MobViGNet: An Explainable AI-based Hybrid Transfer Learning Approach for Eye Disease Classification from External Eye Images

Abu Kowshir Bitto

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

APRIL 2025

DECLARATION OF THESIS AND COPYRIGHT

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Abu Kowshir Bitto
Date of Birth	:	10/04/1999
Title	:	Master of Science
Academic Session	:	2024-2025

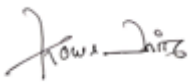
I declare that this thesis is classified as:

☒	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☒	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☒	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

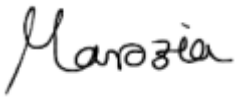
1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:



_____
(Student's Signature)

241-44-008
_____
Student ID
Date: 24/04/2025

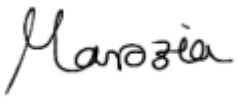


_____
(Supervisor's Signature)

Dr. Marzia Ahmed
_____
Name of Supervisor
Date: 24/04/2025

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science.



(Supervisor's Signature)

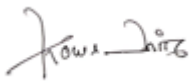
Full Name : Dr. Marzia Ahmed

Position : Assistant Professor

Date : 24/04/2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.



(Student's Signature)

Full Name : Abu Kowshir Bitto

ID Number : 241-44-008

Date : 24/04/2025

APPROVAL

APPROVAL

This thesis titled on "MobViGNet: An Explainable AI-based Hybrid Convolutional Neural Network Approach for Eye Disease Classification from External Eye Images," submitted by Abu Kowshir Bitto, ID: 241-44-008 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Masters of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

09.07.25

Chairman

Dr. S.M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Daffodil International University

9-May-25

Internal Examiner 1

Afsana Begum
Assistant Professor
Department of Software Engineering
Daffodil International University

9-5-2025

Internal Examiner 2

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Daffodil International University

09-05-25

External Examiner

ISHTIAK HUSSAIN
Project Manager
Dynamic Solution Innovators (DSi)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Almighty Allah for granting me the strength, patience, and determination to complete this thesis. My deepest gratitude to my supervisor, Dr. Marzia Ahmed, forer guidance, encouragement, and continuous support throughout this research process. Her comments and constructive criticism greatly enhanced the quality of this work. My heartfelt thanks are due to the faculty and staff of Department of Software Engineering, Daffodil International University, for the stimulating academic environment and the facilities provided for conducting this research. I am also greatly indebted to my friends and colleagues for their encouragement, thought-provoking discussions, and emotional support, which sustained me and kept me going. I special note of appreciation to my family especially my parents for their love, prayers, and unwavering belief in me. Their encouragement has been the cornerstone of my academic endeavors. Finally, to everyone who contributed directly or indirectly to the completion of this thesis thank you.

DEDICATION

This thesis is dedicated to my loving parents, whose limitless love, unshakable support, and ongoing prayers have been the foundations of every one of my achievements. To my teachers and mentors, who inspired me to question, to search, and to grow. And to all dreamers who are brave enough to pursue knowledge in the face of adversity this is for you.

ABSTRACT

The application of artificial intelligence (AI) in medical diagnostics, particularly in ophthalmology, holds great promise for enhancing early detection and improving patient outcomes. However, the complexity of diagnosing common eye diseases from external eye images, along with issues such as limited labeled datasets and the need for model transparency, pose significant challenges. To address these, we propose MobViGNet, an innovative hybrid transfer learning model for multi-categorical detection of common eye diseases, such as cataracts, conjunctivitis, and normal conditions. Our propose MobViGNet, integrating the strengths of MobileNetV2 and VGG19 to achieve efficient and effective eye disease classification from external eye images. The lightweight, depth wise separable convolutions of MobileNetV2 and deep hierarchical feature extraction of VGG19 are integrated through parallel feature learning and subsequent concatenation and custom dense layers to construct a compact yet robust feature representation. Both networks are initialized with pre-trained ImageNet weights and frozen base layers to allow prior knowledge, while the additional layers are fine-tuned with the Adam optimizer and sparse categorical cross-entropy loss. MobViGNet outperforms individual and state-of-the-art networks like Xception, InceptionV3, and VGG19, achieving 99.85% accuracy, 99.77% F1-score, and 0.15% error rate on 5-fold cross-validation. Explainability is incorporated into the model with LIME and Grad-CAM, with interpretable predictions that are crucial for clinical trust. The model also showed excellent performance when tested on an external dataset, confirming its ability to generalize beyond the training data.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Prompt identification and accurate diagnosis of ocular diseases are key to effectively preventing irreversible vision loss and improving patient outcomes. Eye-related diseases, such as cataracts, conjunctivitis, diabetic retinopathy, and glaucoma, stand as some of the foremost causes of visual impairment worldwide. Their burden lies beyond health, as they devastate the quality of life, economic productivity, and even educational welfare, with the estimated global reduction in work productivity due to vision impairment reaching US$410.7 billion in 2020 (World Health Organization, 2008). This represents approximately 2.2 billion people on the globe, over half of whom are experiencing conditions that could be prevented or treated with diagnostic competence (SV et al., 2025).

Cataracts, a significant cause of blindness, result in the eye's lens becoming cloudy, increasingly compromising vision if untreated (Canatan, 2024). Conjunctivitis, or "pink eye," is an extremely contagious infection of the conjunctiva, usually caused by viral, bacterial, or allergic infections (Satish et al., 2024). Though in most cases non-lethal, its high incidence means that correct diagnosis is imperative. Diabetic retinopathy is a complication of diabetes resulting from damage to the retinal blood vessels, causing loss of vision if not detected early (Safi et al., 2018). Glaucoma is a collection of eye diseases in which optic nerve degeneration is the common etiology, associated with a rise in intraocular pressure mainly; it is a leading cause of preventable blindness (Allison et al., 2020). All these conditions need precise early diagnosis for proper management.

Traditional diagnostic methods are based on either clinical expertise or manual evaluations, which are not only time- and resource-consuming but also introduce variability between evaluators. Deep learning, especially convolutional neural networks (CNNs, for short), redefined the very approach by which medical image analysis used to be done (Sarvamangala & Kulkarni, 2022). CNNs have been excellent at extracting

complex patterns embedded in images, hence resetting the state-of-the-art performance levels in classifying many medical conditions (Ersavas et al., 2024). Bottlenecks do occur, though, in the requirement for enormous, well-annotated datasets and in the model's ability to interpret the prediction.

1.2 Problem Statement

Problem Statement 1: Enhancing Classification Performance with Limited Datasets - Classifying eye diseases like cataract, conjunctivitis, and normal eye are difficult when there is not enough data to train models. Many traditional deep learning models struggle to generalize well with small and imbalanced datasets. This means we need new hybrid models that combine the strengths of pre-trained architectures, allowing us to improve accuracy and performance even when we have limited labelled data to work with.

Problem Statement 2: Clinical Trust and Decision-Making with Interpretable AI - Deep learning models have been extremely promising, but their "black-box" nature makes clinicians uncomfortable trusting their decisions. Physicians in clinical settings need to understand why a model predicts something in a particular way, especially when diagnosing diseases like cataract, conjunctivitis, and diabetic retinopathy. In order to gain universal acceptance in medicine, AI must be comprehensible and appealing to medical thought patterns doctors use in the process of arriving at their conclusions.

Problem Statement 3: Generalization and Scalability of Models for Real-World Deployment - Deep learning models work best in controlled research settings but fail to deliver in real-world clinical settings. Differences in image quality, patient populations, and disease presentations render models unreliable. For practical application, models need to be robust, scalable, and adaptable, and must be capable of dealing with diverse datasets and handling real-time applications like telemedicine and mobile health platforms.

1.3 Research Objectives

Objective 1: To create a hybrid deep learning model that combines the strengths of pre-trained architectures with cross validation techniques, improving the classification

of external eye diseases like cataract, conjunctivitis, and normal eye, especially when working with limited and unbalanced data.

Objective 2: To make the hybrid model more transparent by incorporating Explainable AI (XAI) techniques, helping clinicians understand how the model makes its predictions and building trust in its decision-making process for diagnosing eye diseases.

Objective 3: To test the hybrid model on internal dataset that is collected and validate by eye expert and as well external datasets and ensure it can be scaled for real-world use, so it can effectively handle diverse data and work well in clinical environments, including telemedicine and mobile health applications.

1.4 Research Question

Four research questions (RQs) are considered for this study and these are given below:

- RQ1: How does a hybrid deep learning approach enhance the classification performance of eye diseases, especially with limited and imbalanced datasets?
- RQ2: What is the impact of fine-tuning pre-trained models (e.g., MobileNetV2, VGG19) on classification accuracy, precision, and inference time?
- RQ3: How can Explainable AI (XAI) techniques improve clinical trust by highlighting image regions influencing the model's predictions?
- RQ4: How well does the proposed hybrid model generalize across external datasets and real-world clinical conditions?

1.5 Contribution

Four This study presents several key contributions toward advancing AI-driven eye disease diagnosis and these are:

- Collected and validated a high-quality internal dataset in collaboration with ophthalmology experts, ensuring clinical relevance and data reliability.

- Developed a novel hybrid deep learning model by combining multiple pre-trained architectures, significantly improving classification accuracy for eye diseases such as cataract, conjunctivitis, and normal eye.

- Incorporated advanced cross-validation techniques to enhance the model's robustness, generalizability, and resistance to overfitting across various data splits.

- Integrated Explainable AI (XAI) methods to provide interpretable, visual explanations of model predictions, fostering trust and transparency in clinical decision-making.

- Performed comprehensive validation on both internal and external datasets, demonstrating the model's scalability and effectiveness for real-world applications, including telemedicine and mobile health platforms.

CHAPTER 2

RELATED WORK

2.1 Literature Review

The art of eye disease diagnosis has seen unprecedented strides in the last few years thanks to machine learning and deep learning. Scientists have been trying to refine devices that not only identify ailments like cataracts and conjunctivitis with laser-like accuracy but also provide results faster than ever before—allowing physicians to respond more rapidly and provide patients with a brighter path to recovery. It's like giving medicine a pair of sharpers, quicker eyes.

2.2 Deep Learning in Medical Imaging

Diagnosis and grading of ocular diseases have been enhanced significantly with the introduction of newer imaging modalities and machine learning (ML) technology. Optical Coherence Tomography Angiography (OCTA) has been a promising imaging modality in this context, offering high-resolution images that allow detailed assessment of the retinal vascular network. The recent literature emphasizes the growing overlap between deep learning techniques and OCTA imaging, marking a new era in the computerized detection and classification of numerous ocular pathologies, including diabetic retinopathy (DR) and choroidal neovascularization Le et al. (2020) and Yao et al. (2020). Deep learning algorithms, including an important example of Convolutional Neural Networks (CNNs), have been shown to be very efficient at classifying different retinal diseases based on OCT images.

For instance, Umer et al. (2023) reported a deep feature fusion model that achieved an accuracy rate of 93.25% in retinal eye disease classification from OCT images, demonstrating the potential of deep learning in this field. Similarly, Ahmed and Hameed's research emphasized the efficiency of the back-propagation algorithm in achieving as much as an accuracy of 98.79% in the diagnosis of various eye diseases, further confirming the efficiency of convolutional neural networks for high-risk diagnosis

conditions (Ahmed & Hameed 2021). Additionally, it has been demonstrated through a systematic review that deep learning enhances not just the diagnosis of single-class disease but also multi-class classification performance, thus enhancing clinical decision-making (Zhao et al. 2022). Additionally, image preprocessing is also a critical factor in eye disease detection, affecting the classification result.

Sarki et al. (2021) emphasized the importance of preprocessing techniques aimed at noise removal and image enhancement that subsequently enhances the overall classification model accuracy. Feature fusion and hybrid feature selection have also been employed to enhance performance metrics on different algorithms, thereby elevating the detection of diseases such as diabetic eye diseases (Shamsan et al. 2023). The combination of advanced imaging methods, deep learning techniques, and intense preprocessing strategies is very essential in improving diagnosis and classification of eye diseases. This innovation has great promise for future advancements and applications in the field of healthcare, which will eventually lead to better patient outcomes and improved healthcare services.

2.3 Deep Learning in Ophthalmology

Deep learning has emerged as an enabler in the development of medical imaging accuracy and efficiency, including in the field of ophthalmology. Deep learning methods in the detection of eye diseases have helped a lot, especially with the application of CNNs, which have shown success in the analysis of retinal images. Several studies have described success in applying deep learning models to diagnose ocular disease, such as cataract, conjunctivitis, diabetic retinopathy (DR), age-related macular degeneration (AMD), and glaucoma. Cen et al. (2021) research demonstrated the capacity of a deep learning system to automatically diagnose several retinal diseases at similar accuracy levels to retina specialists, indicating the prospects of overall diagnosis systems that rely on deep learning. Gulshan et al. (2016) trained a deep learning model to detect diabetic retinopathy with very good sensitivity and specificity, which goes a long way to describe the effectiveness of such models in the clinical environment.

The effectiveness of deep learning for the detection and classification of ocular abnormalities is also affirmed by Abbas et al. (2023) who presented a transfer learning strategy with self-attention mechanisms in CNNs for improving the recognition of

disease. Their results show continued novelty in model architectures for improved recognition capability in different ocular diseases. In addition, research like that of Kang et al. (2021) demonstrates the use of multimodal imaging methods, combining several imaging modalities with deep learning to enhance diagnostic precision for retinal vascular disease. Deep learning not only assists in the identification of the disease but also increases classification precision in other diseases of the eye. For example, the use of hybrid models, as suggested by Umer et al., (2023) integrates deep feature extraction and selection techniques and reports impressive performance in retinal eye disease detection, which is important for clinical decision-making. Deep learning has revolutionized medical imaging, particularly ophthalmology, with its potential improving detection and grading of various eye diseases. The merits of sophisticated algorithms and novel methods in multimodal imaging and transfer learning are examples of continuing developments that can produce even broader and more effective eye care in the not-too-distant future.

2.4 Explainable AI in Medical Imaging

Explainable AI (XAI) is now a critical component in the healthcare sector, primarily addressing the transparency, interpretability, and trust issues of AI solutions. XAI becomes important due to the necessity to provide transparent explanations of AI-generated outcomes, especially about clinical decision-making, and patient treatment, where the consequences are extremely high. Specifically, clinicians must be assured that AI systems are not only "black boxes," but instead deliver explainable findings that can inform patient treatment protocols and diagnostic decisions. The addition of XAI to medical AI applications has been widely advocated as a means of enhancing user confidence and facilitating adoption by clinicians (Thakur 2024) (Yang 2022) (Zhang et al. 2022)

A significant challenge to medical artificial intelligence is the interpretability of its models. As AI algorithms grow more complex, it is crucial that the decision-making is not obscured from healthcare professionals so that they can effectively understand the suggestions being made by the AI (Thakur 2024). Methods such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) have been identified as essential methods for producing interpretable output from black-box models, thus allowing medical professionals to interact with AI-generated outputs with

increased trust (Madi et al. 2024). Among various healthcare applications, including diagnosis and treatment planning, XAI's capability to offer insights into the way conclusions are reached is vital for clinical decision-making that is informed and for building patient trust (Patel et al. 2024).

Practically, the use of XAI is far broader than for diagnosis; it has tremendous potential for predictive modelling and for precision medicine as well. For instance, XAI can be utilized to translate predictive results into actionable clinical recommendations so that personalized treatment can be provided that is specific to the precise needs of the patient. These developments accentuate the demand for end-to-end frameworks that optimize explanation in addition to the trade-off between model performance and interpretability, given the heterogeneity of healthcare data and diverse requirements of clinicians (Özalp & Sevil 2023) (Rahmati 2025)

The integration of explainable AI in medicine is critical to enhancing transparency, trust, and accountable medical decision-making. The recently innovated XAI methods, combined with the momentum for cross-disciplinary research, have the potential to transform the health delivery system through more explainable and trustworthy AI systems.

2.5 Explainable AI in Ophthalmology

Explainable AI (XAI) has grown more important in the case of eye disease detection and diagnosis, particularly with its ability to improve diagnostic precision, enhance clinician awareness, and ultimately improve patient outcomes. Incorporating XAI into AI systems, particularly in the case of ophthalmology, allows for greater transparency into how algorithms arrive at certain diagnostic conclusions, enhancing trust between healthcare providers and patients.

In diabetic retinopathy (DR), it has been established through research that AI systems can identify and detect such and other eye diseases with high sensitivity and specificity. One such example is a multi-task AI classifier, which had a high sensitivity of 94.84% in detecting patients who are at risk of vision loss due to retinopathies, indicating the clinical utility of AI tools by Alam et al. (2019). XAI approaches are particularly pertinent in the scenario of explaining such AI systems' decision-making in diagnosing eye conditions. Not only does this enhance interpretability of findings, but it

also allows clinicians to validate and associate AI findings with their clinical understanding, thereby developing trust in AI-assisted diagnosis (Alsadoun et al. 2024). The successful implementation of XAI in retinal disorder classification paves the way for an AI-friendly future where AI is seamlessly integrated into clinical practice, making scalable screening programs efficient enough to effectively counter eye diseases (Li et al. 2021).

2.6 Existing Research in Eye Disease Classification and Detection

Recent advancements in deep learning have significantly contributed to disease classification, particularly in ophthalmology. Convolutional Neural Networks (CNNs) have shown promising results; for instance, Ganokratanaa et al. (2023) implemented LeNet-CNN and Support Vector Machine (SVM) for cataract detection, with LeNet-CNN reaching 96% accuracy. Transfer learning approaches have also been widely explored. Bitto & Mahmud (2022) employed pre-trained models including Inception-v3, ResNet-50, and VGG-16 to classify cataracts, conjunctivitis, and normal eye images, with Inception-v3 performing best at 97.08% accuracy. Similarly, Niloy et al. (2024) utilized MobileNetV2, ResNet50, and EfficientNet-B7 for detecting uveitis and eyelid disorders, with MobileNetV2 achieving a recognition rate of 96%. Hybrid and ensemble learning models further enhanced performance in eye disease detection. Yadav & Yadav (2023) developed custom CNN model and achieving an accuracy of 92.50% for cataract detection. Appiahene et al. (2023) integrated CNN with Logistic Regression and Gaussian Blur in a conjunctiva imaging system for non-invasive anemia detection, attaining 92.5% accuracy. These studies highlight the growing effectiveness of CNN-based, transfer learning, and hybrid models in medical image classification tasks.

The field of eye disease classification has advanced significantly with the application of machine learning (ML) and deep learning (DL), improving diagnostic accuracy for conditions such as cataracts, conjunctivitis, and diabetic retinopathy. This literature review highlights recent methodologies, key findings, and ongoing research gaps. While deep learning models have shown promising results, many studies remain limited to specific conditions and fail to generalize across multiple external eye diseases. Most existing approaches rely on single-modality imaging and small, imbalanced datasets, often augmented synthetically. In contrast, our study introduces a hybrid model that combines lightweight spatial and deep hierarchical features to enhance classification

accuracy. The dataset used is carefully curated, manually validated by medical experts, and preprocessed using perceptual hashing to ensure data integrity. Furthermore, although CNN-based models are widely used, few incorporate explainability methods such as LIME to interpret model decisions—an essential feature for clinical acceptance. Our work addresses these gaps by proposing robust, interpretable hybrid architecture with strong generalization capabilities, validated through extensive cross-validation and tested on both internal and external datasets, outperforming existing methods in both accuracy and transparency.

CHAPTER 3

METHODOLOGY

3.1 Proposed Methodology

The diagnosis and classification of ocular pathologies through transfer learning have become an important facet in modern-day medical practice. In our work, represented in Figure 1, data aggregated of several sources, including both our carefully compiled dataset and publicly shared datasets, culminated in a single, overall data bank. Dataset preprocessing involved Python-based methodologies for identifying duplicates in images, and then a manual checking process for its quality assurance. In the whole process, it was assured that no duplicates existed in the dataset, and for the final version, a medical expert validated it. The dataset involves three categories: cataract, conjunctivitis, and healthy eye cases. For efficient model performance, a balanced distribution in all categories was assured through adopted data augmentation techniques. In addition, five pre-trained Convolutional Neural Networks (CNNs), InceptionV3, VGG19, MobileNetV2, Xception, and EfficientNet-B7, were taken for feature extraction and subsequently for classification of these ocular pathologies through external eye photographs. For best performance, a composite model, developed through combining MobileNetV2 and VGG19 with adding and frozen layers, performed best, and all model validation was performed through 5-fold cross-validation for reliable reporting of accuracy, recall, F1-score, and confusion matrices. We have even incorporated LIME (Local Interpretable Model-agnostic Explanations) for interpretability, identifying regions in an image impacting model prediction. Hybrid model performance was even validated through a publicly shared Kaggle dataset, consisting of four categories: cataract, diabetic retinopathy, glaucoma, and healthy eyes.

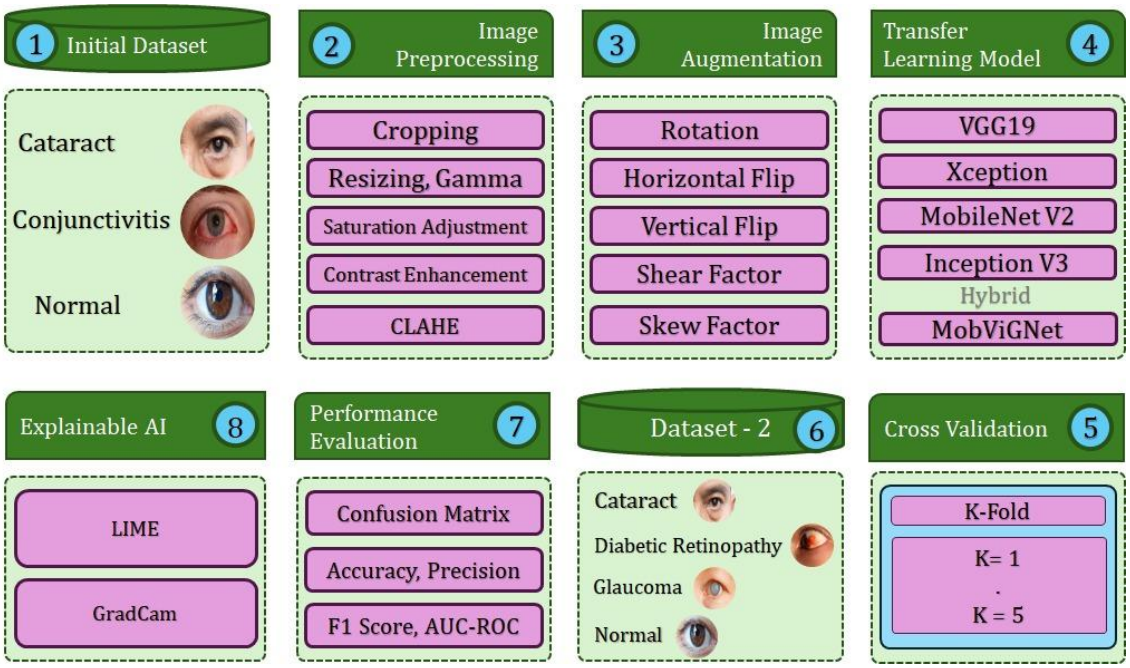


Figure 3.1 Step by Step Methodology Diagram

3.2 Data Description and Pre-Processing

3.2.1 Dataset - 1

The dataset used in this study includes external eye images from curated datasets, such as Shutterstock and Google Images. The dataset originally contained 750 images each of cataract, conjunctivitis, and normal eyes. To make the dataset even more representative, in addition to the 750 images collected, another 50 images from each class were gathered, resulting in a total of 800 images per class. Figure 2 shows the example of a dataset and Table 1 shows symptoms of the disease. To avoid possible data losses, each image was checked manually for duplicates. In addition, to identify and eliminate potential duplicates, the duplicate image detection algorithm was implemented in Python using a perceptual hashing approach. No duplicate images could be found upon verification. The final dataset was manually reviewed, with its labels validated and authenticated by an ophthalmology expert to ensure reliability and validity. Our final dataset consists of 2400 images shows in the figure where's to test purpose we use 480 data and for train purpose 1920.

(A) Cataract (B) Conjunctivitis (C) Normal

Figure 3.2 Sample images from the dataset represent different eye pathologies (a) Cataract, (b) Conjunctivitis, and (c) Normal.

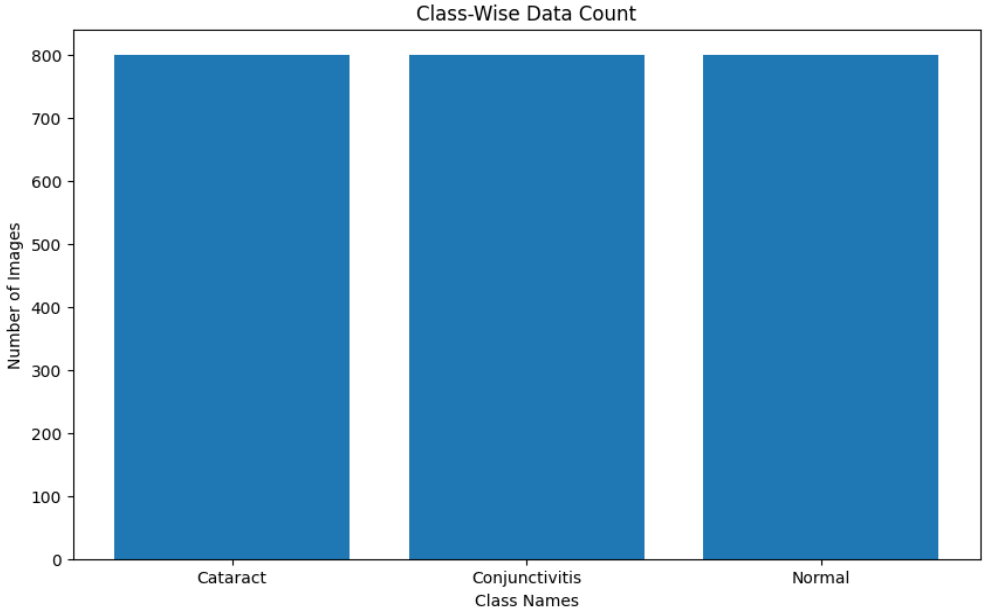


Figure 3.3 Class-wise data distribution of Dataset-1.

Table 3.1 Disease with its symptoms of Dataset-1.

Disease Name	Symptoms
Cataract	Cloudy or blurred vision, difficulty seeing at night, sensitivity to glare.
Conjunctivitis	Redness, itching, tearing, discharge, and crusting of the eyelids.
Normal	No abnormalities, clear vision, no redness or swelling.

A perceptual hashing algorithm was applied to plot to eliminate dataset redundancy and maintain the quality of the dataset. For this, we initialize an empty hash table and dataset to store unique images, described through Algorithm step shows in Algorithm 1. Each image is converted into gray scale and resized to a fixed dimension before calculating its perceptual hash value using the pHash function. The image is then inserted into the hash table and the deduplicated dataset if no hash value currently exists in the hash table. If not, the image is one of duplicate and is dismissed. This step ensures that there are no duplicates in our dataset and improves the quality of our dataset for machine learning purposes. This deduplicated dataset is then used to train and validate machine learning models. The algorithm below describes the method for identifying and eliminating duplicate images through a hashing technique. This process ensures the dataset is devoid of redundancy, enhancing its quality and suitability for machine learning applications.

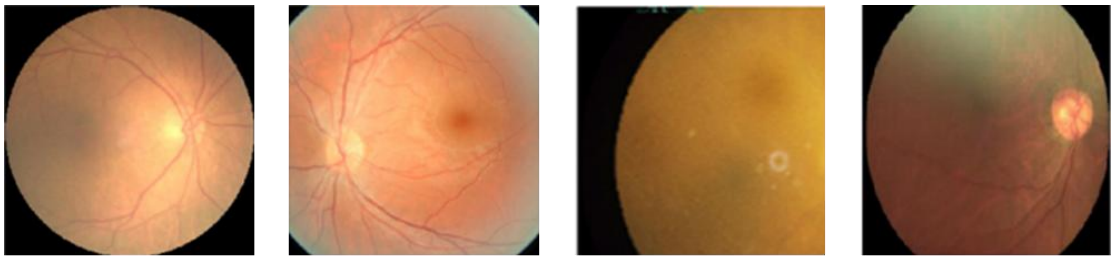
Algorithm 1 Duplicate Image Detection and Removal Using Hashing

Input:	Image dataset $D = \{1, 2, ..., I_n\}$ containing n images
Output:	Deduplicated dataset D'
Step-1:	Initialize an empty hash table $H \leftarrow \{\}$
Step-2:	Initialize an empty list $D' \leftarrow \{\}$
Step-3:	**FOR** each image $I_i \in D$
Step-4:	Convert I_i to grayscale
Step-5:	Resize I_i to a fixed dimension (h, w)
Step-6:	Compute perceptual hash h_i for I_i using *pHash*
Step-7:	**IF** $h_i \notin H$
Step-8:	Add h_i to the hash table H
Step-9:	Add I_i to D'
Step-10:	**ELSE**
Step-11:	Mark I_i *as a duplicate*
Step-12:	**ENDIF**
Step-13:	**ENDFOR**
Step-14:	Return D'

3.2.2 Dataset -2

The external Kaggle dataset [40] consists of 4 classes, Normal, Diabetic Retinopathy, Cataract and Glaucoma retinal images where each class have approximately 1000 images. These images are collected from various sorces like IDRiD, Oculur recognition, HRF etc. Figure shows the example of dataset and Table 2 shows symptoms of the disease. Our final dataset consists of 4217 images showing in the figure where is

to test purpose, we use 841 data and for train purpose 3376. Figure 5 depicts the class-wise distribution of Dataset-2.



(a) (b) (c) (d)

Figure 3.4 Sample images of Kaggle dataset consists of 4 classes, (a) Normal, (b) Diabetic Retinopathy, (c) Cataract and (d) Glaucoma retinal.

Table 3.2 Disease with its symptoms of Dataset-2.

Disease Name	**Symptoms**
Normal	No abnormalities, clear vision, no redness or swelling.
Diabetic Retinopathy	Blurred vision, floaters, dark spots, impaired color vision, vision loss.
Cataract	Cloudy or blurred vision, difficulty seeing at night, sensitivity to glare.
Glaucoma	Gradual vision loss, tunnel vision, eye pain, halos around lights, nausea.

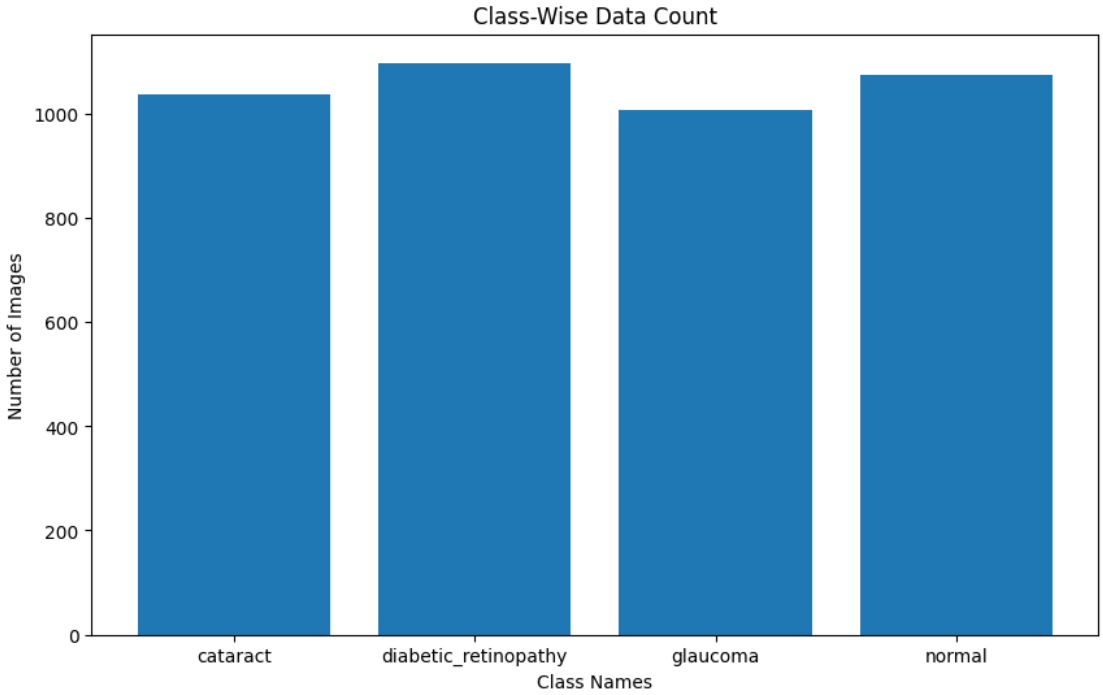


Figure 3.5 Class-wise data distribution of Dataset-2.

3.3 Data Partitioning

One of the most crucial steps in developing a good machine learning model is effectively partitioning the data such that we can reliably evaluate and generalize against the learned concept. To evaluate the performance of the 4 models and porposed hybrid model the dataset was separated using a 5-fold cross-validation strategy. The whole dataset is randomly split into five equal parts or folds. One of the folds is used as the test set and the other 4 for training. It's done this way for five folds, with each serving as a testing set once. By applying 5-fold cross-validation, the data samples' positions are adequately balanced between training and testing sets to capture the model's performance. This also protects against overfitting as the model performance is evaluated on several splits of the data. In each iteration, InceptionV3, VGG19, MobileNetV2, Xception and hybrid model were trained on the training folds and validated on the associated test fold. This was repeated for all five iterations, resulting in five accuracy scores for each model. These scores have been averaged to give a robust estimate of the model's overall performance. By comparing performance on each data split, this formulation ensures that only the fair evaluations are impacted, while also showing the consistency of all the model applied through the process. This partitioning strategy is also optimal for our study as the size of our dataset is relatively small. 5-fold

cross-validation also provides high accuracy and generalization in finding the best model due to the large number of samples being used for training, while maintaining an adequate number of trained samples for testing.

3.4 Model Development

In this work, four state-of-the-art pre-trained Convolutional Neural Networks (CNNs) have been used to classify external eye diseases: InceptionV3, VGG19, MobileNetV2, and Xception. These models were chosen for their demonstrated performance over a variety of image classification problems. Transfer learning was used to take advantage of pre-trained weights from ImageNet and help improve performance on our dataset, where we froze the base layers and fine-tuned each model. By fine-tuning, the models learned to adjust for domain-specific accents in the external eye disease dataset. MobileNetV2 and VGG19 outperformed the others, and they were Willene together to build a hybrid model where for the hybrid model we did many experiments in context we froze many layers and add some custom layers for each model and after concept both model we add some more layers. The model brief description is given below:

- InceptionV3: Inception is a type of neural network that has inception modules, which apply multiple filters of various sizes in the same layer. It works well with images where there are small but significant details, for example, identifying small changes in external eye conditions.
- VGG19: The VGG19 has a total of 19 layers, out of which there are 16 for convolutional and 3 are fully connected layers, and its architecture is a deep and sequential model. This process leads to hierarchical learning of features, from low-level patterns such as edges to higher levels of complexity, allowing it to excel in identifying intricate patterns of eye pathologies.
- MobileNetV2: A lightweight model designed for mobile and edge devices, using depthwise separable convolutions to reduce computation with minimal loss in accuracy. This architecture works well in contexts where available computing resources are scarce and accurate classification is essential
- Xception: Utilising depthwise separable convolutions, Xception takes this idea further by adding cross-channel correlations. Its adaptive design captures intricate patterns in images, boosting performance on difficult datasets.

3.4.1 Proposed Model (MobViGNet)

The hybrid structure combines the MobileNetV2 and VGG19 structures in such a way that they share each other's strengths to do a better job of classifying eye disease. Both MobileNetV2 and VGG19 are applied to the input image first, where the base levels of both networks are pre-trained on the ImageNet dataset and their weights froze to retain the learned features. The results from the two models are flattened and then passed through single custom dense layers, each of which is followed by Batch Normalization and Dropout to reduce overfitting and improve generalization. The features are concatenated together, and additional dense layers are added to further increase the model's capacity for learning rich representations. The output layer consists of a softmax layer that produces the input into one of the classes defined. The model is trained using the Adam optimizer and sparse categorical cross-entropy loss so that it is trained efficiently yet with high accuracy. The hybrid approach not just improves the performance of classification but also makes the predictions robust and accurate by combining the strengths of both MobileNetV2 and VGG19. The architecture of the proposed model is presented in Figure 6.

Depthwise separable convolutions are used in MobileNetV2 to minimize computational complexity. For an input tensor X with shape $H \times W \times C$, depthwise separable convolution breaks into:

$$X' = (X \times K_d) \times K_p \quad (1)$$

Where K_d is a depthwise kernel and K_p is a pointwise kernel. This already reduces the number parameters from $(K^2 \cdot C \cdot C')$ to $(K^2 \cdot C + C \cdot C')$ as C is input channels and C' is output channels.

VGG19 employs a very deep architecture with standard convolutions. Each convolution layer l is applied to the input X through:

$$X_l' = \{ReLU\}(W_l \cdot X + b_l) \quad (2)$$

Where W_l is for the weight matrix and bl is for the bias. These deep architectures allow the model to learn features in hierarchy. To merge the advantages of both models MobileNetV2 & VGG19, the extracted features are concatenated into a single finalized feature vector based on the representation below:

$$F_{\{combined\}} = [F_M; F_V] \quad (3)$$

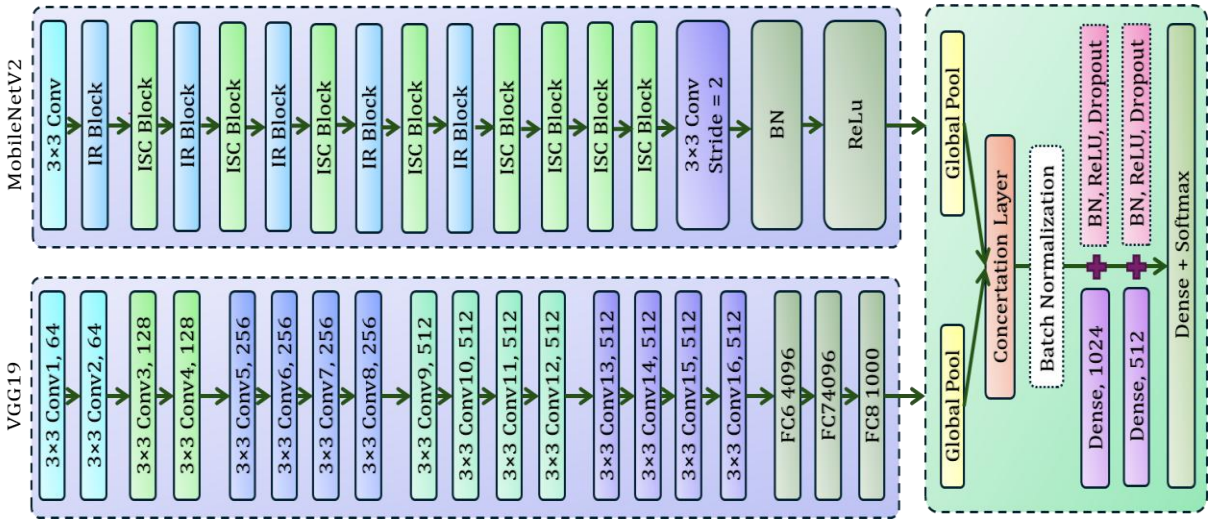


Figure 3.6 The proposed architecture of hybrid model.

Where F_M and F_V are the flattened outputs of MobileNetV2 and VGG19, respectively. The $F_{combined}$ has dimensions of d_M+d_V, where d_M and d_V are the dimensions of F_M and F_V, respectively.

Dense layers execute the linear transformation + nonlinear activation functions. For a dense layer (with input F):

$$z = W.F + b \quad (3)$$

$$a = ReLU(z) \quad (4)$$

Where W is the weight matrix and b is the bias. The non-linearity introduced by ReLU helps in capturing complex patterns.

The final layer uses the softmax function to produce class probabilities. For an output vector z:

$$p_i = \{e^{\{z_i\}}\}\{\sum_j e^{\{z_j\}}\} \quad (5)$$

Where pi is the probability for class *i*, and z_i is the score for class *i*. This ensures that the output is a valid probability distribution over the classes.

The loss function used is *sparse_categorical_crossentropy*, which for true label y and predicted probabilities y^, is:

$$L(y, \hat{y}) = -\log(\hat{y}_y) \tag{6}$$

Where $\hat{y}_y$ is the predicted probability for the true class *y*. This loss function is appropriate for multi-class classification.

The Adam optimizer updates the model parameters using:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t \tag{7}$$

Where η is the learning rate, m_t is the first moment estimate, v_t is the second moment estimate, and ϵ is a small constant to avoid division by zero. This adaptive learning rate helps in converging faster.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Experiment and Result

To evaluate the effectiveness of the proposed hybrid deep learning model for classifying external eye diseases, a series of experiments were conducted using both internally validated and publicly available datasets. The model's performance was assessed through rigorous cross-validation techniques, and comparisons were made with existing state-of-the-art architectures. This section presents the experimental setup, followed by quantitative results and analysis based on standard evaluation metrics.

4.1.1 Experimental Setup

All models were trained using the Adam optimizer, selected for its adaptive learning rate capabilities that promote faster and more stable convergence. A fixed learning rate of 0.0001 was employed across all experiments to ensure consistent and gradual updates, minimizing abrupt parameter shifts. Cross-Entropy Loss was used as the objective function, which is standard for multi-class classification tasks. Although training was conducted in a CPU environment rather than a GPU, the computational setup was adequate for effectively training the models given the dataset size and architecture. Non-linear transformations were introduced through ReLU activation functions across various layers to capture complex patterns within the data. To ensure robustness and generalizability, a 5-fold cross-validation strategy was applied with shuffled splits and a fixed random state (42). Each fold was trained for 20 epochs with a batch size of 32, balancing between training efficiency and convergence stability. A summary of the model training configurations is provided in Table 4.1.

Table 4.1 Experimental setup for model training parameter.

Configuration	InceptionV3	VGG19	MobileNetV2	Xception	MobViGNet

Model Architecture	48-layer CNN with Inception modules	19-layer CNN (16 Conv + 3 FC)	Depthwise Separable CNN (53 layers)	Depthwise Separable CNN (36 Conv layers)	MobileNetV2 + VGG19
Pretrained Weights	Yes (ImageNet)	Yes (ImageNet)	Yes (ImageNet)	Yes (ImageNet)	Yes (ImageNet)
Optimizer	Adam	Adam	Adam	Adam	Adam
Learning Rate	0.0001	0.0001	0.0001	0.0001	0.0001
Loss Function	CrossEntropyLoss	CrossEntropyLoss	CrossEntropyLoss	CrossEntropyLoss	CrossEntropyLoss
Batch Size	32	32	32	32	32
Epochs	20	20	20	20	20
Activation Function Hidden Layer	ReLU	ReLU	ReLU	ReLU	ReLU
Activation Function Output Layer	Softmax	Softmax	Softmax	Softmax	Softmax
Final Layer	Fully connected (3 classes)	Fully connected (3 classes)	Fully connected (3 classes)	Fully connected (3 classes)	Fully connected (3/4 classes)
Device	CPU	CPU	CPU	CPU	CPU

4.1.2 Performance Metrices

We used test data to quantify the models' performance after training. Here are some of the metrics that were calculated for performance evaluation. We discovered the most accurate model to forecast using these parameters. Using Eqs (8–14), several percentage performance measures have been created depending on the model's given confusion matrix.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ Images} \times 100\% \quad (8)$$

$$True\ Positive\ Rate\ (TPR) = \frac{True\ Positve}{True\ Positive + False\ Negative} \times 100\% \quad (9)$$

$$True\ Negative\ Rate\ (TNR) = \frac{True\ Negative}{False\ Positive + True\ Negative} \times 100\% \quad (10)$$

$$False\ Positive\ Rate\ (FPR) = \frac{False\ Positive}{False\ Positive + True\ Negative} \times 100\% \quad (11)$$

$$False\ Negative\ Rate\ (FNR) = \frac{False\ Negative}{False\ Negative + True\ Positive} \times 100\% \quad (12)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100\% \quad (13)$$

$$F1\ Score = 2\ \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (14)$$

4.1.3 Performance of Xception

Table 4.2 presents the fold-wise confusion matrices for the Xception model across five cross-validation folds. The model consistently demonstrates strong performance in classifying Cataract, Conjunctivitis, and Normal cases. Across all folds, Cataract and Normal classes show high true positive counts with minimal misclassifications, indicating reliable detection. Conjunctivitis also shows solid results, though with slightly higher confusion, particularly with Cataract in fold 3. Table 4.3 presents a detailed fold-wise performance evaluation of the Xception model across three classes: Cataract, Conjunctivitis, and Normal. Among all the folds, Fold 5 demonstrates the most superior and consistent performance. It achieved the highest accuracy across all classes, with 98.84% for Cataract, 98.38% for Conjunctivitis, and 99.07% for Normal. Additionally, Fold 5 recorded the highest F1 Scores—98.38% for Cataract, 97.23% for Conjunctivitis, and 98.68% for Normal—alongside the lowest error rates, indicating a minimal number of misclassifications.

Table 4.2 Fold-wise confusion matrix for Xception model.

Fold	Class	Cataract	Conjunctivitis	Normal
1	Cataract	142	2	2
	Conjunctivitis	4	122	2
	Normal	2	3	153
2	Cataract	159	0	0
	Conjunctivitis	5	135	2
	Normal	2	2	127
	Cataract	121	0	0

3	Conjunctivitis	8	171	2
	Normal	4	4	122
4	Cataract	135	0	3
	Conjunctivitis	1	139	5
	Normal	2	6	141
5	Cataract	152	2	1
	Conjunctivitis	2	123	2
	Normal	0	1	149

Table 4.3 Fold-wise confusion matrix for Xception model.

Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate
1	Cataract	97.69	97.26	2.74	2.10	97.90	95.95	96.60	2.31
	Conjunctivitis	97.45	95.31	4.69	1.64	98.36	96.06	95.69	2.55
	Normal	97.92	96.84	3.16	1.46	98.54	97.45	97.14	2.08
2	Cataract	98.38	100.00	0.00	2.56	97.44	95.78	97.85	1.62
	Conjunctivitis	97.92	95.07	4.93	0.69	99.31	98.54	96.77	2.08
	Normal	98.61	96.95	3.05	0.66	99.34	98.45	97.69	1.39
3	Cataract	97.22	100.00	0.00	3.86	96.14	90.98	95.28	2.78

	Conjunctivitis	96.76	94.48	5.52	1.59	98.41	97.71	96.07	3.24
	Normal	97.69	93.85	6.15	0.66	99.34	98.39	96.06	2.31
4	Cataract	98.61	97.83	2.17	1.02	98.98	97.83	97.83	1.39
	Conjunctivitis	97.22	95.86	4.14	2.09	97.91	95.86	95.86	2.78
	Normal	96.30	94.63	5.37	2.83	97.17	94.63	94.63	3.70
5	Cataract	98.84	98.06	1.94	0.72	99.28	98.70	98.38	1.16
	Conjunctivitis	98.38	96.85	3.15	0.98	99.02	97.62	97.23	1.62
	Normal	99.07	99.33	0.67	1.06	98.94	98.03	98.68	0.93

While Fold 2 also showed strong results—particularly achieving 100% TPR for Cataract and high precision scores—it fell slightly short in balancing the performance across all classes compared to Fold 5. Fold 3, in contrast, exhibited a drop in performance, particularly for the Normal class, where the TPR was 93.85% and the FNR rose to 6.15%, reflecting more misclassifications. Fold 4 showed moderate results but experienced slightly higher FNRs and error rates for the Normal and Conjunctivitis classes. Fold 1 provided satisfactory accuracy levels but was outperformed by the later folds in terms of

F1 Scores and overall balance. Figure 4.1 illustrates the fold-wise training and validation performance of the model. Specifically, Figure 4.1 (a) presents a comparison between training accuracy and validation accuracy across all folds, while Figure 4.1 (b) displays the corresponding training loss and validation loss trends.

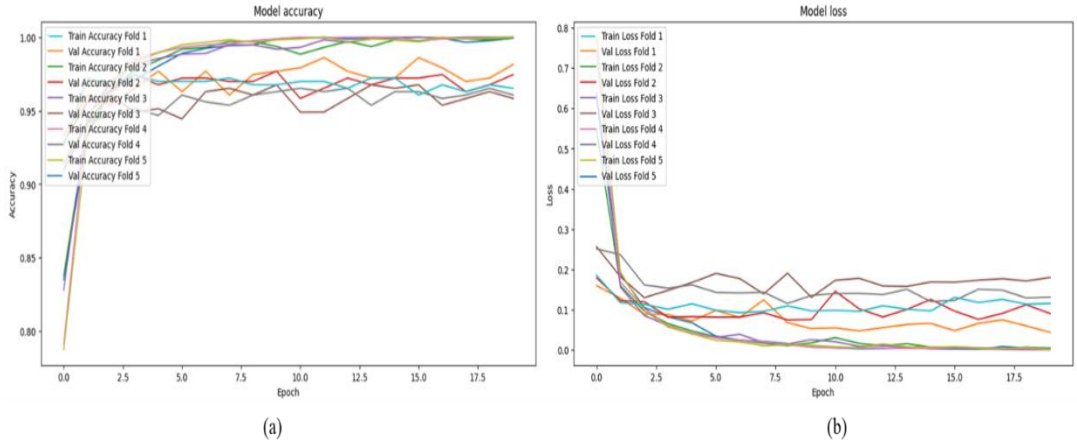


Figure 4.1 The (a) model accuracy, and (b) loss curve of Xception.

4.1.4 Performance of MobileNet-V2

The fold-wise confusion matrix for the MobileNet-V2 model (Table 4.4) reveals consistent performance across all five folds in correctly classifying Cataract, Conjunctivitis, and Normal eye images. Cataract cases were generally well-identified, with minimal misclassifications, particularly in folds 3 and 4 where the model achieved perfect classification (zero misclassifications). Conjunctivitis and Normal cases also demonstrated strong performance, though minor confusion occurred between these two classes, especially in folds 1 and 2. Overall, the model maintained a high level of classification accuracy across folds, indicating strong generalization capability and

robustness of the MobileNet-V2 model. The fold-wise performance metrics of the MobileNet-V2 model demonstrate consistently high accuracy, precision, and F1 scores across all five folds, indicating stable and effective classification performance. Notably, Fold 4 achieved the best overall performance, with class-wise accuracies of 99.07% for Cataract, 99.54% for Conjunctivitis, and 99.54% for Normal. These were accompanied by 100% True Negative Rate (TNR) and 100% precision for both Conjunctivitis and Normal, and 100% True Positive Rate (TPR) for Cataract and Normal. The error rate in this fold was minimal, just 0.46% for both Conjunctivitis and Normal, making it the most optimal among all. Fold 3 also performed exceptionally well, with class-wise accuracies of 99.07% for Cataract, 98.84% for Conjunctivitis, and 99.31% for Normal. The model achieved 100% TPR for Cataract and 100% TNR for Normal, with overall high precision and F1 scores across the board. The error rate was as low as 0.69% for Normal and remained below 1.2% for other classes. The remaining folds (1, 2, and 5) also yielded strong results.

For instance, Fold 1 had class-wise accuracies of 97.92% for Cataract and

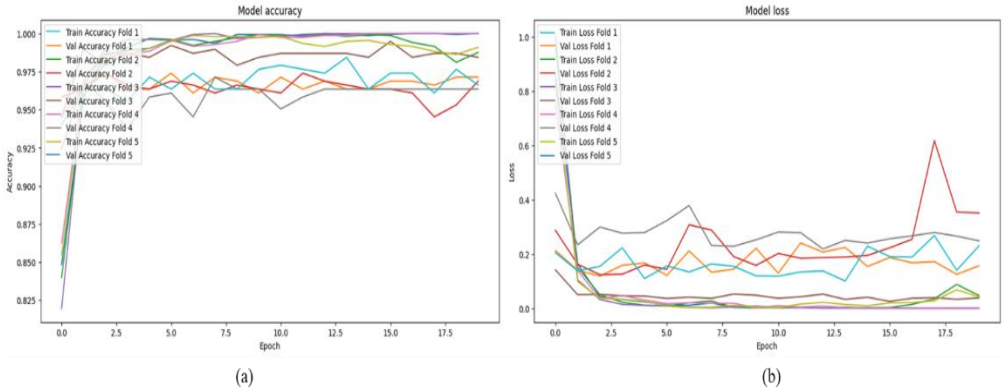


Figure 4.2 The (a) model accuracy, and (b) loss curve of MobileNetV2

Conjunctivitis, and 98.15% for Normal. In Fold 5, the Normal class reached an accuracy of 98.38%, while Cataract and Conjunctivitis achieved 98.15% and 98.38%, respectively. However, these folds exhibited slightly higher error rates and marginally lower F1 scores compared to Folds 3 and 4. Specifically, Figure 4.2 (a) presents a comparison between training accuracy and validation accuracy across all folds, while Figure 4.2 (b) displays the corresponding training loss and validation loss trends.

Table 4.4 Fold-wise confusion matrix for MobileNet-V2 model.

Fold	Class	Cataract	Conjunctivitis	Normal
1	Cataract	131	2	1
	Conjunctivitis	3	118	2
	Normal	3	2	122
2	Cataract	126	4	0
	Conjunctivitis	1	110	2
	Normal	3	4	134
3	Cataract	130	0	0

	Conjunctivitis	3	136	0
	Normal	1	2	112
4	Cataract	133	0	0
	Conjunctivitis	2	118	0
	Normal	2	0	118
5	Cataract	115	1	0
	Conjunctivitis	3	131	2
	Normal	4	1	127

Table 4.5 Fold-wise performance metrices for MobileNet-V2 model.

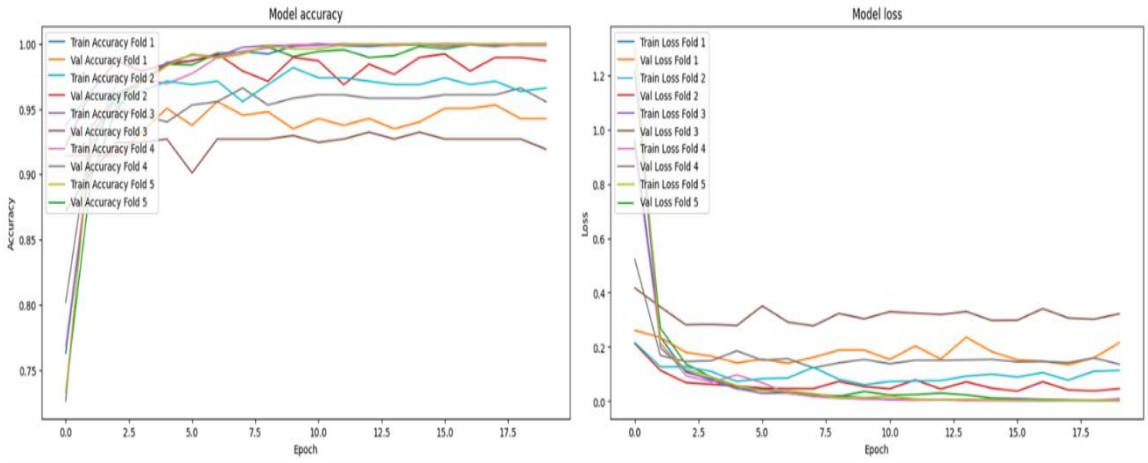
Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate
1	Cataract	97.92	97.76	2.24	2.01	97.99	95.62	96.68	2.08
	Conjunctivitis	97.92	95.93	4.07	1.29	98.71	96.72	96.33	2.08
	Normal	98.15	96.06	3.94	0.98	99.02	97.60	96.83	1.85
2	Cataract	98.15	96.92	3.08	1.32	98.68	96.92	96.92	1.85
	Conjunctivitis	97.45	97.35	2.65	2.51	97.49	93.22	95.24	2.55
	Normal	97.92	95.04	4.96	0.69	99.31	98.53	96.75	2.08
3	Cataract	99.07	100.00	0.00	1.32	98.68	97.01	98.48	0.93
	Conjunctivitis	98.84	97.84	2.16	0.68	99.32	98.55	98.19	1.16

	Normal	99.31	97.39	2.61	0.00	100.00	100.00	98.68	0.69
4	Cataract	99.07	100.00	0.00	1.34	98.66	97.08	98.52	0.93
	Conjunctivitis	99.54	98.33	1.67	0.00	100.00	100.00	99.16	0.46
	Normal	99.54	98.33	1.67	0.00	100.00	100.00	99.16	0.46
5	Cataract	98.15	99.14	0.86	2.22	97.78	94.26	96.64	1.85
	Conjunctivitis	98.38	96.32	3.68	0.68	99.32	98.50	97.40	1.62
	Normal	98.38	96.21	3.79	0.67	99.33	98.45	97.32	1.62

4.1.5 Performance of Inception-V3

The fold-wise confusion matrix of the Inception-V3 model reveals generally strong classification performance across all five folds, with some variation (see Table 4.6). Fold 1 shows accurate predictions for Cataract and Conjunctivitis, though a noticeable number of Normal cases were misclassified as Conjunctivitis. Fold 2 improved overall, especially with perfect Cataract classification, though minor confusion existed between Conjunctivitis and Normal. Fold 3 showed relatively weaker performance, particularly with misclassified Cataract and some overlap between Conjunctivitis and Normal classes. Fold 4 maintained a balanced prediction across all categories with only minimal errors. Fold 5 demonstrated the best performance, achieving near-perfect classification for all three classes, indicating the highest model consistency and accuracy among the folds. The fold-wise performance metrics for the Inception-V3 model indicate strong and consistent classification capabilities, with slight performance variations across the five folds (see Table 4.7). In Fold 1, the model achieved an overall good accuracy, with Cataract and Conjunctivitis classes performing well (98.38% and 95.83%, respectively), though the Normal class accuracy dropped to 95.60% due to higher false negatives. Fold 2 demonstrated improved stability, achieving the highest accuracy for Cataract (98.61%), perfect TPR (100%), and solid performance for the other two classes as well, suggesting this fold handled all classes with high sensitivity and precision. Fold 3 showed a slight decline in overall accuracy across all classes (Cataract: 96.30%, Conjunctivitis: 95.60%, Normal: 93.75%), mainly due to higher false negative and false positive rates, indicating this fold had more classification confusion. In Fold 4, performance recovered with balanced accuracy scores across the classes (Cataract: 97.92%, Conjunctivitis: 96.99%, Normal: 97.22%) and better F1 scores, showcasing the model's improved reliability. However, Fold 5 stands out as the best performing fold, with the highest accuracy scores for all three classes. Figure 4.3 (a) presents a comparison

between training accuracy and validation accuracy across all folds, while Figure 4.3 (b) displays the corresponding training loss and validation loss trends of Inception-V3.



(a) (b)

Figure 4.3 The (a) model accuracy, and (b) loss curve of Inception-V3.

4.1.6 Performance of VGG-19

The fold-wise confusion matrices for the VGG-19 model illustrate a consistently strong classification performance across all five folds (see Table 4.8). In Fold 1, the model perfectly identified all Cataract cases (130/130) and misclassified only a few instances in the Conjunctivitis and Normal classes, indicating high reliability. Fold 2 shows slightly more confusion, particularly in the Conjunctivitis and Normal classes, but Cataract classification remained very strong (124/125 correctly identified). Fold 3 exhibits flawless Cataract recognition (133/133) with minimal confusion in other classes, maintaining high precision. Fold 4 also shows robust performance across all categories, with only minor misclassifications. However, Fold 5 stands out with near-perfect classification, correctly identifying nearly all instances across all classes—Cataract (116/117), Conjunctivitis (122/122), and Normal (145/145), with zero misclassification in Conjunctivitis and Normal classes, and only one instance of Cataract misclassified.

The fold-wise performance metrics of the VGG-19 model demonstrate consistent and high accuracy across all five folds, with overall accuracy values ranging from 98.15% to 98.84% (see Table 4.9). Notably, the True Positive Rate (TPR) remained high for all classes in every fold, particularly for Cataract, which achieved a perfect 100% TPR in Folds 1, 3, and 5. Similarly, Precision and F1 scores for all classes were above 95%,

underscoring the model's balanced performance in terms of both sensitivity and specificity.

Among all folds, Fold 5 slightly outperformed others, achieving near-perfect classification scores across all three classes. Both Conjunctivitis and Normal classes reached an F1 Score of 98.01, supported by high precision and minimal false rates. The Error Rate was consistently low throughout—no higher than 1.85% in any fold. Overall, VGG-19 demonstrated exceptional robustness and generalizability, with Fold 5 slightly edging ahead as the top-performing fold based on its balanced and accurate predictions, minimal error, and consistent high scores across all evaluated metrics. Figure 4.4(a) presents a comparison between training accuracy and validation accuracy across all folds, while Figure 4.4(b) displays the corresponding training loss and validation loss trends of VGG-19.

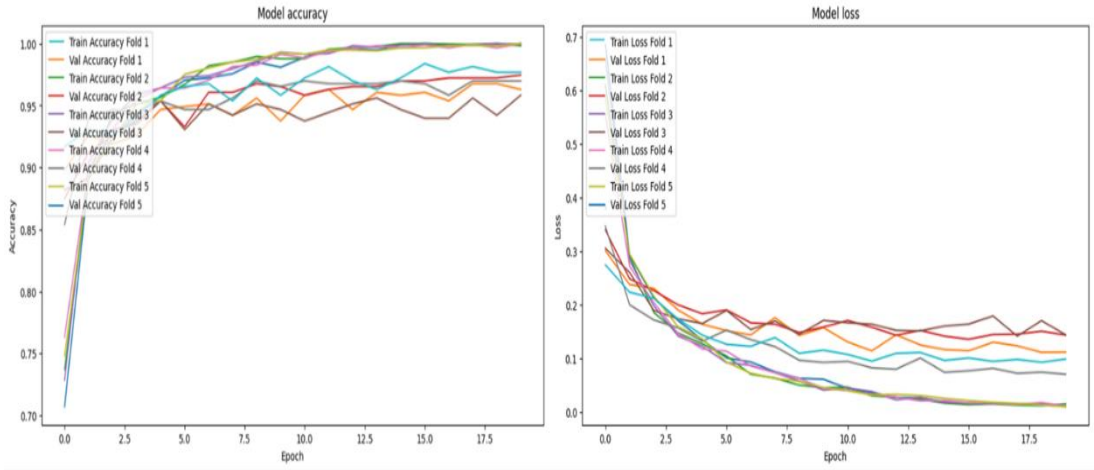


Figure 4.4 The (a) model accuracy, and (b) loss curve of VGG-19

Table 4.6 Fold-wise confusion matrix for Inception-V3 model.

Fold	Class	Cataract	Conjunctivitis	Normal
1	Cataract	118	3	1
	Conjunctivitis	0	119	1
	Normal	3	14	115
2	Cataract	125	0	0
	Conjunctivitis	5	127	6
	Normal	1	1	119
3	Cataract	117	0	6

	Conjunctivitis	4	121	10
	Normal	6	5	115
4	Cataract	128	1	3
	Conjunctivitis	4	112	4
	Normal	1	4	127
5	Cataract	123	0	0
	Conjunctivitis	1	123	0
	Normal	2	2	133

Table 4.7 Fold-wise performance metrices for Inception-V3 model.

Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate
1	Cataract	98.38	96.72	3.28	0.97	99.03	97.52	97.12	1.62
	Conjunctivitis	95.83	99.17	0.83	5.45	94.55	87.50	92.97	4.17
	Normal	95.60	87.12	12.88	0.67	99.33	98.29	92.37	4.40
2	Cataract	98.61	100.00	0.00	1.95	98.05	95.42	97.66	1.39
	Conjunctivitis	97.22	92.03	7.97	0.34	99.66	99.22	95.49	2.78
	Normal	98.15	98.35	1.65	1.93	98.07	95.20	96.75	1.85
3	Cataract	96.30	95.12	4.88	3.24	96.76	92.13	93.60	3.70
	Conjunctivitis	95.60	89.63	10.37	1.68	98.32	96.03	92.72	4.40

	Normal	93.75	91.27	8.73	5.23	94.77	87.79	89.49	6.25
4	Cataract	97.92	96.97	3.03	1.67	98.33	96.24	96.60	2.08
	Conjunctivitis	96.99	93.33	6.67	1.60	98.40	95.73	94.51	3.01
	Normal	97.22	96.21	3.79	2.33	97.67	94.78	95.49	2.78
5	Cataract	99.31	100.00	0.00	0.97	99.03	97.62	98.80	0.69
	Conjunctivitis	99.31	99.19	0.81	0.65	99.35	98.40	98.80	0.69
	Normal	99.07	97.08	2.92	0.00	100.00	100.00	98.52	0.93

Table 4.8 Fold-wise confusion matrix for VGG-19 model.

Fold	Class	Cataract	Conjunctivitis	Normal

1	Cataract	130	0	0
	Conjunctivitis	3	123	1
	Normal	3	1	123
2	Cataract	124	1	0
	Conjunctivitis	2	121	1
	Normal	3	1	131
3	Cataract	133	0	0
	Conjunctivitis	2	127	3
	Normal	4	1	114
4	Cataract	131	2	0

	Conjunctivitis	2	132	0
	Normal	2	3	112
5	Cataract	116	0	1
	Conjunctivitis	0	122	0
	Normal	0	0	145

Table 4.9 Fold-wise performance metrices for VGG-19 model.

Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate
1	Cataract	98.61	100.00	0.00	1.99	98.01	95.59	97.74	1.39

	Conjunctivitis	98.84	96.85	3.15	0.33	99.67	99.19	98.01	1.16
	Normal	98.84	96.85	3.15	0.33	99.67	99.19	98.01	1.16
2	Cataract	98.61	99.20	0.80	1.63	98.37	96.12	97.64	1.39
	Conjunctivitis	98.84	97.58	2.42	0.65	99.35	98.37	97.98	1.16
	Normal	98.84	97.04	2.96	0.34	99.66	99.24	98.13	1.16
3	Cataract	98.61	100.00	0.00	2.01	97.99	95.68	97.79	1.39
	Conjunctivitis	98.61	96.21	3.79	0.33	99.67	99.22	97.69	1.39
	Normal	98.15	95.80	4.20	0.96	99.04	97.44	96.61	1.85
4	Cataract	98.61	98.50	1.50	1.34	98.66	97.04	97.76	1.39
	Conjunctivitis	98.38	98.51	1.49	1.68	98.32	96.35	97.42	1.62

	Normal	98.84	95.73	4.27	0.00	100.00	100.00	97.82	1.16
5	Cataract	98.61	100.00	0.00	1.99	98.01	95.59	97.74	1.39
	Conjunctivitis	98.84	96.85	3.15	0.33	99.67	99.19	98.01	1.16
	Normal	98.84	96.85	3.15	0.33	99.67	99.19	98.01	1.16

4.1.7 Performance of MobViGNet (Proposed)

The fold-wise confusion matrix for the MobViGNet model reveals consistently high classification performance across all five folds (see Table 4.10). In the case of the Cataract class, the model performed exceptionally well, achieving perfect classification in Folds 2, 3, and 4, where all 162 instances were correctly identified without any misclassifications. Even in Folds 1 and 5, only a single misclassification occurred, underscoring the model's robustness. Similarly, for the Conjunctivitis class, the model maintained a high level of precision, with Folds 2, 4, and 5 showing perfect classification (130/130 or 129/129), and only minor errors noted in Folds 1 and 3. The Normal class followed the same trend, with perfect results in Folds 2, 4, and 5, while only one or two instances were misclassified in Folds 1 and 3. Overall, the MobViGNet model demonstrated exceptional consistency and accuracy, with Folds 2, 4, and 5 standing out as the best-performing ones, exhibiting near-perfect or flawless classification across all three classes.

From Table 4.11, the fold-wise performance metrics of the MobViGNet model further affirm its remarkable classification ability, demonstrating consistent and near-perfect results across all folds and classes. In Fold 1, although the performance was already high—with accuracy values hovering around 98.38% to 98.61% across the three classes, the model's true potential is revealed in Folds 2 through 5. Fold 2 exhibited perfect performance for the Cataract class with 100% accuracy, TPR, and precision, while Conjunctivitis and Normal also attained very high metrics (Accuracy: 99.54%, F1 Scores: 99.24 and 99.28 respectively), indicating very minimal misclassification.

Table 4.10 Fold-wise confusion matrix for MobViGNet model.

Fold	Class	Cataract	Conjunctivitis	Normal
1	Cataract	160	1	1
	Conjunctivitis	3	125	2
	Normal	2	1	137
2	Cataract	162	0	0
	Conjunctivitis	0	130	2
	Normal	0	0	138
3	Cataract	161	0	0

	Conjunctivitis	0	129	1
	Normal	0	1	140
4	Cataract	162	0	0
	Conjunctivitis	0	129	1
	Normal	0	0	140
5	Cataract	162	1	0
	Conjunctivitis	0	130	0
	Normal	0	0	140

Table 4.11 Fold-wise performance metrices for MobViGNet model.

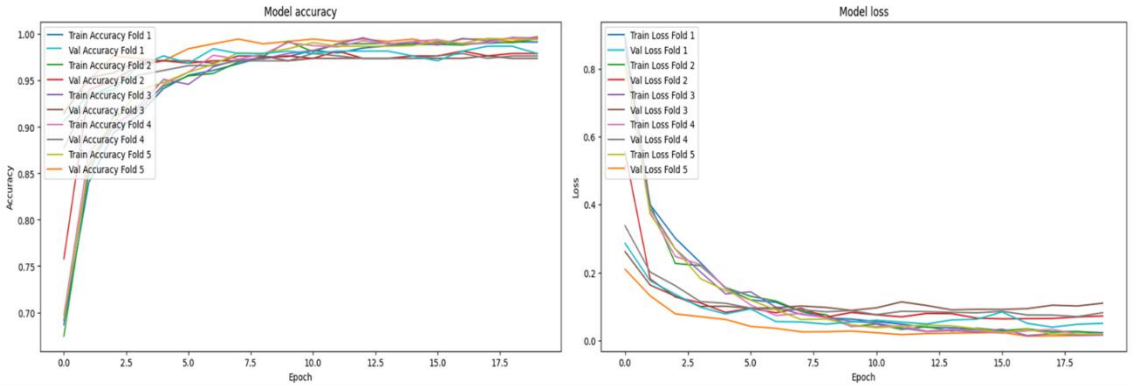
Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate
1	Cataract	98.38	98.77	1.23	1.85	98.15	96.97	97.86	1.62
	Conjunctivitis	98.38	96.15	3.85	0.66	99.34	98.43	97.28	1.62
	Normal	98.61	97.86	2.14	1.03	98.97	97.86	97.86	1.39
2	Cataract	100.00	100.00	0.00	0.00	100.00	100.00	100.00	0.00
	Conjunctivitis	99.54	98.48	1.52	0.00	100.00	100.00	99.24	0.46
	Normal	99.54	100.00	0.00	0.68	99.32	98.57	99.28	0.46
3	Cataract	100.00	100.00	0.00	0.00	100.00	100.00	100.00	0.00
	Conjunctivitis	99.54	99.23	0.77	0.33	99.67	99.23	99.23	0.46

	Normal	99.77	99.29	0.71	0.00	100.00	100.00	99.64	0.23
4	Cataract	100.00	100.00	0.00	0.00	100.00	100.00	100.00	0.00
	Conjunctivitis	99.77	99.23	0.77	0.00	100.00	100.00	99.61	0.23
	Normal	99.77	100.00	0.00	0.34	99.66	99.29	99.64	0.23
5	Cataract	99.77	99.38	0.62	0.00	100.00	100.00	99.69	0.23
	Conjunctivitis	99.77	100.00	0.00	0.33	99.67	99.24	99.62	0.23
	Normal	100.00	100.00	0.00	0.00	100.00	100.00	100.00	0.00

Table 4.12 Overall best performance of applied models (Average).

Model	Best Fold	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	Error Rate

Xecption	4	99.69	99.49	0.51	0.17	99.83	99.43	99.54	0.31
MobileNetV2	5	99.46	99.19	0.81	0.23	99.77	99.18	99.15	0.54
Inception-V3	3	99.39	98.84	1.16	0.28	99.72	99.06	98.94	0.61
VGG-19	5	98.76	97.9	2.1	0.88	99.12	97.99	97.92	1.24
MobViGNet	5	99.85	99.79	0.21	0.11	99.89	99.75	99.77	0.15



(a) (b)

Figure 4.5 The (a) model accuracy, and (b) loss curve of MobViGNet.

From Fold 3 onwards, the model began achieving an even more impressive consistency. In Fold 3, all three classes achieved nearly perfect or perfect values across all metrics, including 100% Accuracy for Cataract and nearly 99.77% for Normal with a flawless TNR of 100%. Fold 4 matched this performance, showing identical 100% Accuracy for Cataract and Normal, with extremely high precision and recall for Conjunctivitis as well. By Fold 5, the model managed to reach perfect scores (100%) in both Accuracy and F1 Score for the Normal class, and nearly perfect metrics for the other two classes, particularly Conjunctivitis which achieved 100% TPR and 99.62% F1 Score. Overall, Folds 3, 4, and 5 stand out as the top-performing folds, with all classes consistently achieving Accuracy and F1 Scores above 99.5% or even 100%, and Error Rates nearing zero. These results suggest that MobViGNet provides exceptional stability, generalizability, and robustness in classifying Cataract, Conjunctivitis, and Normal eye conditions, outperforming traditional models across multiple evaluation metrics. Figure 4.5 (a) presents a comparison between training accuracy and validation accuracy across all folds, while Figure 4.5 (b) displays the corresponding training loss and validation loss trends of MobViGNet.

4.2 Discussions

The comparative analysis of the best-performing folds across all models reveals MobViGNet as the top performer (see Table 4.12). Specifically, Fold 5 of MobViGNet demonstrated outstanding classification capability with an average accuracy of 99.85%,

an exceptionally low error rate of 0.15%, and robust metric scores across the board, including an average TPR of 99.79%, FNR of only 0.21%, FPR of 0.14%, and TNR of 99.86%. It also maintained high values in average precision (99.75%) and F1 score (99.76%), making it the most consistent and reliable model for multiclass classification of eye diseases in this study. Closely following MobViGNet, MobileNet also achieved high performance in Fold 5, with an average accuracy of 99.46%, TPR of 99.47%, precision of 99.24%, and an error rate of just 0.54%. The results suggest that MobileNet, a lightweight architecture, can still compete with more complex models while offering computational efficiency.

The Xception model achieved its best results in Fold 4, recording an impressive average accuracy of 99.62% and an error rate of 0.38%. It maintained excellent class-wise balance, evidenced by an average TPR of 99.52%, FNR of 0.48%, FPR of 0.17%, and a TNR of 99.83%. Its precision and F1 score, 99.43% and 99.47% respectively, position it as a highly dependable model for minimizing both false positives and false negatives. Similarly, InceptionV3 performed well in Fold 3 with an average accuracy of 99.39% and an F1 score of 98.94%. Though slightly behind Xception and MobViGNet in precision (98.93%), its TNR of 99.72% and FPR of just 0.28% affirm its strong specificity. Finally, VGG-19, a classical deep CNN model, showed its highest performance in Fold 5 with an average accuracy of 98.76% and an F1 score of 97.92%. Although its metrics were slightly lower than those of the other models, it still maintained a respectable precision of 97.99% and an error rate of 1.24%, which reaffirms its reliability as a baseline model. In conclusion, MobViGNet stands out as the most efficient model in terms of both accuracy and balanced performance across metrics, followed closely by MobileNet and Xception.

4.2.1 Comparative Study

In the past few years, deep learning has matured the ophthalmic disease diagnosis from external eye images significantly, and numerous studies have proposed models ranging from basic custom CNNs to sophisticated transfer learning and hybrid models. From Table 4.13 we can see comparative literature analysis shows salient differences in classification scope, model architecture, performance, and real-world applicability.

Previous works, such as (Ganokratanaa et al. 2023), (Yadav & Yadav 2023), (Bawa & Koul 2024) , primarily utilized custom CNN structures to discriminate between two or more eye diseases like cataract, conjunctivitis, and anemia-related symptoms. Although the models were found with reasonable accuracy (between 88.80% and 96%), their simplicity limited the performance against more sophisticated newer transfer learning approaches. For instance, (Niloy et al. 2024) and (Albelaihi & Ibrahim 2024) utilized Inception-v3 and EfficientNetB0, respectively, using pre-trained weights on large data sets for better generalization and performance. These models were more accurate, with Albelaihi & Ibrahim (2024) achieving 98.76% for four diseases, including diabetic retinopathy and macular edema, highlighting the strength of transfer learning in multi-class classification. Some studies like Niloy et al. (2024) and Appiahene et al. (2023) emphasized model efficiency and deployment in resource-constrained environments. Niloy et al. (2024) used a MobileNetV2 variant to balance performance and computational cost, achieving 96% accuracy across three external eye disease classes, while Appiahene et al. (2023) uniquely combined CNN with logistic regression and Gaussian blur preprocessing, resulting in an ensemble model suitable for mobile application deployment.

Our works introduces MobViGNet, a specifically developed hybrid model architecture. This model gone through advanced cross-validation techniques to enhance the model’s robustness, generalizability, and resistance to overfitting across various data splits. The present work outperforms all the current literature by achieving a state-of-the-art accuracy of 99.85% for a three-class classification task (Normal, Conjunctivitis, Cataracts). In addition to the performance gain, MobViGNet's key contributions include generatability test by a different dataset for cross-dataset robustness analysis and usage of explainable AI (XAI) techniques such as LIME and Grad-CAM to make visual interpretability key step towards clinical acceptability and trust. The analysis reflects a steep improvement in model efficacy and complexity. Custom CNNs pioneered disease classification, but now transfer learning and hybrid models are supreme for precision as well as usability in practical application.

Table 4.13 Comparative study on eye disease classification using deep learning.

Ref.	Year	Classes	Target Disease(s)	Best Model	Accuracy	Key Contribution
Ganokratanaa et al. (2023)	2023	2	Cataract, Normal	LeNet-CNN	96.00%	Developed a custom CNN model; improved performance in binary classification
Bitto & Mahmud (2022)	2024	3	Normal, Conjunctivitis, Cataracts (External Eye Images)	Inception-v3 (Transfer Learning)	97.08%	Used custom dataset; showed effectiveness of Inception-v3 for external eye image classification
Niloy et al. (2024)	2024	3	=Uveitis, Eyelid Diseases, Healthy (External Eye Images)	MobileNetV2 (EfficientNet variant)	96.00%	Efficient model suited for deployment in low-resource settings
Yadav & Yadav (2023)	2023	4	Cataract (Severity Levels)	Custom CNN	93.10%	Built custom CNN for severity prediction of cataracts
Appiahene et al. (2023)	2023	2	Anemia and Non-eye conditions	CNN + Logistic	92.50%	Developed an ensemble model and integrated it into a mobile application

				Regression + Gaussian Blur		
Bawa & Koul (2024)	2024	2	Conjunctivitis and Normal	Custom CNN	88.80%	Proposed ensemble model for binary conjunctivitis detection
Albelaihi & Ibrahim (2024)	2024	4	Diabetic Retinopathy, Macular Edema, Glaucoma, Cataract	EfficientNetB0	98.76%	Achieved high performance using transfer learning on a multi-class dataset
This Study	2025	3	Normal, Conjunctivitis, Cataracts (External Eye Images)	Hybrid Model (MobViGNet)	99.85%	Custom hybrid model and performance improvement; interpretability through Explainable AI; cross-dataset generatability assessment and performance improvement

4.2.2 Model Generatability

The generalizability of the MobViGNet model was further evaluated using a second dataset, which consisted of images representing multiple common eye diseases, including Cataract, Diabetic Retinopathy, Glaucoma, and Normal classes. The fold-wise confusion matrices in Table 4.14 and 4.15 demonstrate the model's ability to consistently perform across all five folds, maintaining strong prediction capabilities for all four classes. The model's generalizability on Dataset-2 highlights its robustness, with high classification accuracy and strong performance in detecting each disease. For instance, in Fold 1, MobViGNet demonstrated excellent performance in classifying Cataract (145 true positives) with a small number of false negatives and false positives. Diabetic Retinopathy was classified almost perfectly with 171 true positives and minimal errors. Similarly, Glaucoma and Normal classes were identified with high precision, despite some misclassifications, particularly within the Glaucoma class where minor confusion occurred between the Glaucoma and Normal categories (17 false positives and 16 false negatives). The results indicate that MobViGNet maintains robust performance across all five folds. The model's overall accuracy for detecting different eye diseases is high, with Cataract achieving an average accuracy ranging from 95.71% to 96.45%, Diabetic Retinopathy demonstrating near-perfect performance (99.56% to 99.85%), Glaucoma ranging from 91.12% to 94.67%, and Normal category achieving accuracy values between 93.05% and 96.30%. These figures suggest that MobViGNet is highly capable of detecting different eye conditions with minimal misclassification. In terms of performance metrics, Diabetic Retinopathy consistently shows the highest true positive rates (TPR), reaching 100% in several folds, indicating that the model can accurately classify this disease. Cataract also performs strongly with a TPR of over 85% in all folds, while Glaucoma and Normal show slightly lower TPR values, particularly for Glaucoma, which shows a higher false negative rate (FNR) compared to other categories. Overall, MobViGNet demonstrates significant generalizability on the second dataset, showing consistent performance across different eye diseases.

Table 4.14 Fold-wise confusion matrix for MobViGNet on Dataset-2.

Fold	Class	Cataract	Diabetic Retinopathy	Glaucoma	Normal
1	Cataract	145	0	17	3
	Diabetic Retinopathy	0	171	1	1
	Glaucoma	7	0	131	16
	Normal	2	1	12	169
2	Cataract	160	0	14	1
	Diabetic Retinopathy	0	162	0	0

	Glaucoma	12	2	125	21
	Normal	2	0	11	166
3	Cataract	152	0	4	1
	Diabetic Retinopathy	0	172	0	0
	Glaucoma	16	0	143	16
	Normal	3	1	4	164
4	Cataract	156	0	10	6
	Diabetic Retinopathy	0	182	0	2
	Glaucoma	5	0	130	17
	Normal	2	0	5	161

5	Cataract	160	0	14	1
	Diabetic Retinopathy	0	162	0	0
	Glaucoma	12	2	125	21
	Normal	2	0	11	166

Table 4.15 Fold-wise performance metrices for MobViGNet on Dataset-2.

Fold	Class	Accuracy	TPR	FNR	FPR	TNR	Precision	F1 Score	rror Rate
1	Cataract	95.71	87.88	12.12	1.76	98.24	94.16	90.91	4.29
	Diabetic Retinopathy	99.56	98.84	1.16	0.20	99.80	99.42	99.13	0.44

	Glaucoma	92.16	85.06	14.94	5.75	94.25	81.37	83.17	7.84
	Normal	93.05	91.85	8.15	6.50	93.50	84.08	87.79	6.95
2	Cataract	95.71	91.43	8.57	2.79	97.21	91.95	91.69	4.29
	Diabetic Retinopathy	99.70	100.00	0.00	0.39	99.61	98.78	99.39	0.30
	Glaucoma	91.12	78.13	21.88	4.84	95.16	83.33	80.65	8.88
	Normal	94.82	92.74	7.26	4.43	95.57	88.30	90.46	5.18
3	Cataract	96.45	96.82	3.18	3.66	96.34	88.89	92.68	3.55
	Diabetic Retinopathy	99.85	100.00	0.00	0.20	99.80	99.42	99.71	0.15
	Glaucoma	94.08	81.71	18.29	1.60	98.40	94.70	87.73	5.92

	Normal	96.30	95.35	4.65	3.37	96.63	90.61	92.92	3.70
4	Cataract	96.60	90.70	9.30	1.39	98.61	95.71	93.13	3.40
	Diabetic Retinopathy	99.70	98.91	1.09	0.00	100.00	100.00	99.45	0.30
	Glaucoma	94.53	85.53	14.47	2.86	97.14	89.66	87.54	5.47
	Normal	95.27	95.83	4.17	4.92	95.08	86.56	90.96	4.73
5	Cataract	96.45	96.86	3.14	3.68	96.32	89.02	92.77	3.55
	Diabetic Retinopathy	99.85	100.00	0.00	0.21	99.79	99.48	99.74	0.15
	Glaucoma	94.67	82.32	17.68	1.37	98.63	95.07	88.24	5.33
	Normal	96.30	94.48	5.52	3.12	96.88	90.59	92.49	3.70

Table 4.16 Comparative study on Dataset 2 classification using deep learning.

Ref.	Model	Accuracy
Doddi (2022)	EfficientnetB3	92%
	ResNet-18	90%
	VGG19	89%
	InceptionResNetV2	92%
	Resnet26	92%
	Custom CNN	86%
	DensetNet121	89%
	Custom CNN	90%

	PCA + Random Forest	79.59%
This Study	MobViGNet	94%

Comparative evaluation was performed with Dataset 2 within a Kaggle-based classification contest environment to study the performance difference between deep and machine learning architectures shows in Table 4.16. Some of the most established and popular ones were tried, including EfficientNetB3, ResNet-18, VGG19, InceptionResNetV2, ResNet26, DenseNet121, and bespoke CNNs, together with a standard PCA + Random Forest pipeline base classical machine learning solution. Among the selected deep learning models, EfficientNetB3, ResNet26, and InceptionResNetV2 performed well, each having achieved 92% accuracy, demonstrating the strength of deeper or compound-scaled models for visual recognition. However, other models such as ResNet-18, VGG19, and DenseNet121 performed a bit lower, with accuracy at between 89% and 90%, which indicates that while they are robust, they may not have the depth or efficiency of scaling necessary for optimal performance on this dataset. Custom CNN models, although lightweight and domain-specific, struggled to outperform the larger pre-trained architectures, with accuracies of 86% and 90%, reflecting limitations in generalizability and feature representation capacity when compared to transfer learning approaches. The traditional machine learning pipeline with PCA dimensionality reduction and Random Forest classification yielded a significantly lower accuracy of 79.59%, again affirming the dominance of deep learning models in extracting complex hierarchical features from image data.

In contrast, the proposed MobViGNet hybrid model, which yielded an average fold accuracy of 94%, outperforming all other models in this comparative study. MobViGNet not only achieves outstanding accuracy but also fold robustness, demonstrating its stability and versatility across data splits. This comparative study reveals that while traditional CNN architectures provide good baselines, the hybrid MobViGNet model produces a significant improvement in performance, and thus the top-performing model in this competition. Its success points towards the new trend of model fusion and hybrid deep learning models, particularly when high classification accuracy and deployment-readiness in real-world applications are chief objectives.

4.2.3 Model Interpretation with XAI

Explainable AI not only facilitates the interpretability of the model but also enables transparent decision-making, an essential requirement in high-risk domains such as ophthalmology. The diversity of XAI methods used provides robust evidence that the behavior of the model aligns with domain knowledge, reducing the possibility of bias or spurious correlations.

Figure 4.5 shows side-by-side visualization of two popular explainable AI (XAI) techniques, LIME (Local Interpretable Model-Agnostic Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) on MobViGNet's model for eye disease classifier. Every row corresponds to one sample of three different disease classes, illustrating both the original input image as well as the respective model explanations. Figure 4.5 (A) provides the raw eye images for three eye conditions: the top row is a grossly clouded lens typical of cataract, the middle row is a red sclera indicative of conjunctivitis, and the bottom row shows normal corneal eye. These provide a clinical reference point against which areas of interest for the model can be measured. In Figure 4.5 (B) LIME is used to emphasize the locally significant regions responsible for the model's prediction. In cataract, LIME shows the inner opaque region of the lens, confirming that the model is considering the clinically informative feature. In conjunctivitis, LIME highlights the red inflamed sclera as a salient region, as expected from human expert knowledge. In the third case, LIME reveals anomalies close to the corneal boundary, showing its relevance in classification. In Figure 4.5 (C) shows Grad-CAM heatmaps for class-discriminative sites in the convolutional layers of the deep model. In all three cases, Grad-CAM activations are focused and localized. In the first case, especially, lens region in the center shows strong activation, affirming the relevance of lens opacity. In the second case, attention within the model is distributed diffusely over the red sclera, as it must be for conjunctival inflammation detection. For the third case, the majority of attention is focused on the upper-right part of the cornea, perhaps due to structural normal eye beneficial for class separation.

The complementarity between LIME and Grad-CAM in this image presents both fine-grained, pixel-level interpretation by LIME and overall spatial regions learned by the deep neural network by Grad-CAM. Together, they affirm that the model is learning

and focusing on clinically meaningful features, which is highly crucial in developing trust in AI-aided diagnostic systems.

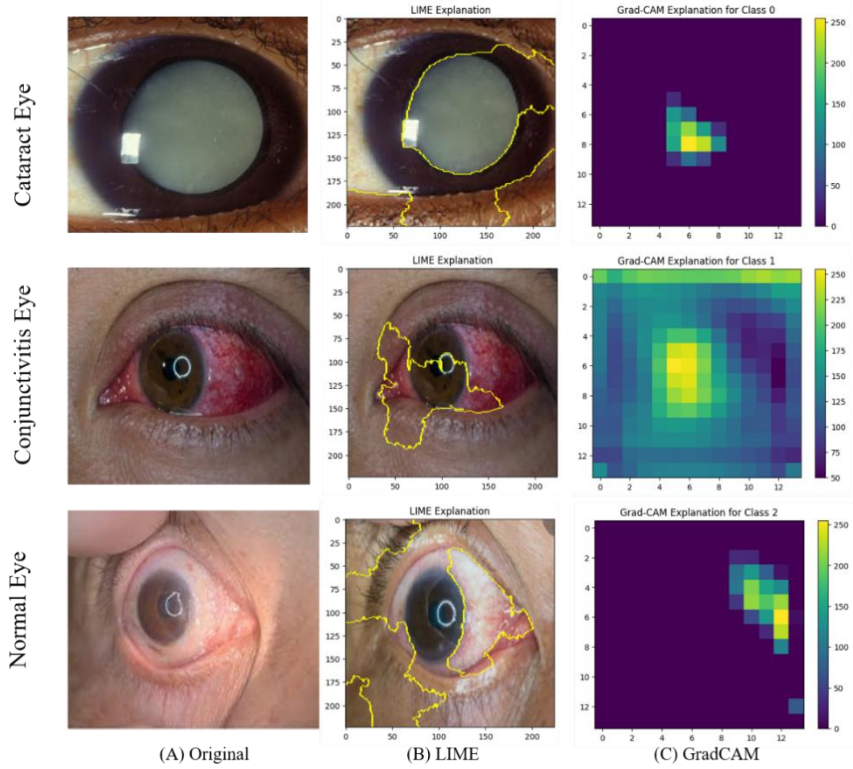


Figure 4.6 Explainable AI Model with Proposed MoViGNet.

4.2.4 Limitation and Future Scope

Although MobViGNet demonstrates high accuracy and generalizability across benchmark and external datasets, there are several limitations that must be addressed in order to expand its reach at the level of real-world clinical applications. To begin with, the model currently detects only three eye conditions from external eye images. However, many other diseases like pterygium, blepharitis, styes, and uveitis also possess visible external features that can be detected by ophthalmologists. The current dataset is not as varied, and in the future, work should be done to obtain real-world, annotated, and balanced datasets with higher diversity in ocular diseases to expand the diagnostic application of the system. The system was trained and evaluated under relatively controlled conditions. Real-world deployment scenarios, especially on mobile devices, will subject variations in lighting, background, image quality, and patient diversity.

Future models will, therefore, need to focus on domain adaptation, real-time inference resilience, and cross-environment and hardware generalization.

Although MobViGNet does include explainability with LIME and Grad-CAM, deeper clinical interpretability—e.g., natural language explanations and expert-conformant feature saliency—is needed to establish practitioner trust. Adding federated learning would also preserve data privacy during collaborative multi-center training. The future directions should focus on developing a mobile-based real-time triage and screening application. This can greatly assist rural and underserved communities with limited access to ophthalmic care. Regulatory approval and incorporation into Electronic Health Record (EHR) systems will be critical steps to translate clinically.

CHAPTER 5

CONCLUSION

5.1 Conclusion

This study presents MobViGNet, a hybrid model with high performance that fuses MobileNetV2 and VGG19 strategically for classifying common eye diseases from external eye images. By leveraging the efficiency of MobileNetV2 and the depth of VGG19, and combining their learned features through a carefully designed concatenated architecture, the model achieves state-of-the-art classification performance with 99.85% accuracy and excellent precision, recall, and F1-scores across folds. Not only does our model outperform single CNN baselines and intricate architectures like InceptionV3 and Xception, but it also shows good generalizability to external datasets, a testament to its strength. Utilization of Explainable AI (XAI) techniques like LIME and Grad-CAM enable decision-making transparency that is crucial for clinical adoption. The lightweight architecture and fast inference time also make MobViGNet an attractive option for deployment in low-resource environments and real-time mobile devices. Although the system currently manages a limited number of eye diseases, it has a scalable and extensible design for broader application. Future work needs to focus on increasing the number of diseases, making the dataset more diverse, enhancing interpretability, and clinically validating the system. This research offers a strong foundation for creating equitable, AI-driven eye health solutions that are accessible, accurate, and clinician-trusted.

REFERENCES

Abbas, Q., Albathan, M., Altameem, A., Almakki, R. S., & Hussain, A. (2023). Deep-Ocular: Improved Transfer Learning Architecture Using Self-Attention and Dense Layers for Recognition of Ocular Diseases. Diagnostics, 13(20), 3165.

Ahmed, H. M., & Hameed, S. R. (2021). Eye Diseases Classification Using Back Propagation with Welch Estimation Based-Learning Rate. Al-Qadisiyah Journal of Pure Science, 26(1), 22-30.

Alam, M., Le, D., Lim, J. I., Chan, R. V., & Yao, X. (2019). Supervised machine learning based multi-task artificial intelligence classification of retinopathies. Journal of clinical medicine, 8(6), 872.

Albelaihi, A., & Ibrahim, D. M. (2024). DeepDiabetic: an identification system of diabetic eye diseases using deep neural networks. IEEE Access, 12, 10769-10789.

Allison, K., Patel, D., & Alabi, O. (2020). Epidemiology of glaucoma: the past, present, and predictions for the future. Cureus, 12(11).

Alsadoun, L., Ali, H., Mushtaq, M. M., Mushtaq, M., Burhanuddin, M., Anwar, R., ... & Ahmed, F. (2024). Artificial Intelligence (AI)-Enhanced Detection of Diabetic Retinopathy From Fundus Images: The Current Landscape and Future Directions. Cureus, 16(8).

Appiahene, P., Arthur, E. J., Korankye, S., Afrifa, S., Asare, J. W., & Donkoh, E. T. (2023). Detection of anemia using conjunctiva images: A smartphone application approach. Medicine in novel technology and devices, 18, 100237.

Bawa, R. K., & Koul, A. (2024). Automated detection of conjunctivitis using convolutional neural network. In Applied Data Science and Smart Systems (pp. 91-97). CRC Press.

Bitto, A. K., & Mahmud, I. (2022). Multi categorical of common eye disease detect using convolutional neural network: a transfer learning approach. Bulletin of Electrical Engineering and Informatics, 11(4), 2378-2387.

Canatan, A. N. (2024). Restoring sight: exploring cataracts as the leading treatable cause of blindness: a narrative review.

Cen, L. P., Ji, J., Lin, J. W., Ju, S. T., Lin, H. J., Li, T. P., ... & Zhang, M. (2021). Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nature communications, 12(1), 4828.

Doddi, G. V. (2022, August 28). Eye_diseases_classification. Kaggle. https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data

Ersavas, T., Smith, M. A., & Mattick, J. S. (2024). Novel applications of Convolutional

Neural Networks in the age of Transformers. Scientific Reports, 14(1), 10000.

Ganokratanaa, T., Ketcham, M., & Pramkeaw, P. (2023). Advancements in cataract detection: The systematic development of LeNet-convolutional neural network models. Journal of Imaging, 9(10), 197.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. jama, 316(22), 2402-2410.

Kang, E. Y. C., Yeung, L., Lee, Y. L., Wu, C. H., Peng, S. Y., Chen, Y. P., ... & Lai, C. C. (2021). A multimodal imaging–based deep learning model for detecting treatment-requiring retinal vascular diseases: model development and validation study. JMIR Medical Informatics, 9(5), e28868.

Le, D., Alam, M., Yao, C. K., Lim, J. I., Hsieh, Y. T., Chan, R. V., ... & Yao, X. (2020). Transfer learning for automated OCTA detection of diabetic retinopathy. Translational Vision Science & Technology, 9(2), 35-35.

Li, Z., Jiang, J., Chen, K., Chen, Q., Zheng, Q., Liu, X., ... & Chen, W. (2021). Preventing corneal blindness caused by keratitis using artificial intelligence. Nature communications, 12(1), 3738.

Madi, I. A. E., Redjdal, A., Bouaud, J., & Seroussi, B. (2024). Exploring Explainable AI Techniques for Text Classification in Healthcare: A Scoping Review. Digital Health and Informatics Innovations for Sustainable Health Care Systems, 846-850.

Niloy, G. M., Bitto, A. K., Biplob, K. B. M. B., Sammak, M. H., Das, A., & Hridoy, G. G. (2024, March). MobileNet-Eye: An Efficient Transfer Learning for Eye Disease Classification. In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) (pp. 01-05). IEEE.

ÖZALP, U., & SEVLİ, O. Classification of Classical Music Composers Using Explainable Artificial Intelligence. Available at SSRN 5044239.

Patel, A. U., Gu, Q., Esper, R., Maeser, D., & Maeser, N. (2024). The crucial role of interdisciplinary conferences in advancing explainable AI in healthcare. BioMedInformatics, 4(2), 1363-1383.

Rahmati, M. (2025). Federated Learning and Explainable AI for Personalized Healthcare in Resource-Limited Settings.

Safi, H., Safi, S., Hafezi-Moghadam, A., & Ahmadieh, H. (2018). Early detection of diabetic retinopathy. Survey of ophthalmology, 63(5), 601-608.

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., & Wang, K. (2021). Image preprocessing in classification and identification of diabetic eye diseases. Data Science and Engineering, 6(4), 455-471.

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. Evolutionary intelligence, 15(1), 1-22.

Satish, R. T., Sayali, B., Gajanan, A. T., & Kirti, S. (2024). A SYSTEMATIC REVIEW OF CONJUNCTIVITIS.

Shamsan, A., Senan, E. M., & Shatnawi, H. S. A. (2023). Automatic classification of colour fundus images for prediction eye disease types based on hybrid features. Diagnostics, 13(10), 1706.

SV, K., OS, Z., & NV, P. (2025). A current view of the issue of rehabilitation of patients with ocular pathology. Journal of Ophthalmology (Ukraine)/Oftal′mologičeskij Žurnal, (1).

Thakur, R. (2024). Explainable AI: developing interpretable deep learning models for medical diagnosis. Int J Multidisciplinary Res, 6.

Umer, M. J., Sharif, M., Raza, M., & Kadry, S. (2023). A deep feature fusion and selection-based retinal eye disease detection from oct images. Expert Systems, 40(6), e13232.

World Health Organization. (2008). Diagnosis and treatment (Vol. 4). World Health Organization.

Yadav, S., & Yadav, J. K. P. S. (2023). Automatic cataract severity detection and grading using deep learning. Journal of Sensors, 2023(1), 2973836.

Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. Journal of healthcare informatics research, 6(2), 228-239.

Yao, X., Alam, M. N., Le, D., & Toslak, D. (2020). Quantitative optical coherence tomography angiography: a review. Experimental biology and medicine, 245(4), 301-312.

Zhang, Y., Weng, Y., & Lund, J. (2022). Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics, 12(2), 237.

Zhao, J., Lu, Y., Zhu, S., Li, K., Jiang, Q., & Yang, W. (2022). Systematic bibliometric and visualized analysis of research hotspots and trends on the application of artificial intelligence in ophthalmic disease diagnosis. Frontiers in Pharmacology, 13, 930520.

Plagiarism Report

241-44-008

ORIGINALITY REPORT

19%	14%	14%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	thesis.eur.nl Internet Source	1%
4	dokumen.pub Internet Source	1%
5	www.frontiersin.org Internet Source	<1%
6	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1%

Account Clearance