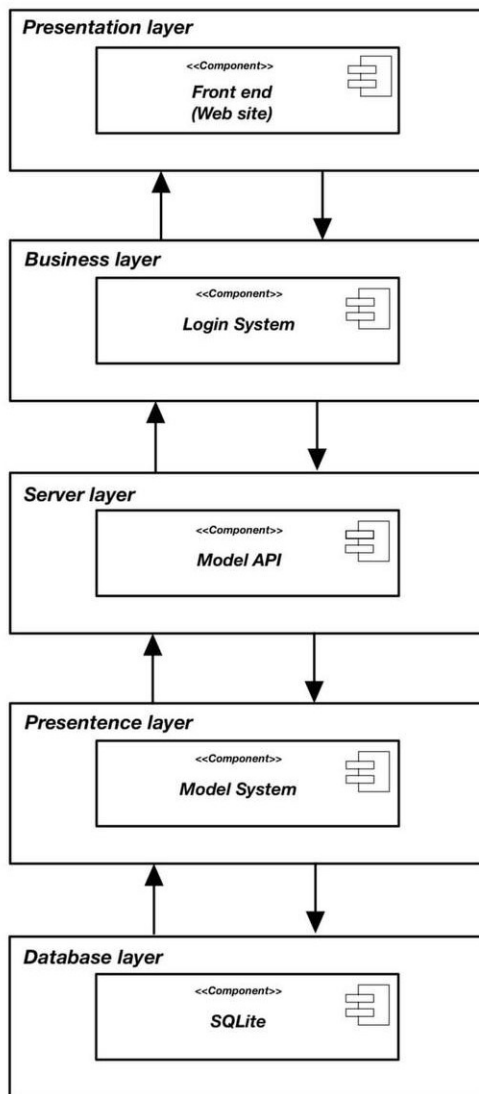


Figure 21:System components

3.7.6. Package diagram:

The system has three main components:

Classes: This layer handles the business logic of the system. It defines the data model, implementing the business rules and offering APIs for all the other components that interact with the system.

Front end: The layer displays the user interface and tackling all the user interactions. It will render the user interface, processing user input and requesting for resources from the backend part.

Backend: The backend is responsible for storing and managing data. It will also process all the business logic requests from the front-end layer.

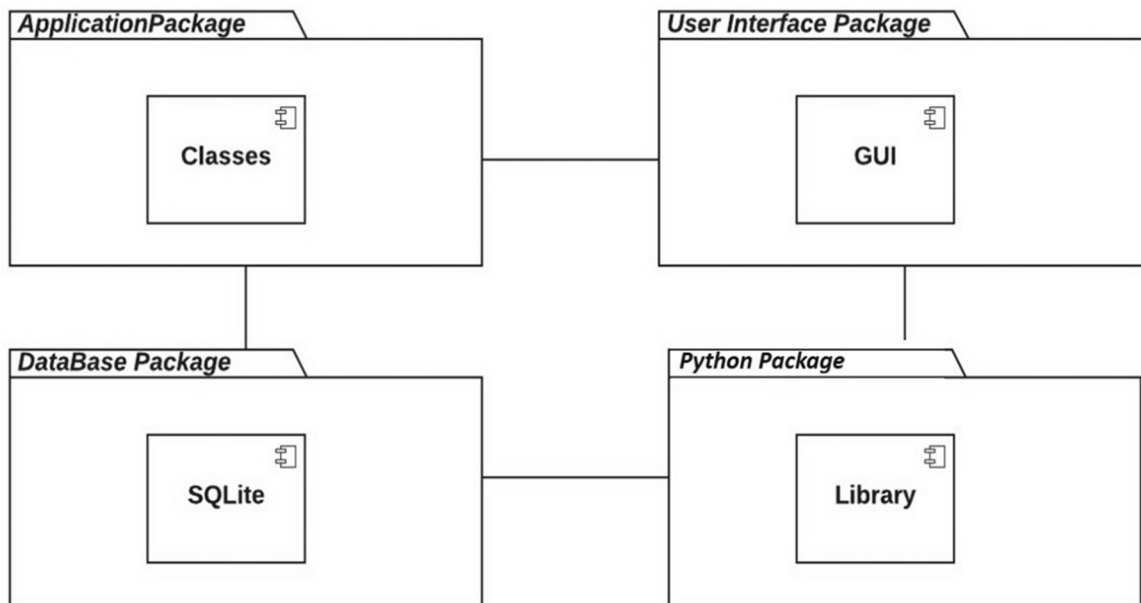


Figure 22:Package Diagram

3.7.7. Class Diagram

- The class diagram shows the details of the system and its components and how they interact with the different users.
- The users remain the core of the system. They will be interacting with the system through login and diagnostic features. Users are authenticated to securely access system features.
- The diagnostic component analyses the retinal images
- The developer maintains and updates the system. The developer interacts with the database and ensures that all the system components work as expected. The database is the repository for all the system's data, containing user information, details of the model and the outcomes of the system productions.
- The Train() function is responsible for updating the system by learning from new data. When new data comes, the function adjusts the model to improve the future predictions to enhance accuracy.
- The Predict() function analyzes the data from the users. This gives the diagnostic outcome based on the patterns the system has learned.

- The Use() function applies the prediction results to carry out the required tasks like displaying a diagnosis or supporting clinical decision-making.
- The SaveImage() function stores the output of the analysis as an image file. This permits the results to be preserved for future reference, comparison, or record keeping.

It is important to note that the system has been designed to be flexible. In the future additional components can be added to the architecture to ensure the scalability of the system as the number of users increase

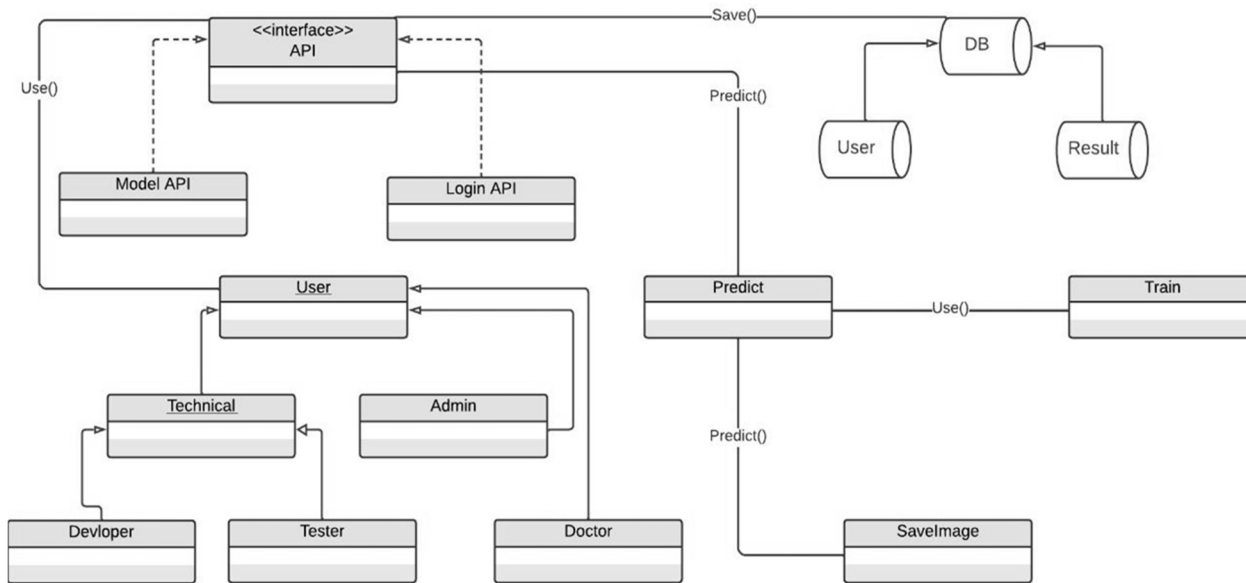


Figure 23:Class Diagram

3.7.8. Sequence Diagram:

The sequence diagram has the following components:

1. Login and image upload:
 - The doctor logs into the system.
 - The system validates the doctor's login information.
 - The doctor uploads an image.
2. Prediction and save:
 - The system sends the image to the ML model for prediction.
 - The ML model returns the prediction results to the system.
 - The system saves the prediction results to the database.

The following messages are exchanged as different objects interact with the system.

- Doctor to System: Login
- System to Doctor: Validate login information.
- System to Doctor: Display login success message
- Doctor to System: Upload image
- System to ML Model: Predict diagnosis.
- ML Model to System: Return prediction results
- System to Database: Save prediction results.

The diagram below visualized the doctor sequence diagram

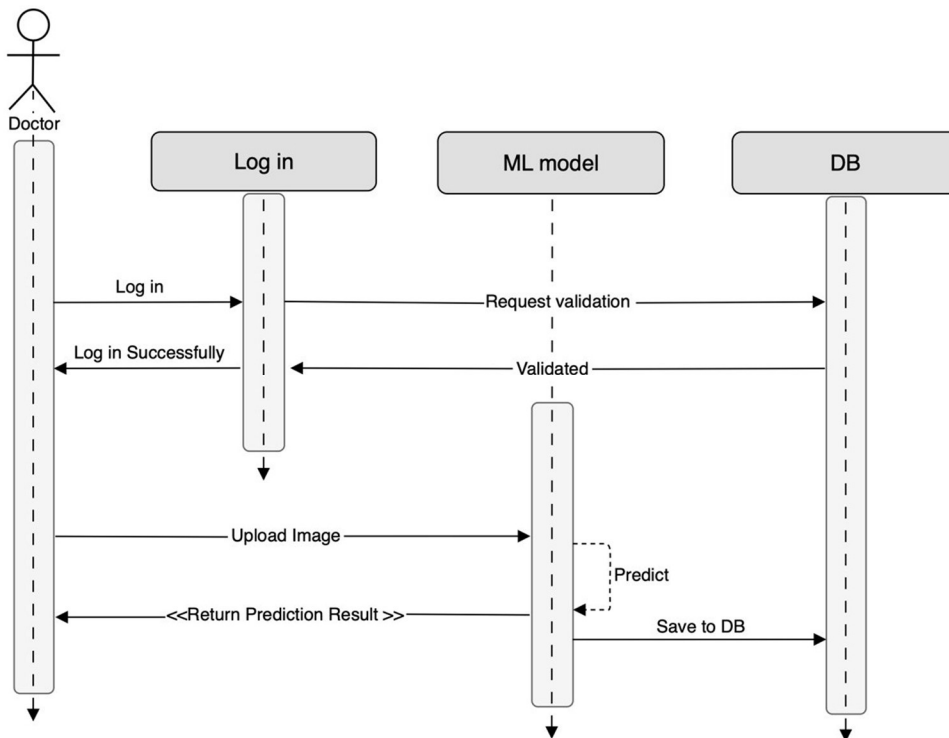


Figure 24: Doctor Sequence Diagram showing interactions among Admin, Register, and DB

The process commences when the Admin object sends a Register message to the Register object. After the validation of the registration information, the Register object sends a Validate Register info to the DB object. The information is saved to the database. A successful registration allows the Register object to send a Register Successful message back to the Admin object. The diagram has the following control measures.

- The Validate Register info message from the Register object to the DB object is a synchronous message, which means that the Register object will wait for a response from the DB object before continuing. allows the Register object to establish if the registration was successful before proceeding.

- The Register Successfully message from the Register object to the Admin object is an asynchronous message. Hence, the Register object does not wait for a response from the Admin object before continuing.

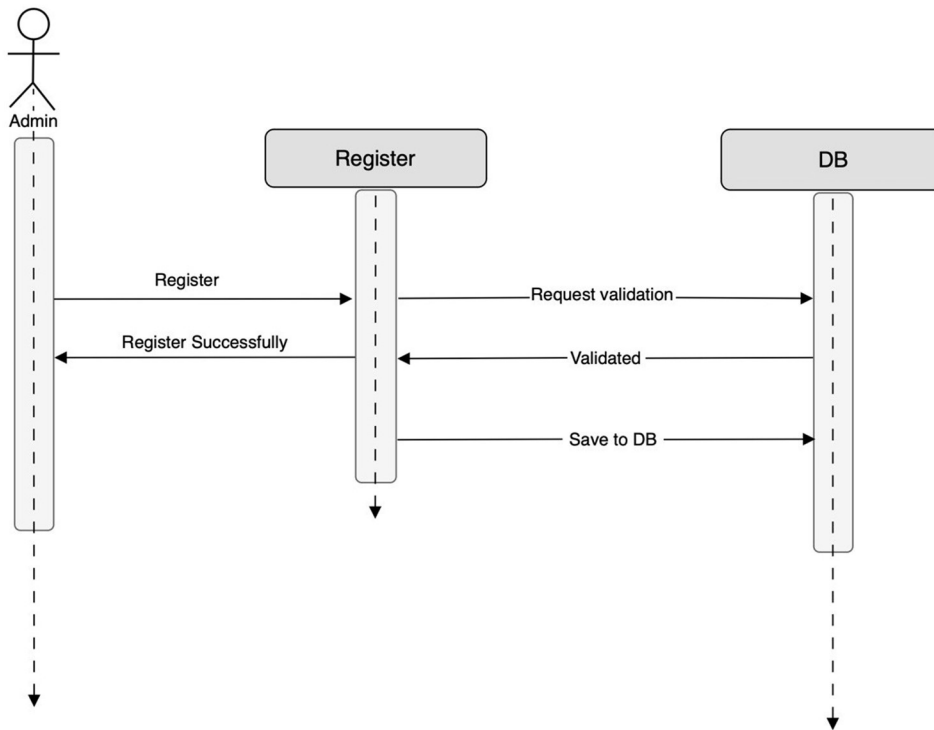


Figure 25:A Simple and Straightforward Admin Sequence Diagram

3.7.9. The Relational Database Schema

The relational database schema shows a table called Doctor Model Table. This table has the following columns:

- Doctor ID: A unique identifier for the doctor.
- Image ID: A unique identifier for the image.
- Prediction Class: Indicates the predicted diagnosis for the image.
- Prediction Percentage: Represents the probability that the predicted diagnosis is correct.
- Timestamp: Records the date and time of the prediction.

This table can be used to implement various features in a medical system, such as:

- Displaying the predicted diagnosis for an image to the doctor.
- Tracking the doctor's accuracy over time.
- Generating reports on the most common diagnoses for different types of medical images.
- Indicating trends in the prevalence of various diseases

Overall, the relational database schema provided is suitable for storing predictions from the ML model on medical images. The following considerations are necessary when designing a relational database schema suitable for a medical system.

- **Normalization:** Normalization is a process of organizing data in a database to minimize redundancy and improve data integrity. Different levels of normalization ensure various levels of data integrity., depending on the needs of the system.
- **Security:** A database schema should be secure to protected the data stored. The appropriate data types and constraints should be used, while encrypting any sensitive data.
- **Performance:** The database schema should be designed to support the expected workload of the system. This includes optimizing the schema for the most common queries and operation

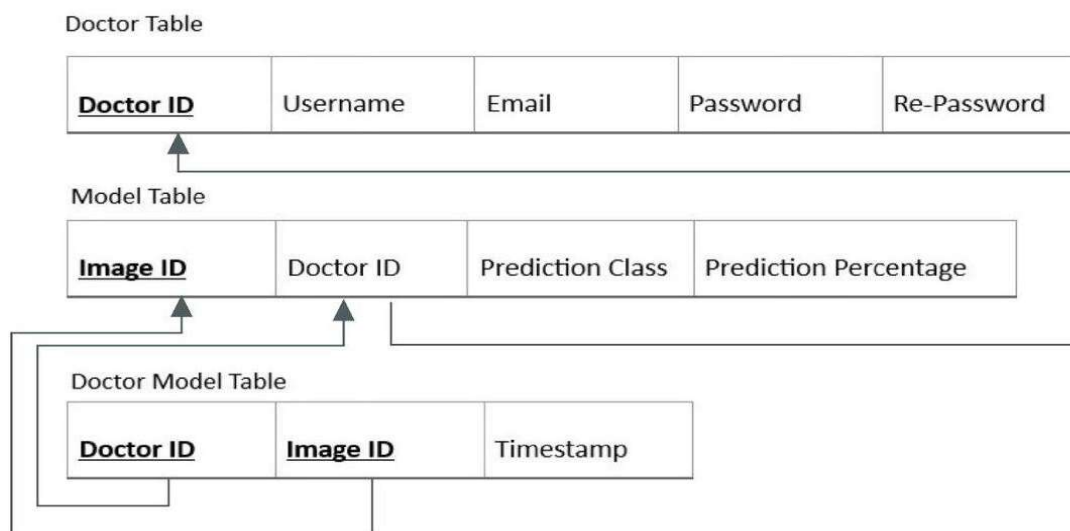


Figure 26:Relational Database Schema

3.7.3 System implementation:

3.7.3.1. Build HTML pages:

HTML pages were created to form the foundation of web-based systems.

DME.html

HOMECONTACT USABOUT US

Diabetic Macular Edema (DME)



Diabetic Macular Edema (DME)

Diabetes is the leading cause of new blindness in the United States, with DME contributing greatly to this vision loss. DME may affect up to 10% of people with diabetes. DME is a complication of diabetes caused by fluid accumulation in the macula that can affect the fovea. The macula is the central portion in the retina which is in the back of the eye and where vision is the sharpest. Vision loss from DME can progress over a period of months and make it impossible to focus clearly.

What Causes DME?

DME is an eye condition which can occur in people living with diabetes – both type 1 and type 2. Consistently high blood sugar due to poor glucose control over time can damage small blood vessels in the body, including the eye. Diabetic retinopathy is a disease that damages the blood vessels in the retina, resulting in vision impairment. Left untreated, fluid can leak into the center of the macula, called the fovea, the part of the eye where sharp, straight-ahead vision occurs. The fluid makes the macula swell, blurring vision. This condition is called DME. It can occur at any stage of diabetic retinopathy, although it is more likely to occur as the disease progresses.

What Causes DME?

- Blurred vision
- Double vision
- Sudden increase in eye floaters

What Increases Risk for DME?



- Poor glucose control over a long period of time
- Increased length of time living with diabetes- The lifetime risk of developing DME for people with diabetes is about 10%.
- Type of diabetes – After living with diabetes for 20 years, nearly all with type 1 and 60% with type 2 diabetes will have some diabetic retinopathy. It usually does not affect vision in the early stages.
- Very high blood pressure
- Fluid retention
- Kidney disease

Useful Links

- Home
- Useful Articles

Where We Are

Riyadh,
King Abudallah Road
Contact-no 00966
Email: info@VivekNetralaya.com

Figure 27:DME disease explanation page.html

Dursen.html



[HOME](#)
[CONTACT US](#)
[ABOUT US](#)

What Is DRUSEN?



Disc drusen are composed of small proteinaceous material that become calcified with advancing age. These deposits can be considered small tumors that develop within the optic nerve head, and may lead to an elevated disc (and therefore this condition is sometimes referred to as pseudopapilledema).

They can be associated with retinitis pigmentosa, angioid streaks, Usher syndrome, Noonan syndrome and Alagille syndrome. They also may lead to a loss of visual field or, in rare cases, central acuity or anterior ischemic optic neuropathy.

Disease Entity

- Consist of refractive, hyaline-like calcified nodules made of mucoproteins and mucopolysaccharides that are located within the optic nerve head
- Prevalence 3.4-24/1000
- Males = Females
- Males = Females
- >70% bilateral, but asymmetric
- Autosomal dominant inheritance with incomplete penetrance is suspected

Pathophysiology

Pathogenesis is thought to be related to possibly inherited dysplasia of the optic disc with blood supply compromise, causing slowed axoplasmic flow and leading to the formation of calcific excrescences.

Diagnosis

Diagnosis is made by clinical examination of the optic nerve head in addition to ancillary imaging studies (see below). Pseudopapilledema vs. true papilledema must be distinguished.

History

Patients are usually asymptomatic. Rarely, they can have transient visual obscurations (approximately 9% of patients). Visual field loss may also be seen gradually, and it has been estimated that a quarter of patient of patients with optic disc drusen have a visual field defect other than an enlarged blind spot.

Useful Links

- Home
- Useful Articles

Where We Are

Riyadh,
King Abudallah Road
Contact-no 00966
Email: info@VivekNetralaya.com

Figure 28:Drusen Disease explanation page .html

ForgetPassword.html



Name

Email

Forget

Already have an account? [Log In here!](#)

Don't have an account? [Sign up here!](#)

Useful Links

[Home](#)
[Useful Articles](#)

Where We Are

Riyadh,
 King Abudallah Road
 Contact-no 00966
 Email: info@VivekNetralaya.com

Figure 29:ForgetPassword page to reset account password.html

Index.html

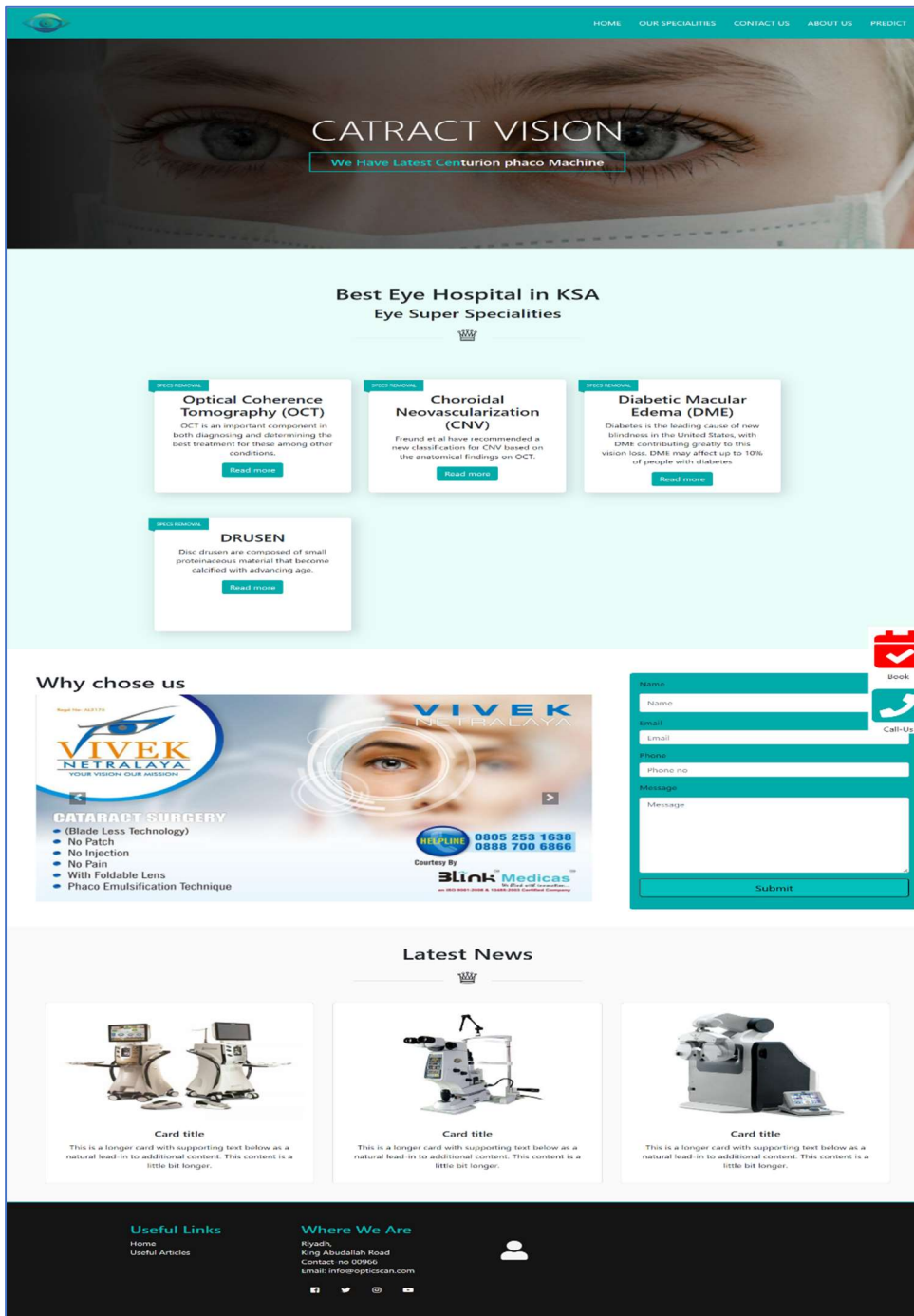


Figure 30:Index main page information.html

Login.html



Doctor only

Email

Password

[Log in](#)

[Forget ID or Password here!](#)

[Don't have an account? Sign up here!](#)

Useful Links

[Home](#)

[Useful Articles](#)

Where We Are

Riyadh,
King Abudallah Road
Contact-no 00966
Email: info@VivekNetralaya.com



[f](#) [t](#) [@](#) [v](#)

Figure 31:Login page for the doctor to use the prediction.html

OCT.html

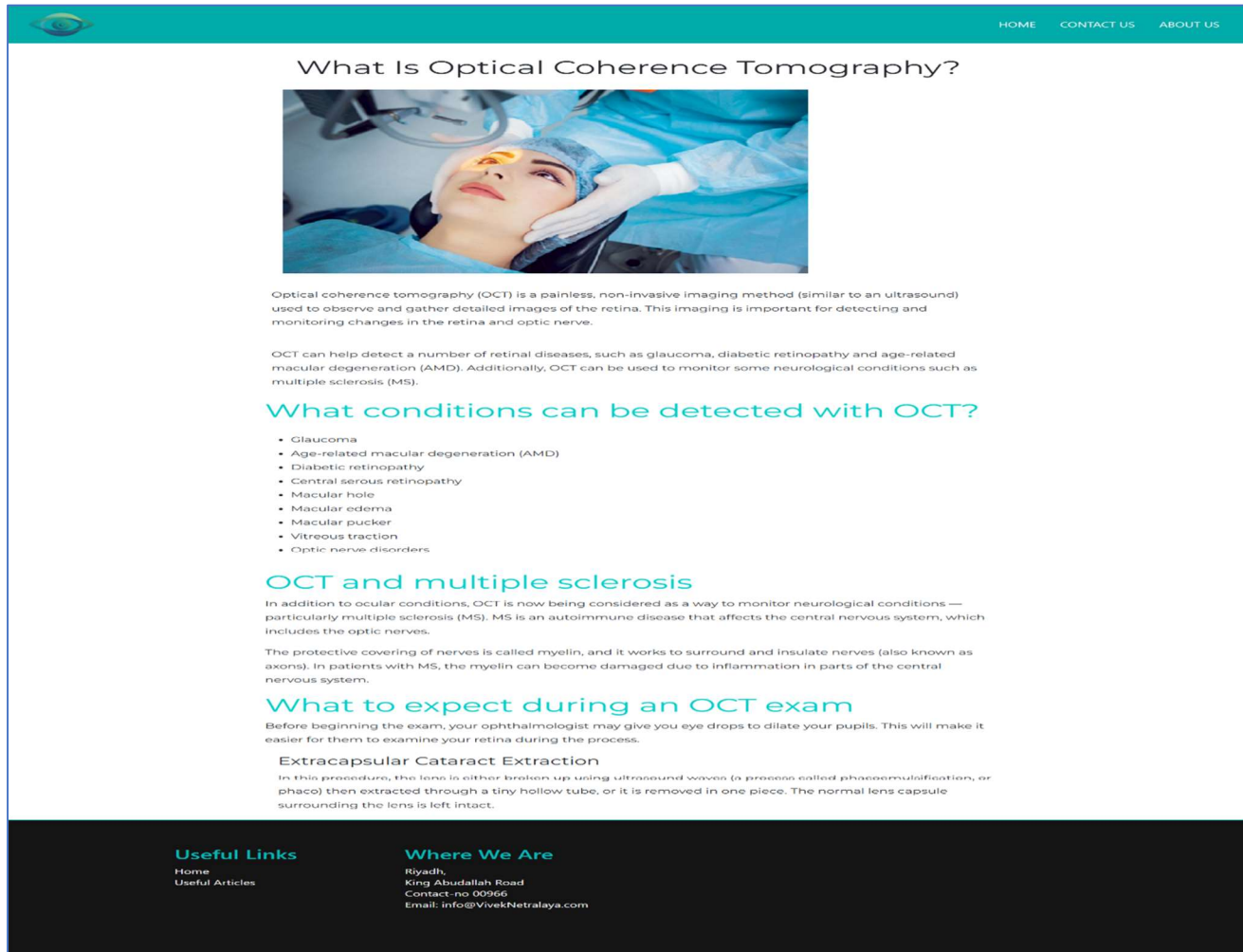


Figure 32:OCT disease explanation page.html

Predict.html

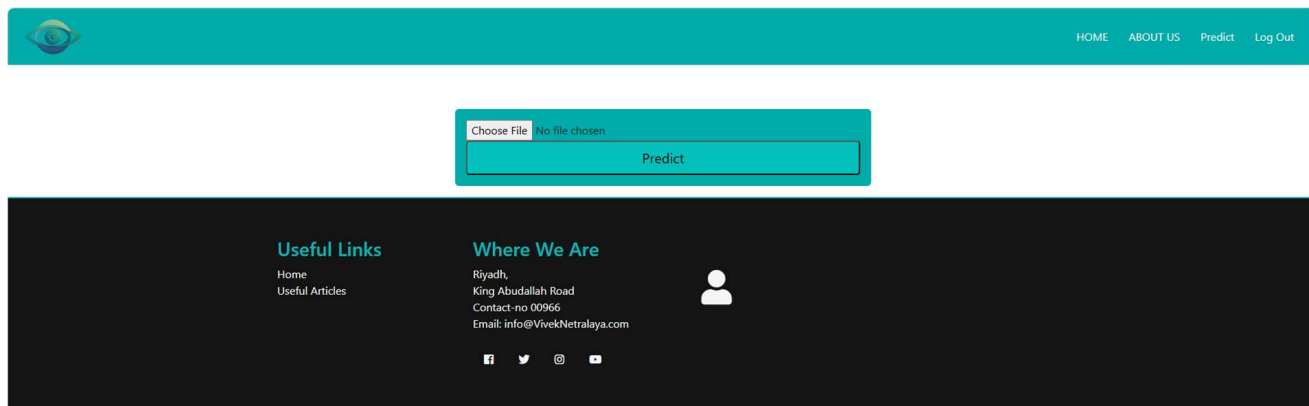


Figure 33:Prediction page where the doctor chose the oct image.html

Results.html

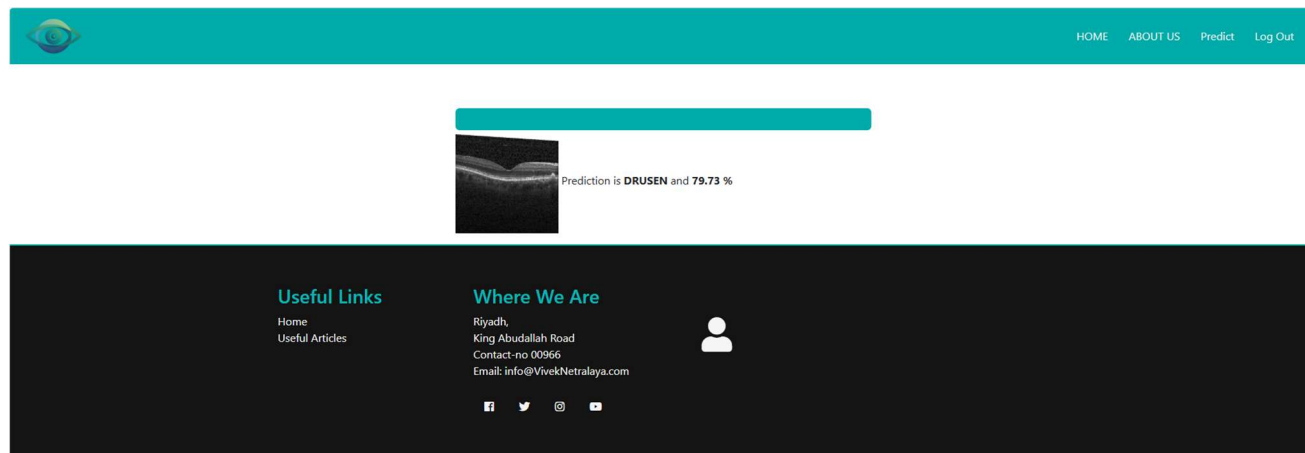


Figure 34: Results of the prediction.html

Signup.html

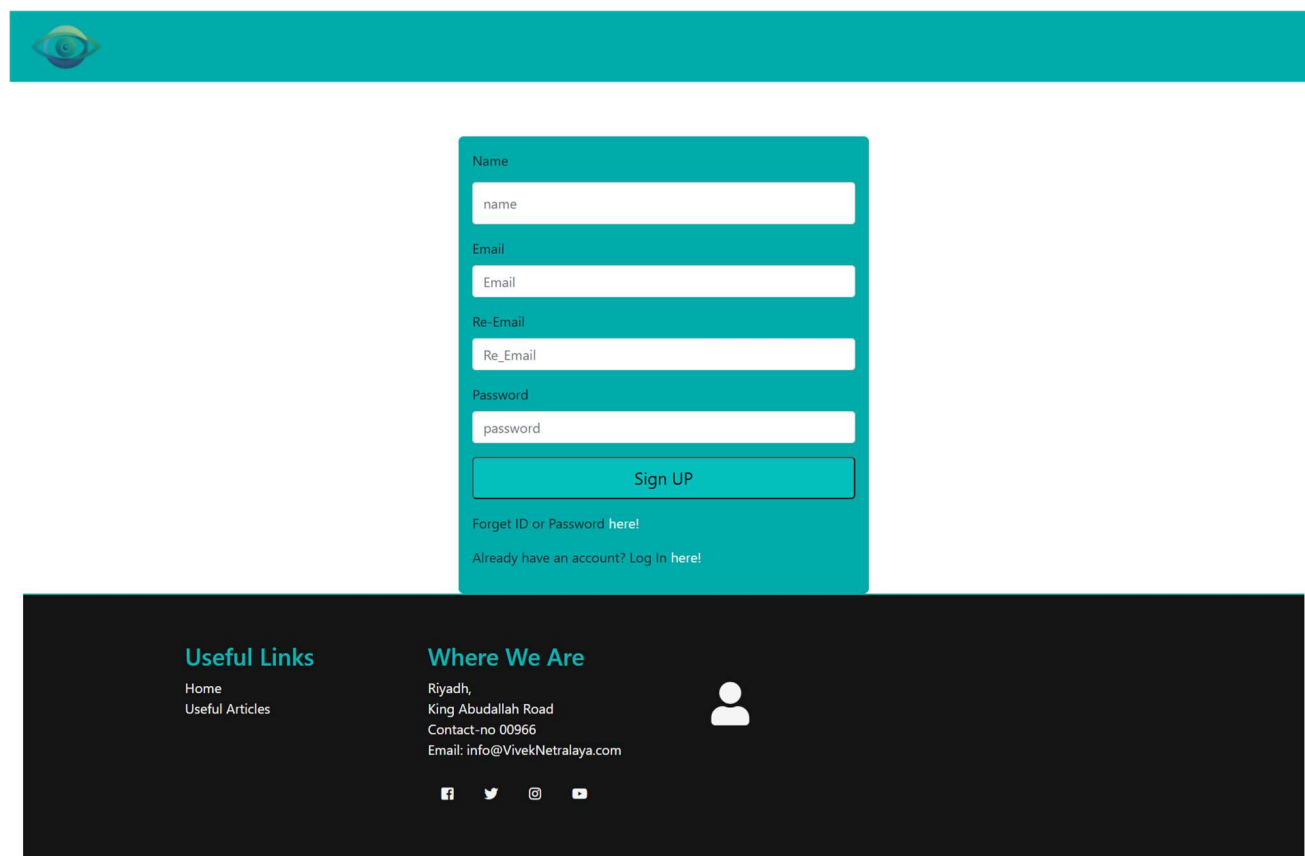


Figure 35: Signup page for the doctor.html

3.7.3.2. Build DB:

SQLite was used because it is lightweight and suitable for smaller projects especially in the local development. Since it has a smaller footprint, it does not need a separate server process. This makes it easier to embed within an application in the local environment.

3.7.3.3. Build System App

User Authentication Module (Login–Signup–Password Recovery)

The need for security prompted the creation of a user account management module. The module allows the creation of accounts, authenticated users and facilitates password recovery using the SQLite backend.

(1) Signup Method – Create New User Account

The creation of a new user follows the following steps:

- Inputs include: **name, email, repeated email, and password.**
- On execution, the function opens a secure connection to the database and attempts to store the provided user details.
- If the insertion succeeds, the changes are committed and the method returns success; otherwise, it returns failure.

(2) Login Method – User Authentication

The Login method checks if the credentials entered match the entries already stored in the database.

- If no matching record exists, access is denied.
- A valid match indicates successful authentication, granting the user access to their dashboard. This approach enables controlled entry and maintains system security through identity verification.

(3) Forget-Password Method – Account Recovery

The password-recovery method retrieves stored credentials when a user provides their registered **name and email address.**

- If both parameters match a database entry, account details are returned to allow recovery.
- Otherwise, an empty response indicates invalid credentials. This ensures continuity of access without requiring full account re-registration.

Appointment Management Module

Apart from the user login, another appointment-handling component was implemented using SQLite.

(1) Appointment Booking Function

- Stores user appointment requests including **name, contact number, and selected date.**
- Data is inserted into the Appointment database, and successful submissions return a confirmation.

(2) Opinion Submission Function

- Allows users to submit feedback/opinions through a similar database-insert procedure.
- Records include name, email, phone contact, and message text.

Both operations use parameterized queries to improve data safety and prevent SQL injection.

Flask API Integration

The integration of the model with the web interface was done using Flask-based API as shown in the table below.

Table 5: Flask API Integration

Component	Purpose
Model Loader	Loads the trained OCT classification model into memory
Prediction Handler	Processes uploaded images and returns predicted class
User Authentication Routes	Manages login, signup and password-reset requests
Appointment Routes	Stores appointment requests and user opinions
Web Page Rendering	Serves HTML templates to users via the browser

When an OCT image is uploaded, the API preprocesses it. Then it runs inference using the deep-learning model, and displays output back to the clinician along with diagnosis probability.

3.8 Ethical Considerations

3.8.1. Data Privacy

- To enhance data privacy, de-identification of data and the use of publicly available dataset was employed.
- No patient identifiers stored or transmitted
- Compliance with GDPR and HIPAA principles

3.8.2. Bias Mitigation

- Stratified sampling across demographic variables was done to ensure that the sample was representative of the dataset
- Regular fairness audits using disparate impact analysis
- Transparency in model limitations

3.8.3. Clinical Safety

- The system was designed as decision support. Hence, it does not operate as an autonomous diagnosis system. A professional must validate the outcomes of the model
- Clear disclaimers on prediction confidence
- Escalation pathways for uncertain cases

3.9 Methodological Limitations

Despite the robustness of the model, it has some limitations. Firstly, having the primacy data from a single source limits the generalizability of the model. Secondly, the performance of the model may differ in prospective clinical use. Lastly, there are computational constraints, limiting the model to 2D B-scans. As a result, 3D volume images were not analyzed.

3.10 Summary

The chapter outlined the methodology used to design OptiScan retinal disease classification system. The study employed the CRISP-DM framework. The chapter also examined OCT dataset and used it to refine model to ensure it met clinical performance measures. The steps used establish a solid foundation for the next chapter, which builds, tests and is applied in the practical diagnostic setting.

CHAPTER 4: RESULTS AND IMPLEMENTATION

4.1 Model Performance Evaluation

This chapter shows the results from training and testing of the five deep learning models for the classification of OCT mages. The models include a 3-layer CNN, a 7-layer CNN, VGG16, ResNet50, and DenseNet121. Each model was trained for 50 epochs with early stopping applied. Performance was assessed on the training, validation, and test datasets using standard evaluation metrics.

4.2 Training Performance

The models demonstrated reduced loss in the training process, showing that the learning was effective. For instance, the 7-layer CNN showed the most stable convergence with minimal overfitting. Pre-trained models showed faster initial convergence but needed careful fine-tuning to avoid catastrophic forgetting.

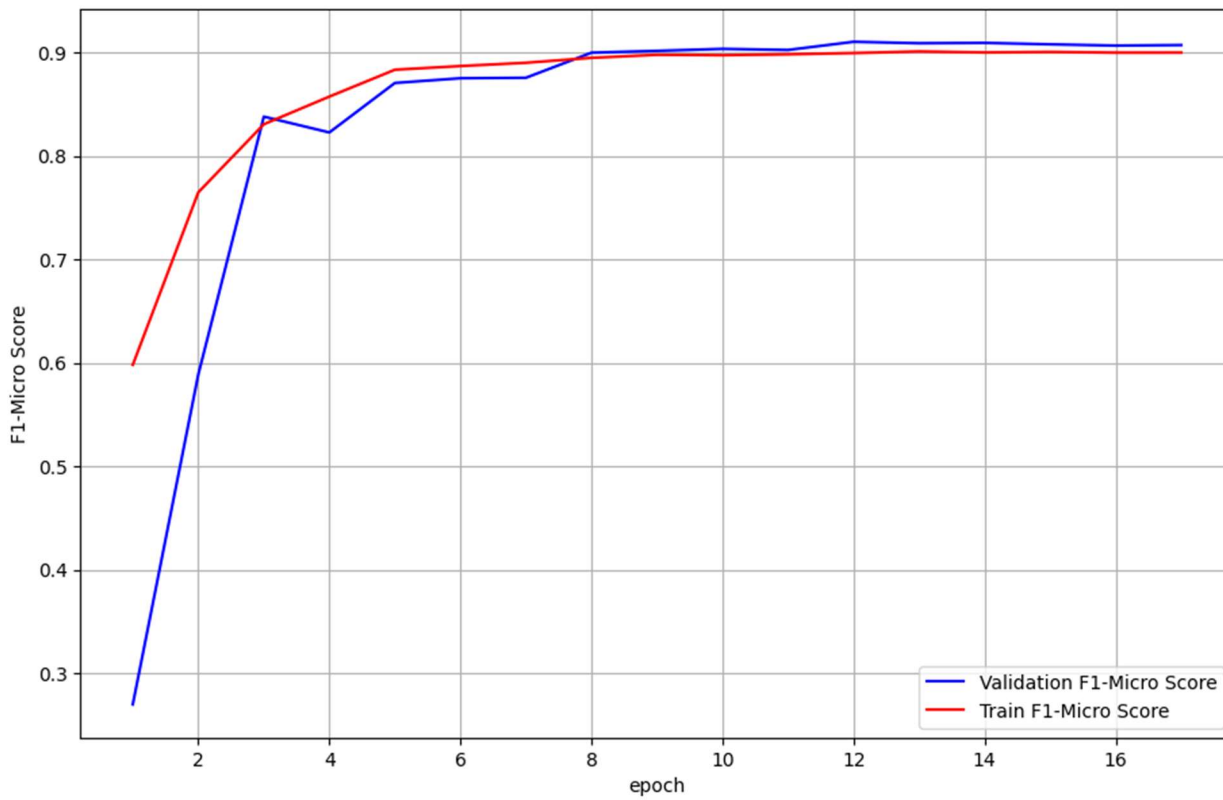
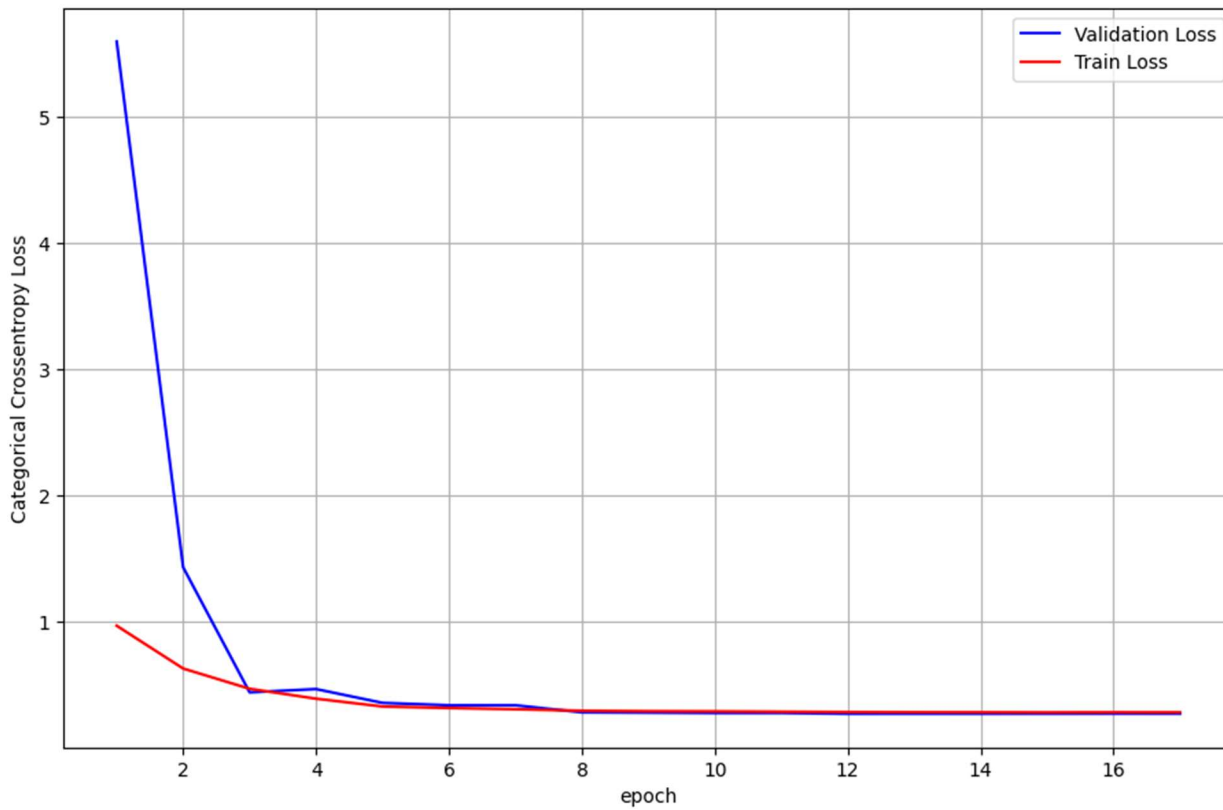


Figure 36:3 -Layer CNN

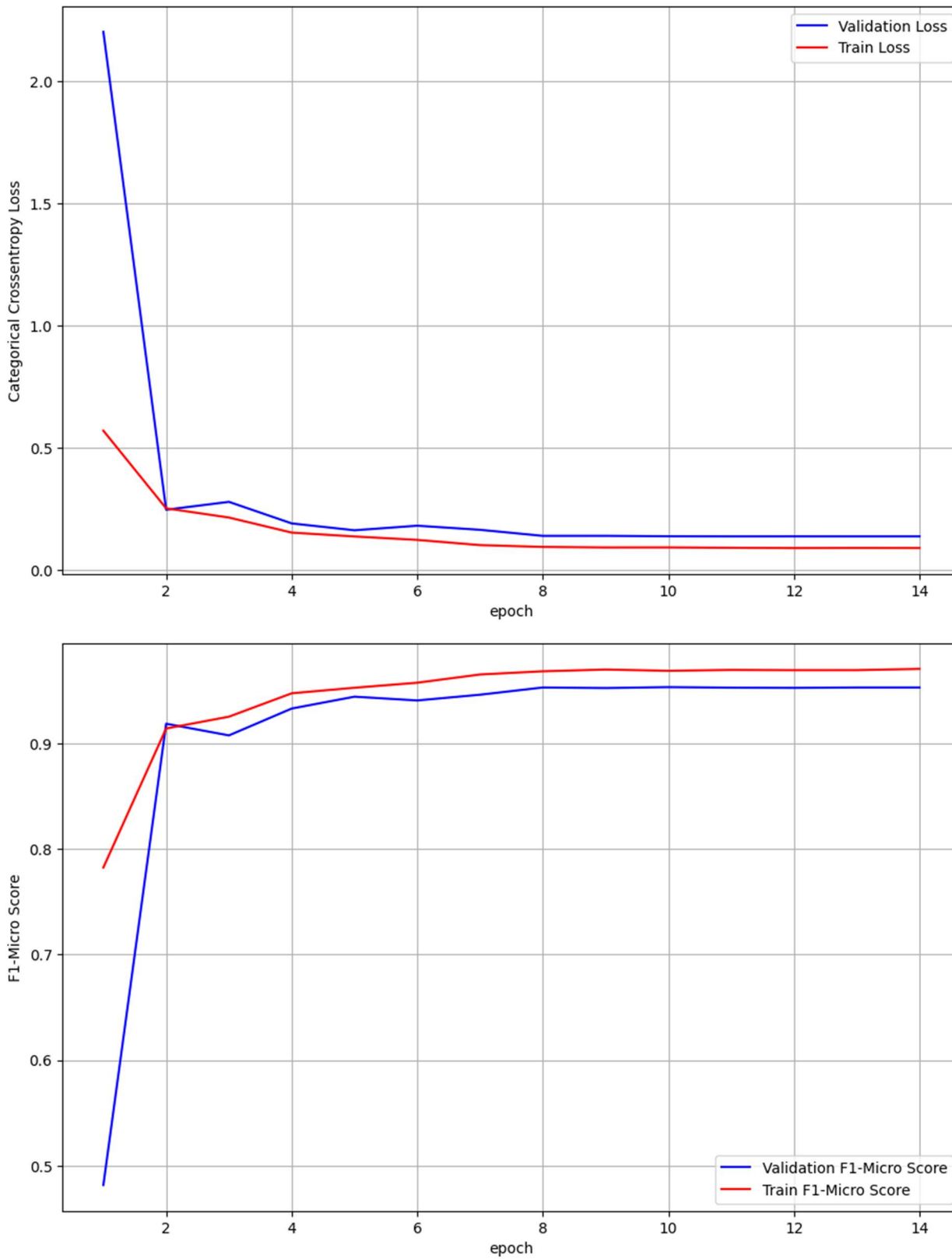


Figure 37: 7-Layer CNN

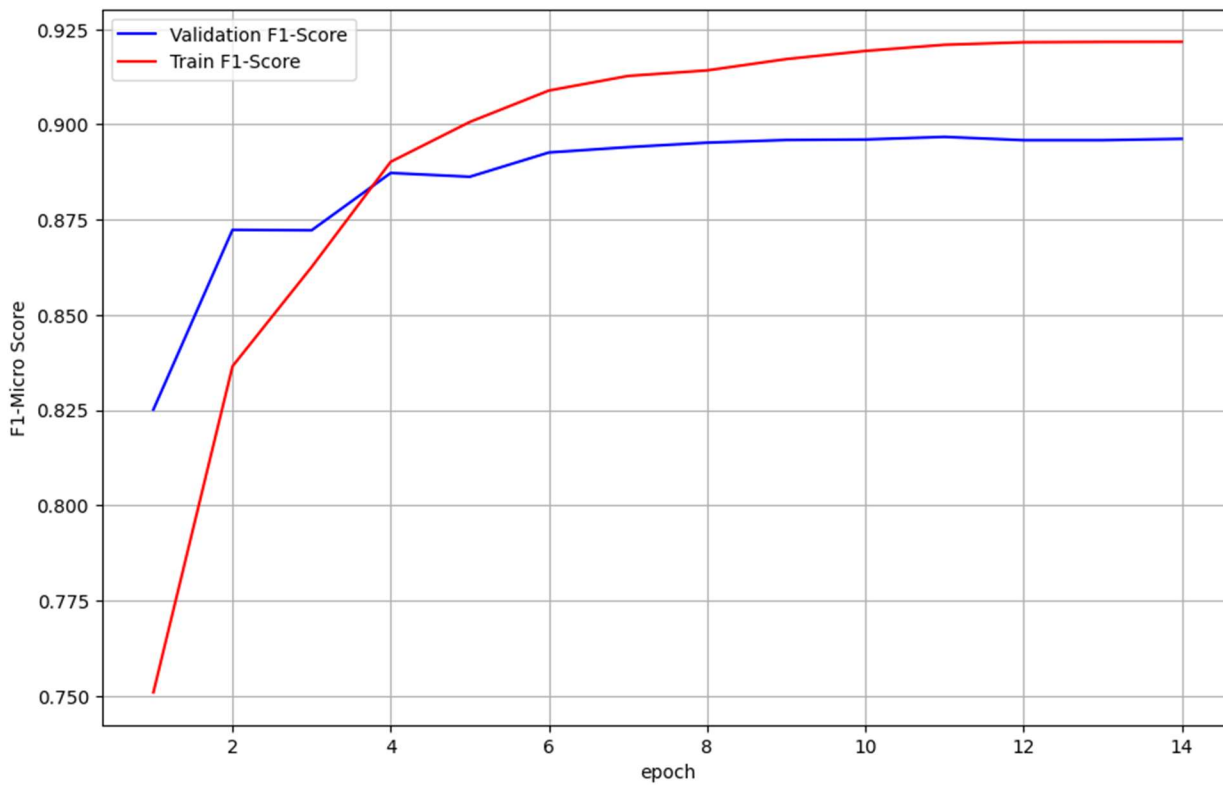
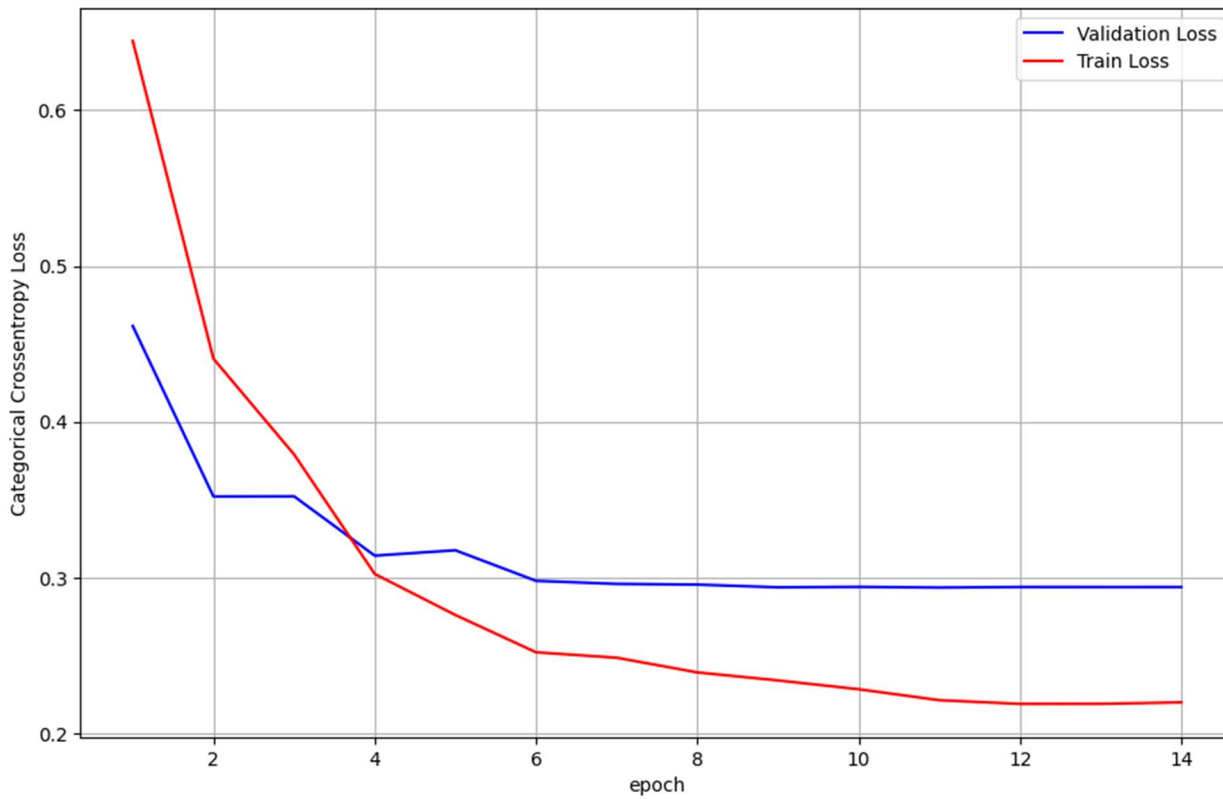


Figure 38: VGG Model

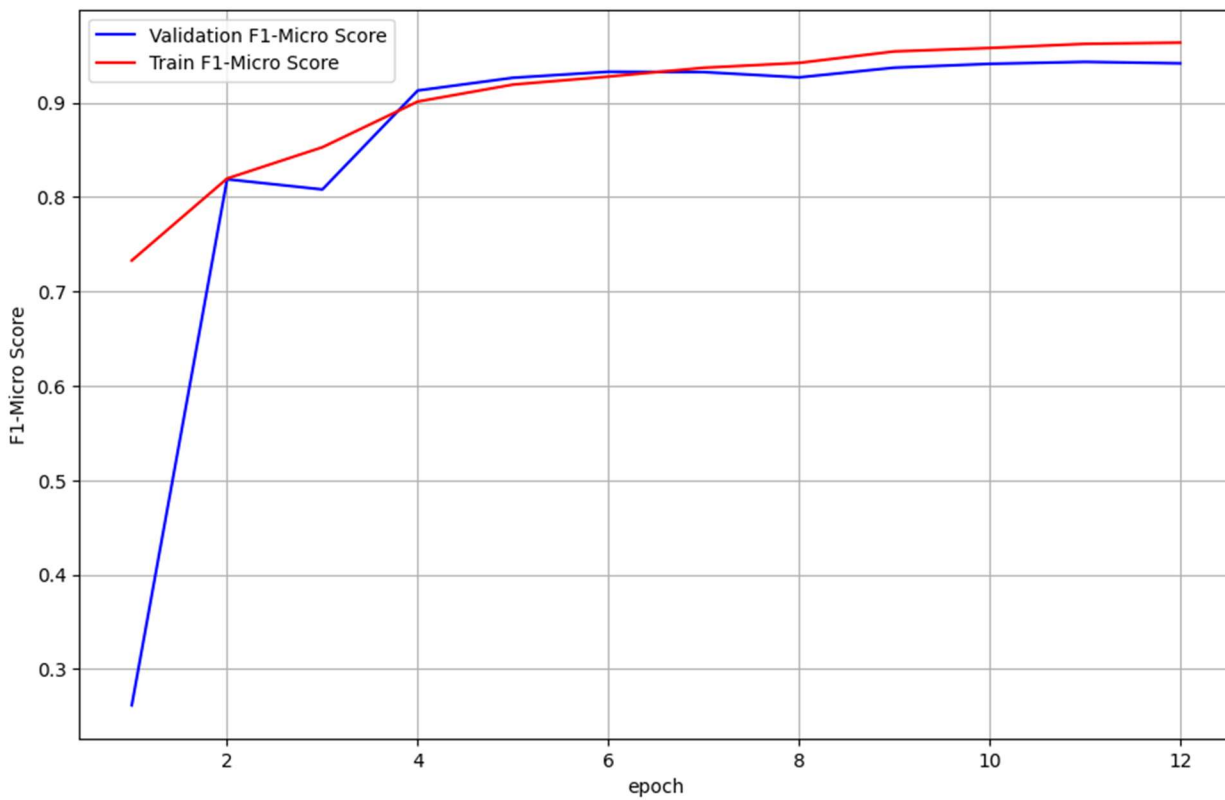
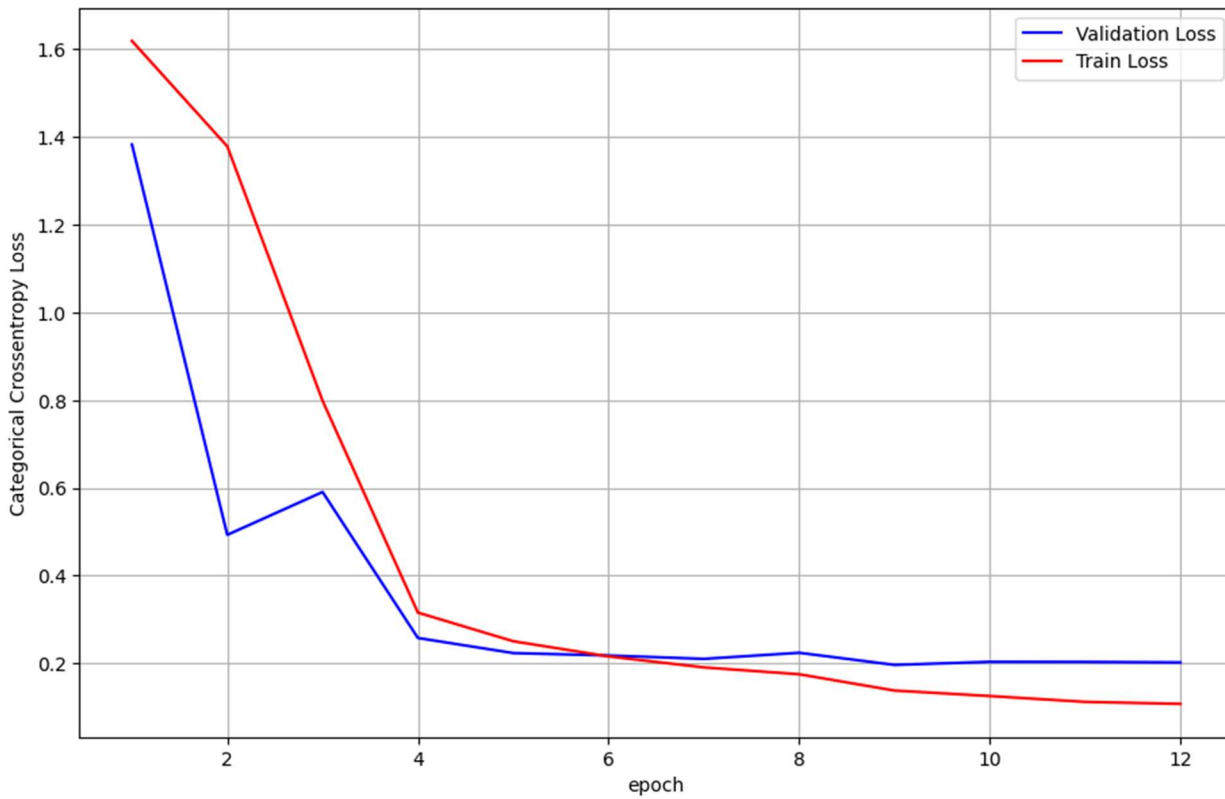


Figure 39: ResNet50 Transfer Learning

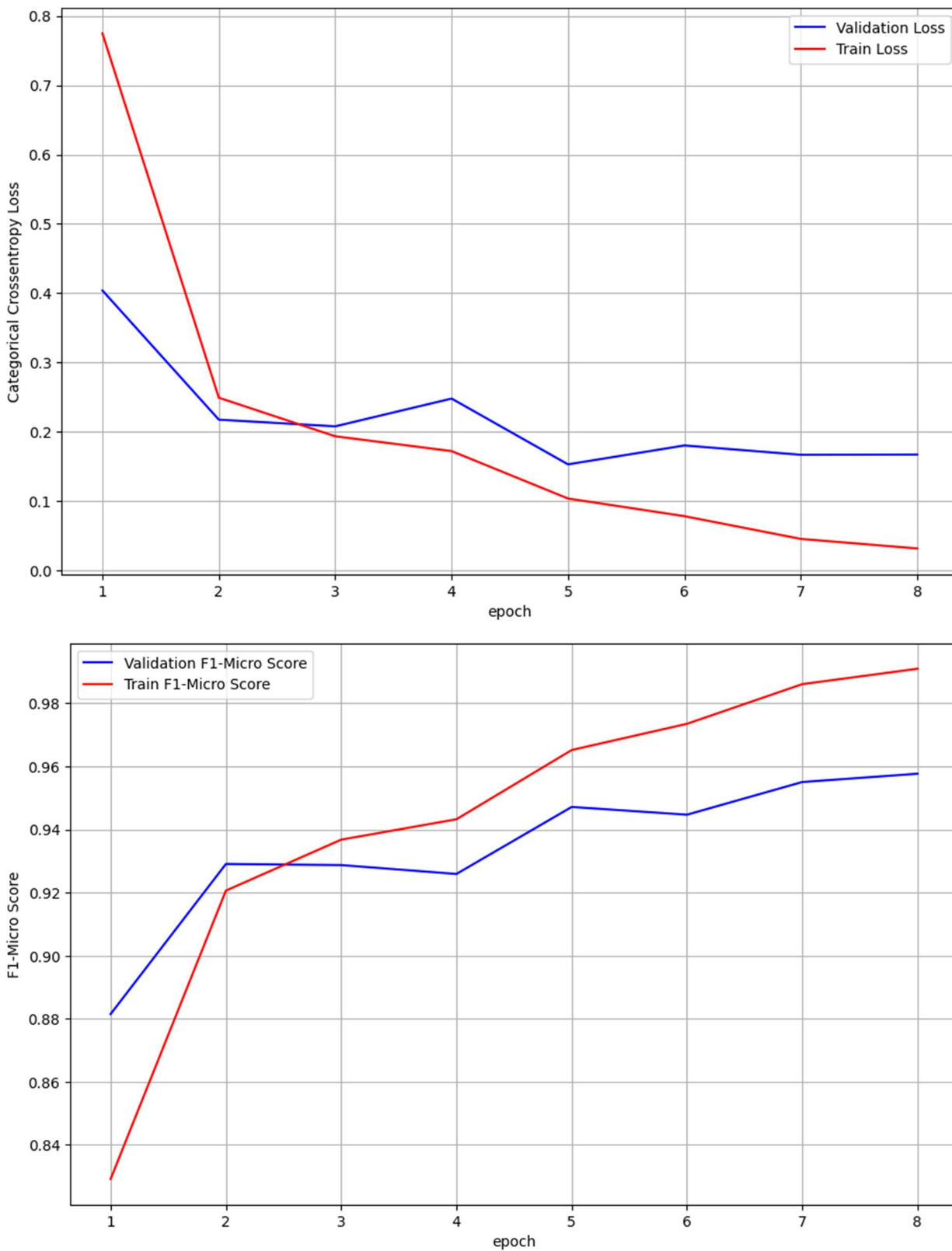


Figure 40: DenseNet121

4.2.1. Performance Metrics

Table 6: Comparative Performance of All Models on Test Set

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
3-layer CNN	0.89	0.88	0.87	0.87	0.92
7-layer CNN	0.95	0.94	0.93	0.93	0.96
VGG16	0.91	0.90	0.89	0.89	0.93
ResNet50	0.94	0.93	0.92	0.92	0.95
DenseNet121	0.96	0.95	0.94	0.94	0.97

Key Findings:

DenseNet121 had the best overall results. It achieved an accuracy of 96% AUC-ROC. The model demonstrated its ability to reuse features more efficiently even with limited parameters. Despite the overall performance of DenseNet121, the 7-layer CNN displayed better performance than VGG16, where it showed that a carefully designed task-specific model can outperform general transfer-learning approaches. Further, ResNet50 displayed consistent results across all evaluation metrics, signifying its suitability for stable and dependable clinical use.

4.2.2. Per-Class Performance Analysis

Table 7: Class-Specific F1-Scores

Model	CNV	DME	Drusen	Normal
3-layer CNN	0.88	0.86	0.85	0.90
7-layer CNN	0.94	0.92	0.91	0.95
VGG16	0.90	0.88	0.87	0.92
ResNet50	0.93	0.91	0.90	0.94
DenseNet121	0.95	0.93	0.92	0.96

Observations:

- CNV was the most accurately classified condition across all models. This is likely because of distinct hyperreflective features in OCT scans.
- Drusen showed the lowest performance consistently, indicating challenges in detecting subtle sub-RPE deposits.
- DenseNet121 demonstrated the most balanced performance across all classes, minimizing bias toward majority categories.

4.2.3. Learning Dynamics and Overfitting Analysis

All models had a degree of overfitting as the training performance exceeding validation performance. The 3-layer CNN showed the largest generalization gap. However, DenseNet121 maintained the smallest difference between training and validation metrics. Early stopping at approximately 15-25 epochs effectively prevented excessive overfitting in all models.

4.3 Statistical Analysis

4.3.1. McNemar's Test Results

Pairwise comparison using McNemar's test revealed statistically significant differences ($p < 0.05$) between:

- DenseNet121 vs. 3-layer CNN
- DenseNet121 vs. VGG16
- 7-layer CNN vs. 3-layer CNN

No significant difference was found between:

- DenseNet121 vs. 7-layer CNN
- DenseNet121 vs. ResNet50

4.3.2. Confidence Intervals

Table 8: 95% Confidence Intervals for Model Accuracy

Model	Lower Bound	Point Estimate	Upper Bound
3-layer CNN	0.877	0.89	0.903
7-layer CNN	0.941	0.95	0.959
VGG16	0.898	0.91	0.922
ResNet50	0.931	0.94	0.949
DenseNet121	0.952	0.96	0.968

4.4 Computational Efficiency

Table 9: Computational Requirements

Model	Parameters (M)	Training Time (min)	Inference Time (ms)	Memory Usage (GB)
3-layer CNN	0.05	45	12	1.2
7-layer CNN	0.50	78	18	1.8
VGG16	138	120	35	3.5
ResNet50	25	95	22	2.5
DenseNet121	8	85	20	2.2

Analysis:

- **3-layer CNN was the most efficient** but sacrificed accuracy for speed and memory savings.
- **DenseNet121 offered the best accuracy-efficiency tradeoff**, achieving highest performance with moderate computational requirements.
- **VGG16 was computationally intensive** without commensurate performance gains.

4.5 Model Explainability with Grad-CAM

Grad-CAM visualizations provided clear insight into how the models reached their decisions. For images classified as CNV, the highlighted regions consistently corresponded to hyperreflective material beneath the retina and areas of pigment epithelial detachment. In DME cases, the models focused mainly on intraretinal cystic spaces and regions of increased retinal thickness, which are well-known clinical indicators of the condition. Drusen images showed attention concentrated beneath the retinal pigment epithelium, where deposits of varying reflectivity are typically found. On the contrary, images labeled as normal had a more uniform distribution of attention in the different retinal layers. Hence, the visual patterns closely matched established clinical knowledge, supporting confidence in the model outputs.

4.6 Web Application Testing Results

4.6.1. Functional Testing

Table 10: Functional Test Results

Test Case	Description	Expected Result	Actual Result	Status
TC-1	Upload OCT image	Return JSON prediction	JSON returned	Pass
TC-2	Upload invalid file	Return 404 error	404 returned	Pass

TC-3	Login with valid credentials	Successful login	Login successful	Pass
TC-4	Login with invalid credentials	Failed login	Login failed	Pass

From the table, all core functionalities passed testing with 100% success rate as shown in the table above.

4.6.2. Usability Testing

The usability testing process entailed nine healthcare professionals and they participated as follows: \

- **Completion Rate:** 100% of assigned tasks completed
- **Average Task Time:** 2.3 minutes per diagnostic case
- **System Usability Scale (SUS) Score:** 82.5 (Excellent)
- **Net Promoter Score (NPS):** +55 (Strong)

Key Feedback:

- Positive: Intuitive interface, fast results, helpful visual explanations
- Suggestions: Add batch processing, integrate with EMR systems, improve mobile responsiveness

4.6.3. Acceptance Testing Using Nielsen's Heuristics

Table 11:Heuristic Evaluation Results

Heuristic	Score (1-5)	Comments
Visibility of System Status	4	Clear indicators but could show processing progress
Match with Real World	5	Clinical terminology and workflow alignment
User Control & Freedom	4	Good navigation with undo/redo options
Consistency & Standards	5	Consistent UI patterns throughout
Error Prevention	4	Good validation with improvement suggestions
Recognition Over Recall	5	Intuitive iconography and layout
Flexibility & Efficiency	3	Could add shortcuts for expert users
Aesthetic & Minimalist Design	4	Clean design with minor optimization needed
Help Users Recognize Errors	4	Clear error messages with guidance
Help & Documentation	3	Basic help; could add context-sensitive help

4.6.4. System Performance Benchmarks

Table 12:System Performance Metrics

Metric	Value	Target	Status
Average Response Time	1.8 seconds	<2 seconds	Met
Concurrent Users Supported	25	>20	Met
System Uptime	99.7%	>99%	Met
Database Query Time	120ms	<200ms	Met
Image Processing Time	0.8 seconds	<1 second	Met

4.6.5. Comparative Analysis with Previous Studies

Table 13:Comparison with State-of-the-Art

Study	Model	Accuracy	Dataset	Novelty
Kermany et al. (2018)	Transfer Learning	96.6%	Kermany	First large-scale application
Li et al. (2020)	ResNet50 Ensemble	97.3%	Kermany	Ensemble approach
Current Study	DenseNet121	96.0%	Kermany	Explainability + Clinical usability

Current Study	7-layer CNN	95.0%	Kermany	Custom architecture performance
---------------	-------------	-------	---------	---------------------------------

The accuracy of the model is no doubt slightly lower than some of the ensembled methods. Despite this challenge, the study makes contributions. To begin with, this study facilitates comprehensive comparison of different architectures under ideal conditions. This assists in identifying the most appropriate model in the clinical setting. Secondly, the model ensures explainability which facilitates a high level of transparency. Thirdly, the model offers a complete clinical workflow which ahs usability validation. Lastly, there is a high level of transparency.

4.7 Limitations and Error Analysis

4.7.1. Error Patterns

To enhance model transparency, error analysis was done, which revealed misclassified cases. There were instances where Drusen misclassified as Normal due to subtle deposits which were below the detection threshold. Also, early DME was misclassified as Normal, showing minimal edema without distinct cystic spaces. Lastly, there was Quiescent CNV misclassified as Drusen. This was due to instances of small, but inactive neovascular membranes.

4.7.2. Failure Cases

About 4% of the test cases displayed inconsistent predictions across all the models. These exhibited poor image quality, co-existing pathologies and early or subtle disease manifestations.

4.8 Chapter Conclusion

This chapter compared the performances of different models using validation and test results. The 7-layer CNN achieved the lowest validation loss and strong test-set F1 scores. While this showed a good ability to learn, there were noticeable traces of overfitting. The pre-trained models like DenseNet121 and ResNet50 demonstrated consistent stable results. They had lower losses and balanced F1 in both validation and test data. This shows a higher level of generalization when the models are subjected to unseen images. Despite the higher performance of VGG16 model, the lower validation and test scores presents limitations when using the model. Overall, DenseNet121 was the most reliable model. It achieved a higher accuracy and consistency which was critical in the clinical environment where misdiagnosis can be costly. Apart from the model evaluation, the chapter tested the system to ensure it worked as required. Through the functional testing, the API was checked to establish how it handles errors. Usability testing established that users could access the system and navigate to the different components easily. The acceptance testing confirmed that the system met the practical needs of the users. These results show that the system is not only technically feasible, but also suitable for the real-world use in healthcare facilities.

CHAPTER 5: CONCLUSION AND FUTURE WORKS

5.1. Research Summary and Conclusions

This study delves into the growing demand for a reliable diagnostic system which can support ophthalmologists in implementing an automated system for detecting retinal disease by classifying OCT images. In particular, the study responds to the need to move OCT classification from the manual and specialist-dependent approach, to a more automated one. The results show that the system is capable of supporting the identification of retinal diseases. It offers aid to clinicians to screen eye and facilitate early detection while minimizing reliance on time-consuming manual image assessment.

5.2. Key Achievements and Findings

A comparison of the five models evaluated in this study show several key observations:

1. DenseNet121 displayed a superior performance among the models tested. With an accuracy of 96%, DenseNet121 had a strong F1-score and AUC-ROC values.
2. Custom-designed networks proved to be effective. While transfer learning models had a better performance, the custom 7-layer CNN had a high accuracy of 95%. This shows that carefully designed task-specific architectures can perform well even without large-scale pre-training. They only need more optimization and data preparation
3. There was a variation in how each model identified the different diseases. All the models were accurate in classifying CNV, then DME and lastly Drusen. This is because CNV has clearer structural changes in OCT scans which allow AI models to identify them. However, early-stage DME and Drusen tend to be more subtle and difficult to distinguish, making it more challenging to detect.
4. The use of early stopping, image augmentation, and class-weight loss functions limited overfitting in the different models. In particular, DenseNet121 had the smallest difference between training and validation performance, indicating good generalization to unseen data.

5.3. Answering the Research Questions

The following research questions were answered.

RQ1: How effectively can deep learning models classify OCT images into the four target categories?

The models had a good performance in classifying normal eyes from diseases ones. The high level of performance is suitable in screening and early detection settings.

RQ2: Which CNN architecture yields the highest diagnostic performance?

DenseNet121 gave the best results across accuracy, F1-score, and overall stability. It needed less parameters than the other models. This makes DenseNet121 a more practical model in clinical settings with less computing resources.

RQ3: To what extent does data augmentation improve model generalization?

Data augmentation enabled the reduction of overfitting by about 30%. Techniques like rotation, flipping, and intensity changes assisted models to handle differences in images.

RQ4: Can a trained model be successfully integrated into a user-friendly, web-based diagnostic tool? OptiScan system proved to be successful when implemented as a web-based application. The model has a response time of about 1.8 seconds per predictability. The usability testing indicated that clinicians had an easy time using the system. In particular, visual explanations embedded in the system improved confidence in the model's outputs.

5.4. Clinical and Technical Contributions

The paper demonstrates that computer-aided diagnosis is applicable in ophthalmology. By comparing the different CNN architectures under the same experimental conditions, the paper emphasized practical application with usability, security and workflow integration. The system has Grad-CAM visualizations to allow clinicians to interpreted results. Using the CRISP-DM framework and rigorous statistical validation ensured that the development process was structured.

5.5. Practical Implications for Clinical Practice

The OpticScan system offers benefits for the clinical ophthalmology in several ways.

5.5.1. Workflow Enhancement

- **Reduced Diagnostic Latency:** The 1.8-second inference time enables rapid screening, potentially reducing patient wait times and increasing clinic throughput.
- **Consistent Second Opinion:** The system provides standardized, objective assessments to complement subjective clinical judgment, potentially reducing inter-observer variability.
- **Triage Capability:** In high-volume settings, the system could prioritize cases requiring urgent specialist attention, optimizing resource allocation.

5.5.2. Accessibility Improvements

- **Resource-Limited Settings:** The web-based deployment model enables access in regions with limited specialist availability, potentially expanding screening coverage.
- **Primary Care Integration:** The intuitive interface allows non-specialist healthcare providers to conduct preliminary screenings, facilitating earlier referrals.

5.5.3. Educational Value

- **Training Tool:** The visual explanations generated by Grad-CAM can serve as educational aids for ophthalmology trainees, highlighting diagnostically relevant retinal features.
- **Quality Assurance:** The system provides a consistent benchmark for monitoring diagnostic performance and identifying areas for skill development.

5.6. Limitations and Constraints

5.6.1. Dataset Limitations

- **Single-Source Data:** The exclusive use of the Kermany dataset, while large and well-annotated, may not fully capture the demographic and pathological diversity encountered in global clinical practice.

- **Retrospective Design:** All evaluations used retrospectively collected data; prospective clinical validation is required to assess real-world performance.

5.6.2. Technical Constraints

- **2D B-Scan Focus:** The analysis was limited to 2D cross-sectional images; incorporating 3D volumetric OCT data could provide additional diagnostic information.
- **Computational Requirements:** While DenseNet121 offered good efficiency, GPU acceleration remains necessary for real-time inference, potentially limiting deployment in some settings.
- **Class Imbalance Persistence:** Despite mitigation strategies, the minority classes (particularly Drusen) showed slightly lower performance, suggesting opportunities for further refinement.

5.6.3. Clinical Translation Challenges

- **Regulatory Pathway:** The system requires formal regulatory approval (e.g., FDA, CE marking) before clinical use, a process involving extensive validation studies.
- **Integration Complexity:** Full integration with existing Electronic Medical Record systems presents technical and interoperability challenges not addressed in this prototype.
- **Liability Considerations:** The appropriate medico-legal framework for AI-assisted diagnosis requires careful consideration and professional guideline development.

5.7. Recommendations for Future Research

5.7.1. Technical Enhancements

In the future, the system can be enhanced by having different types of imaging like fundus photography and FA. These can be improved with relevant information like patient history and genetic markers to allow for a more comprehensive assessment. The current study relied on 2D B-scans, but these are limiting. In the future, the model could also target 3D OCT volumes to allow for better detection of spatial disease pattern. Also, approaches to estimate uncertainty in predictions would allow clinicians to identify cases that require more evaluation.

5.7.2. Clinical Expansion

The research covered only a limited number of retinal conditions. In the future, the model could cover more conditions like macular holes, retinal vein occlusions, and epiretinal membranes to improve its clinical application. The model in the future should track disease progression to support personalized treatment plans. Further, the system can be adapted for large-scale screening programs. This will require access to large and diverse retinal datasets, while ensuring patient data remains protected from unauthorized third parties.

5.7.3. Implementation Pathways

The interpretation of the system is several pathways. It starts with a multi-center validation across different clinics and patient populations to establish its application in diverse environments. Next, the tool needs to be integrated into everyday clinical workflows. This needs clear guidelines on how clinicians should treat the model outputs. Lastly, regulators need to understand the potential of the system in enhancing healthcare. They will need to understand that the system meets safety requirements and can be employed without harming patients.

References

- Akça, S., Garip, Z., Ekinci, E., & Atban, F. (2024). Automated classification of choroidal neovascularization, diabetic macular edema, and drusen from retinal OCT images using vision transformers: a comparative study. *Lasers in Medical Science*, 39(1), 140.
- Alawi, Z. B. (2025). A Comparative Survey of PyTorch vs TensorFlow for Deep Learning: Usability, Performance, and Deployment Trade-offs. *arXiv preprint arXiv:2508.04035*.
- Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., ... & Albekairy, A. M. (2023). Revolutionising healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1), 689.
- Avanzo, M., Stancanella, J., Pirrone, G., Drigo, A., & Retico, A. (2024). The evolution of artificial intelligence in medical imaging: from computer science to machine and deep learning. *Cancers*, 16(21), 3702.
- Baba, S., Kumari, P., & Saxena, P. (2024). Retinal Disease Classification Using Custom CNN Model From OCT Images. *Procedia Computer Science*, 235, 3142-3152.
- Braik, M., Al-Betar, M. A., Mahdi, M. A., Al-Shalabi, M., Ahamad, S., & Saad, S. A. (2024). Enhancement of satellite images based on CLAHE and augmented elk herd optimizer. *Artificial Intelligence Review*, 58(2), 38.
- Chew, E. Y. (2020). The landmark Beaver Dam eye study ushered in modern epidemiologic research in age-related macular degeneration. *Ophthalmology*, 127(4), S120-S121.
- Chiu, T. M., & Chu, S. Y. (2022). Hypersensitivity reactions to iodinated contrast media. *Biomedicines*, 10(5), 1036.
- Cui, Y., Zhu, Y., Wang, J. C., Lu, Y., Zeng, R., Katz, R., ... & Miller, J. B. (2019). Imaging artifacts and segmentation errors with wide-field swept-source optical coherence tomography angiography in diabetic retinopathy. *Translational vision science & technology*, 8(6), 18-18.
- Cui, Y., Zhu, Y., Wang, J. C., Lu, Y., Zeng, R., Katz, R., ... & Miller, J. B. (2019). Imaging artifacts and segmentation errors with wide-field swept-source optical coherence tomography angiography in diabetic retinopathy. *Translational vision science & technology*, 8(6), 18-18.
- Deepak, G. D., & Bhat, S. K. (2024). Deep learning-based CNN for multiclassification of ocular diseases using transfer learning. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 12(1), 2335959.
- Dhamodaran, S., Phukan, P. P., Singh, M., & Nandakumar, S. (2024). Review on Computer Vision Using OpenCV. *International Journal of Research Publication and Reviews*. Vol 5, no 5, pp 8608-8621
- Du, K., Nair, A. R., Shah, S., Gadari, A., Vupparaboina, S. C., Bollepalli, S. C., ... & Vupparaboina, K. K. (2024). Detection of Disease Features on Retinal OCT Scans Using RETFound. *Bioengineering*, 11(12), 1186.

- Du, K., Shah, S., Bollepalli, S. C., Ibrahim, M. N., Gadari, A., Sutharahan, S., ... & Vupparaboina, K. K. (2024). Inter-rater reliability in labeling quality and pathological features of retinal OCT scans: A customized annotation software approach. *PLoS One*, 19(12), e0314707.
- Falavarjani, K. G., Anvari, P., Alemzadeh, S. A., Moghaddam, M. M. J., Habibi, A., & Ashrafkhorasani, M. (2022). Segmentation error correction of the optical coherence tomography angiography images in peer-reviewed studies. *Journal of current ophthalmology*, 34(3), 273-276.
- Gianfredi, V., Nucci, D., Pennisi, F., Maggi, S., Veronese, N., & Soysal, P. (2025). Aging, longevity, and healthy aging: the public health approach. *Aging Clinical and Experimental Research*, 37(1), 1-12.
- Gholami, P., Roy, P., Parthasarathy, M. K., & Lakshminarayanan, V. (2020). OCTID: Optical coherence tomography image database. *Computers & Electrical Engineering*, 81, 106532.
- Goździewska, E., Wichrowska, M., & Kocięcki, J. (2023). Early optical coherence tomography biomarkers for selected retinal diseases—A review. *Diagnostics*, 13(14), 2444.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, 12(1), 5979.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- Jadeja, V., Rao, A. L. N., Srivastava, A., Singh, S., Chaturvedi, P., & Bhardwaj, G. (2023, September). Convolutional neural networks: a comprehensive review of architectures and application. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)* (Vol. 6, pp. 460-467). IEEE.
- Karen, S. (2014). Very deep convolutional networks for large-scale image recognition. *arxiv preprint arxiv:1409.1556*.
- Kaur, A., Kukreja, V., Kumar, M., Choudhary, A., & Sharma, R. (2024, March). RetDense: A fine-tuned DenseNet121 model for retinal eye disease detection. In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* (Vol. 2, pp. 1-5). IEEE.
- Kermany D, Goldbaum M, Cai W et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*. 2018; 172(5):1122-1131. doi:10.1016/j.cell.2018.02.010.
- Khan, S. M., Liu, X., Nath, S., Korot, E., Faes, L., Wagner, S. K., ... & Denniston, A. K. (2021). A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digital Health*, 3(1), e51-e66.
- Kocak, B., Klontzas, M. E., Stanzione, A., Meddeb, A., Demircioğlu, A., Bluethgen, C., ... & Cuocolo, R. (2025). Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations. *European Journal of Radiology Artificial Intelligence*, 100030.

- Kráľová, J. Š., Kolář, P., Kapounová, Z., Veselý, P., & Brázdová, Z. D. (2024). Lifestyle factors associated with age-related macular degeneration: Case-control study. *European Journal of Ophthalmology*, 34(5), 1548-1554.
- Krichen, M. (2023). Convolutional neural networks: A survey. *Computers*, 12(8), 151.
- Kulyabin, M., Zhdanov, A., Nikiforova, A., Stepichev, A., Kuznetsova, A., Ronkin, M., ... & Maier, A. (2024). Octdl: Optical coherence tomography dataset for image-based deep learning methods. *Scientific data*, 11(1), 365.
- Leitgeb, R. A., Szkulmowski, M., Blatter, C., & Wojtkowski, M. (2015). Doppler Fourier domain optical coherence tomography for label-free tissue angiography. In *Optical Coherence Tomography* (pp. 1321-1352). Springer, Cham.
- Li, F., Chen, H., Liu, Z., Zhang, X. D., Jiang, M. S., Wu, Z. Z., & Zhou, K. Q. (2019). Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomedical optics express*, 10(12), 6204-6226.
- Linse, C., Barth, E., & Martinetz, T. (2023, June). Convolutional neural networks do work with pre-defined filters. In *2023 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Mandal, M. K. (2003). The human visual system and perception. In *Multimedia Signals and Systems* (pp. 33-56). Boston, MA: Springer US.
- Mauschitz, M. M., & Finger, R. P. (2022). Age-related macular degeneration and cardiovascular diseases: revisiting the common soil theory. *The Asia-Pacific Journal of Ophthalmology*, 11(2), 94-99.
- Meyer, A. N., Giardina, T. D., Khawaja, L., & Singh, H. (2021). Patient and clinician experiences of uncertainty in the diagnostic process: current understanding and future directions. *Patient Education and Counseling*, 104(11), 2606-2615.
- Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. In *Computer science on-line conference* (pp. 15-25). Cham: Springer International Publishing.
- Nasef, D., Nasef, D., Sawiris, V., Weinstein, B., Garcia, J., & Toma, M. (2025). Integrating artificial intelligence in clinical practice, hospital management, and health policy: literature review. *Journal of Hospital Management and Health Policy*, 9.
- Nikolopoulou, E., Lorusso, M., Micelli Ferrari, L., Cicinelli, M. V., Bandello, F., Querques, G., & Micelli Ferrari, T. (2018). Optical coherence tomography angiography versus dye angiography in age-related macular degeneration: sensitivity and specificity analysis. *BioMed research international*, 2018(1), 6724818.
- Puneet, Kumar, R., & Gupta, M. (2022). Optical coherence tomography image based eye disease detection using deep convolutional neural network. *Health Information Science and Systems*, 10(1), 13.
- Quintana-Quintana, O. J., Aceves-Fernández, M. A., Pedraza-Ortega, J. C., Alfonso-Francia, G., & Tovar-Arriaga, S. (2025). Deep Learning Techniques for Retinal Layer Segmentation to Aid Ocular Disease Diagnosis: A Review. *Computers*, 14(8), 298.
- Raj, R., & Kos, A. (2025). An extensive study of convolutional neural networks: applications in computer vision for improved robotics perceptions. *Sensors*, 25(4), 1033.

- Rasheed, A. A. (2023). Improving prediction efficiency by revolutionary machine learning models. *Materials today: proceedings*, 81, 577-583.
- Rudnicka, A. R., Kapetanakis, V. V., Jarrar, Z., Wathern, A. K., Wormald, R., Fletcher, A. E., ... & Owen, C. G. (2015). Incidence of late-stage age-related macular degeneration in American whites: systematic review and meta-analysis. *American journal of ophthalmology*, 160(1), 85-93.
- Salimiaghdam, N., Riazi-Esfahani, M., Fukuhara, P. S., Schneider, K., & Kenney, M. C. (2019). Age-related macular degeneration (AMD): A review on its epidemiology and risk factors. *The Open Ophthalmology Journal*, 13(1).
- Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, 27(4S), 4023-4031.
- Sheng, B., Chen, X., Li, T., Ma, T., Yang, Y., Bi, L., & Zhang, X. (2022). An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. *Frontiers in Public Health*, 10, 971943.
- Sheng, M., Cai, H., Cheng, M., Li, J., Zhang, J., & Liu, L. (2021). Identification of Novel Choroidal Neovascularization-Related Genes Using Laplacian Heat Diffusion Algorithm. *BioMed Research International*, 2021(1), 2295412.
- Shourie, P. (2024, November). Optimizing Traffic Sign Detection with MobileNetV2: A Lightweight Deep Learning Approach. In *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 272-276). IEEE.
- Shreffler, J., & Huecker, M. R. (2020). Diagnostic testing accuracy: Sensitivity, specificity, predictive values and likelihood ratios.
- Singh, R. K., & Gorantla, R. (2020). DMENet: diabetic macular edema diagnosis using hierarchical ensemble of CNNs. *Plos one*, 15(2), e0220677.
- Song, S., Xu, H., Ma, J., Li, S., Peng, L., Wan, Q., ... & Yu, J. (2025). How to Alleviate Catastrophic Forgetting in LLMs Finetuning? Hierarchical Layer-Wise and Element-Wise Regularization. *arXiv preprint arXiv:2501.13669*.
- Subban, V., & Raffel, O. C. (2020). Optical coherence tomography: Fundamentals and clinical utility. *Cardiovascular Diagnosis and Therapy*, 10(5), 1389.
- Sundaram, J., Gowri, K., Devaraju, S., Gokuldev, S., Jayaprakash, S., Anandaram, H., ... & Thenmozhi, M. (2023). An Exploration of Python Libraries in Machine Learning Models for Data Science. In *Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science* (pp. 1-31). IGI Global.
- Uchida, A., Manjunath, D., Singh, R. P., Rachitskaya, A. V., Kaiser, P. K., Srivastava, S. K., ... & Ehlers, J. P. (2018). Optical coherence tomography angiography in eyes with indeterminate choroidal neovascularization: results from the AVATAR Study. *Ophthalmology Retina*, 2(11), 1107-1117.
- Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... & Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, 2, e453.
- VanDenLangenberg, A. M., & Carson, M. P. (2020). Drusen bodies.

Vujosevic, S., Parra, M. M., Hartnett, M. E., O'Toole, L., Nuzzi, A., Limoli, C., ... & Nucci, P. (2023). Optical coherence tomography as retinal imaging biomarker of neuroinflammation/neurodegeneration in systemic disorders in adults and children. *Eye*, 37(2), 203-219.

Wang, Y., Li, Y., Song, Y., & Rong, X. (2020). The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, 10(5), 1897.

Wojciuk, M., Swiderska-Chadaj, Z., Siwek, K., & Gertych, A. (2024). Improving classification accuracy of fine-tuned CNN models: Impact of hyperparameter optimization. *Heliyon*, 10(5).

Xu, H. H., & Ge, D. Y. (2020). A novel image edge smoothing method based on convolutional neural network. *International Journal of Advanced Robotic Systems*, 17(3), 1729881420921676.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4), 611-629.

Zeppieri, M., Marsili, S., Enaholo, E. S., Shuaibu, A. O., Uwagboe, N., Salati, C., ... & Musa, M. (2023). Optical coherence tomography (OCT): a brief look at the uses and technological evolution of ophthalmology. *Medicina*, 59(12), 2114.

Zhang, S., Ren, J., Chai, R., Yuan, S., & Hao, Y. (2024). Global burden of low vision and blindness due to age-related macular degeneration from 1990 to 2021 and projections for 2050. *BMC Public Health*, 24(1), 3510.