

Fusion MobileNetV2 for OCT Disease Detection with Grad-CAM and LLM Interpretation

BY

SAJIB BIN MAMUN
ID: 242-25-033

This Report Presented in Partial Fulfillment of the Requirements for
The Degree of Masters of Science in Computer Science and Engineering

Supervised By

Dr. Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

Dr. Abdus Sattar
Associate Professor
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH

APPROVAL

This Project/Thesis titled “Fusion MobileNetV2 for OCT Disease Detection with Grad-CAM and LLM Interpretation”, submitted by Sajib Bin Mamun, ID No: 242-25-033 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13-09-2025.



BOARD OF EXAMINERS


Chairman

Dr. S.M Aminul Haque
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



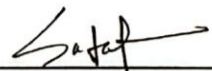
Internal Examiner

Ms. Nazmun Nessa Moon
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Dr. Md Alamgir Kabir
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



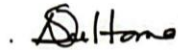
External Examiner

Mr. Sadat Hossain
Data Scientist
Risk Management Division
BRAC Bank Limited

DECLARATION

I hereby declare that this research has been done by me under the supervision of **Dr. Naznin Sultana, Associate Professor, Department of CSE, Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



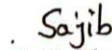
Dr. Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Dr. Abdus Sattar
Associate Professor
Department of CSE
Daffodil International University

Submitted by:



Sajib Bin Mamun
ID: 242-25-033
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartfelt thanks and gratitude to Almighty Allah for His divine blessing, which makes it possible to complete the final year project/internship successfully.

I am grateful and wish to express my profound indebtedness to **Dr. Naznin Sultana, Associate Professor**, Department of CSE, Daffodil International University, Dhaka, deep knowledge & keen interest in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartfelt gratitude to **Dr. Sheak Rashed Haider Noori, Head** of the Department of CSE, for his kind assistance in completing our project, as well as to the other faculty members and staff of the CSE department at Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Early and accurate detection of retinal disease is highly important in order to avoid losing sight as well as in the treatment of patients. This study focused to integrating deep learning models with large language models (LLMs) to improve the classification of retinal diseases and diagnostic interpretation. The OCTDL dataset was preprocessed through augmentation, noise removal, sharpening, contrast enhancement, and outlier removal to improve image quality and model generalization. Four state-of-the-art convolutional neural networks (CNNs), Such as MobileNetV2, ResNet50, VGG16, and DenseNet121, were evaluated. Among them, MobileNetV2 outperformed others. To improve performance, we proposed a Fusion MobileNetV2 model, which combines global and local feature extraction using a fusion mechanism. Among 10 folds, the model achieved 100 % training accuracy, 97.27% validation, and 99.57% testing accuracy on 5-fold. We utilize Grad-CAM to visualize the model predictions. The Grad-CAM outputs with predicted classes were validated by GPT-4o, while GPT-4 responded to user questions from its knowledge base using a retrieval-augmented generation (RAG) pipeline. Finally, the overall framework was deployed as a web application, providing an accessible tool for assisting in early retinal disease diagnosis.

Keywords: Retinal disease, OCTDL, Fusion MobileNetV2, Grad-CAM, LLM, RAG.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	1-4
1.1 Introduction	1
1.2 Motivation	2
1.3 Research Objectives	2
1.4 Research Questions	3
1.5 Expected Output	3
1.6 Project Management and Finance	4
1.7 Report Layout	4
CHAPTER 2: BACKGROUND	5-13
2.1 Preliminaries/Terminologies	5
2.2 Related Works	5
2.3 Research Gap	13
2.4 Challenges	13
CHAPTER 3: RESEARCH METHODOLOGY	14-29
3.1 Proposed Methodology	14
3.2 Data Collection Procedure	15
3.3 Image pre-processing	17
3.3.1 Data Augmentation	17
3.3.2 Data Enhancement	19
3.4 Model selection	21
3.5 Grad-CAM (Gradient-Weighted Class Activation Mapping)	28
3.6 RAG (Retrieval Augmented Generation)	28
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	30-48
4.1 Results Evolution Methods	30

4.2 Experimental Results & Analysis	31
4.3 Model Performance Comparison	44
4.4 Grad-CAM Analysis	44
4.5 LLM Interpretation	45
4.6 Discussion	47
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	49-50
5.1 Impact on Society	49
5.2 Impact on Environment	49
5.3 Ethical Aspects	50
5.4 Sustainability Plan	50
CHAPTER 6: CONCLUSION AND FUTURE WORK	51-31
6.1 Summary of the Study	51
6.2 Conclusions	51
6.3 Implication for Further Study	52
REFERENCES	53-56

LIST OF FIGURES

FIGURES	PAGE NO
Fig 3.1: The working process of overall studies	14
Fig. 3.2: Augmented data of OCTDL Dataset	18
Fig. 3.3: Enhanced Images of OCTDL Dataset	21
Fig. 3.4: Architecture of MobileNetV2	22
Fig. 3.5: Architecture of ReseNet50	23
Fig. 3.6: Architecture of VGG16	24
Fig. 3.7: Architecture of DenseNet121	25
Fig. 3.8: Architecture of proposed model (Fusion MobileNetV2)	27
Fig. 3.9: How Grad-CAM highlights predicted class regions	28
Fig. 3.10: RAG working process	29
Fig. 4.1: Confusion Matrix (MobileNetV2)	34
Fig. 4.2: Confusion Matrix (VGG16)	34
Fig. 4.3: Confusion Matrix (ResNet50)	34
Fig. 4.4: Confusion Matrix (DenseNet-121)	34
Fig. 4.5: Training and Validation accuracy and loss curve of MobileNetV2	35
Fig. 4.6: Training and Validation accuracy and loss curve of VGG16	35
Fig. 4.7: Training and Validation accuracy and loss curve of ResNet50	36
Fig. 4.8: Training and Validation accuracy and loss curve of DenseNet-121	36
Fig. 4.9: Confusion Matrix Fusion MobileNetV2	39
Fig. 4.10: Training and Validation accuracy and Loss curve (Fusion MobileNetV2)	39
Fig. 4.11: Training and validation accuracy across 10 folds (Fusion MobileNetV2)	40
Fig. 4.12: Training Loss Across Folds (Fusion MobileNetV2)	41
Fig. 4.13: Validation Loss Across Folds (Fusion MobileNetV2)	41
Fig. 4.14: Confusion Matrix of (5-folds Fusion MobileNetV2)	43
Fig. 4.15: Grad-CAM analysis of test images	45
Fig 4.16. RetinoAID Web APP (gpt-4o) validation	46
Fig 4.17: RetinoAID Web APP (gpt-4) user question answering	46

LIST OF TABLES

TABLES	PAGE NO
Table 2.1: Research matrix	9
Table 3.1: The details representation of OCTDL dataset	16
Table 4.1: Classification reports of four CNN models	32
Table 4.2: Performance Comparison of four CNN Models	37
Table 4.3: The classification report of Fusion MobileNetV2	38
Table 4.4: K-Fold Cross Validation Results (Fusion MobileNetV2)	41
Table 4.5: Classification report of Fold 5 (Fusion MobileNetV2)	42
Table 4.6: All model performance comparison	44

CHAPTER 1

INTRODUCTION

1.1 Introduction

Retinal diseases, including diabetic retinopathy, age-related macular degeneration, and glaucoma, are leading causes of vision impairment globally [1]. Optical Coherence Tomography (OCT) has become a useful imaging modality in the evaluation of retinal conditions, offering high-resolution cross-sectional images of the retina. Nevertheless, manual analysis of OCT images is time-consuming and it is necessary to emphasize the necessity of automated diagnostic systems [2].

Convolutional Neural Networks (CNNs) and deep learning in general have shown great promise in medical image analysis because it can automatically extract hierarchical features of images. Recent studies have applied various CNN architectures to retinal disease classification tasks [3]. For instance, MobileNetV2 has been utilized for retinal disease detection, achieving high accuracy in classifying OCT images [4]. Similarly, ResNet50, VGG16, and DenseNet121 have been employed for retinal OCT image classification. Each of these models showing different levels of accuracy and effectiveness in OCT image classification [5].

Although these models perform well, their internal structure is complex to understand. These makes the decision-making processes are difficult. Healthcare professionals need clear explanations to trust and use the model predictions. To address this, visualization techniques like Grad-CAM have been applied to CNNs, enhancing the interpretability of model predictions by highlighting important regions in the input images [6]. Furthermore, large language models (LLMs) such as GPT-4, GPT-4o have shown promise in medical image interpretation. Integrating LLMs with deep learning models can provide a comprehensive diagnostic system that not only classifies retinal diseases but also explains the reasoning behind its predictions [7].

The purpose of the present research is to further improve the classification and diagnosis of early retinal disease, with the help of deep learning models and LLMs. Firstly, we introduce the Fusion MobileNetV2 models that combines the global and local feature extraction processes. Secondly, we combine the Fusion MobileNetV2 and LLMs to justify and confirm diagnostic results. The proposed detection process is implemented in a web application and it presents an interactive tool to aid in early detection of retinal diseases.

1.2 Motivation

Despite the high accuracy of deep learning models in the classification of OCT images, the models have a number of challenges. The majority of models need extensive volumes of labeled data, which is challenging and expensive to acquire in the medical field. Moreover, the current models tend to be sensitive to noisy or low-quality images and cannot be generalized to other groups of patients [8]. These constraints lead to the necessity of better models that are robust and reliable. Moreover, Clinicians tend to be reluctant to embrace systems that give predictions without clear explanations. Large language models (LLMs) provide a chance to fill this gap by justifying and interpreting model outputs in natural language. Lastly, the combination of deep learning and LLMs into a convenient web application opens the prospect of scalable diagnostic support tools [9].

1.3 Research Objectives

The primary goal of the research is to enhance the early retinal disease classification and diagnosis through the combination of deep learning models and large language models (LLMs). Firstly, the study will preprocess the OCTDL data with augmentation, noise elimination, sharpening, contrast, and outlier to improve image quality and strength. Secondly, the paper will compare four state-of-the-art CNN models, including MobileNetV2, ResNet50, VGG16, and DenseNet121, to find the most effective baseline model. Third, the research will propose to develop a Fusion MobileNetV2 architecture

through the integration of a global and local feature extraction process that will be more effective in capturing the intricate retinal patterns. Fourthly, the efficiency of the Fusion MobileNetV2 will be evaluated with the help of k-fold cross-validation and compared with the original MobileNetV2. Fifthly, Grad-CAM will be used to give visual interpretability of the model predictions. Sixth, CNN predictions will be integrated with LLMs, with GPT-4o validating predictions and GPT-4 generating user-facing diagnostic explanations via a retrieval-augmented generation (RAG) pipeline. Lastly, the research will implement the general framework as a web-based application to offer an interactive and convenient tool to assist clinicians in the early diagnosis of retinal diseases.

1.4 Research Questions

- RQ1: How effective are the state-of-the-art CNN models (MobileNetV2, ResNet50, VGG16, and DenseNet121) in classifying retinal diseases from the OCTDL dataset?
- RQ2: Can a Fusion MobileNetV2 model improve classification accuracy compared to the original MobileNetV2?
- RQ3: How can Grad-CAM visualizations and LLM validation (enhance the interpretability of retinal disease predictions)?
- RQ4: Is it feasible to deploy the integrated deep learning and LLM framework as a web application to provide interactive support for clinicians in early retinal disease diagnosis?

1.5 Expected Output

- Evaluation of state-of-the-art CNN models (MobileNetV2, ResNet50, VGG16, DenseNet121) to determine their effectiveness in classifying retinal diseases from OCTDL images.
- Development of a Fusion MobileNetV2 model to improve classification accuracy and capture complex retinal patterns compared to the original MobileNetV2.

- Integration of Grad-CAM visualizations and LLM-based validation (GPT-4o and GPT-4) to provide reliable diagnostic explanations.
- Deployment of a web-based framework combining deep learning and LLMs to assist clinicians in early retinal disease detection and provide an interactive decision-support tool.

1.6 Project Management and Finance

This research did not receive any funding from individuals or organizations. The research was conducted independently by the author.

1.7 Report Layout

Chapter 1 is the introduction, which contains the motivation, research objectives, research questions, expected outcomes, and project management. Chapter 2 presents the background and preliminaries, a literature review in detail, and the research matrix, with the research gaps and challenges being pointed out. Chapter 3 explains the proposed methodology in detail, the model architecture and its components. Chapter 4 presents and interprets the experimental findings and their conclusions. Chapter 5 focuses on the sustainability plan, societal and environmental impacts, and ethical considerations. Lastly, Chapter 6 gives the conclusion of the study, its limitations, and future research directions.

CHAPTER 2

BACKGROUND

2.1 Preliminaries

Diagnosis of retinal diseases is essential, and automated. Early detection of this retinal disease is significant as it will lessen the likelihood of the disease progression. The effectiveness of detection of retinal disease has already been proven by recent developments in computer-aided diagnosis. Specifically, computer vision and image processing techniques have worked well in identifying these diseases. Nonetheless, numerous researches have been conducted in this field. This consideration renders it significant in medical research as well as clinical practice. These current methods still require more research and experimentation to enhance them. It is possible to improve the state of things by discovering the constraints and deficiencies of the earlier work. The chapter will contain the literature review, identify the gaps in the research, as well as address the main difficulties in this area.

2.2 Related works

Şahin et al. [10] presented a fine-tuned convolutional neural network (CNN) model for classifying retinal disorders from Optical Coherence Tomography (OCT) images. The authors optimized the convolution, max pooling, dropout, and ReLU layers to detect retinal diseases while minimizing overfitting. The model's performance was evaluated using 5-fold cross-validation, achieving a classification accuracy of 94%. Yang et al. [11] proposed a lightweight MobileNetV2 Retinopathy Classification (MobileNetV2-RC) to classify retinopathy on the basis of OCT images. The paper has included spatial and channel reconstruction convolution, and an advanced CBAM attention mechanism to minimize redundancy of features and maximize feature representation, which improves the classification performance. Their model was classified with an accuracy of 98.96.

To diagnose hypertensive retinopathy (HR) based on fundus images, Sajid et al. [12] came up with a model that combined a pretrained MobileNet architecture with dense blocks. Data augmentation was also used by the authors to address the weakness of a small dataset. Their method had a 99 percent accuracy. Ogundokun et al. [13] proposed the MobileNetV2-SVM model to achieve effective classification of ocular diseases on mobile devices. In order to carry out classification, the authors quantized a pretrained MobileNetV2 model using Support Vector Machine (SVM). The proposed model had a general accuracy of 92.59. Islam et al. [14] suggested an optimized version of DenseNet-201-ANN to classify diabetic retinopathy. The conv5 block32 layer of DenseNet-201 was used to extract features and fed them into an Artificial Neural Network (ANN) that was trained using scaled conjugate gradient backpropagation. The accuracy of the proposed method was high at 98.6%.

Sunija et al. [15] proposed a deep neural network-based classifier to identify retinal diseases (DME, drusen and CNV) using OCT images. The experiment used a CNN that had six convolutional blocks, including downsampling, weight sharing, and class activation mapping to explain the results. The accuracy of the proposed model was 99.69%. Awais et al. [16] suggested a way of classifying normal and abnormal OCT images with a pre-trained CNN model. The paper used VGG16 architecture to obtain features at various layers and carried out classification using several classifiers. The suggested model was accurate 87.5 percent with a sensitivity of 93.5 and specificity of 81.

Rasti et al. [17] created a multi-scale convolutional mixture of experts (MCME) ensemble model to classify OCT diseases. The model employed a novel cost function that allowed learning of features quickly and discriminatively. It included concomitant interactions between expert modules with correlated multivariate components. The four-scale-dependent experts MCME model had the average precision (AP) of 98.86 and an AUC of 99.85. Gour et al. [18] developed a lightweight convolutional neural network (CNN) with class weight balancing (CWB) to classify OCT images. The authors trained the CNN with the CWB method, which punishes the classes with more samples. The suggested approach

reached the accuracy of 99.17 on the ZhangLab dataset and 98.46 on the Duke dataset. Das et al. [19] proposed Deep Multi-scale Fusion CNN (DMF-CNN) to learn OCT disease features at a variety of scales. This method makes use of several CNNs having different receptive fields to isolate scale-specific features, which are subsequently joined to create cross-scale discriminative information. A multi-loss training approach is used jointly to learn individual scale and combined scale characteristics. The designed DMF-CNN model obtained 96.03% and 99.60% accuracy on the UCSD and NEH datasets, respectively. Das et al. [20] in another study designed a multi-scale deep feature fusion (MDFF) convolutional neural network (CNN) to classify macular pathologies. The methodology derives feature at different scales to obtain complementary and discriminating information. A cost-sensitive loss function was used to manage the imbalance in the dataset of classes during training. This technique had a mean sensitivity of 99.6, specificity of 99.87 and the overall accuracy of 99.6. Chan et al. [21] designed a deep learning model to identify Diabetic Macular Edema (DME) at its early stages with Spectral Domain Optical Coherence Tomography (SD-OCT) images. BM3D filtering was used to remove noise in the images and the images were cropped to concentrate on the region of interest. AlexNet was then used to extract features and an SVM was used to classify. Their method was accurate more than 96 percent. Yang et al. [22] proposed an explainable ensemble method that relies on transfer learning to identify retinal diseases. The experiment compared Individual CNNs using and without pre-trained ImageNet weights, and combined them through majority soft polling. Grad-CAM and CAM were used to perform feature visualization to highlight lesion areas. The ensemble method suggested had a 100 % accuracy.

In order to categorize retinal diseases, Adel et al. [23] created a deep learning model based on two transfer learning models, Xception and InceptionV3. The framework used categorical hinge loss, which served as SVM instances, to enhance the performance of classification. Patient-level 10-fold cross-validation experiments were conducted. The suggested method was able to reach the overall accuracy of 98 % with Xception and 93 percent with InceptionV3. To achieve the correct retinal disease detection, Subramanian et

al. [25] investigated the different CNN architectures, including VGG16, VGG19, DenseNet201, and InceptionV3. The research used preprocessing methods like histogram equalization, image enhancement and augmentation on the retinal images. These CNN architectures were then used to train the processed data. The VGG16 model was the best among them with a classification accuracy of 97%. Tayal et al. [26] suggested a deep learning model that is used to classify four eye diseases. The framework involved the preprocessing of images like noise removal, contrast enhancement, and retinal layer extraction to enhance the quality of the images. The images were analyzed using three CNN models with five, seven and nine layers to extract the relevant features. The proposed approach had 96.5 % accuracy, 96.0 % sensitivity and 98.6 % specificity.

A CNN-based stacking ensemble learning (EL) retinal disease detection method, suggested by Kayadibi et al. [27]. OCT images were classified by extracting features with fine-tuned AlexNet (FT-CNN). Two publicly available Duke and UCSD OCT datasets were subjected to both homogeneous and heterogeneous ensemble methods. The heterogeneous stacking ensemble model had a classification accuracy of more than 99.6%. Abirami et al. [28] used DenseNet and VGG16 deep learning models to diagnose Choroidal Neovascularization (CNV). The authors optimized the hyperparameters of the model and processed images, and segmented them with the Python OpenCV library. VGG16 was more successful than DenseNet, and its accuracy was near 97.53. Kim et al. [29] proposed an ensemble model based on three ResNet152 networks to classify retinal diseases correctly. Image classification was done using several CNNs, where Fully Convolutional Networks (FCN) were employed to eliminate noise and a projection approach was employed to repair tilted retinal layers. The ensemble learning approach further enhanced performance. The three-ResNet152 ensemble was found to be more accurate (98.9) than the single models. Table 2.1 shows the general research grid of the literature.

Table 2.1: Research matrix

Authors	Dataset	Model	Results	Contribution
Şahin et al. [10]	OCT [30]	Custom CNN	94% (Accuracy)	Optimized the CNN layers to minimize overfitting and accurately classify retinal diseases.
Yang et al. [11]	OCT	MobileNetV2-RC	98.96% (Accuracy)	A lightweight model that integrates spatial and channel reconstruction convolution, and an enhanced CBAM attention mechanism.
Sajid et al. [12]	Retinal Fundus Image	Mobilenet Hypertensive Retinopathy (Mobile-HR)	99% (Accuracy)	Introduced the Mobile-HR model by utilizing MobileNet with dense blocks for hypertensive retinopathy diagnosis.
Ogundokun et al. [13]	ODIR5K	MobileNetV2-SVM	92.59% (Accuracy)	Combined the MobileNetV2 model with a Support Vector Machine (SVM) to improve retinal disease classification.

Islam et al. [14]	OCT	DenseNet-201-ANN	98.6% (Accuracy)	Utilized an optimized DenseNet-201 as a feature extractor, combined with an ANN.
Sunija et al. [15]	OCT	CNN	99.69% (Accuracy)	An efficient CNN-based OCT classifier with fewer parameters
Awais et al. [16]	SD-OCT	VGG16	87.5% (Accuracy), 93.5% (Sensitivity) , and 81% (Specificity)	The VGG16 model was used to extract features to improve the classification performance.
Rasti et al. [17]	OCT	MCME	98.86% (AP), 99.85% (AUC)	Created a multi-scale convolutional mixture of experts (MCME) ensemble model to classify accurately.
Gour et al. [18]	ZhangLab dataset, Duke dataset	CNN	99.17% (accuracy, ZhangLab), and 98.46% (accuracy,	Integrated a CNN with class weight balancing (CWB) to successfully address class imbalance in OCT image classification.

Das et al. [19]	UCSD dataset, NEH dataset	DMF-CNN	96.03% (accuracy, UCSD), and 99.60% (accuracy, NEH)	Created a Deep Multi-scale Fusion CNN (DMF-CNN) that is effective at capturing and combining multi-scale OCT features.
Das et al. [20]	OCT	MDFFF-CNN	99.6% (sensitivity), 99.87% (specificity), and 99.6% (accuracy)	Trained a multi-scale deep feature fusion CNN to classify macular pathologies accurately.
Chan et al. [21]	SD-OCT	AlexNet-SVM	96% (accuracy)	Utilize AlexNet and SVM for accurate early detection of Diabetic Macular Edema
Yang et al. [22]	OCT	Ensemble Model	100% (accuracy)	Proposed an ensemble model for accurately classifying OCT disease.
Adel et al. [23]	OCT	Xception, InceptionV3	98% (accuracy, Xception), and 93% (accuracy, InceptionV3)	Developed a deep learning framework using Xception and InceptionV3 models.

Bhowmik et al. [24]	OCT	InceptionV3	94% (accuracy)	Utilize the InceptionV3 model for accurately classifying OCT disease.
Subramanian et al. [25]	OCT	VGG16	97% (accuracy)	Presented a VGG16 model and preprocessing techniques to improve retinal disease classification
Tayal et al. [26]	OCT	CNN	96.5% (accuracy), 96.0% (sensitivity) , and 98.6% (specificity)	Developed a deep learning framework with multi-layer CNNs.
Kayadibi et al. [27]	Duke and UCSD dataset	FT-CNN	99.6% (accuracy)	Developed a CNN-based stacking ensemble learning system using fine-tuned AlexNet.
Abirami et al. [28]	OCT	VGG16	97.53% (accuracy)	Optimized DenseNet and VGG16 models with image segmentation.
Kim et al. [29]	OCT	three-ResNet152	98.9% (accuracy)	Developed three-ResNet152 ensemble model

				to accurately classify retinal diseases.
--	--	--	--	--

2.3 Research Gap

Several studies have been conducted to detect retinal disease from OCT images. However, there are still some significant gaps in the existing research. While most of the studies primarily focused only on model accuracy, few studies have explored integrating these models with real-time applications for real-world clinical scenarios. These proposed models also required high computational resources to provide optimal performance. Additionally, there has been limited work on integrating deep learning models with large language models (LLMs) to enhance the diagnosis support. Though LLMs have the potential to provide better diagnosis suggestions.

In this study, we aimed to address these gaps and optimize a CNN-based model, especially MobileNetv2. Furthermore, we integrate the optimized model with an LLM model to validate its output and provide detailed disease diagnosis interpretations.

2.4 Challenges

To conduct this study, we faced several challenges. Firstly, one of the major challenges was obtaining primary data. Although secondary data are available on Kaggle and Mendeley. However, they are highly imbalanced; some classes have few samples, and the data quality was poor. Thirdly, we faced limitations in computational resources. Deep learning and image classification tasks require high-performance GPUs and a well configured system.

Moreover, we faced another challenge in integrating deep learning models with LLMs. There were version mismatches and compatibility issues with TensorFlow and other Python frameworks.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Proposed Methodology

The figure 3.1 represents the overall working process of this study to accurately detect and diagnosis retinal disease.

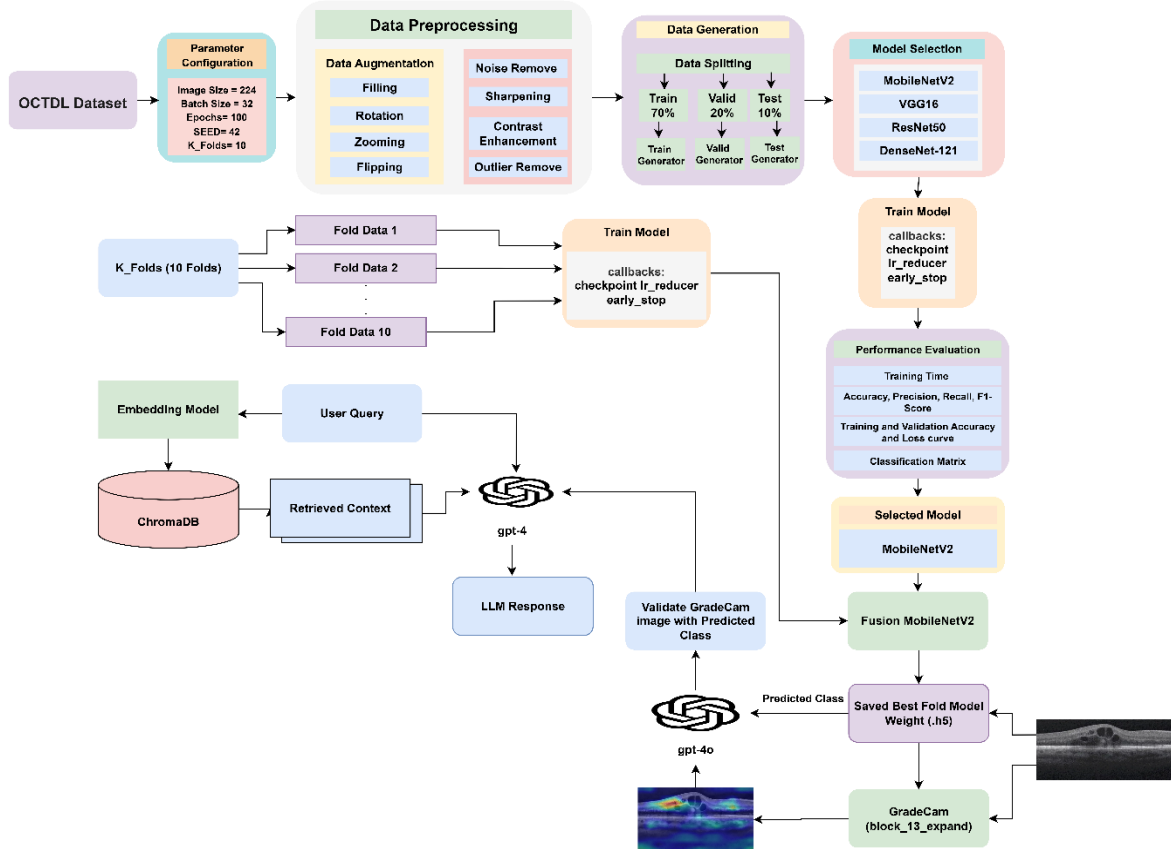


Fig 3.1: The experimental process of overall studies.

To experiment, a secondary dataset has been used, collected from Mendeley. Primarily, the dataset was highly imbalanced and had a lower number of images. Firstly, we performed augmentation on this dataset to increase the dataset size and class balancing. Then enhanced the data quality by performing noise removal, sharpening, contrast enhancement, and outlier removal. Then the dataset was split into train, validation, and test. Then we create training, testing, and validation data generation for model training. Secondly, we

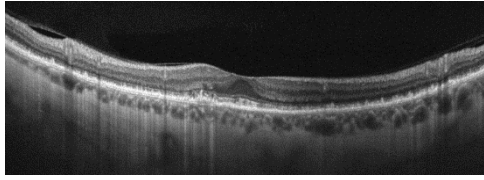
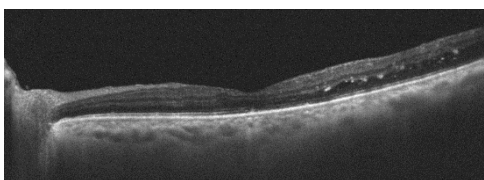
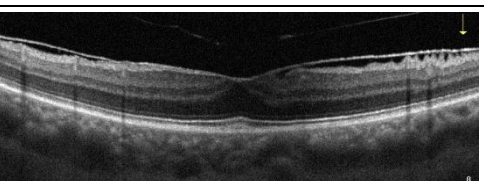
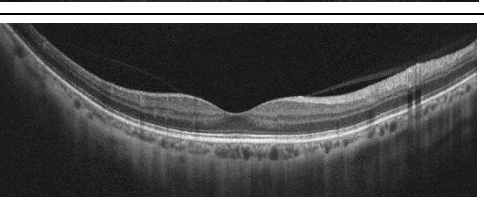
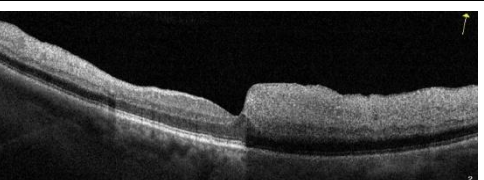
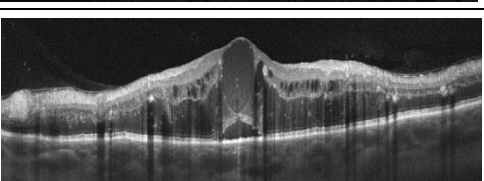
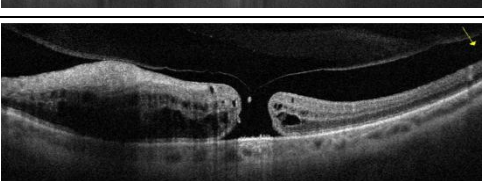
selected four CNN models (MobileNetV2, VGG16, ResNet50, and DenseNet-121) for the experiment. The model selection was based on previous studies that performed well on the OCT dataset. All experiments were conducted on the same parameter settings (Image size = 224, Batch Size = 32, Epoch 100). To prevent overfitting, we used callback functions, such as early stopping, reducing the learning rate, etc. Thirdly, the model performance is evaluated through training time, accuracy, precision, recall, f1-score, training and validation accuracy and loss curve, and classification matrix and report. The MobileNetV2 model outperforms others. Fourthly, we optimized the MobileNetV2 and performed feature fusion to improve the model performance. 10 K-Folds were applied to the Fusion MobileNetV2 to validate its performance. The Fusion MobileNetV2 performed well on each fold with better classification results. The best fold result was on the 5th fold, which is greater than 97%. Then Grad-CAM is used to check which region the model predicts the class from the test data. Then the predicted class and the Grad-CAM image were passed into an LLM model (gpt-4o), to validate whether the predicted class appropriately matched with the Grad-CAM image or not. The LLMs provide a summary of the matching. Then this summary was passed with a system prompt to another LLM (gpt-4). These LLMs take user input for questioning about the disease and provide responses. The LLMs tried to provide answers from their knowledge base first. If the information is not available in its knowledge base, then provide the answer from reasoning. To provide the knowledge base to the LLM model, we followed the retrieval augmented generation (RAG).

3.2 Data Collection Procedure

The study used a publicly available dataset called OCTDL: Optical Coherence Tomography Dataset to Image-Based Deep Learning Methods [31] on Mendeley. The dataset includes 1,600 OCT images that are divided into seven classes: Age-Related Macular Degeneration (885), Diabetic Macular Edema (143), Epiretinal Membrane (133), Normal (284), Retinal Artery Occlusion (22), Retinal Vein Occlusion (93), and

Vitreomacular Interface Disease (58). Table 3.1 below gives the detailed representation of the dataset.

Table 3.1: The details representation of OCTDL dataset

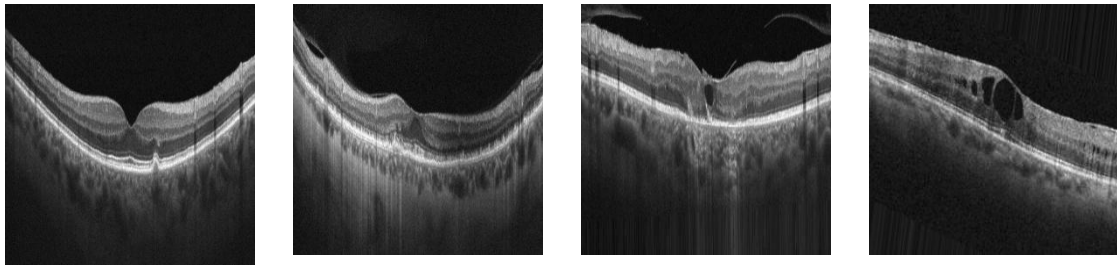
Class Name	Image
Age-Related Macular Degeneration	
Diabetic Macular Edema	
Epiretinal Membrane	
Normal	
Retinal Artery Occlusion	
Retinal Vein Occlusion	
Vitreomacular Interface Disease	

3.3 Image Pre-processing

The original OCTDL dataset had issues of class imbalance, low image quality, and a limited number of samples. To overcome these challenges and improve model performance, we applied data preprocessing, enhancement, and augmentation techniques. Preprocessing included resizing, shifting, rotation, brightness adjustment, zooming, and filling to standardize the images. For quality improvement, we performed noise removal, sharpening, contrast enhancement, and outlier removal.

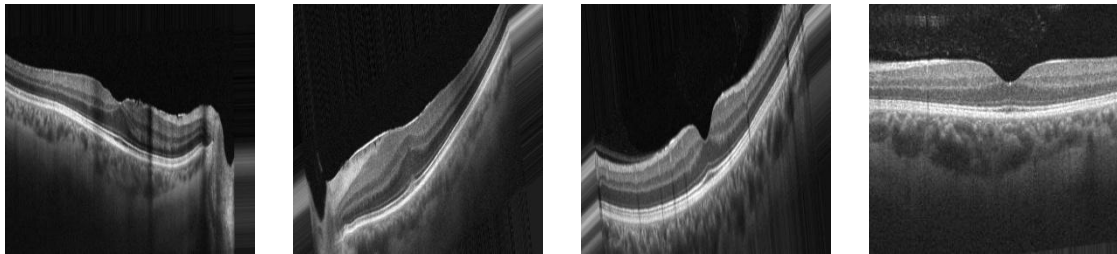
3.3.1 Data Augmentation

To perform data augmentation, we applied several techniques. Firstly, we resized the images to a 224×224 dimension. Secondly, we applied a 20-degree rotation followed by horizontal flipping. Additionally, we used nearest-neighbor filling, applied height and width shifting by a factor of 0.1, and adjusted the brightness within the range of 0.8 to 1.2 to improve image diversity. Data augmentation was applied based on class size: classes with less than 100 images were increased to 800, those with between 100 and 300 images were increased to 1,000, and those with more than 300 images were increased to 1,200. This ensured data consistency across all classes, resulting in a final dataset of 6,600 images after augmentation. The details of augmented data are presented in figure 3.2.



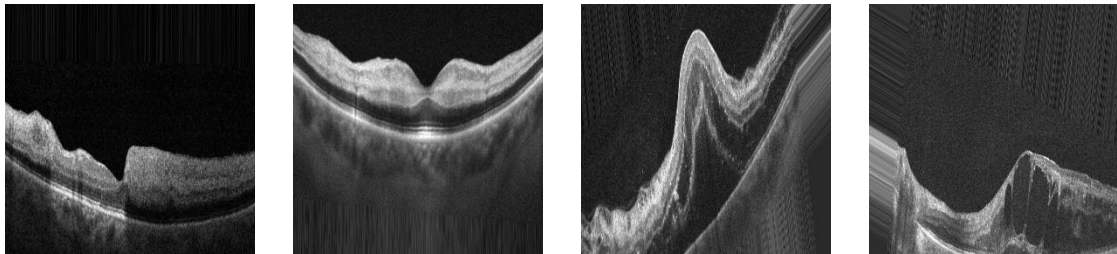
Age-Related Macular Degeneration (1200)

Diabetic Macular Edema (1000)



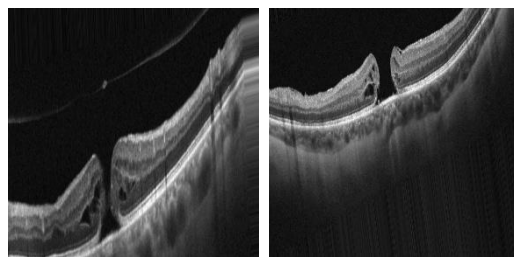
Epiretinal Membrane (1000)

Normal (1000)



Retinal Artery Occlusion (800)

Retinal Vein Occlusion (800)



Vitreomacular Interface Disease (800)

Fig 3.2: Augmented data of OCTDL dataset.

3.3.2 Data Enhancement

The quality of the dataset images is relatively low. Therefore, we applied noise removal, sharpening, contrast enhancement, and outlier removal to improve them. The details of the enhancement process are discussed in Sections 3.3.2.1, 3.3.2.2, 3.3.2.3, and 3.3.2.4.

3.3.2.1 Noise Removal

To remove the noise, we reduced random noise from the images by utilizing non-local means (NLM). These methods preserve important image details and also average information from similar pixel neighborhoods. In this study, the image dataset was a grayscale image. For this instance, we set a higher filtering strength $h=10$. Secondly, template window size 7, which is the size of the patch around each pixel used for measuring similarity. Finally, a search window size of 21, which is the size of the region to search for similar patches.

3.3.2.2 Sharpening

To obtain more detailed image edges, performed sharpening by applying a convolutional kernel. This method helps to enhance the image edges and makes details more distinct. The kernel filter utilizes the center pixel with a weight of 5 and subtracts the influence of its neighbors with weights of -1. The kernel details are presented in Equation (1).

$$K = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (1)$$

3.3.2.3 Contrast Enhancement

To enhance the image contrast, CLAHE was applied to the sharpened image. Firstly, apply a clip limit of 2, which prevents over-amplification of noise by limiting contrast. Secondly, we divided the image into small tiles (8 X 8) and applied histogram equalization to each tile. Finally, applied CLAHE to improve the local contrast and make fine details without over-saturating.

3.3.2.4 Outlier Removal

To remove the outlier of the images, firstly, the average pixel value of the grayscale images was calculated. Then set the min_mean and max_mean limits. Secondly, if an image's average brightness is too low, then it is considered very dark. If it is very high, then it is considered bright. Too dark or too bright was considered an outlier, and it was removed. Finally, the grayscale image's single channel is duplicated 3 times to make it a 3-channel RGB. This process helps to learn pretrained models well. The detailed representation enhanced the image presented in figure 3.3.

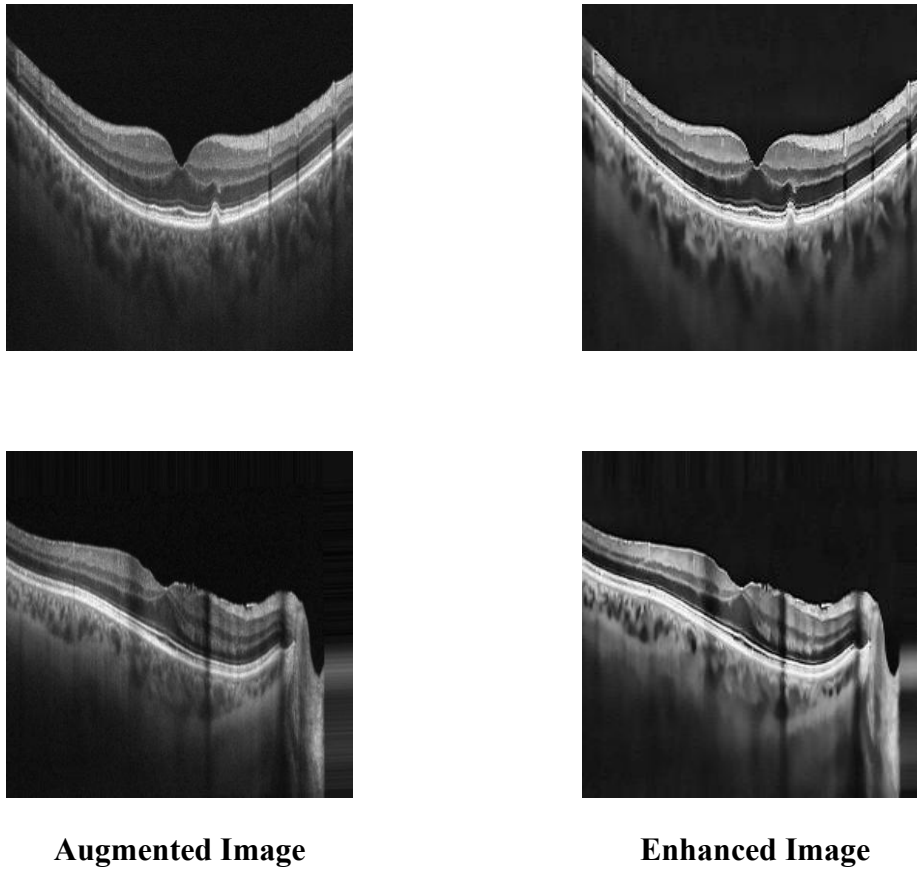


Fig 3.3: Enhanced Images of OCTDL Dataset

3.4 Model Selection

To conduct the experiment, four CNN models (MobileNetV2, VGG16, ResNet50, and DenseNet50) were selected based on previous studies demonstrating strong performance on OCT datasets. In our experiments, MobileNetV2 outperformed the other models. Consequently, we customized MobileNetV2 and proposed the Fusion MobileNetV2 model.

3.4.1. MobileNetV2

The MobileNetV2 [32] is a lightweight deep learning architecture designed for mobile and edge devices. It uses Inverted Residuals and Linear Bottlenecks to improve its performance over the MobileNetV1 [33] architecture which uses only Depthwise and Pointwise convolutions. The Inverted Residuals are opposite of how traditional ResNet Bottlenecks operate. The dimensions are expanded first, and then reduced back to low-dimension features. The Linear Bottlenecks use linear activation function instead of ReLU at the last 1x1 projection, preventing loss of information due to excessive non-linearity in low dimensional space. This allows the architecture demonstrate great results on ImageNet dataset and COCO dataset in comparison to state-of-the-art models while being very lightweight. Figure 3.4 presents details of MobileNetV2 architecture.

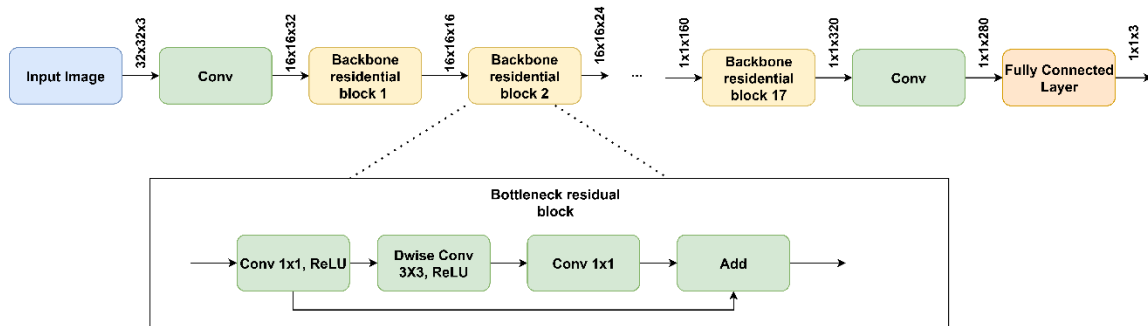


Fig 3.4: Architectures of MobileNetV2

3.4.2 ResNet50

Deep neural networks suffer from Vanishing Gradient problem and Degradation problem. The ResNet [34] uses skip connections to overcome this issue and allow models to have deeper layers for extensive feature extraction. The ResNet-50 is a 50-layer deep learning network using bottleneck residual blocks instead of regular residual blocks, where each bottleneck residual block consists of 3 layers. ResNet-50 achieves a 20.74% top-1 error with only 3.8 billion FLOPs on the ImageNet dataset, retaining a close position with ResNet-101 and ResNet-152 with significantly lower parameters. The balance between

performance and efficiency makes it a perfect candidate model for our research. Figure 3.5 presents details of ResNet50 architecture.

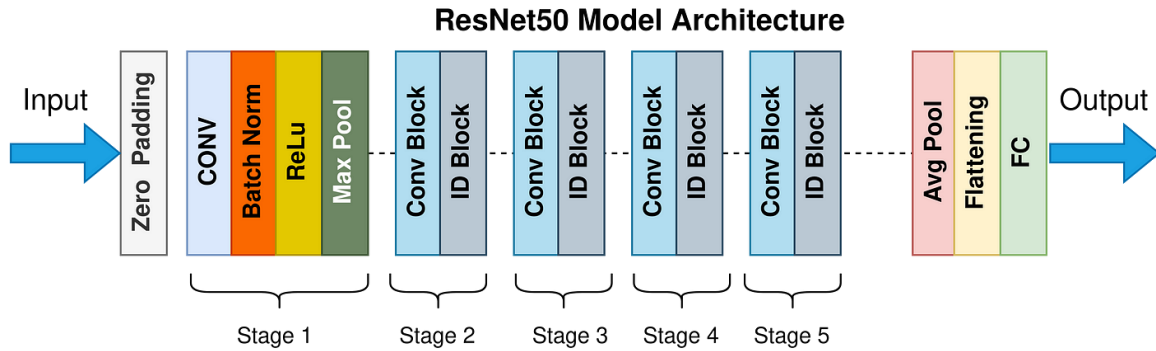


Fig. 3.5: Architecture of ResNet50

3.4.3 VGG16

One of the most traditional CNN architectures is the VGG16 [35], which has 16 deep layers with the first 13 layers being convolutional layers and the remaining 3 fully connected layers. The authors show that 3x3 convolutions are much more efficient than 7x7 or 11x11 convolutions in terms of the number of parameters, deeper network construction, and feature extraction. In the 2014 ImageNet Large-Scale Visual Recognition Challenge, the VGG-16 won the localization and second place in classification. Transfer learning is also commonly applied to the model, with its trainable weights being far more compatible with more recent datasets. Figure 3.6 shows the details architecture of VGG16.

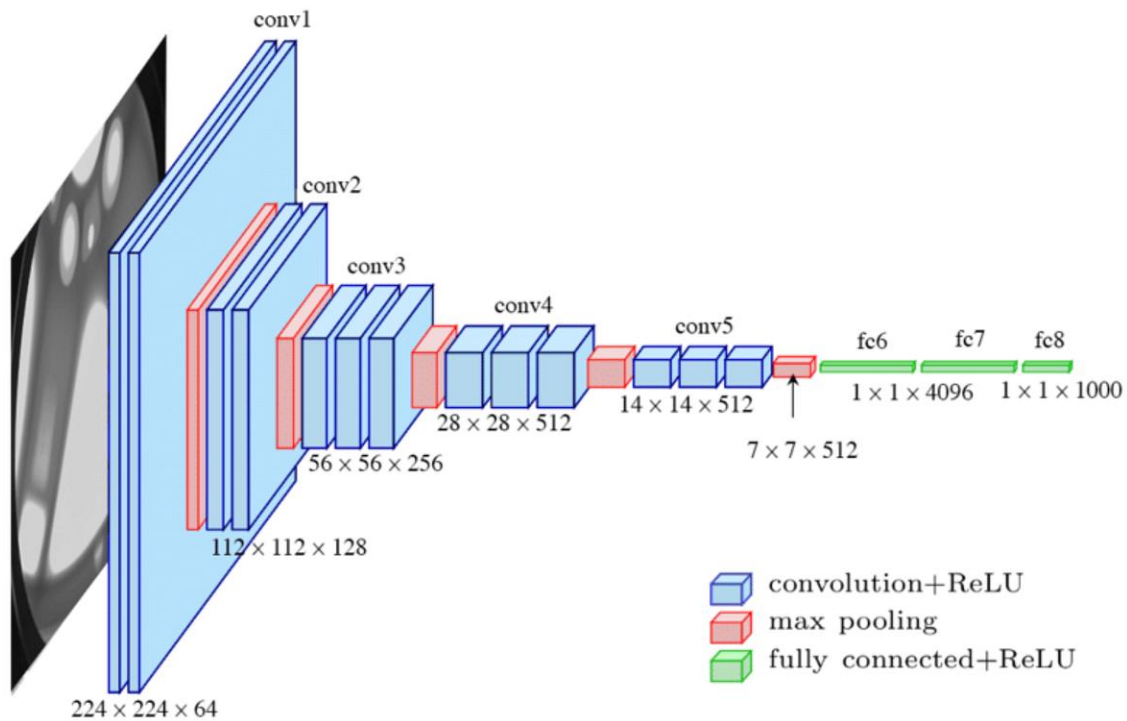


Fig 3.6: Architecture of VGG16

3.4.4 DenseNet121

The DenseNet [36] is a special architecture in which every layer is linked to all its previous layers, with input in a feed-forward method. The backbone of the DenseNet-121 consists of 121 convolutional neural network layers. The network is partitioned into four dense blocks with 6, 12, 24 and 16 bottleneck layers, respectively. In between dense blocks are transition layers, which do a 1×1 convolution (to reduce channels) and then 2×2 average pooling (to down sample space). The model employs the bottleneck and compression factor (usually 0.5) to minimize the model parameters. This assists the model to attain the top-1 and top-5 error rates on ImageNet with much fewer parameters than its counterparts, which makes it another ideal candidate to our research. The details architecture of DenseNet121 is shown in Figure 3.7.

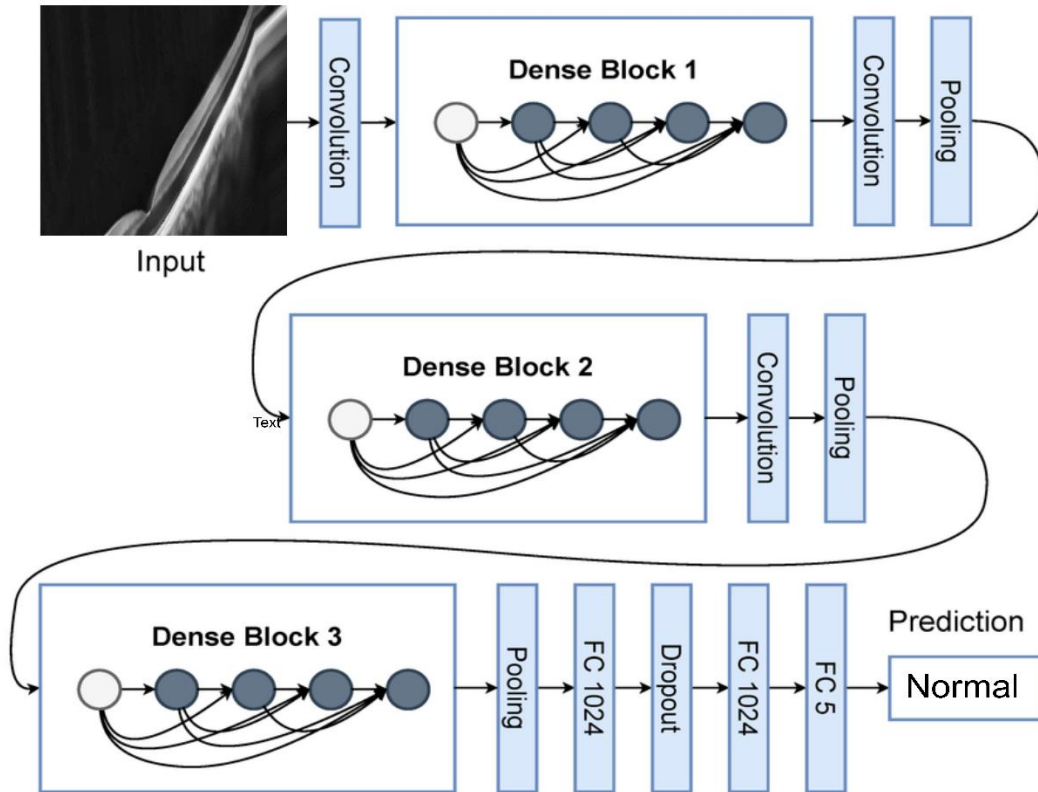


Fig 3.7: Architecture of DenseNet121

3.4.5 Fusion MobileNetV2

In this study, we propose a fusion MobileNetV2 model, which enhances the feature representation capability of the original MobileNetV2 architecture. The input image is represented as:

$$\mathbf{X} \in \mathbb{R}^{H \times W \times C} \quad (2)$$

Where H, W, and C denote the height, width, and number of channels. As the backbone of the model, the fusion MobileNetV2 model utilizes the original MobileNetV2, which was pretrained on the ImageNet dataset. The model utilized the first 100 layers frozen to preserve low-level features. The remaining layers are fine tuned for domain specific learning. From the MobileNetV2 feature map:

$$\mathbf{F} = \mathbf{f}_{MobileNetV2}(\mathbf{X}) \quad (3)$$

We extract two complementary feature representations here. The Global Contextual Features are extracted through Global Average Pooling (GAP) layer. Which can be denoted:

$$\mathbf{G} = \frac{1}{H.W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{F}_{ij} \quad (4)$$

Then, the local structural features are extracted using a 3X3 convolution followed by Batch Normalization and Global Average Pooling (GAP). Which can be presented as:

$$\mathbf{L} = \mathbf{GAP}(\sigma(\mathbf{W}_c * \mathbf{F} + \mathbf{b}_c)) \quad (5)$$

Where \mathbf{W}_c and \mathbf{b}_c are convolutional weights and biases, * denoted convolution and σ is the ReLU activation function.

After getting extracted Global (G) and Local (L) features, fused using concatenation. Which can be represent as:

$$\mathbf{Z} = [\mathbf{G} || \mathbf{L}] \quad (6)$$

The fused feature vector Z is passed through a fully connected layer with L2 regularization. To prevent overfitting a dropout added. These can be represented as:

$$\mathbf{H} = \sigma(\mathbf{W}_d \mathbf{Z} + \mathbf{b}_d) \quad (7)$$

Finally, the SoftMax output layer perform the multi class classification. Which can be represent as:

$$\hat{\mathbf{y}}_k = \frac{\exp(H_k)}{\sum_{i=1}^C \exp(H_i)}, \quad k = 1, 2, \dots, C \quad (8)$$

where C is the total number of classes and $\hat{\mathbf{y}}_k$ is the predicted probability for class k.

By combining global and local feature extraction with a fusion mechanism, the model can learn more complex and more descriptive feature representations compared to the original

MobileNetV2. The details representations of model architecture presented below figure 3.8.

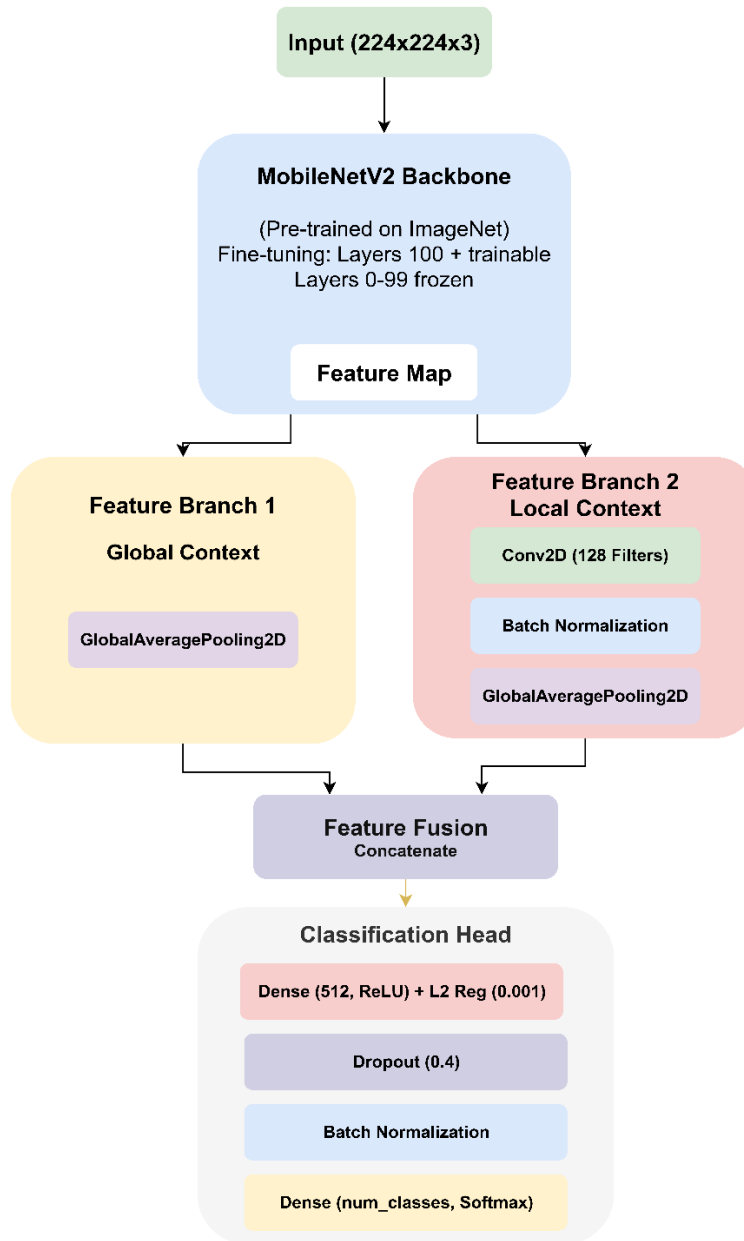


Fig 3.8: Architecture of proposed model (Fusion MobileNetV2)

3.5 Grad-CAM (Gradient- Weighted Class Activation Mapping)

Gradient-Weighted Activation Mapping (Grad-CAM) is applied to explain deep learning model choices. In particular, Convolutional Neural Networks [37]. Grad-CAM calculates the gradient of the target classification with respect to the feature maps of the last convolutional layer [38]. Grad-CAM compares the gradient of the predicted class using the feature maps of the final convolutional layer. These gradient techniques imply the significance of every feature map in the decision-making. The sum of this feature maps weighted is then computed and subjected to a ReLU activation to produce the heatmap. Which points out the areas that primarily contribute to the prediction [38]. The prediction regions are highlighted by Grad-CAM, as shown in Figure 3.9.

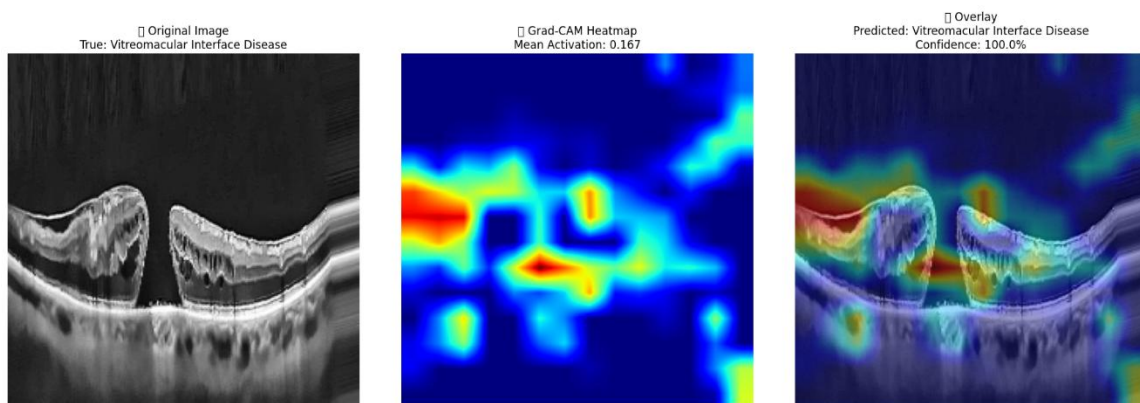


Fig 3.9: How Grad-CAM highlights predicted class regions.

3.6 RAG (Retrieval Augmented Generation)

Retrieval Augmented Generation (RAG) is a technique to provide knowledge to the LLM models from external sources (PDF, JSON, Text, etc.). This technique combines the retrieval-based methods with generative models. To provide the knowledge to the LLMs, the input data is divided into several chunks. Then perform these chunks' vector embeddings with vector embedding models. The vector data is saved into a vector database

(Pinecone, chromadb, etc.). When a user provides a prompt to the LLM. Firstly, the prompt performs vector embedding, then the retrieval search in the vector database for similar chunks by using the cosine similarity index. The similarity information is then combined with the original prompt. These augmented prompts were passed to the LLM. Finally, LLM generates the enhanced response. Figure 3.10 illustrates the RAG working diagram in detail.

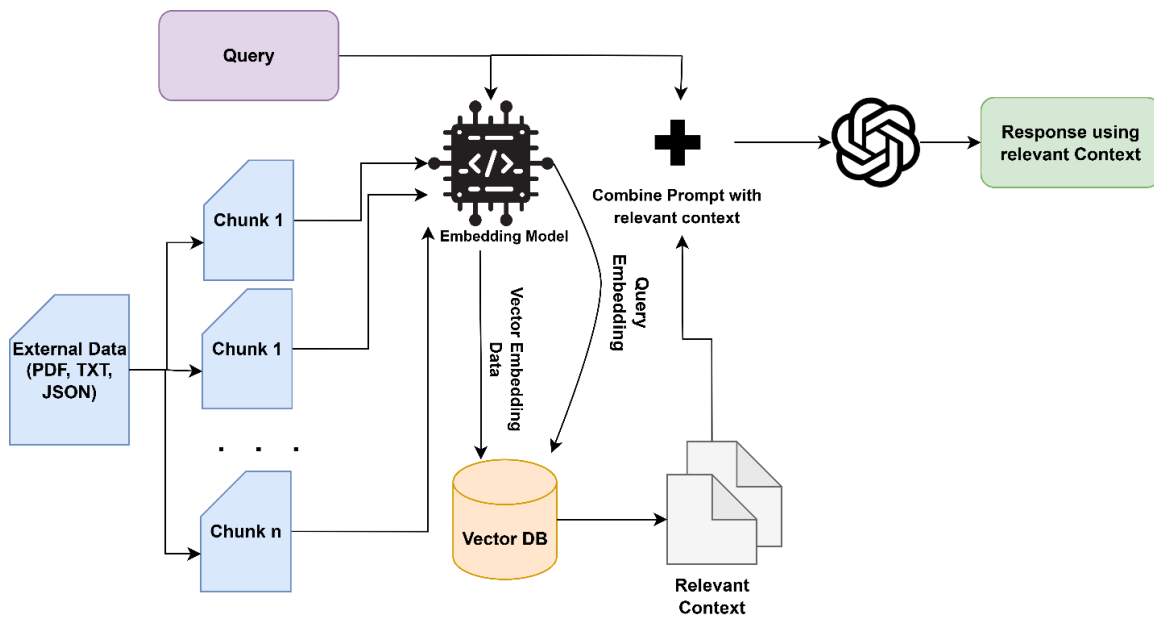


Fig 3.10: RAG working process

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Result Evolution Methods

To evaluate the result of the study, several performance metrics were employed. These include Accuracy, Precision (P), Recall (R), F1-Score, Classification Report, and Confusion Matrix. Each of these metrics is discussed in detail in this chapter.

Accuracy: The performance of a classification model is measured by accuracy which is the ratio of images which are correctly classified to the total numbers of the images. The formula to compute it is as below.

$$\mathbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

Precision: Precision indicates how many positive predictions are made correctly of all the predicted cases. It is especially notable in the situations, where false positives (FP) are more vital than false negatives (FN). The following formula may be used to compute precision.

$$\mathbf{Precision} = \frac{TP}{TP+FP} \quad (10)$$

Recall: The ratio of actual positive instances which are correctly recognized by the model is also known as recall, and is of particular significant in a situation where errors on the false negatives (FN) have serious consequences than those on false positives (FP). The following formula used to compute recall.

$$\mathbf{Recall} = \frac{TP}{TP+FN} \quad (11)$$

F1-Score: F1-Score is the harmonic mean of recollection and correctness. The formulaic calculation of F1- Score as presented below.

$$F1 - Score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (12)$$

4.2 Experimental Results & Analysis

This section will discuss in detail the experimental results of this study. Firstly, the 4.2.1 analyses the classification reports of the four CNN models. Second, the discussion of the 4 models (section 4.2.2) covers the confusion matrices of the 4 models. Thirdly, section 4.2.3 examines their training and validation loss curves. Besides, section 4.2.4 provides a comparison of the performance of the four models and section 4.2.5 contains results of proposed Fusion MobileNetV2 model. Lastly, the Section 4.2.6 shows the K-Fold cross-validation results of the proposed Fusion MobileNetV2 model.

4.2.1 Classification Reports of Four CNN Models

This section compares the classification results of four CNN models: MobileNetV2, VGG16, ResNet50, and DenseNet121. Table 4.1 presents the detailed classification report for each model.

Table 4.1: Classification reports of four CNN models

MobileNetV2					VGG16			
Class Name	Precision	Recall	F1 Score	Support	Precision	Recall	F1 Score	Support
Age Related Macular Degeneration	0.92	0.86	0.89	180	0.94	0.84	0.89	180
Diabetic Macular Edema	0.86	0.84	0.85	150	0.85	0.89	0.87	150
Epiretinal Membrane	0.83	0.84	0.84	150	0.84	0.83	0.84	150
Normal	0.84	0.88	0.86	150	0.77	0.89	0.83	150
Retinal Artery Occlusion	0.92	0.97	0.94	120	0.98	0.98	0.98	120
Retinal Vein Occlusion	0.88	0.80	0.84	120	0.86	0.74	0.79	120
Vitreomacular Interface Disease	0.89	0.97	0.93	120	0.89	0.93	0.91	120
ResNet50					DenseNet121			
Class Name	Precision	Recall	F1 Score	Support	Precision	Recall	F1 Score	Support
Age Related Macular Degeneration	0.60	0.59	0.60	180	0.89	0.86	0.87	180
Diabetic Macular Edema	0.50	0.43	0.46	150	0.85	0.83	0.84	150
Epiretinal Membrane	0.42	0.47	0.44	150	0.84	0.85	0.84	150
Normal	0.49	0.65	0.56	150	0.83	0.87	0.85	150
Retinal Artery Occlusion	0.57	0.68	0.62	120	0.94	0.98	0.96	120
Retinal Vein Occlusion	0.42	0.26	0.32	120	0.82	0.76	0.79	120
Vitreomacular Interface Disease	0.42	0.36	0.39	120	0.90	0.93	0.91	120

From table 4.1, among of four CNN models, the MobileNetV2, VGG16, and DenseNet121 perform superior in per-class classification. MobileNetV2 and DenseNet121 perform similarly. However, in terms of per-class detection, MobileNetV2 provides the reliable performance than DenseNet121 (see Table 4.1).

4.2.2 Confusion Metrics of Four CNN models

This section will compare the confusion metrics of each model to evaluate the model performance. This metrics will present how well the model's predicated labels match the actual labels.

Age Related Macular Degeneration = 0
 Diabetic Macular Edema = 1
 Epiretinal Membrane = 2
 Normal = 3

Retinal Artery Occlusion = 4
 Retinal Vein Occlusion = 5
 Vitreomacular Interface Disease = 6

	0	1	2	3	4	5	6
0	154	7	4	9	0	2	4
1	2	126	8	2	0	6	6
2	1	3	126	10	2	4	4
3	7	2	4	132	4	1	0
4	0	1	2	0	116	0	1
5	4	7	5	4	4	96	0
6	0	0	2	1	0	0	117

Fig 4.1: Confusion matrix (MobileNetV2)

	0	1	2	3	4	5	6
0	152	5	3	14	0	6	0
1	3	134	4	2	0	5	2
2	1	6	125	11	1	1	5
3	3	1	5	133	0	2	6
4	0	1	0	0	118	1	0
5	0	11	8	9	2	89	1
6	2	0	3	3	0	0	112

Fig 4.2: Confusion matrix (VGG16)

	0	1	2	3	4	5	6
0	106	17	10	23	8	8	8
1	27	65	21	4	12	12	9
2	12	10	70	27	13	5	13
3	3	4	25	98	6	1	13
4	2	8	9	9	82	7	3
5	12	18	12	11	23	31	13
6	14	8	20	26	0	9	43

Fig 4.3: Confusion matrix (ResNet50)

	0	1	2	3	4	5	6
0	155	5	6	8	0	3	3
1	6	124	5	2	0	12	1
2	2	4	127	10	2	3	2
3	5	1	5	131	2	2	4
4	0	0	0	1	118	0	1
5	3	12	5	4	3	91	2
6	4	0	3	2	0	0	111

Fig 4.4: Confusion matrix (DenseNet-121)

Figures 4.1, 4.2, 4.3, and 4.4 show the confusion matrices of MobileNetV2, VGG16, ResNet50, and DenseNet-121, respectively. From these figures, it can be observed that MobileNetV2 outperforms the other models in classification. This superior performance is also reflected in the classification report (see Table 4.1).

4.2.3 Training Validation Accuracy and Loss Curve of Four CNN Models

This section presents the training and validation accuracy curves for each model during training. It also shows the loss curves of the models, which indicate how well each model is trained and validated. Additionally, these curves help identify whether any overfitting or underfitting has occurred.

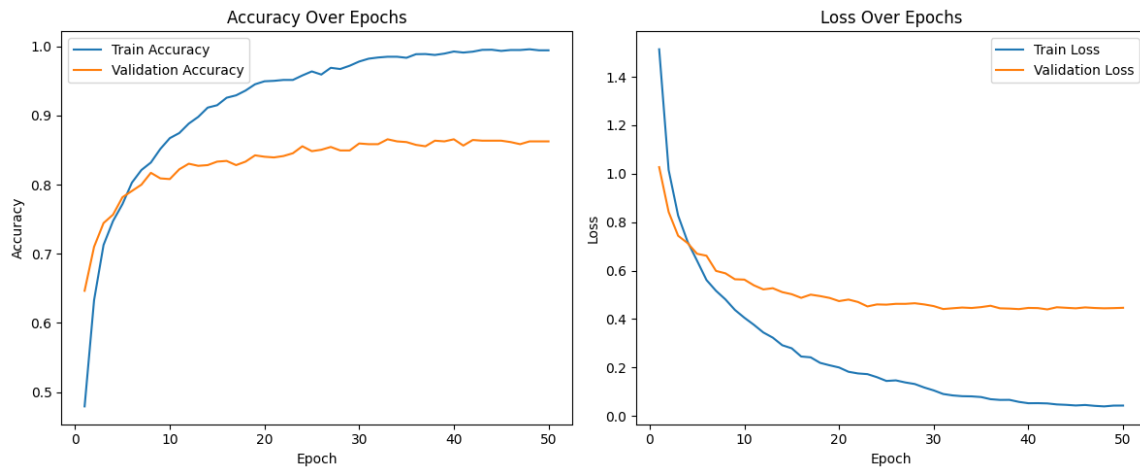


Fig 4.5: Training and validation accuracy and loss curve of MobileNetV2

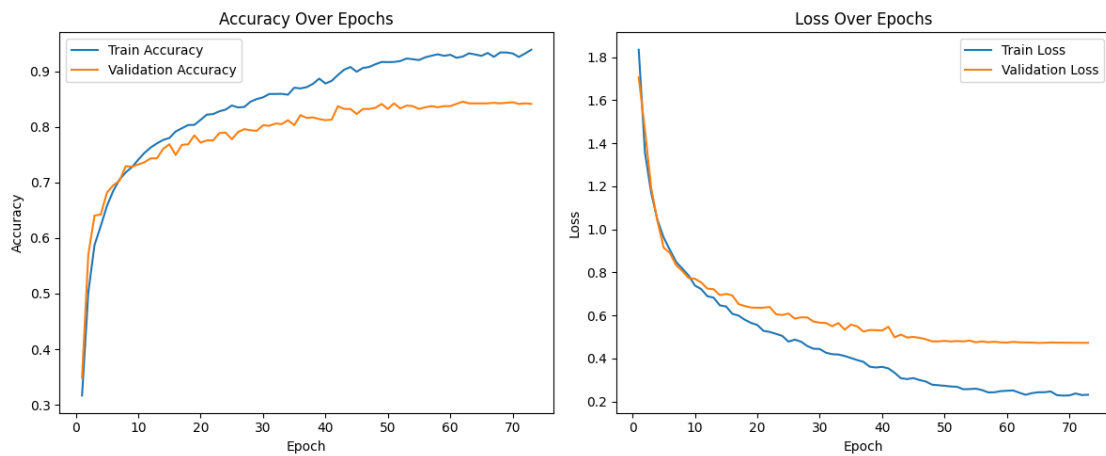


Fig 4.6: Training and validation accuracy and loss curve of VGG16

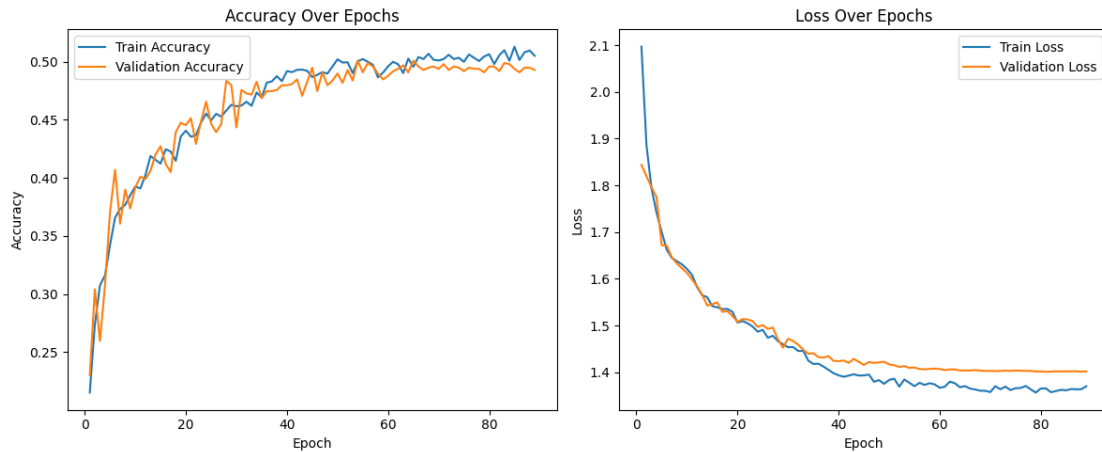


Fig 4.7: Training and validation accuracy and loss curve of ResNet50

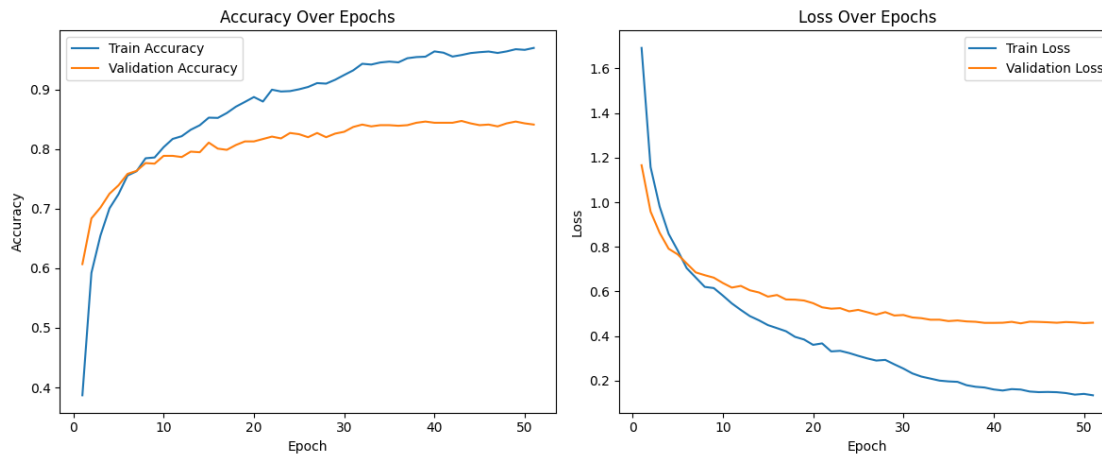


Fig 4.8: Training and validation accuracy and loss curve of DenseNet-121

The accuracy and loss curves of the four models are presented in figures 4.5, 4.6, 4.7, and 4.8. Figure 4.7 shows that the training and validation accuracy and loss of ResNet50 are quite similar to one another, but the training and validation accuracy of ResNet50 is rather low. Conversely, both DenseNet121 and MobileNetV2 are highly performing (see Figures 4.5 and 4.7). DenseNet121 exhibits a smaller difference between training and validation accuracy and loss than MobileNetV2. However, according to the findings in Table 4.1 and Figure 4.1, MobileNetV2 can be regarded as the better model in general.

4.2.4 Performance Comparison of Four Models

Table 4.2: Performance Comparison of four CNN Models

Model Name	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Time (s)	Best Epoch
MobileNetV2	0.9998	0.8657	0.8758	387	33
VGG16	0.9721	0.8455	0.8717	2341	65
ResNet50	0.5429	0.5010	0.5000	1426	81
DenseNet-121	0.9933	0.8475	0.8657	763	43

From the experiment, it is notable that MobileNetV2 performs exceptionally well in OCT image analysis. After extensive testing, it clearly outperforms the other CNN models (ResNet50, VGG16, and DenseNet121) in terms of accuracy, training time, and overall efficiency. As shown in Table 4.2, MobileNetV2 achieves the highest performance with minimal training time, reaching 99.98% training accuracy, 86.57% validation accuracy, and 87.58% testing accuracy. Based on these results, MobileNetV2 was selected as the base model for our proposed fusion model.

4.2.5 Result of Proposed Fusion Mobilenetv2

This section presents the performance of the Fusion MobileNetV2 model on the OCTDL dataset. It includes the classification report, confusion matrix, and the training and validation accuracy and loss curves.

Table 4.3: The classification report of Fusion MobileNetV2

Fusion MobileNetV2				
Class Name	Precision	Recall	F1 Score	Support
Age Related Macular Degeneration	0.97	0.94	0.96	180
Diabetic Macular Edema	0.94	0.96	0.95	150
Epiretinal Membrane	0.96	0.96	0.95	150
Normal	0.92	0.98	0.95	150
Retinal Artery Occlusion	1.00	1.00	1.00	120
Retinal Vein Occlusion	0.94	0.88	0.91	120
Vitreomacular Interface Disease	0.98	0.98	0.98	120

From Table 4.3, it is notable that the Fusion MobileNetV2 performs better in per-class classification under the same settings compared to ResNet50, VGG16, and DenseNet121.

Age Related Macular Degeneration = 0
 Diabetic Macular Edema = 1
 Epiretinal Membrane = 2
 Normal = 3

Retinal Artery Occlusion = 4
 Retinal Vein Occlusion = 5
 Vitreomacular Interface Disease = 6

	0	1	2	3	4	5	6
0	170	3	0	5	0	1	1
1	1	144	0	0	0	5	0
2	2	2	142	3	0	0	1
3	1	0	1	147	0	1	0
4	0	0	0	0	120	0	0
5	1	5	4	4	0	106	0
6	0	0	1	1	0	0	118

Fig 4.9: Confusion Matrix of Fusion MobileNetV2

From Figure 4.9, it can be observed that the number of misclassifications is much lower compared to ResNet50, VGG16, and DenseNet121. The model performs well across all class classifications.

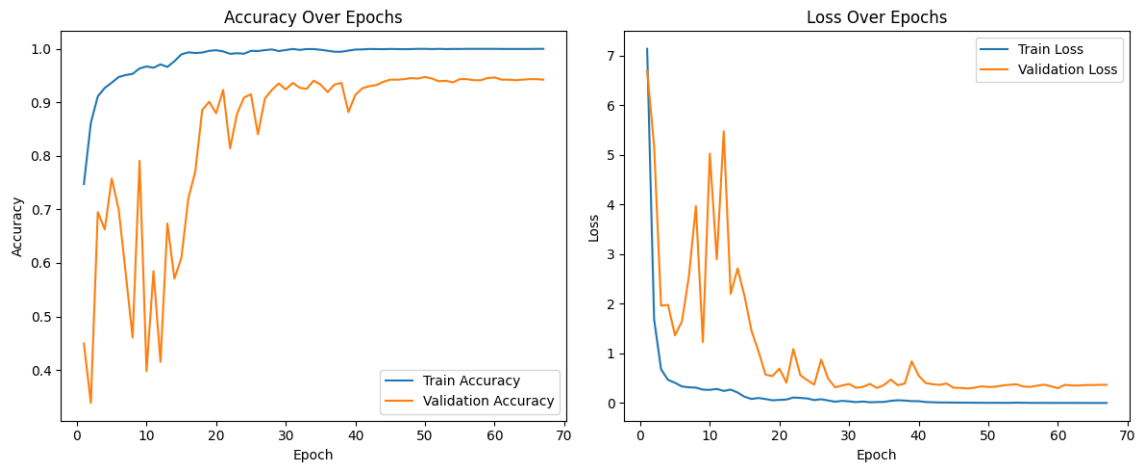


Fig 4.10: Training and Validation accuracy and loss curve (Fusion MobileNetV2)

Figure 4.10 shows the training and validation accuracy and loss curves of the Fusion MobileNetV2 model. The curves indicate that the model achieves 100% training accuracy, 94.75% validation accuracy, and 95.66% testing accuracy. These results demonstrate highly efficient classification performance for OCT image analysis compared to the other models.

4.2.6 K_Folds Results of Fusion Mobilenetv2

To validate the results of the Fusion MobileNetV2, we applied K-fold cross-validation to ensure the model's reliability. A 10-fold cross-validation was performed, and this section presents the detailed results of the K-fold evaluation.

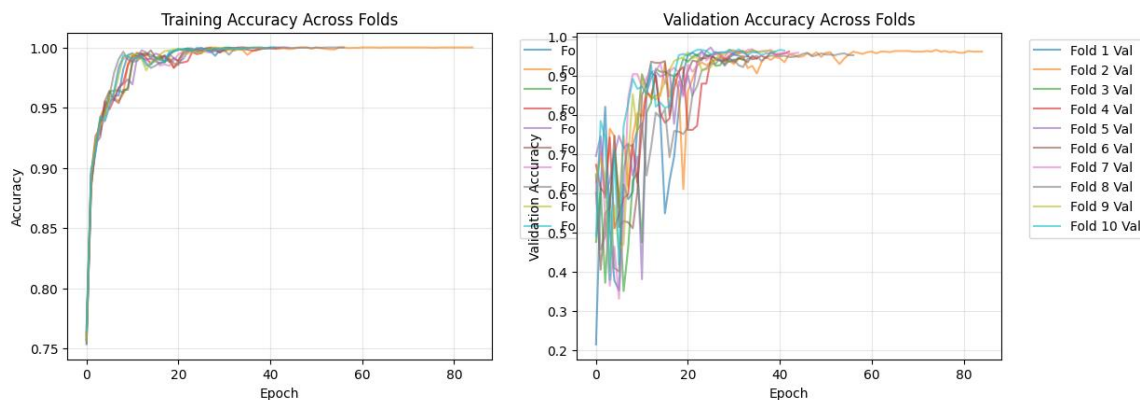


Fig 4.11: Training and validation accuracy across 10 folds (Fusion MobileNetV2)

Figure 4.11 shows that the Fusion MobileNetV2 provides efficient results across all folds. The accuracy in each fold greater than 95%. The results are consistently close across all folds, indicating stable and reliable model performance.

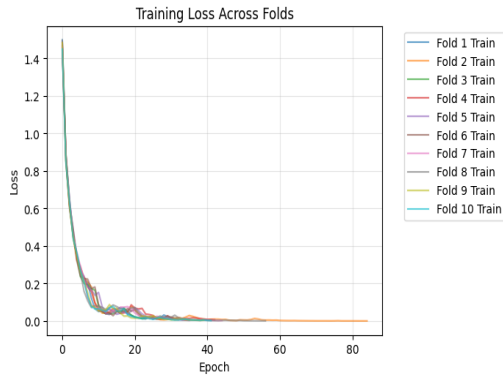


Fig 4.12: Training Loss Across Folds (Fusion MobileNetV2)

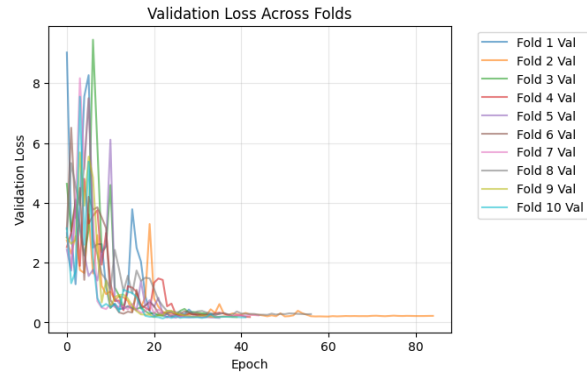


Fig 4.13: Validation Loss Across Folds (Fusion MobileNetV2)

Figures 4.12 and 4.13 show the training and validation loss across all folds. The low loss values for both training and validation indicate a minimal chance of model overfitting.

Table 4.4: K-Fold Cross Validation Results (Fusion MobileNetV2)

K-Fold Cross Validation Results (Fusion MobileNetV2)	
Folds Name	Validation Accuracy
Fold 1	96.21%
Fold 2	96.67%
Fold 3	95.45%
Fold 4	96.36%
Fold 5	97.27%
Fold 6	95.45%
Fold 7	96.82%

Fold 8	95.91%
Fold 9	96.67%
Fold 10	96.82%

From Table 4.4, it is notable that Fusion MobileNetV2 provides stable results across all folds, with accuracy greater than 95% in each fold. This indicates that the model is highly efficient for OCT image analysis. The best-performing fold is Fold 5, achieving a validation accuracy of 97.27%.

Table 4.5: Classification report of Fold 5 (Fusion MobileNetV2)

Fusion MobileNetV2				
Class Name	Precision	Recall	F1 Score	Support
Age Related Macular Degeneration	0.99	1.00	1.00	240
Diabetic Macular Edema	0.99	0.99	0.99	200
Epiretinal Membrane	1.00	1.00	1.00	200
Normal	1.00	0.98	0.99	200
Retinal Artery Occlusion	1.00	1.00	1.00	160
Retinal Vein Occlusion	0.99	0.99	0.99	160
Vitreomacular Interface Disease	0.99	0.99	0.99	160

Table 4.5 demonstrates strong classification results. In Fold 5, the model achieves nearly 100% accuracy in per-class classification, indicating excellent performance for medical diagnosis.

Age Related Macular Degeneration = 0

Retinal Artery Occlusion = 4

Diabetic Macular Edema = 1

Retinal Vein Occlusion = 5

Epiretinal Membrane = 2

Vitreomacular Interface Disease = 6

Normal = 3

	0	1	2	3	4	5	6
0	240	0	0	0	0	0	0
1	0	199	1	0	0	0	0
2	0	0	200	0	0	0	0
3	1	0	0	197	0	1	1
4	0	0	0	0	160	0	0
5	0	1	0	0	0	159	0
6	1	0	0	0	0	0	159

Fig 4.14: Confusion Matrix of (5-folds Fusion MobileNetV2)

Figure 4.14 shows the confusion matrix across the 5 folds. The matrix indicates a very low number of misclassifications, with only 6 errors. These results demonstrate the model's high efficiency. The testing accuracy achieved is 99.55%.

4.3 Model performance Comparison

Table 4.6 summarizes the overall experimental results. Across extensive experiments, Fusion MobileNetV2 outperforms the other models under all parameter settings.

Table 4.6: All model performance comparison

Model Name	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Time (s)	Best Epoch
MobileNetV2	0.9998	0.8657	0.8758	387	33
VGG16	0.9721	0.8455	0.8717	2341	65
ResNet50	0.5429	0.5010	0.5000	1426	81
DenseNet-121	0.9933	0.8475	0.8657	763	43
Fusion MobileNetV2	1.00	0.9475	0.9566	653	47
Fusion MobileNetV2 (5-Fold)	1.00	0.9727	0.9955	420	26

To further evaluate its performance, K-fold cross-validation was applied. The model demonstrates stable and consistently improving results across all folds. As highlighted in Table 4.6, Fusion MobileNetV2 (5-Fold) achieves the highest performance, with 100% training accuracy, 97.27% validation accuracy, and 99.55% testing accuracy, while requiring minimal training time, outperforming all other models.

4.4 Grad-CAM analysis

After getting the best model weights, we tested the model using various test images. Grad_CAM was then applied to validate the model prediction. The Grad_CAM utilized the “best_5_fold_model.h5” and target layer “block_13_expand” to highlight the regions

which influenced the model's prediction. The output of Grad-CAM is presented in figure 4.15.

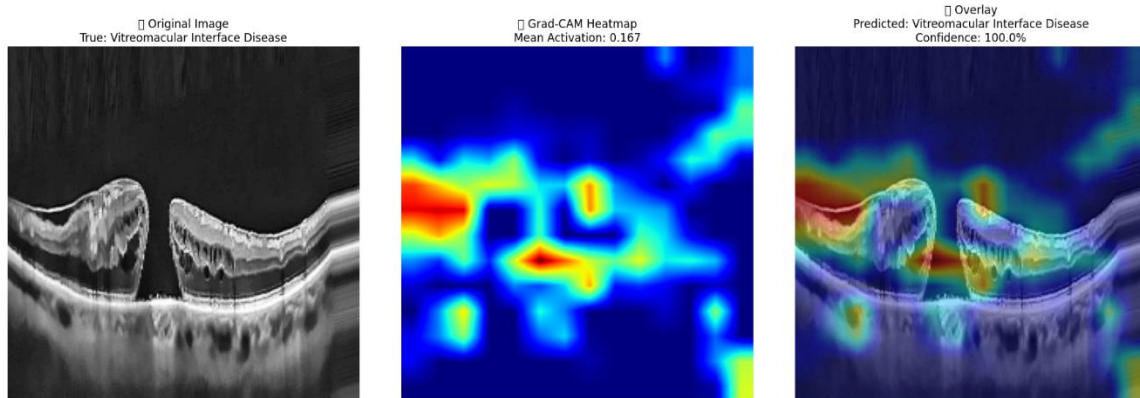


Fig 4.15: Grad-CAM analysis of test images.

4.5 LLM Interpretation

After getting the Grad-CAM image, we passed it to an LLM model (gpt-4o) with the predicted class. The LLM model that validates the predicted class with the Grad-CAM image highlighted region. If a match was found, the LLM responded with a summary explaining why it matched, with reasoning. Then this reasoning and summarization passed with the system prompting another LLM (gpt-4) to understand the personalization of the user. When a user gives a prompt to the LLM Model, then the LLM provides a response from its given knowledge base, which is provided from the OCT disease-related book using the RAG pipeline (see section 3.6). If no answer is found answer, then it provides an answer from its own reasoning. We deploy these with a web app (Retino AID) using FastAPI. Figures 4.16 and 4.17 represent the details of the web app (Retino AID) for real-time analysis.

OCT Retinal Image Diagnosis

Choose File No file chosen

Analyze Image

Model Prediction

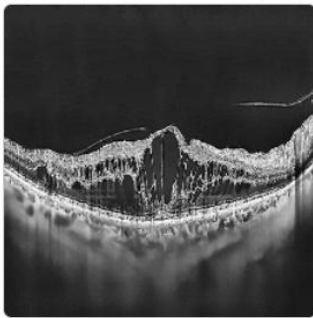
Class: Diabetic Macular Edema

Confidence: 100.0%

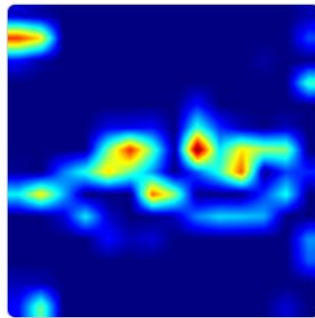
Mean Activation: 0.0975

Focused Area Ratio: 0.0406

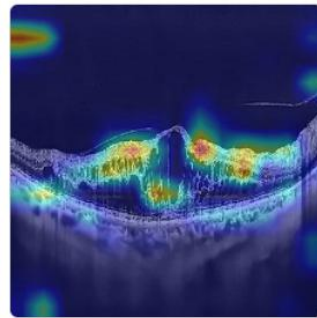
Original



Grad-CAM Heatmap



Overlay



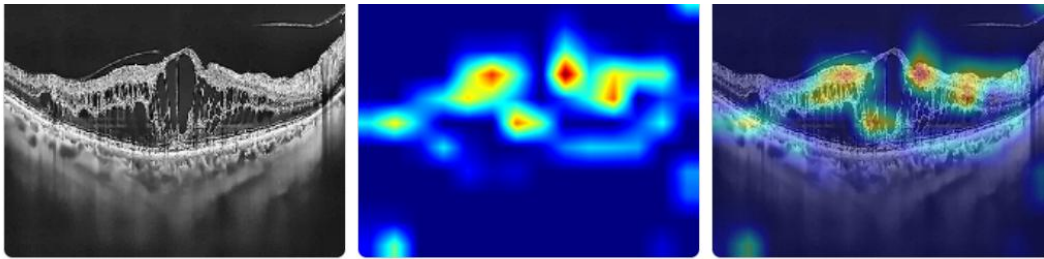
✓ GPT-4o Validation

The Grad-CAM heatmap highlights regions in the macula, focusing particularly on the areas that appear thickened and swollen. This is consistent with Diabetic Macular Edema, which is characterized by fluid accumulation in the macula. Therefore, the explanation is plausible as it targets the relevant pathological features associated with the condition.

💬 Ask Medical Question

e.g. What treatment is suggested for Diabetic Macular Edema?

Fig 4.16: Retino AID Web APP (gpt-4o) validation



✓ **GPT-4o Validation**

The Grad-CAM heatmap highlights regions in the retinal layers where fluid accumulation or cyst-like spaces are common in Diabetic Macular Edema (DME). The focus on these areas aligns with the typical presentation of DME, indicating that the explanation provided by the model is plausible and clinically relevant.

💬 **Ask Medical Question**

what is the treatment of Diabetic Macular Edema

Ask GPT-4

🔴 GPT-4 Answer:

The treatment for Diabetic Macular Edema (DME) typically involves a combination of approaches. This may include laser photocoagulation, which helps to seal off leaking blood vessels in the retina. Intravitreal injections of anti-VEGF drugs like Avastin, Lucentis, or Eylea can also be used to reduce swelling in the macula. Steroid injections or implants may also be used in some cases. In addition, controlling blood sugar levels and blood pressure is crucial in managing DME. In severe cases, a vitrectomy surgery might be required. It's important to note that the treatment plan is individualized based on the patient's specific condition and needs.

Fig 4.17: Retino AID Web APP (gpt-4) user question answering

4.6 Discussion

There was a need to bridge the deep learning model with the LLM model to improve medical diagnosis, especially in retinal disease classification. This study tried to provide a solution for retinal disease classification. An extensive experiment was conducted on four CNN models (MobileNetV2, ResNet50, VGG16, and DenseNet121). MobileNetV2 outperforms the others and is selected for fusion. The fusion MobileNetV2 model, by combining global and local feature extraction with a fusion mechanism, can learn more complex and more descriptive feature representations. The proposed Fusion MobileNetV2 model can accurately classify the retinal disease with 100% training accuracy, 97.27%

validation accuracy, and 99.55% testing accuracy (see table 4.6). Grad-CAM is used to validate the prediction by highlighting the regions of the image that are responsible for the model's decision. An LLM validates the Grad-CAM with the predicted class and provides summarization. Another LLM provides answers to user questions based on the knowledge base and validation summarization. The results of the Fusion MobileNetV2 model and Grad-CAM with LLM interpretation ensure that this study will be highly valuable for real-time retinal disease classification and diagnosis.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on society

Computer vision, image processing, and large language models (LLMs) have proven significant efficiency and effectiveness in the field of medical imaging. They help in the early diagnosis of diseases, facilitating timely prevention and treatment. This study will help accurately identify retinal diseases and help in real-time diagnosis. Early and accurate disease identification and diagnosis can prevent vision loss. The proposed high-performance model ensures reliable prediction. Moreover, remote or underserved regions have limited diagnostic tools. The RetinoAID web can also help in this area to identify disease from OCT images at an early stage. The Grad-CAM visualization will also help ophthalmologists to understand why a particular disease diagnosis was made. By combining model predictions with LLM-powered explanations, doctors can make more informed decisions. As well as cross-validate AI-generated results with clinical expertise. This study will enhance diagnostic precision and efficiency. It will also support timely medical decision-making and improve patient care through early detection.

5.2 Impact on the environment

The integration of computer vision, image processing, and LLM in OCT disease diagnosis not only improves medical efficiency but also benefits the environment. By enabling accurate and real-time diagnosis, it reduces unnecessary medical tests and lowers energy consumption. Training deep learning models (MobileNetV2, VGG16, ResNet50, DenseNet121) requires significant computational power, increasing electricity usage and carbon footprint. Using lightweight architectures, feature fusion, early stopping, and optimized batch sizes reduces energy consumption during training. AI-driven remote screening allows wider deployment without significantly increasing environmental impact.

It promotes efficient computing, reuse of datasets, and green AI practices within the medical research community.

5.3 Ethical Aspects

The use of AI and computer vision in OCT disease diagnosis raises several ethical considerations. This study ensure that patient OCT images and medical data are securely stored and used only for experiment purposes. This study focused on preventing biases in AI models to ensure accurate diagnosis for all patient groups, regardless of age, gender, or ethnicity. This study clearly defined the process of uses the models to avoid misdiagnosis or misuse of medical information. We also ensure transparency so doctors and patients understand AI decisions. This study ensures to monitor the model performance regularly to detect and correct errors. This study also respects ethical standards in research and clinical practice.

5.4 Sustainability Plan

To ensure that the OCT disease classification and diagnosis system will remain efficient in the long run, a well-defined sustainability plan should be developed. The AI models will be retrained periodically to remain accurate and new diseases will be added to the training data as they are identified. Quality datasets will be gathered in an ethical manner giving the patient privacy and balanced and reliable data to achieve improved model performance. The knowledge base of the system will be updated on a regular basis with the current medical information and physicians will offer advice on new diseases and treatment. Such updates will assist the Large Language Model (LLM) to provide useful and correct answers to questions posed by the users. The system is currently in the form of a web application; however, a mobile app will be created so that users can upload OCT images on their phones and receive real-time diagnostic support. This plan guarantees that the system is accurate, up to date, simple to operate and accessible, and it keeps on improving with new medical knowledge and technology.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the Study

This work is a combination of image classification, feature fusion, explainable AI, and Large Language Models (LLMs) interpretations. This study covers important issues related to detection, diagnosis, and treatment of retinal diseases in their early stages. Four CNN models were analyzed in this study, who demonstrated good results on the OCT disease classification. Based on the experiments, the highest-performing model was chosen. Next, feature fusion was carried out on the chosen model in order to improve its performance. Secondly, K-Fold cross-validation was used to assess the model performance and the model weights of the best-performing fold were stored. Third, the predictions of the best-weighted model were assessed by Grad-CAM. Grad-CAM image and the predicted class were then used as input to an LLM model to confirm the highlighted regions in Grad-CAM with the predicted class. The system prompted the summarization was then passed to another LLM. This LLM gives the answers in terms of equations associated with retinal diseases, based on its knowledge base and logic. Lastly, these assessment procedures were incorporated in a web-based app called RetinoAID.

6.2 Conclusions

This study examines the effectiveness of image processing models with large language models (LLMs) for the early classification and diagnosis of retinal diseases. Four recent CNN architectures (MobileNetV2, ResNet50, VGG16, and DenseNet121) were evaluated on the secondary OCTDL dataset. MobileNetV2 is showing superior performance. To further enhance its capability, a Fusion MobileNetV2 model was proposed by combining global and local feature extraction through a fusion mechanism. This approach enabled the model to better capture complex patterns in OCT images, outperforming the original

MobileNetV2. The 10-fold cross-validation applied to evaluated proposed model performance, achieving the highest performance in the 5th fold with 100% training accuracy, 97.27% validation accuracy, and 99.57% testing accuracy. Furthermore, the best-performing model weights were applied to test data, and Grad-CAM was employed to interpret model predictions. The Grad-CAM outputs with predicted classes were validated by GPT-4o, while GPT-4 responded to user questions from its knowledge base using a retrieval-augmented generation (RAG) pipeline. Finally, the entire framework was deployed in a web application, providing an accessible and interactive tool for assisting in retinal disease diagnosis. By integrating the deep learning model with LLMs, this study enhances early retinal disease classification and diagnosis, while also opening new directions for future research.

6.3 Implication for Further Study & Limitations

This study was conducted using a secondary OCTDL dataset. In future, the work can be extended by utilizing primary retinal disease datasets to ensure broader applicability. The performance of the proposed Fusion MobileNetV2 can also be improved through further layer customization and parameter tuning. Due to time constraints, the LLMs used in this study were limited to knowledge bases derived mainly from medical books related to retinal diseases. However, incorporating expert feedback from ophthalmologists and clinical practitioners could provide more specific and reliable data for fine-tuning the LLM. Additionally, future research could extend this work by evaluating the LLM responses against verified medical knowledge bases to further enhance accuracy and trustworthiness. Third, multi-modal data integration, such as combining OCT images with fundus images or patient demographic data, could enhance disease classification.

REFERENCES

- [1] Zhou, C., Li, S., Ye, L., Chen, C., Liu, S., Yang, H., ... & Xu, X. (2023). Visual impairment and blindness caused by retinal diseases: A nationwide register-based study. *Journal of Global Health*, 13, 04126.
- [2] Chua, J., Tan, B., Wong, D., Garhofer, G., Liew, X. W., Popa-Cherecheanu, A., ... & Schmetterer, L. (2024). Optical coherence tomography angiography of the retina and choroid in systemic diseases. *Progress in Retinal and Eye Research*, 103, 101292.
- [3] Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information*, 16(3), 195.
- [4] Huang, C., Sarabi, M., & Ragab, A. E. (2024). MobileNet-V2/IFHO model for Accurate Detection of early-stage diabetic retinopathy. *Heliyon*, 10(17).
- [5] Dai, H., Yang, Y., Yue, X., & Chen, S. (2024). Improving retinal OCT image classification accuracy using medical pre-training and sample replication methods. *Biomedical Signal Processing and Control*, 91, 106019.
- [6] Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. *Informatics in Medicine Unlocked*, 51, 101587.
- [7] Urooj, B., Fayaz, M., Ali, S., Dang, L. M., & Kim, K. W. (2025). Large Language Models in Medical Image Analysis: A Systematic Survey and Future Directions. *Bioengineering*, 12(8), 818.
- [8] Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in public health*, 11, 1273253.
- [9] Atf, Z., Safavi-Naini, S. A. A., Lewis, P. R., Mahjoubfar, A., Naderi, N., Savage, T. R., & Soroush, A. (2025). The challenge of uncertainty quantification of large language models in medicine. *arXiv preprint arXiv:2504.05278*.
- [10] Şahin, M. E. (2022). A Deep Learning-Based Technique for Diagnosing Retinal Disease by Using Optical Coherence Tomography (OCT) Images. *Turkish Journal of Science and Technology*, 17(2), 417-426.
- [11] Yang, B., Zhang, Z., Yang, P., Zhai, Y., Zhao, Z., Zhang, L., ... & Kuang, C. (2024). MobilenetV2-RC: a lightweight network model for retinopathy classification in retinal OCT images. *Journal of Physics D: Applied Physics*, 57(50), 505401.
- [12] Sajid, M. Z., Qureshi, I., Abbas, Q., Albathan, M., Shaheed, K., Youssef, A., ... & Hussain, A. (2023). Mobile-Hr: An ophthalmologic-based classification system for diagnosis of hypertensive retinopathy using optimized MobileNet architecture. *Diagnostics*, 13(8), 1439.

- [13] Ogundokun, R. O., Awotunde, J. B., Akande, H. B., Lee, C. C., & Imoize, A. L. (2024). Deep transfer learning models for mobile-based ocular disorder identification on retinal images. *Computers, materials and continua.*, 80(1), 139-161.
- [14] Islam, K. T., Wijewickrema, S., & O'Leary, S. (2019, June). Identifying diabetic retinopathy from oct images using deep transfer learning with artificial neural networks. In *2019 IEEE 32nd international symposium on computer-based medical systems (CBMS)* (pp. 281-286). IEEE.
- [15] Sunija, A. P., Kar, S., Gayathri, S., Gopi, V. P., & Palanisamy, P. (2021). Octnet: A lightweight cnn for retinal disease classification from optical coherence tomography images. *Computer methods and programs in biomedicine*, 200, 105877.
- [16] Awais, M., Müller, H., Tang, T. B., & Meriaudeau, F. (2017, September). Classification of sd-oct images using a deep learning approach. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 489-492). IEEE.
- [17] Rasti, R., Rabbani, H., Mehridehnavi, A., & Hajizadeh, F. (2017). Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE transactions on medical imaging*, 37(4), 1024-1034.
- [18] Gour, N., & Khanna, P. (2022). Ocular diseases classification using a lightweight CNN and class weight balancing on OCT images. *Multimedia Tools and Applications*, 81(29), 41765-41780.
- [19] Das, V., Dandapat, S., & Bora, P. K. (2021). Automated classification of retinal OCT images using a deep multi-scale fusion CNN. *IEEE sensors journal*, 21(20), 23256-23265.
- [20] Das, V., Dandapat, S., & Bora, P. K. (2019). Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images. *Biomedical signal processing and Control*, 54, 101605.
- [21] Chan, G. C., Muhammad, A., Shah, S. A., Tang, T. B., Lu, C. K., & Meriaudeau, F. (2017, September). Transfer learning for diabetic macular edema (DME) detection on optical coherence tomography (OCT) images. In *2017 IEEE international conference on signal and image processing applications (ICSIPA)* (pp. 493-496). IEEE.
- [22] Yang, J., Wang, G., Xiao, X., Bao, M., & Tian, G. (2024). Explainable ensemble learning method for OCT detection with transfer learning. *Plos one*, 19(3), e0296175.
- [23] Adel, A., Soliman, M. M., Khalifa, N. E. M., & Mostafa, K. (2020, December). Automatic classification of retinal eye diseases from optical coherence tomography using transfer learning. In *2020 16th International computer engineering conference (ICENCO)* (pp. 37-42). IEEE.
- [24] Bhowmik, A., Kumar, S., & Bhat, N. (2019, May). Eye disease prediction from optical coherence tomography images with transfer learning. In *International Conference on Engineering Applications of Neural Networks* (pp. 104-114). Cham: Springer International Publishing.
- [25] Subramanian, M., Shanmugavadivel, K., Naren, O. S., Premkumar, K., & Rankish, K. (2022, January). Classification of retinal oct images using deep learning. In *2022 international conference on computer communication and informatics (ICCCI)* (pp. 1-7). IEEE.

- [26] Tayal, A., Gupta, J., Solanki, A., Bisht, K., Nayyar, A., & Masud, M. (2022). DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases. *Multimedia systems*, 28(4), 1417-1438.
- [27] Kayadibi, I., & Güraksın, G. E. (2023). An early retinal disease diagnosis system using OCT images via CNN-based stacking ensemble learning. *International Journal for Multiscale Computational Engineering*, 21(1).
- [28] Abirami, M. S., Vennila, B., Suganthi, K., Kawatra, S., & Vaishnava, A. (2022). Detection of choroidal neovascularization (CNV) in retina OCT images using VGG16 and DenseNet CNN. *Wireless Personal Communications*, 127(3), 2569-2583.
- [29] Kim, J., & Tran, L. (2020, July). Ensemble learning based on convolutional neural networks for the classification of retinal diseases from optical coherence tomography images. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 532-537). IEEE.
- [30] Kermany D, Zhang K, Goldbaum M, 2018, Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images, Mendeley Data.
- [31] Kulyabin, Mikhail; Zhdanov, Aleksei; Nikiforova, Anastasia; Stepichev, Andrey; Kuznetsova, Anna; Borisov, Vasili; Ronkin, Mikhail; Bogachev, Alexander; Korotkich, Sergey; Maier, Andreas (2023), "OCTDL: Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods", Mendeley Data, V1, doi: 10.17632/sncdhf53xc.1
- [32] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
- [33] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [34] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [35] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [36] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- [37] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[38] Kumar, K., & Jyoti, K. (2024). Recent Advancements in Grad-CAM and variants: Enhancing Brain Tumor Detection, Segmentation, and Classification.

242-25-033

ORIGINALITY REPORT

14%	9%	9%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dharendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 3", CRC Press, 2025 Publication	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	link.springer.com Internet Source	1%
5	www.nature.com Internet Source	1%
6	web.archive.org Internet Source	1%
7	www.mdpi.com Internet Source	1%
8	Vineeta Das, Samarendra Dandapat, Prabin Kumar Bora. "Automated Classification of Retinal OCT Images using a Deep Multi-Scale Fusion CNN", IEEE Sensors Journal, 2021 Publication	<1%
9	ijcs.stmikindonesia.ac.id Internet Source	<1%