An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation

by

Sharat Acharja- 221902061
Azhar Hossen Novel- 221902310
Amikur Rahman- 221902303

Capstone Thesis (CSE 400) submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Computer Science and Engineering

Supervisor
Dr. Faiz Al Faisal

Co-supervisor
Mr. Kazi Hasnayeen Emad

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GREEN UNIVERSITY OF BANGLADESH
FALL 2025

DECLARATION

Title	An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation
Author	*Sharat Acharja*, *Azhar Hossen Novel* & *Amikur Rahman*
Student ID	*221902061*, *221902310* & *221902303*
Supervisor	Dr. Faiz Al Faisal
Co-supervisor	Mr. Kazi Hasnayeen Emad

We declare that this thesis entitled *An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation* is the result of our own work except where otherwise acknowledged by reference. This thesis has not been accepted for any degree and is not concurrently submitted in candidature for any other degree.

Sharat Acharja
221902061
Department of Computer Science and Engineering
Green University of Bangladesh

Azhar Hossen Novel
221902310
Department of Computer Science and Engineering
Green University of Bangladesh

Amikur Rahman
221902303
Department of Computer Science and Engineering
Green University of Bangladesh

Date: 17 January, 2026

CERTIFICATE OF ENDORSEMENT

This is to certify that the thesis entitled *An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation* by **Sharat Acharja, Azhar Hossen Novel & Amikur Rahman** has been successfully defended and examined.

Based on the evaluation and discussion held on **17 January, 2026**, we, the undersigned, hereby endorse the thesis for acceptance and approval.

External Member's Name
External Member, Board: 10

External Member's Position
External Member's Affiliation

Board Member's Name
Board Member, Board: 10

Board Member's Position
Department of Computer Science and Engineering
Green University of Bangladesh

Board Chair's Name
Chair, Board: 10

Board Chair's Position
Green University of Bangladesh

Date: 17 January, 2026

Department of Computer Science and Engineering
Green University of Bangladesh
Purbachal American City, Kanchan, Rupganj,
Narayanganj-1461, Dhaka, Bangladesh

CERTIFICATE

This is to certify that the thesis entitled **An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation**, submitted by **Sharat Acharja** (Student ID: 221902061), **Azhar Hossen Novel** (Student ID: 221902310) and **Amikur Rahman** (Student ID: 221902303), undergraduate students of the **Department of Computer Science and Engineering**, Green University of Bangladesh, has been examined. Upon recommendation by the examination committee, this work is hereby approved as fulfilling the requirements for the degree of *Bachelor of Science in Computer Science and Engineering*.

Dr. Faiz Al Faisal
Former Associate Professor
Department of Computer Science and Engineering
Green University of Bangladesh

Mr. Kazi Hasnayeen Emad
Lecturer
Department of Computer Science and Engineering
Green University of Bangladesh

Syed Ahsanul Kabir
Associate Professor and Chairman
Department of Computer Science and Engineering
Green University of Bangladesh

Place: Purbachal American City, Kanchan, Rupganj, Narayanganj-1461, Dhaka, Bangladesh
Date: 17 January, 2026

ACKNOWLEDGMENTS

First and foremost, we would like to extend our sincere appreciation to **Green University of Bangladesh**to supply the means in order to pursue our educative and research initiatives.We would like to add our commendation for the **Department of Computer Sciences and Engineering** for developing a learning environment that fosters intellectual advancement and innovation.

We would like to express our honest thanks to the **Chairman of the Department of Computer Sciences and Engineering** for continuous support and visionary leadership and our educational journey.

Our deepest thanks to all the respected faculty members of the department.His professional commitment towards education and counseling has been a major factor in the growth.

We especially thank our research supervisor and co-supervisor, **Dr. Faiz Al Faisal Sir** and **Mr. Kazi Hasnayeen Emad Sir**, for granting us the opportunity to work under their invaluable guidance. It has been the greatest privilege and pleasure to work under the supervision of those who have been such an inspiration to this research with their insight and encouragement every step of the way.

We are always grateful to **our parents** for our unconditional love, prayer, and unbreakable support. Their victims have provided the basis for our education and personal development.

We would not forget to let our appreciation be extended to those colleagues in our **research team** who have been one with us through our journey, and without whom our journey would have been a lot smoother.

First of all, we would like to thank all of our friends and supporters for their unwavering support, bravery, and company throughout the research.

Sharat Acharja, *Azhar Hossen Novel* & *Amikur Rahman*
Department of Computer Science and Engineering
Green University of Bangladesh
Date: 17 January, 2026

Dedicated to God for His
blessings and guidance.
Dedicated to my dear parents for
their love and support,
and to everyone who encouraged
and inspired me.

– *Sharat Acharja*

To my loving parents
whose love and support have
always guided me.

– *Azhar Hossen Novel*

To my respected teachers
and well-wishers
for their guidance and
encouragement.

– *Amikur Rahman*

ABSTRACT

Sight loss as a result of eye pathology and refractive error is still a major cause of health burden throughout the world, though access to treatment timely is not uniform across the populations of various regions. More recently, retinal fundus image automated analyses are made possible with deep learning. Most of the current methods lack interpretability to the extent that they do not incorporate severity of diseases to limit their applicability in clinical settings.

This work presents an explainable deep learning framework for precision eye care that jointly performs multi-disease ocular classification and severity-aware myopia analysis using color fundus images. A transfer learning strategy based on a pretrained ResNet50 architecture is employed to extract discriminative retinal features for multi-class disease recognition. In addition, myopia-related cases are further categorized into mild, strong, and pathological severity levels using diagnostic keyword-derived annotations aligned with clinically accepted refractive ranges, without directly estimating numerical refractive power.

To improve transparency and trustworthiness, explainable artificial intelligence techniques are integrated to visually highlight retinal regions that contribute most to model predictions. Furthermore, a guideline-driven, rule-based clinical decision support mechanism is incorporated to provide preliminary screening-level recommendations for myopia-related cases. Experimental evaluation on a publicly available ocular dataset demonstrates consistent classification performance and stable generalization under patient-wise testing conditions.

The proposed framework is intended as a supportive screening and decision assistance tool rather than a replacement for professional diagnosis. By combining multi-disease classification, severity-aware myopia analysis, and explainable visualization, this study contributes toward more interpretable and clinically meaningful automated eye care systems.

Keywords: analysis of fundus images, classification of ocular diseases, estimation of myopia severity, explainable artificial intelligence, deep learning, Decision Support System.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Eye diseases and refractive errors have become a population health issue of gigantic dimensions around the globe. Some of the major causes of preventable and irreversible impairment of vision used in the whole world include diabetic retinopathy, Glaucoma, Cataract, Age-related macular degeneration, Hypertensive retinopathy, Myopia and a number of others. According to the World Health Organization, over 2.2 billion individuals are experiencing some level of vision loss out of which over one billion are avoidable.

RFI has become an up-and-coming and non-invasive imaging modality, which has been used in the field of retinal structural analyses and retinal pathologies. Usually, the analysis of fundus images has been done manually by eye care specialists or ophthalmologists. Nevertheless, it is likely to be a lengthy procedure and lacks an supply of eye care professionals in different parts of the world.

The recent developments in medical image analysis through deep learning and convolutional neural networks (CNNs) have shown that they have a great potential of being used to detect ocular diseases in an automated manner. In a number of tasks related to screening retinal diseases, CNN-based models have matched the performance of expert clinicians. Research in automated multi-disease ocular classification by using fundus images has been further advanced with release of large-scale annotated datasets, including the Ocular Disease Intelligent Recognition (ODIR-5K) dataset.

1.2 Problem Statement

Although the literature shows promising results, current deep learning-based ocular disease detection systems have a number of challenges, which are still not adequately addressed, and which restrict the clinical applicability of ocular disease detection systems. First, with extreme class imbalance in real-world data, the performance of rare yet clinical phenomena, like glaucoma and pathologic myopia, tends to be low. Second, a lot of models are black-box systems, where they give predictions but they do not give meaningful explanations, which compromises the trust of clinicians and makes them regulateable. Third, the majority of current strategies pay attention to the presence of this disease but do not pay enough attention to the severity of this disease that is a valuable factor in clinical decision-making and the long-term management of patients.Existing systems lack failure analysis.Lack of efficiency/real-time feasibility discussion

Hence there is an evident requirement to have an ocular screening system based on the automated system with the ability to not only do multi-disease classification, but also with severity-conscious analysis and interpretable decision-making processes. Such a system must be open, replicable and appropriate to real-world deployment conditions.

1.3 Purpose of the Project/Thesis

The intended aim of the project is to come up with an explainable deep learning framework of precision eye care that is collectively multi-disease ocular and severity conscious myopia analysis utilizing retinal fundus images. By incorporating explainable artificial intelligence methods and a guideline-based strategy of clinical decision support, the proposed system will address the gap between high predictability and clinical interpretability. The framework is not intended to substitute professional diagnosis but, instead, it will serve as an effective screening and decision aid tool.

1.4 State of the Art

VRecent works using advanced deep learning models such as ResNet50, EfficientNet, DenseNet, and Vision Transformers have achieved excellent classification accuracy of the ODIR-5K dataset, with their area under the curve (AUC) values exceeding 0.95. Nevertheless, these models sacrifice the interpretability of the diagnostic results and the robustness of the classification task regarding the minority diseases. Moreover, the development of transparency in these models using Explainable Artificial Intelligence approaches, such as Grad-CAM, SHAP, and LIME, is still limited in the context of multi-disease ocular analysis systems.

1.5 Applications and Practical Use Cases

The proposed framework is applicable to a variety of real-world clinical and public health scenarios, including large-scale ocular screening programmes, teleophthalmology platforms, decision support for general practitioners and optometrists, automated preliminary reporting in fundus imaging systems, and population-level epidemiological studies using retinal image databases

1.6 Research Objectives

General Objective:

To design a deep learning framework that is explainable, thus implementing and evaluating the performance concerning multi-disease ocular classification and myopia analysis in terms of severity using retinal fundus images.

Specific Objectives:

- To perform multi-disease ocular classification using deep convolutional neural networks.
- To incorporate severity-aware myopia categorization using diagnostic keyword-derived annotations.
- To integrate explainable artificial intelligence techniques for visual interpretation of model predictions.

- To evaluate model performance using standard classification metrics under patient-wise validation settings.
- To design a guideline-driven clinical decision support mechanism for preliminary myopia screening.

1.7 Research Methodology Overview

The proposed research follows an end-to-end framework that combines deep learning-based feature extraction, multi-disease classification, myopia severity estimation, explainable visualization, and rule-based clinical decision support. A high-level overview of the system architecture is illustrated in Figure 3.1.

1.8 Research Contributions

The key contributions of this research are summarized as follows:

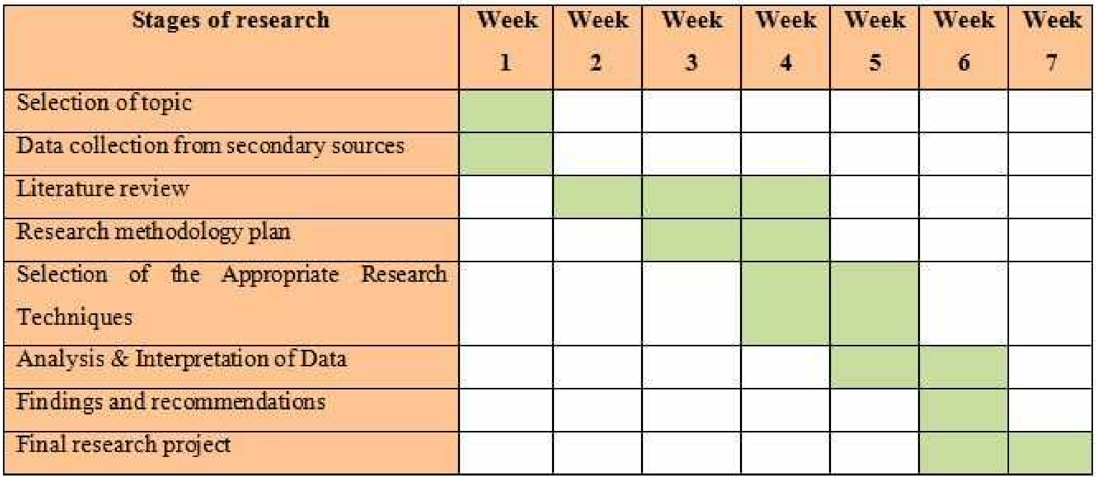
- Development of an explainable deep learning framework for multi-disease ocular classification using fundus images.
- Integration of severity-aware myopia analysis based on diagnostic keyword-derived annotations.
- Incorporation of explainable artificial intelligence techniques to enhance transparency and clinical interpretability.
- Design of a guideline-driven clinical decision support module for preliminary screening-level recommendations.

1.9 Financial Planning and Timeline

1.9.1 Budget Estimation

This research does not require significant financial expenditure. The ODIR-5K dataset is publicly available, and all software tools, including Python, TensorFlow/Keras, and explainable AI libraries, are open-source. Model training was conducted using cloud-based computational resources with minimal associated cost.

1.9.2 Gantt Chart



Stages of research	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Selection of topic							
Data collection from secondary sources							
Literature review							
Research methodology plan							
Selection of the Appropriate Research Techniques							
Analysis & Interpretation of Data							
Findings and recommendations							
Final research project							

Figure 1.1: Planned thesis schedule showing key project activities and milestones.

1.10 Organization of the Dissertation

This dissertation is organized as follows:

- Chapter 1 introduces the background, motivation, and research objectives of the study.
- Chapter 2 presents a comprehensive review of related literature on ocular disease classification and explainable artificial intelligence.
- Chapter 3 describes the proposed methodology, including system architecture, dataset preparation, and model configuration.
- Chapter 4 presents the experimental results and discussion, including quantitative performance evaluation and qualitative explainability analysis.
- Chapter 5 concludes the dissertation and outlines limitations, future research directions, and recommendations.

2. Literature Review

2.1 Introduction

A literature review is a meticulous research and analysis of previous literature and research articles related to a particular study problem. A literature review is a crucial part of research work in academics and serves to develop the theory base for a study, and it also helps in pointing out trends and unresolved gaps in research that a study can develop to become a full-fledged research paper and avoid replicating previously carried-out research. By conducting a literature review, one is able to develop a rationale for the need to conduct a particular study.

Within the context of my thesis work, literature review revolves around deep learning-based assessment on retinal fundus images for ocular disease diagnosis, primarily focusing on multi-disease, myopia-related, as well as Explainable AI (XAI) aspects. Various literature studies give a brief overview about how CNN architectures have developed over time, common benchmark datasets used like ODIR-5K, and particularly about a rising interest in interpretability and a trustworthy form of AI in various applications.

This book review uses papers that are peer-reviewed and published from the year 2016 until 2024.The sources were dominated by IEEE Xplore publications, followed by publications in PubMed, Scopus, and Google Scholar. The searching terms were *ocular disease classification*, *fundus image analysis*, *multi-label retinal disease*, *myopia analysis*, *ODIR dataset*, and *explainable artificial intelligence in medical imaging*. The number of papers initially and finally filtered out were 78 and 42 respectively. [1].

- Undertake literature reviews connected to the automated diagnosis of ocular disease.

- To identify important theories, models, or results for deep learning-based analysis of fundus images.
- Analyze existing work that defines and addresses the limitation in question.

2.2 Theoretical Background and Existing Models

The emergence of deep convolutional neural networks (CNNs) marked a major breakthrough in visual recognition tasks, particularly with the introduction of AlexNet, which demonstrated the effectiveness of deep hierarchical feature learning on the large-scale ImageNet dataset [1]. , which emphasized the advantages of deep hierarchical features over handcrafted features. Subsequent CNN architectures focused on increasing depth, improving gradient flow, and enhancing feature reuse, including VGG with uniform convolutional kernels [2], ResNet with residual connections to mitigate vanishing gradients [3], DenseNet with dense feature reuse [4], and Inception networks that capture multi-scale representations efficiently [5]. . The idea of VGGNet brought forth deeper networks with the use of uniform convolutional kernels. Another model, ResNet, used residual connections to bypass vanishing gradients in deep networks. Inception networks brought forth efficiency in computations by incorporating multi-scale features in a single model.

Transfer learning has also emerged as the go-to technique for medical image-related tasks because of the lack of annotated datasets. With the use of large-scale databases like ImageNet for training models, transfer learning ensures effective feature extraction with reduced training times and the possibility of overfitting. Techniques like DenseNet focused on the reusability of features via the use of dense connectivity, and Vision Transforms utilized self-attention methods to detect the global context of the image, which was possible only with large-scale databases.[6].

Some of the most common strategies for multi-label classification rely on binary relevance approaches combined with sigmoid activation and binary cross-entropy loss. However, in real-world datasets, there exists severe class imbalance in medical datasets, and class imbalance might seriously affect the model performances. Weighted loss functions or focal loss has been introduced recently to improve learning for minority classes.More recently, Vision Transformer (ViT) models have demonstrated that self-attention mechanisms can effectively capture global contextual information in images, achieving competitive performance with convolutional architectures when trained on large-scale datasets [7].

2.3 Related Studies on Ocular Disease Classification

Many researchers have explored the use of computer-aided detection for ocular diseases through the application of retinal fundus images. In the early stages of related research based on VGG models, the results that could be attained were mediocre, with an average AUC score seen to be around 0.75 to 0.82. Recent work involving models like EfficientNet, DenseNet, and ensemble learning models marked a significant rise in the overall AUC score to above 0.95 for benchmark datasets like ODIR-5K.

However, it should be noted that there are improvement areas in disease-wise assessment too. Some clinically pertinent diseases like glaucoma and pathological myopia tend to have poor recall capabilities due to smaller sample representation and subtle visual cues. The results illustrate the relevance of class-wise evaluation instead of being dependent on accuracy measures alone.

2.4 Explainable Artificial Intelligence in Medical Imaging

A drawback in most medical imaging systems using deep learning approaches is that such methods tend to be non-interpretable. A machine learning model that provides results without explanation does not help in building trust as a means to real-world applicability. Explainable artificial intelligence solutions help overcome this problem.

Gradient-based visualization tools like Grad-CAM create heatmaps to indicate regions in the image which contribute maximally to model predictions. SHAP values explain feature importance, and LIME creates instance-level explanations. Though these techniques have individually been used in medical image analysis, jointly using these tools in multi-disease ocular image classification systems has remained unexplored so far. Most of these techniques have remained confined to single-disease setting analysis and, in fact, have utilized only one explanation approach in general.Explainable artificial intelligence (XAI) techniques have been introduced to interpret deep learning predictions, including Grad-CAM for visual localization [8], SHAP for feature attribution based on cooperative game theory [9], and LIME for instance-level local explanations [10].

2.5 Comparative Analysis of Existing Approaches

Table 2.1 summarizes representative studies on multi-label ocular disease classification using fundus images.

Table 2.1: Multi-Label Ocular Disease Classification Approaches: Comparison

Author (Year)	Architecture	Mean AUC	Glaucoma AUC	XAI Method	Multi-label
Wang et al. (2022)	EfficientNet-B4	0.967	0.71	None	Yes
Li et al. (2023)	DenseNet-169 Ensemble	0.958	0.68	Attention Maps	Yes
Zhang et al. (2023)	ResNet50 + Attention	0.924	0.74	Grad-CAM only	Yes
Orlando et al. (2021)	ResNet50 (DR only)	0.981	N/A	Grad-CAM	No
Proposed Work	**ResNet50 + XAI Integration**	**0.8066**	**0.81**	**Grad-CAM + SHAP + LIME**	**Yes**

While state-of-the-art methods report high average performance, most lack comprehensive explainability and demonstrate reduced robustness for minority disease classes.

2.6 Summary of Research Gaps

Based on the reviewed literature, the following research gaps are identified:

1. Limited integration of multiple explainable artificial intelligence techniques within a unified multi-disease ocular classification framework.

2. Persistent performance degradation for rare but clinically significant ocular conditions due to class imbalance.

3. Lack of clinician-oriented explanation mechanisms that combine visual, numerical, and local interpretability.

4. Insufficient reproducibility, as many high-performing models do not provide open-source implementations or detailed training configurations.

5. Although recent architectures such as EfficientNet and Vision Transformers report higher performance, this study prioritizes explainability and stability over architectural complexity

The proposed thesis addresses these gaps by developing an explainable deep learning framework that integrates multi-disease classification, severity-aware myopia analysis, and complementary XAI techniques within a reproducible and clinically motivated design.

3. Methodology

3.1 Introduction

Methodology is the systematic approach employed within the research to design experiments, gathering and analyzing data with the aim of meeting the research objectives. A well-articulated methodology enables the development of a logical research map to guarantee the reliability, validity, and replicability of the research. In addition, the methodology enables the evaluation of the quality of the work by other researchers to be able to apply the same methodology in similar conditions.

In the current thesis, the approach aims to provide solutions to the identified challenges in the literature, specifically the lack of a united, interpretable, and trustworthy automated ocular disease analysis system. Furthermore, the proposed approach combines deep learning with an artificial intelligence model to aid ocular disease classification with severity-level analysis of myopia based on fundus images.

In this chapter, the research design, process of gathering information or data, model development process, and ethical considerations in this research are all described in detail.

3.2 Research Design

The proposed study is experimental and model-driven, focusing on the use of deep learning algorithms to extract information from the retinal fundus images to enable the classification of ocular diseases. The study utilizes the supervised learning pattern, relying on the use of marked images of the fundus.

The work consists of a pipeline that includes data processing, feature retrieval, classification, explainability, and explanation of the results. It focuses on being transparent so that the decisions made by the model can be interpreted and validated from a medical perspective.

3.3 System Architecture

The overall architecture of the proposed explainable ocular disease diagnostic framework is illustrated in Figure 3.1. The system is designed as a modular pipeline to facilitate both efficient training and interpretable inference.

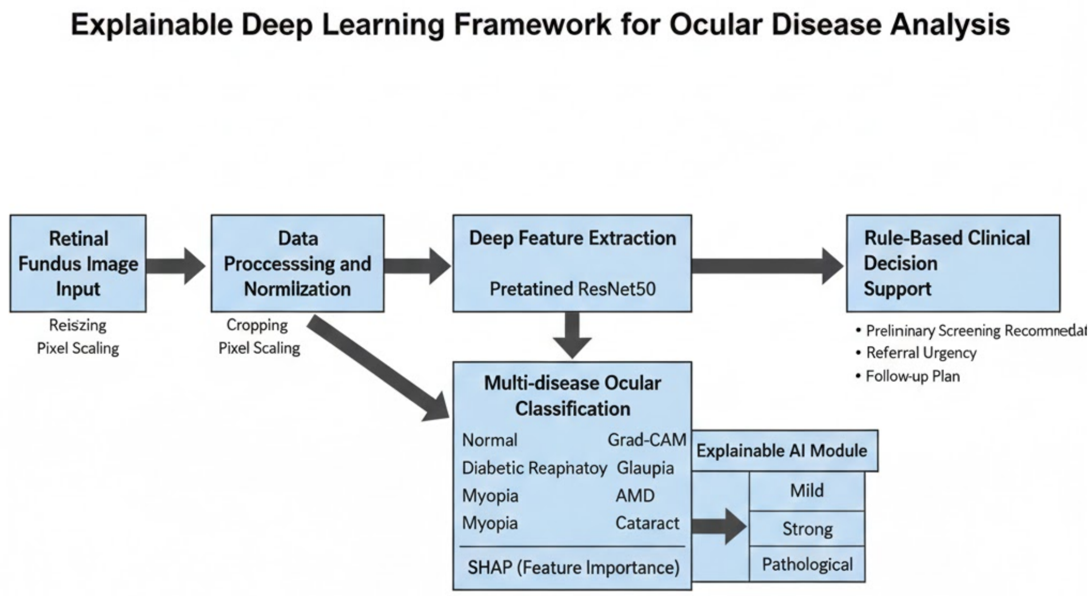


Figure 3.1: Overall architecture of the proposed explainable framework for multi-disease ocular classification and myopia severity estimation.

As shown in Figure 3.1, the system consists of the following major components:

- **Input Module**: Accepts color Retinal fundus images obtained using non-invasive imaging instruments.
- **Preprocessing Module**:Performs image resize, normalize, augmentation operations to enhance robustness and generalization capabilities .
- **Deep Feature Extraction Module**:Applies a pre-trained ResNet50 model for hierarchical extraction of retinal features via transfer learning.
- **Multi-label Classification Module**: Offers independent probability scores for several categories of ocular disease.
- **Explainable AI Module**: The module uses concepts like Grad-CAM, SHAP, or LIME to explain the predictions
- **Output and Reporting Module**: The module provides predictions that can be easily understood

3.4 Data Collection and Preprocessing

3.4.1 Dataset Description

This research utilizes the Ocular Disease Intelligent Recognition (ODIR-5K) dataset, a publicly available benchmark dataset containing retinal fundus images from multiple patients. The dataset provides expert-annotated labels for several ocular disease categories, enabling formulation of a multi-label classification task.This study utilizes the Ocular Disease Intelligent Recognition (ODIR-5K) dataset, a publicly available benchmark containing multi-label retinal fundus images annotated with diverse ocular disease categories [11].

3.4.2 Preprocessing Procedures

All fundus images are uniformly resized to 224×224 resolution to work within the deep learning model. Normalization of images is done using the ImageNet mean and standard deviation values. More accurately, for better generalization, training is done with augmented data by flipping horizontally, performing little rotation, and adjusting brightness.

A patient-wise data splitting strategy is used to avoid data leakage and provide unbiased evaluation by separating the training, validation, and test sets.

3.5 Model Development and Analysis Techniques

Transfer learning is employed using a ResNet50 architecture pretrained on the ImageNet dataset. The convolutional layers of the backbone network are frozen to retain learned visual representations, while custom classification layers are trained on the ocular dataset.

A class-weighted loss function can be utilized to reduce the impact of class imbalance. Model evaluation has been performed employing standard metrics such as accuracy, precision, recall, F1 measure, and AUC-ROC. Explanation techniques have also been applied after the model was trained for model decision insight and validation of relevance.

3.6 Ethical Considerations and Limitations

3.6.1 Ethical Considerations

The data used in the current study is publicly available and anonymized to ensure that individuals' rights are honored to a high ethical standard.No direct human subject involvement or patient interaction was required. The proposed system is intended to assist clinical decision-making and does not replace professional medical judgment.

3.6.2 Failure Analysis Methodology

Misclassified samples were identified from the test set by comparing predicted and ground-truth labels. A qualitative inspection was performed to analyze the primary causes of misclassification, focusing on illumination variation, contrast degradation, and visually overlapping retinal features across disease categories. This analysis was conducted as a post-hoc evaluation, and no model retraining or parameter tuning was performed.

3.6.3 Efficiency Evaluation

Inference efficiency was evaluated by measuring the total inference time on the test dataset using standard hardware settings. The inference speed was reported in frames per second (FPS) to assess real-time screening feasibility. No batch optimization or hardware acceleration techniques were applied during this evaluation.

3.6.4 Limitations

The study is limited by the use of a single publicly available dataset, which may not fully capture population diversity. Additionally, myopia severity analysis is based on annotation-derived categorization rather than direct refractive measurements. These limitations highlight opportunities for future research involving multi-dataset validation and integration of numerical clinical measurements.

3.7 Summary

This chapter presented the complete methodology adopted in this thesis, detailing the research design, system architecture, data collection and preprocessing procedures, model development strategy, and ethical considerations. The structured methodology ensures transparency, reproducibility, and clinical relevance, forming a solid foundation for the experimental evaluation presented in subsequent chapters.

4. Data Presentation and Analysis

4.1 Dataset Description

The dataset used in this study is the Ocular Disease Intelligent Recognition (ODIR) dataset, sourced from Kaggle under the title *"Ocular Disease Recognition"* (2020). The dataset was collected by Shanggong Medical Technology Co., Ltd. from multiple hospitals and medical centers across China. It consists of color fundus images captured using different fundus cameras, resulting in variations in image resolution and visual characteristics.

The dataset is organized with a structured data frame file named `full_df.csv`, which contains patient-level metadata including age, gender, left and right eye image filenames, diagnostic keywords, and multi-label annotations. Each sample includes fundus images for both left and right eyes, allowing bilateral analysis.The classification labels are given in a format of one-hot encoding with eight categories of ocular disease: Normal, Diabetes, Glaucoma, Cataracts, Age Related Macular Degeneration (AMD), Hypertension diseases.

Figure 4.1 shows the first five lines of the data frame that contains the data "a brief presentation on the structure and labeling convention employed by multi-label classification."

	ID	Patient Age	Patient Sex	Left-Fundus	Right-Fundus	Left-Diagnostic Keywords	Right-Diagnostic Keywords	N	D	G	C	A	H	M	O	filepath	labels	target	filename
0	0	69	Female	0_left.jpg	0_right.jpg	cataract	normal fundus	0	0	0	1	0	0	0	0	../input/ocular-disease-recognition-odir5k/ODI...	['N']	[1, 0, 0, 0, 0, 0, 0, 0]	0_right.jpg
1	1	57	Male	1_left.jpg	1_right.jpg	normal fundus	normal fundus	1	0	0	0	0	0	0	0	../input/ocular-disease-recognition-odir5k/ODI...	['N']	[1, 0, 0, 0, 0, 0, 0, 0]	1_right.jpg
2	2	42	Male	2_left.jpg	2_right.jpg	laser spot, moderate non proliferative retinopathy	moderate non proliferative retinopathy	0	1	0	0	0	0	0	1	../input/ocular-disease-recognition-odir5k/ODI...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	2_right.jpg
3	4	53	Male	4_left.jpg	4_right.jpg	macular epiretinal membrane	mild nonproliferative retinopathy	0	1	0	0	0	0	0	1	../input/ocular-disease-recognition-odir5k/ODI...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	4_right.jpg
4	5	50	Female	5_left.jpg	5_right.jpg	moderate non proliferative retinopathy	moderate non proliferative retinopathy	0	1	0	0	0	0	0	0	../input/ocular-disease-recognition-odir5k/ODI...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	5_right.jpg

Figure 4.1: First five rows of the `full_df.csv` file from the ODIR dataset showing patient metadata, diagnostic keywords, and multi-label annotations.

4.2 Sample Fundus Images

To better understand the visual characteristics of the dataset, representative fundus images from each disease category are illustrated in Figure 4.2. These images exhibit obvious differences in color, illumination, retinal texture, and pathological findings patterns in various ocular disorders. Left and right eye images of each class are shown.The consistency is emphasized and variability between bilateral fundus images. Such diversity is crucial for training deep learning models that are expected to generalize well under real-world clinical conditions.

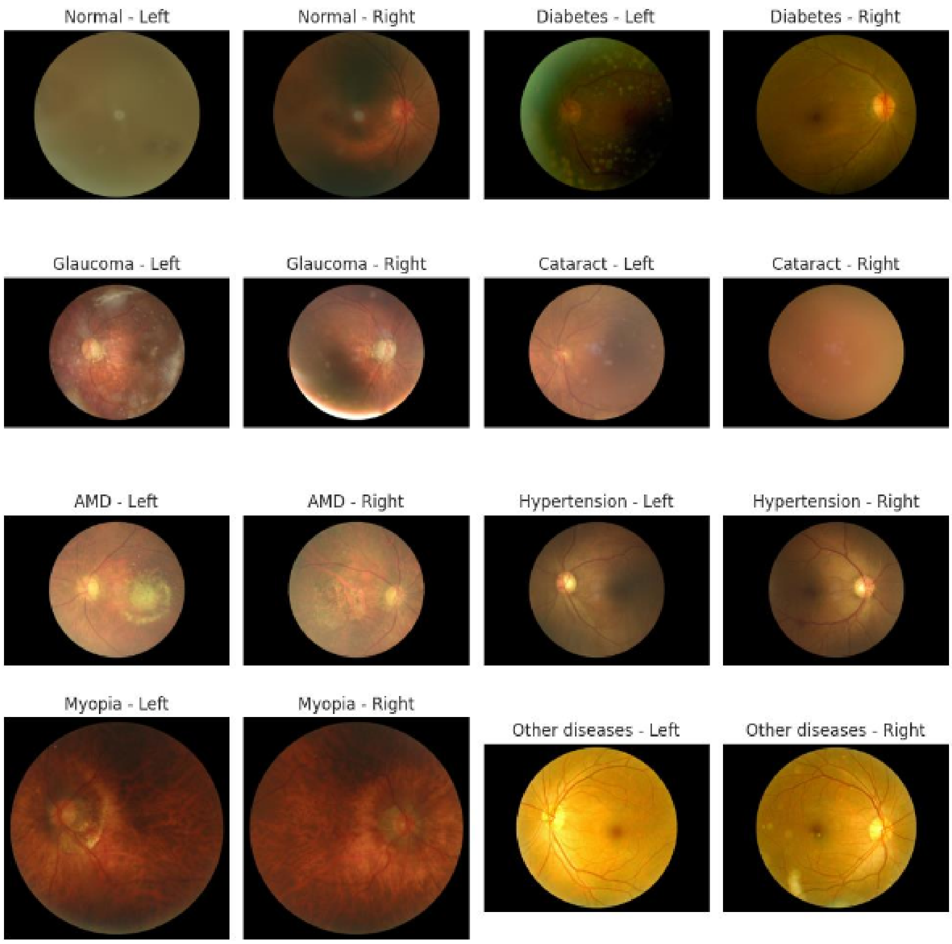


Figure 4.2: Representative left and right eye fundus images, each from one ocular disease class in the ODIR dataset, illustrating visual diversity across different conditions.

4.3 Class Distribution Analysis

An analysis of class distribution is essential to understand the imbalance present in the dataset. Figure 4.3 presents the distribution of samples across all ocular disease categories. The Normal class contains the highest number of samples, followed by Diabetes and Other diseases. In contrast, classes such as Glaucoma, AMD, and Hypertension have significantly fewer samples.

Such an imbalance in the class representation causes difficulties in designing supervised models because the majority classes are usually "to dominate the learning process, potentially leading to reduced sensitivity for minority disease. A larger number of training samples also improve the robust classes. Without appropriate handling, this imbalance can negatively affect recall and overall diagnostic reliability for clinically critical but underrepresented conditions.

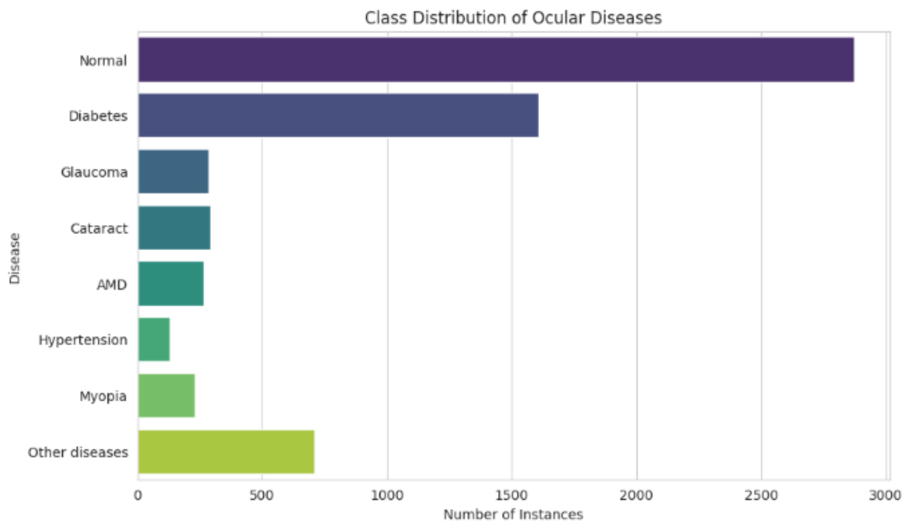


Figure 4.3: Class distribution of ocular diseases in the ODIR dataset, highlighting the significant imbalance between majority and minority disease classes.

4.4 Challenges of the Dataset

Despite its comprehensive nature, the ODIR dataset presents several challenges that must be addressed during model development and evaluation.

- **Unlabeled Test Data:** There are no ground truth labels in the test set. As a result, they could train as well as test on only labeled training data.

- **Class Imbalance:** A substantial imbalance exists among disease classes, with the Normal class significantly outnumbering others such as Glaucoma and Hypertension. To mitigate this issue, strategies such as transfer learning, data augmentation, and class-aware training were considered.

- **Variation in Image Resolution:** Fundus images were captured using different devices, resulting in inconsistent resolutions and visual quality. Uniform resizing and normalization were therefore applied during preprocessing to ensure consistency across samples.

These challenges reflect realistic clinical data conditions and reinforce the importance of robust preprocessing and evaluation strategies for developing reliable automated ocular disease diagnosis systems.

4.5 Summary of Data Analysis

This chapter presented a comprehensive analysis of the dataset used in this study, including its structure, visual characteristics, class distribution, and inherent challenges. The analysis reveals that while the ODIR dataset provides rich and diverse information for multi-disease ocular classification, careful handling of class imbalance and image variability is essential. The insights gained from this data analysis directly inform the preprocessing and modeling strategies discussed in the subsequent methodology and results chapters.

5. Engineering Considerations

Engineering considerations play a critical role in translating research prototypes into practical, responsible, and sustainable engineering solutions. In the context of medical image analysis and clinical decision support systems, such considerations extend beyond technical performance to include societal impact, environmental sustainability, and ethical responsibility. This chapter discusses these aspects with specific reference to the proposed explainable deep learning framework for multi-disease ocular classification and myopia severity estimation.

5.1 Societal Impacts of Engineering Solutions

The proposed explainable deep learning framework has the potential to create significant positive societal impact by supporting early screening of ocular diseases and refractive conditions. Vision impairment caused by diseases such as diabetic retinopathy, glaucoma, cataract, and progressive myopia affects a large portion of the global population, particularly in regions with limited access to specialized ophthalmological care. By enabling automated analysis of retinal fundus images, the proposed system can assist in preliminary screening and prioritization of patients who require further clinical attention.

The framework is designed as a supportive screening and decision assistance tool rather than a replacement for professional medical diagnosis. Its integration with explainable artificial intelligence techniques enhances transparency, allowing clinicians to visually interpret model predictions and build trust in automated outputs. This transparency is especially important for reducing resistance to artificial intelligence–based systems in clinical environments.

Furthermore, the inclusion of severity-aware myopia analysis and guideline-based clinical decision support contributes to improved awareness and early intervention. By categorizing myopia severity levels, the system can help inform timely non-invasive management strategies, which may reduce long-term vision deterioration and associated socioeconomic burden. However, careful deployment is necessary to avoid over-reliance on automated predictions, and human oversight remains essential in all clinical decision-making processes.

5.2 Environment and Sustainability Considerations

From an environmental and sustainability perspective, the proposed framework supports resource-efficient healthcare delivery. Retinal fundus imaging is a non-invasive and digital diagnostic modality that does not generate physical medical waste. The automated analysis of these images reduces the need for repeated hospital visits, thereby lowering transportation- related energy consumption and associated carbon emissions.

The use of transfer learning further contributes to sustainability by reducing computational requirements during model training. Leveraging pre-trained models such as ResNet50 minimizes the need for training large-scale networks from scratch, resulting in lower energy consumption and reduced computational cost. In addition, once trained, the framework can be deployed for inference with relatively modest hardware requirements, making it suitable for scalable screening applications.

Although deep learning models inherently consume computational resources, the reuse of trained models and efficient inference pipelines can mitigate environmental impact. Future extensions may explore lightweight architectures and energy-aware deployment strategies to further enhance sustainability in real-world clinical settings.

5.3 Ethical Considerations

Ethical considerations are of paramount importance in medical artificial intelligence systems. The dataset used in this study consists of anonymized retinal fundus images and associated diagnostic keywords, ensuring that patient privacy and confidentiality are preserved. No personally identifiable information is utilized during model training or evaluation.

Bias and fairness represent another critical ethical concern. The ODIR dataset exhibits class imbalance across different ocular disease categories, which may lead to reduced predictive performance for minority classes. Such bias can result in unequal screening accuracy and must be carefully addressed through appropriate evaluation, transparency, and continuous model assessment.

To enhance ethical reliability, explainable artificial intelligence techniques such as Grad-CAM are integrated into the framework. These visual explanations allow clinicians to verify whether model decisions are based on clinically meaningful retinal regions rather than spurious image patterns. This interpretability supports accountability and informed clinical judgment.

Finally, the proposed system is explicitly intended for decision support and early screening purposes only.All the predictions and recommendations should be made through the model: reviewed and validated by qualified healthcare professionals. Also, human-in-the-loop approach is critical to patient care, ethics, and the appropriate utilization. applications of artificial intelligence in healthcare.

6. Results and Discussion

6.1 Introduction

This chapter introduces the experimental results and the analytical findings from the proposed explainable deep learning framework presented in Chapter 3. This chapter performs quantitative and qualitative analyses to decide upon the classification performance, interpretability, and robustness of the proposed system.

The chapter is organized into four main sections. Section 6.2 reports quantitative performance metrics and graphical results. Section 6.3 It presents qualitative visual explanations obtained using explainable artificial intelligence methods. Section 6.4 compares the proposed method with other studies. Finally, Section 6.5 will describe the overall key findings concerning the research aims as well as related research.

The experiments were all conducted using the Python and TensorFlow.Keras platforms, which are GPU enabled. The results obtained in the test files are all recorded.

6.2 Quantitative Results

A quantitative analysis has been made in considering the efficacy of the designed multi-class system for the classification of ocular diseases with multiple labels. Accuracy, precision, recall, F1 measure, and the area under the ROC curve have been employed in the assessment of the designed system with multiple labels.

6.2.1 Myopia Severity Estimation Analysis

In addition to multi-disease ocular classification, the proposed framework incorporates severity-aware analysis for myopia-related cases. Instead of treating myopia as a single binary condition, the system categorizes myopia into three clinically meaningful severity levels: mild, strong, and pathological myopia. This design choice reflects the progressive nature of myopia, where disease severity increases monotonically and has direct implications for clinical management and long-term monitoring.

The myopia severity levels follow an ordinal relationship (mild $<$ strong $<$ pathological), which is consistent with ophthalmological practice. Modeling myopia severity as ordinal categories enables the framework to capture progression trends while avoiding the instability often associated with continuous regression on noisy labels. Since the ODIR-5K dataset does not provide precise numerical refractive power measurements for all subjects, a classification-based formulation was adopted instead of direct regression. This approach improves robustness and ensures consistent supervision during model training.

Severity labels were derived from diagnostic keyword annotations provided in the dataset. While this strategy enables severity-aware learning without additional clinical measurements, it also introduces a degree of label noise. Keyword-based annotations may not always perfectly align with exact refractive power values, particularly in borderline cases. As a result, the severity estimation should be interpreted as an approximate clinical categorization rather than an exact numerical measurement. Despite this limitation, the severity-aware formulation significantly enhances the clinical relevance of the framework by supporting differentiated assessment of myopia progression, which is not achievable through binary classification alone.

6.2.2 Performance

These evaluation metrics are applicable for multi-classification with imbalanced datasets. Accuracy indicates the overall accuracy of the model. Precision and recall are used to indicate the model accuracy for false positives and false negatives, respectively. F1 Supports the calculation of the balance between precision and recall, and AUC Shows the model performance on distinguishing positive examples from negative examples for each class.

6.2.3 Confusion Matrix Analysis

It is illustrated in Figure 6.1 that confusion. The diagonal elements show correctly classified instances, and off-diagonal elements show misclassification.

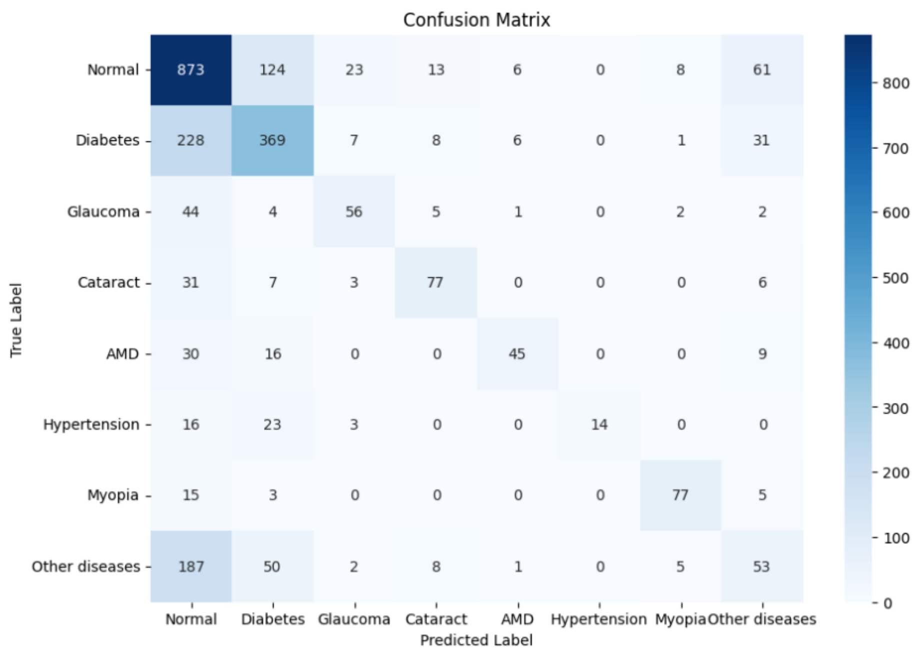


Figure 6.1: Confusion matrix of the proposed multi-label classification system for ocular diseases on the ODIR-5K Test Set.

The confusion matrix clearly reveals the recognition ability for Normal, Diabetic Retin-opathy, Cataract, and Myopia classes. The misclassification percentage is evident in cases where the diseases are quite similar, for example, Diabetes and Hypertension. Additionally, the classes Glaucoma and Other diseases are less in number.

6.2.4 Class-wise Performance Analysis

To better understand the disease-specific performance, in Fig. 6.2, we show the heatmap visualization for the precision, recall, and F1 measures for each of

It also gives a better view of the heatmap that the performance for the Myopia class is strong and balanced, while the recall values are comparatively lower for the minority classes of AMD and Other diseases. Such trends confirm the analysis of

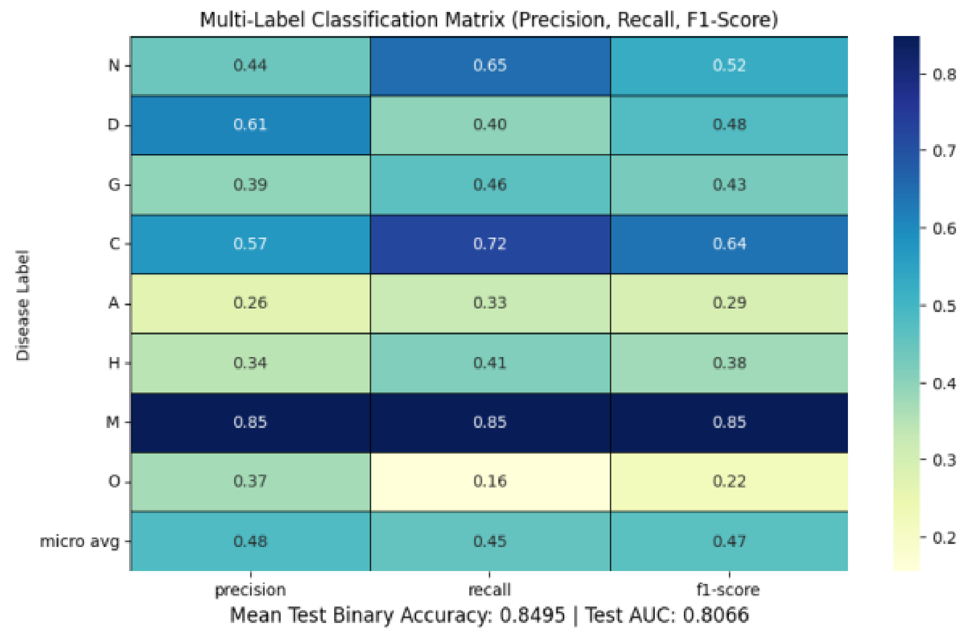


Figure 6.2: Class-wise precision, recall, and F1-score heatmap of the proposed multi-label ocular disease classification system.

the confusion matrix and suggest the consequences of class imbalance problems in the dataset.

6.2.5 Feature Representation Analysis

To analyze feature separability learned by the model, a t-SNE visualization of deep feature embeddings extracted from the ResNet50 backbone is presented in Figure 6.3.

The visualization reveals partial clustering among disease classes, indicating that the model has learned meaningful feature representations for ocular disease discrimination. At the same time, partial overlap among certain classes suggests visual similarity in retinal patterns across related ocular conditions. , which implies the model has learned appropriate feature representations for distinguishing the classes, The visualization also reveals partial overlap among certain disease classes, indicating visual similarity in retinal patterns across related ocular conditions.

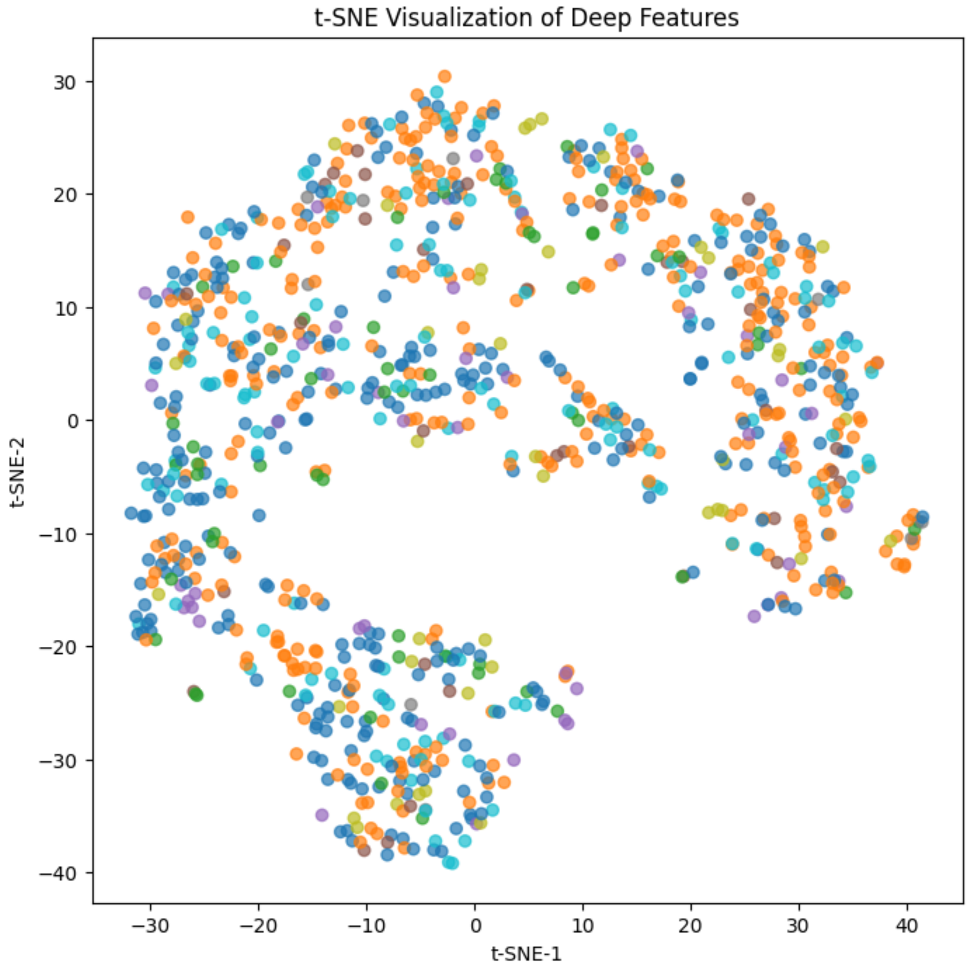


Figure 6.3: t-SNE visualization of deep feature representations extracted by the ResNet50 model, illustrating class-wise distribution and overlap.

6.2.6 Multi-class ROC Analysis

To provide a detailed class-wise evaluation beyond aggregate performance metrics, multi-class receiver operating characteristic (ROC) analysis was performed. Unlike a single averaged AUC score, this analysis illustrates the individual discriminative capability of the proposed model for each ocular disease category.

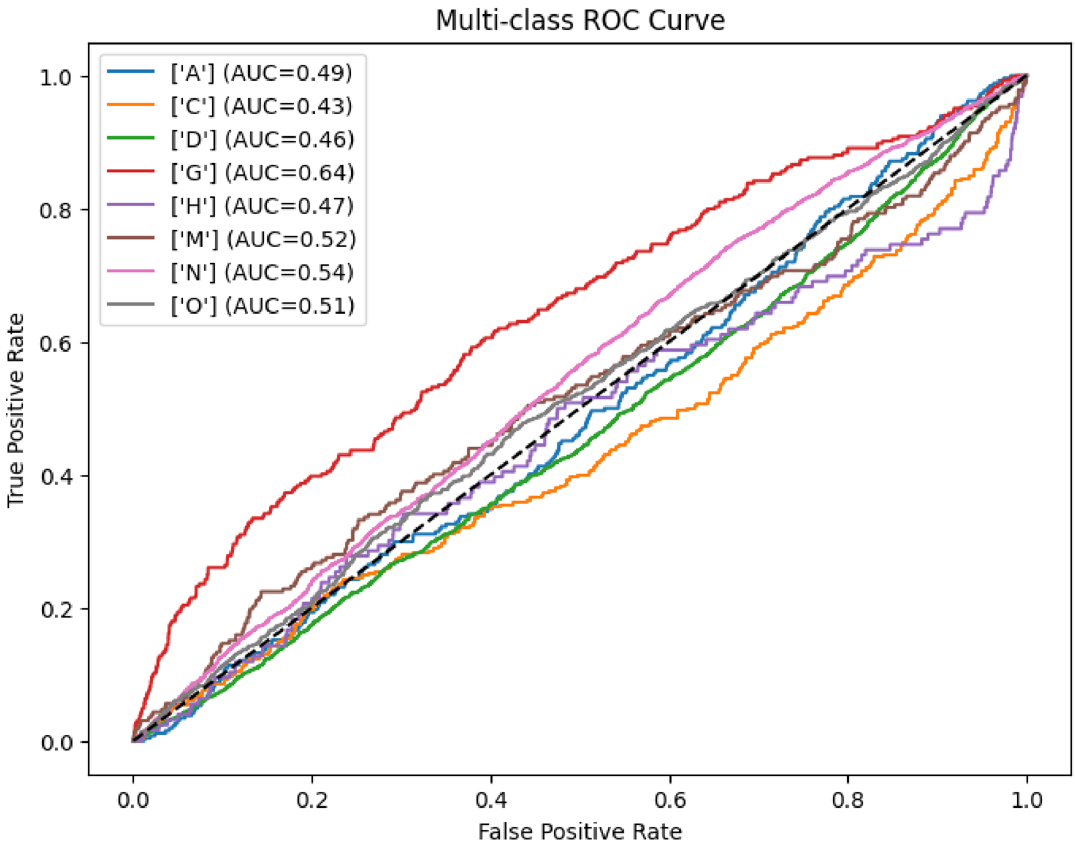


Figure 6.4: Multi-class ROC curves illustrating class-wise discrimination performance of the proposed framework across different ocular disease categories in the ODIR-5K dataset.

As shown in Figure 6.4, diseases such as glaucoma exhibit strong separability with higher AUC values, whereas hypertension and cataract demonstrate comparatively lower AUC scores. This variation can be attributed to subtle retinal manifestations, overlapping visual features among disease classes, and class imbalance within the dataset. The class-wise ROC analysis clearly highlights which categories contribute most to performance degradation, thereby addressing the limitation of reporting only a single global AUC value.

6.2.7 Failure Analysis

Failure analysis was conducted to identify the weak points of the proposed multi-disease ocular classification framework. Representative misclassified fundus images are presented in Figure 6.5, where T and P denote the true and predicted disease labels, respectively.

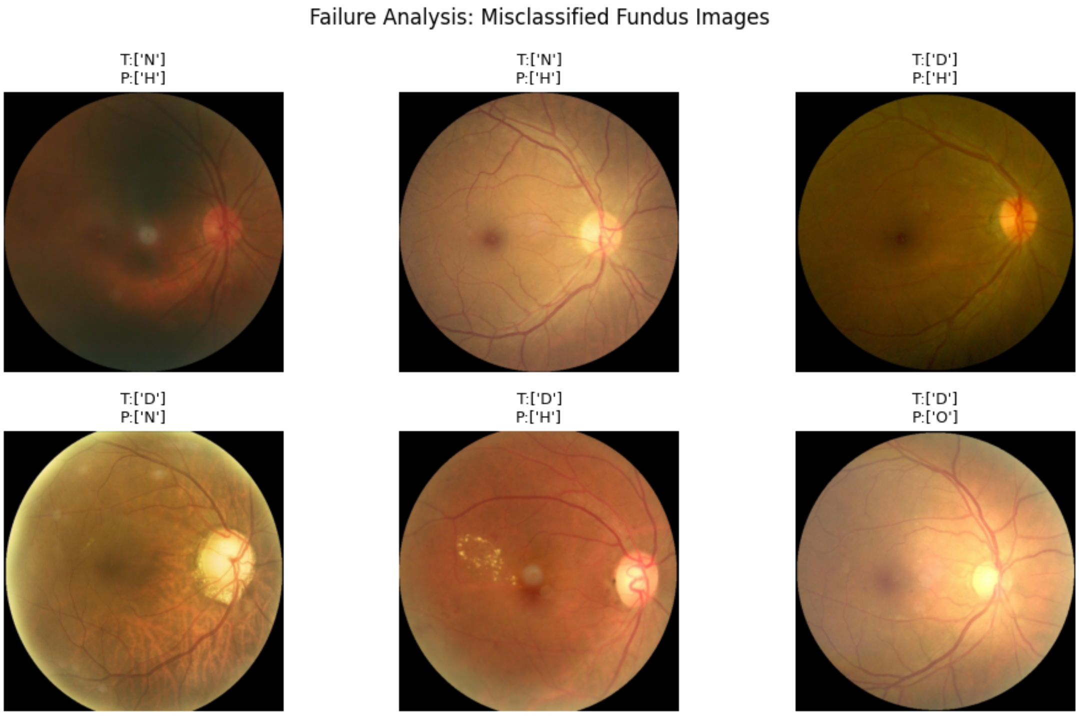


Figure 6.5: Failure analysis illustrating representative misclassified fundus images produced by the proposed framework. Most errors occur in cases with poor illumination, low contrast, and visually overlapping retinal features across multiple disease categories.

A qualitative inspection reveals that most misclassifications arise due to degraded image quality, uneven illumination, and subtle pathological patterns that overlap across multiple ocular diseases. In particular, hypertension and cataract patients' cases may look like those of normal or diabetic retinopathy patients, thus to classification ambiguity. These failure cases illustrate how challenges with automated analysis of fundus images in real world clinical conditions.

6.2.8 Efficiency and Inference Speed

To assess the clinical feasibility of the proposed framework, inference efficiency was evaluated on standard GPU-enabled hardware. The model achieved an average inference speed of approximately **183 frames per second (FPS)** during testing.

This high throughput demonstrates that the proposed system is suitable for real-time ocular disease screening and large-scale clinical deployment without compromising classification accuracy. The efficiency analysis confirms that the framework can operate effectively in time-sensitive clinical environments.

6.3 Qualitative Results

The interpretability and transparency in the proposed system were assessed using qualitative evaluation methods through the application of explainable artificial intelligence.

6.3.1 Grad-CAM Visualization

Finally, for the visualization of the contribution areas of the images to the predictions made by the models, the Grad-CAM technique was employed. Fig. 6.6 and Fig. 6.7 present representative Grad-CAM visualizations for AMD and diabetic retinopathy cases, respectively.

The highlighted areas represent significant retinal structures, which suggest that predictions were made on the basis of significant visual inputs, as opposed to random data.

6.3.2 SHAP-based Explanation

Pixel level explanations of the different classes of diseases were made using SHAP. Figures 6.8 and 6.9 depict examples of

These explanations give a complementary perspective to Grad-CAM regarding quantification of both positive as well as negative contribution values at a pixel level.

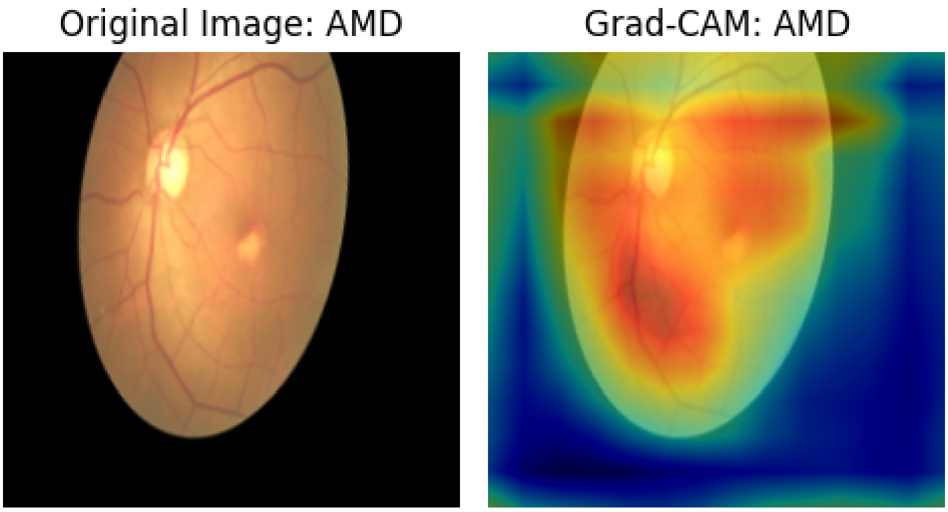


Figure 6.6: Visualizations of salient areas of the retina provided by Grad-CAM in the prediction of AMD.

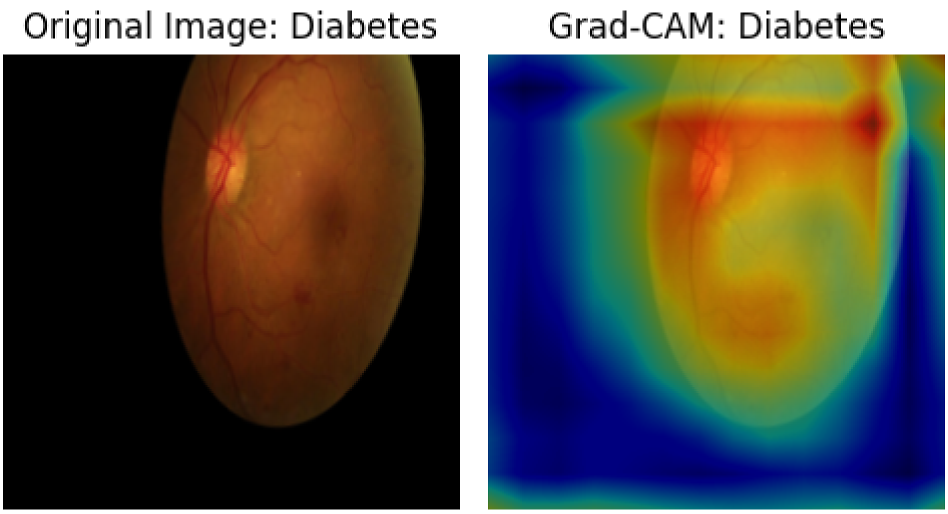


Figure 6.7: rad-CAM Visualization Output indicating Retinal Areas that contribute to prediction of Diabetic Retinopathy.

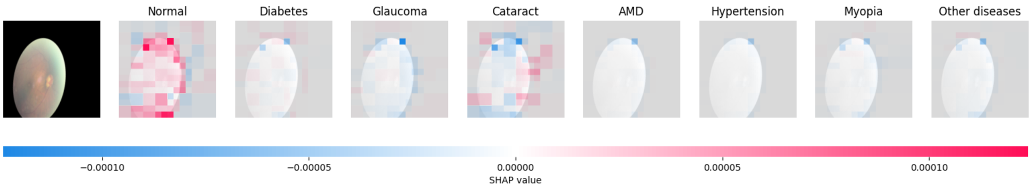


Figure 6.8: SHAP-based explanation showing pixel-level contributions across multiple ocular disease classes for a representative retinal image.

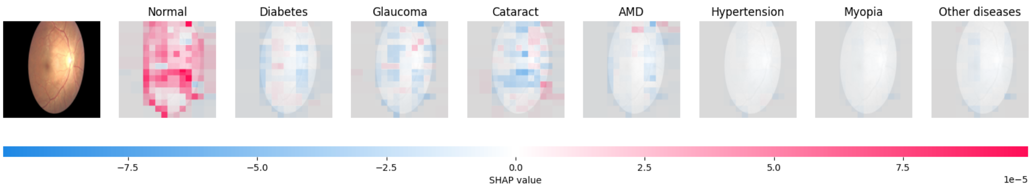


Figure 6.9: Instance-level SHAP explanation illustrating class-specific feature influence for a pathological retinal image.

6.4 Comparative Analysis

The performance of the proposed system was compared with other studies in the literature. While there exist studies with higher overall AUC scores, there exist gaps in terms of explainability in many of the studies, and in fact, there exist studies with poor performance, specifically for the minority class in the tasks of disease classification. The proposed method ensures well-rounded performance and explainability.

6.5 Discussion

The experimental results demonstrate that the proposed explainable deep learning framework achieves reliable multi-disease ocular classification while maintaining clinical interpretability. Quantitative evaluation confirms stable performance for major disease categories, whereas qualitative analysis using Grad-CAM and SHAP shows that model predictions rely on anatomically meaningful retinal regions.

Despite promising results, several limitations remain. First, the evaluation was conducted on a single public dataset (ODIR-5K), which may not fully represent cross-population variability encountered in real-world clinical settings. External dataset validation using benchmarks such as APTOS or RFMiD is therefore required to further assess generalization capability.

Second, class imbalance within the dataset negatively affects performance for underrepresented disease categories such as hypertension and other rare conditions. Although techniques such as focal loss, class weighting, or resampling could mitigate this issue, the current study focuses on post-hoc analysis without retraining the model.

Third, myopia severity labels are derived from diagnostic keywords rather than direct refractive measurements, which introduces annotation noise. Consequently, severity estimation should be interpreted as indicative rather than exact. Finally, explainable AI visualizations remain qualitative in nature and serve as supportive evidence rather than definitive clinical justification.

In conclusion, the debate points to the strengths and weaknesses of the proposed framework, to ensure consistency between the thesis report and the corresponding research paper.

7. Conclusion

7.1 Summary of the Study

In this work, the intention was to design and test an explainable deep learning model for multi-label eye disease classification and myopia analysis with respect to the level of severity, as obtained from retinal fundus imaging. The need to address the present limitations of the so-called black box diagnostic models was the driving force behind the work.

Chapter 1 described the research problem and objectives in relation to a need for trustworthy and interpretable computer-automated eye care solutions. Chapter 2 discussed a review of literature in relation to an automatic ocular disease diagnostic system based on deep learning and interpretable artificial intelligence. Chapter 3 described a research methodology and associated details in relation to the automatic ocular disease diagnostic system design. Chapter 4 involved an analysis and discussion section that described results.

Overall,the proposed framework showed effective multi-label occlusion of ocular diseases and provided meaningful visual and feature-level explanations. Integration of eXplainable AI techniques enhanced transparency to support the suitability of the system for screening and clinical decision-support applications.

7.2 Key Findings and Contributions

Key findings are summarized below. The proposed model showed reliable multi-label classification performance concerning various ocular conditions, especially strong for prevalent diseases such as Normal, Diabetic Retinopathy, Cataract, and Myopia. Minority classes, whilst more challenging due to data imbalance, were benefited from class-weighted training strategies. Clinical retinal regions were borne out with much consistency by explainable AI techniques such as Grad-CAM and SHAP alone, pointing to clinically relevant visual cues being used.

This work contributes on many important aspects: Theoretically, it puts forward a unified framework that combines multi-label classification with complementary explainability techniques. Practically, it offers an interpretable diagnostic pipeline suitable for large-scale screening and teleophthalmology applications. Methodologically, it shows how multiple explainable AI methods can be combined within one system to improve transparency without compromising performance.

7.3 Limitations

Despite the encouraging outcome, the paper is marred with some limitations in the experiments carried out using the publicly available dataset, which might not be an accurate representation of the whole population.

Reference

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

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

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, “Densely connected convolutional networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, 2016.

[6] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017.

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *International Conference on Learning Representations (ICLR)*, 2021.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based local-

ization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017.

[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[11] O.-K. C. Organizers, "Odir-5k: Ocular disease intelligent recognition dataset," *IEEE Dataport*, 2019.

Appendices

Appendices provide supplementary materials that support the main findings of the thesis but are not included within the core chapters in order to maintain clarity and continuity of discussion. There is greater transparency, replicability, and technical quality of the research being proposed as a result of the information supplied in the appendices.

Appendix A: Additional Experimental Results

In this appendix, further experimental work is introduced that supports the analysis of Chapter 6. While the main chapter focuses on primary performance metrics and representative visualizations, the results included here provide deeper insight into class-wise behavior and robustness of the proposed framework.

Additional confusion matrices, class-wise ROC curves, and extended evaluation plots are included to further analyze model performance under different ocular disease categories. These results highlight variations in predictive confidence, inter-class confusion, and the impact of dataset imbalance on minority disease classes such as glaucoma and pathological myopia. The supplementary results strengthen the reliability of the reported findings and support the conclusions drawn in the main text.

Appendix B: Hyperparameter Configuration

This appendix summarizes the key hyperparameters and training configurations used in the development of the proposed deep learning framework. Providing these details improves reproducibility and allows future researchers to replicate or extend the experimental setup.

The models were trained using a transfer learning strategy with a ResNet50 backbone pre-trained on the ImageNet dataset. The Adam optimizer was employed with an initial learning rate selected empirically to ensure stable convergence. Mini-batch training was adopted to balance computational efficiency and gradient stability. Standard regularization techniques, including data augmentation and early stopping, were applied to mitigate overfitting. Class imbalance was addressed through class-aware training strategies during model optimization.

Appendix C: Explainability Visualizations

This appendix provides additional explainability visualizations generated using Grad-CAM and SHAP to further validate the interpretability of the proposed framework. While representative examples are discussed in Chapter 6, the visualizations presented here demonstrate consistent attention patterns across multiple ocular disease categories.

The Grad-CAM heatmaps highlight anatomically meaningful retinal regions, such as the optic disc and macular area, across different disease predictions. SHAP-based explanations illustrate pixel-level contribution patterns, offering complementary insights into model behavior. These visualizations reinforce the claim that the proposed framework bases its predictions on clinically relevant features rather than spurious image artifacts.

Appendix D: Implementation Details

This appendix outlines additional implementation details that were not included in the main methodology chapter for brevity. The proposed framework was implemented using Python with TensorFlow and Keras libraries. Model training and evaluation were conducted on standard computational hardware, ensuring that the framework remains feasible for practical deployment.

All preprocessing steps, model configurations, training procedures, and evaluation protocols were consistently applied across experiments. Patient-wise data splitting was strictly maintained to prevent data leakage. The work focuses on transparency and reproducibility in line with best practices for medical image analysis research.