Evaluating the Effectiveness of CNN-Based Models for Diabetic Retinopathy Detection

by

H.M Niaz
19101421
Nuha Tajrian
19101190
Mohammad Ahsan Ibn Alam
19101434
Md. Shahriar Khan Limon
19101444
Sharnit Saha
19101442

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
June 2023

Declaration

It is hereby declared that

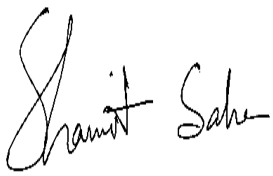
1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



Sharnit Saha
19101442

H.M Niaz
19101421

Nuha Tajrian
19101190

Mohammad Ahsan Ibn Alam
19101434

Md. Shahriar Khan Limon
19101444

Approval

The thesis titled "Evaluating the Effectiveness of CNN-Based Models for Diabetic Retinopathy Detection" submitted by

1. H.M Niaz (19101421)
2. Nuha Tajrian (19101190)
3. Mohammad Ahsan Ibn Alam (19101434)
4. Md. Shahriar Khan Limon (19101444)
5. Sharnit Saha (19101442)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on June 29, 2023.

Examining Committee:

Supervisor:
(Member)

Muhammad Iqbal Hossain, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Associate Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Ethics Statement

In conducting this undergraduate thesis, we pledged to uphold ethical standards by obtaining informed authorization, maintaining confidentiality, adhering to applicable laws and regulations, conducting research with honesty and accountability, and addressing any potential biases or conflicts of interest. This research paper has never been presented in its entirety or part to another university or institution to award a degree.

Abstract

One of the known eye conditions that affect human retinal blood vessels is diabetic retinopathy (DR). People with diabetes are typically at significantly increased risk for this. The blood vessels in the retina are damaged when blood sugar levels in the body increase. Due to the possibility of blindness, people should take precautions and prioritize early detection. As a result, it is a serious condition because it can impair vision. It has several stages, including normal, mild, moderate, severe, and proliferative DR, where it can be quickly determined how severely it has damaged the retinal blood vessels and the area surrounded by the optical disc. Highly qualified specialists typically review the colored fundus photos to diagnose this fatal condition. Clinicians struggle to diagnose this condition accurately, and it takes time. Therefore, several computer vision-based techniques are used to recognize DR and its various stages from retinal scans automatically. These methods, however, can only very roughly categorize the various stages of DR because they are unable to capture the underlying complex properties. However, it is hypothesized that computerized diagnostic systems using intricate Deep Learning (DL) and convolutional neural network (CNN) structures present a compelling approach to learning about different patterns of Diabetic Retinopathy (DR) from fundus images, enabling the precise assessment and categorization of the disease's severity. This study highlights the performance summary of CNN-based models EfficientNetV2B3, EfficientNetV2S, Inception-RestnetV2, MobileNetV2, a fusion model that combines all of these models, and a KNN classifier that uses all of these features that were extracted from each model to improve the classifications of the stages of DR from these optical fundus images. This will consequently give the model's accuracy and a confusion matrix. In addition, we provide an accurate explanation of the performance of the models using ExplainableAI. Here, LIME is used for this purpose.

Keywords: Diabetic Retinopathy (DR), Hybrid model, Fusion model, EfficientNetV2B3, EfficientNetV2S, Inception-ResnetV2, MobileNetV2, feature extraction, KNN classifier, APTOS-2019, DDR_grading, ExplainableAI, LIME.

Dedication

This thesis is dedicated to our parents and friends in recognition of their unwavering support, encouragement, and love throughout our lives. Their sacrifices and unwavering faith in us have been the cornerstone of our accomplishments.

Acknowledgement

We would like to express our heartfelt gratitude to all those who have contributed to the successful completion of this research paper. Their unwavering support, guidance, and encouragement have been invaluable to us throughout this journey.
First and foremost, we would like to extend our deepest appreciation to our supervisor, Muhammad Iqbal Hossain, Ph.D. His expertise, continuous guidance, and insightful suggestions have played a pivotal role in shaping the direction of this research. We are grateful for his patience, unwavering support, and the trust he placed in our abilities. His mentorship has been truly inspiring and has pushed us to achieve excellence.
We are also indebted to our advisors, Rafeed Rahman and Md. Sabbir Ahmed. Their extensive knowledge, critical feedback, and dedication to our project have greatly enriched our research. Their constructive criticism and valuable insights have helped us refine our ideas and methodology, making this paper stronger and more impactful.
Dr. Tamal Kanti Roysarkar, a Senior Consultant in the Ophthalmology department at Bangladesh's ASGAR ALI Hospital, is also contacted for this research. We will be eternally grateful for his guidance. Despite his busy schedule, he makes time for us during this brief period.
Furthermore, we would like to extend our gratitude to our friends and colleagues who have provided valuable assistance and support throughout this endeavor. Their intellectual contributions, brainstorming sessions, and willingness to lend a helping hand have been instrumental in overcoming challenges and pushing the boundaries of our research. Their presence has made this journey enjoyable and memorable.
We are grateful to BRAC University for providing us with the necessary resources and facilities to conduct our research. Access to libraries, laboratories, and other research facilities has been crucial in obtaining the data and information required for this study.
Last but not least, we would like to express our deep appreciation to our families for their unwavering support, understanding, and patience during this research. Their love, encouragement, and belief in our abilities have been the driving force behind our accomplishments.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

In this modern world, people suffer from different types of severe diseases. But most of them are affected by diabetes. And it is considered the cause of multiple illnesses in a single body. As the human body loses control of perfectly protecting the immune system, germs can affect the body. Diabetic retinopathy (DR) is one of these diseases that can harm human health in the long run. It disturbs regular bodily function when the body does not create enough insulin. In this case, it hampers the retinal blood vessels. Blood sugar levels that are excessive disrupt the blood vessels in the retina. Due to the possibility of becoming blind, people should exercise caution. A good number of people are victims of this. Regular retinal screening is necessary for diabetic people to identify and treat DR early on to reduce the risk of blindness[3].

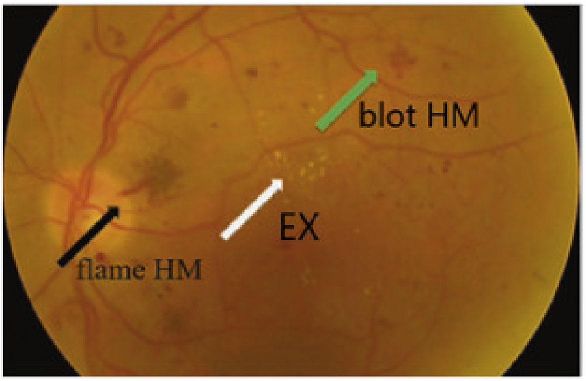


Figure 1.1: Diabetic Retinopathy detection image

Different types of lesions that present on a retinal scan can help identify DR. Microaneurysms (MA), hemorrhages (HM), and soft and hard exudates (EX) are some of these lesions. [4]Due to the fragility of the vessel's walls, microaneurysms (MA), the first symptom of DR, manifest as tiny red circular dots on the retina [1]. Sharp margins and a dimension of less than 125 m are present[8]. Hemorrhages (HM) show up as higher spots on the retina when their size surpasses 125 m and they have an asymmetrical border. Flame (superficial HM) and blot (deeper HM) are the two categories of HM that we are aware of. Hard exudates, typically brought on by plasma leakage, show up as bright yellow patches on the retina. They are located in the outer layers of the retina and have distinct edges. Because the nerve fibers are expanding, soft exudates, often known as cotton wool, appear as white patches on the retina. The form is either oval or circular. Bright lesions are soft and hard exudates (EX), while red lesions are MA and HM. Depending on whether these lesions occur, there are five stages of DR: no DR, mild DR, moderate DR, severe DR, and proliferative DR.

The differences among these stages to detect DR are crucial. Manually diagnosing this condition requires more effort and time. To identify and categorize DR, people are developing automated approaches. Convolutional neural networks, various deep learning techniques, and classifiers built on this foundation have all been the subject of a significant amount of recent research. Numerous fundus and retinal pictures have improved these techniques. They provide a likely positive result on this based on the disease-related factors. Deep learning (DL), a subset of machine learning algorithms, employs hierarchical layers of non-linear processing stages to gather unsupervised features and classify patterns. [5] Classification, segmentation, detection, retrieval, and registration of the pictures are only a handful of the DL applications used in medical image analysis. The detection and classification of DR have recently seen extensive use of DL. Even when numerous disparate sources are combined, it can still successfully learn the characteristics of the incoming data. [6]The CNN design consists of three primary layers: convolution layers (CONV), pooling layers, and fully connected layers (FC). Depending on the author's vision, the CNN has a different size, number of layers, and filters. The CNN architecture has different layers that perform a specific function. Several filters convolve an image to extract the properties of the CONV layers. Usually, the pooling layer comes after the CONV layer to shrink the size of feature maps. The two most popular pooling algorithms are average pooling and maximum pooling. There are many other pooling algorithms as well. The most commonly used categorization function is SoftMax activation. While some studies construct their CNNs from scratch for classification, some investigations transfer learning these pre-trained architectures to accelerate training[9]. The last FC layer, numerous layers, or training all of the pre-trained model's layers are some of the transfer learning methodologies used with these models [13].

The procedure of utilizing DL to recognize and categorize DR images often starts with the dataset that is collected and the necessary preprocessing to augment and refine the images. The DL approach is then used to extract the features and categorize the images from this data.

1.2 Research Problem

Identifying DR manually by ophthalmologists represents a single method, but it is subjective and prone to human error. Consequently, it is necessary to develop and implement a computerized system that outperforms manual detection when it comes to accuracy, effectiveness, and consistency. Detecting DR as early as possible is crucial for preventing eyesight damage. The objective of the research is to highlight the performance of the automated systems that are capable of detecting DR in their earliest phases, thus enabling quick treatment and intervention. The difficulty resides in capitalizing on the distinct advantages of used models and improving their ability to precisely identify DR. This study aims to automate the detection as well as classification of diabetic retinopathy (DR) using a combination of CNN (Convolutional Neural Network) based models EfficientNetV2B3, EfficientNetV2S, Inception-ResNetV2, and MobileNetV2, combined with a K-nearest neighbors (KNN) classifier. Based on retinal and fundus images, the primary objective is to analyze the accuracy of diabetic retinopathy disease detection by introducing two novel hybrid models, EfficentNetV2B3 and EfficientNetV2S, as well as a fusion model with a KNN classifier. The objective is to evaluate the performance of the hybrid models and fusion models and compare them to each other to determine how much better they perform. Moreover, the models EfficientNetV2B3 and EfficientNetV2S are not assessed on the images of diabetic retinopathy, like in the previous works, and there is also no research based on these two models regarding how well they perform on these images. The goal is to merge the results of the CNN models and use the technique of KNN to make the ultimate categorization determination. In addition, the research highlights the importance of preprocessing and data augmentation methods in order to improve the quantity and variety of retinal and fundus images. Effective preprocessing techniques are evaluated and implemented to enhance the overall efficiency of CNN models. To ensure transparency and interpretability, the research analyzes and evaluates the performance of these models using an ExplainableAI approach. The objective is to provide the performance summary and possible reasons for all choices made by the automated system, thereby boosting user confidence and effectiveness. This requires leveraging the abilities of CNN models, incorporating a KNN classifier, improving preprocessing methods, and utilizing ExplainableAI for evaluation of performance. The study seeks to advance the field by evaluating the performance of the accuracy, effectiveness, and interpretability of CNN-based models to detect DR, thus enabling early intervention and preventing vision loss for diabetics.

1.3 Research Objectives

This research revealed that the overall prevalence of VTDR is 4% (range: 3.4–4.8), and diabetic retinopathy is 12.5 percent (95% CI: 11.0–14.2). The prevalence of diabetic retinopathy was 15.5% in people with documented diabetes, compared to 80% in people without the diagnosis. [26] To preserve the patient's eyesight, early detection of DR is crucial. Numerous studies have demonstrated that an early diagnosis of diabetes can prevent DR in 90% of diabetic individuals. [2] An ophthalmologist can manually diagnose DR or use an automated system to do so. Both of these DR detection techniques have advantages and disadvantages. The only advantage of manual detection is that the DR detection procedure does not need computer aid, but it does necessitate that the ophthalmologist be an expert in the field. Sometimes the early symptoms of DR are so subtle that even a skilled ophthalmologist has trouble identifying them. The development of artificial intelligence (AI) has made early disease identification by an automated system more likely and advantageous than manual DR detection. Among the advantages could be a reduction in the ophthalmologist's workload and a lower probability of human error. Additionally, an automated system may be far more effective and easily able to detect lesions and anomalies than is humanly conceivable. Therefore, the automation of DR detection is crucial. These systems can be launched using both deep learning and machine learning techniques. [20] Recently, we have had a better version of these algorithms that has more features at the advanced level. Besides, different types of retinal and fundus images are available on open-source platforms. Some of those image datasets are rarely used to detect and classify DR stages. As a result, we began to analyze these datasets and use the most recent hybrid convolutional neural network models to identify and categorize the various stages of diabetic retinopathy. In this publication, we present a thorough assessment of our research, covering performance analysis, fundus image preparation methods, individual model outcome analysis, and our contribution to the development of our new hybrid model architecture. To improve the readability of our findings, we nourished our research with Explainable AI's (Lime) assistance.

1.4 Thesis Structure

In Section 1.1, we present our motivation for this paper. In Section 1.2, we give our research problem. Section 1.3 will deliver our objectives and contribution. Section 2 will cover the related works. Section 2.1 will add a background, and Section 2.2 will describe the literature review. Section 3 will describe datasets with Section 3.1 (data sample), Section 3.2 (data splitting), and Section 3.3 (data preprocessing). Section 4 will cover the research methodology with Section 4.1 (working progress), and Section 4.2 (used architectures). After that, in Section 5 the result analysis and discussion will cover using models' work and findings, and last but not least, our conclusion part will be added in Section 6.

Chapter 2

Related Works

2.1 Background

Background research is a crucial aspect of any research endeavor, as it provides researchers with valuable insights into past methods and suggestions for future investigations. A literature review enables researchers to comprehend how previous academics approached their research methodologies and what avenues they suggested for further investigation. In addition, it makes it possible for researchers to figure out the different kinds of approaches, algorithms, or extant techniques used in previous studies. This comprehension aids current researchers in developing research plans and anticipating potential challenges associated with the application of these techniques. In the context of our paper, we conducted a thorough literature review and identified various implementations of Convolutional Neural Networks (CNNs) in diabetic retinopathy datasets as described by previous researchers.

2.2 Literature Review

Numerous studies have employed these datasets and models, yielding insightful findings and addressing how they can be used for future research.

In this article [14], the authors discuss the importance of early monitoring for Diabetic Retinopathy (DR) to prevent vision loss, as it is the primary cause of blindness among individuals of working age. They note that automated algorithms based on deep learning have demonstrated positive outcomes in DR screening, attaining a high level of specificity and sensitivity. However, the authors observe that the drawbacks of accessible fundus image datasets limit the efficacy of such models in medical applications. To surmount this limitation, the authors present the DDR dataset, which consists of 13,673 fundus images obtained from 9,598 patients and classified into six classifications depending on the quality of the image and DR level. In addition, 757 images alongside DR-related lesions are annotated for evaluation. To evaluate the efficacy of cutting-edge object detection models, the authors employ the Caffe framework to implement Faster RCNN (Region-based Convolutional Neural Networks), SSD (Single-Shot Multibox Detection), and YOLO (You Only Look Once) architectures. They fine-tune the weights of the models that were trained using the Pascal VOC dataset. While the algorithms achieve an accuracy of 0.8284 on DR classification, they struggle with lesion segmentation and detection, highlighting the difficulty of these aspects. The authors emphasize the need for enhanced segmentation and detection algorithms for lesions. In general, the researchers advance the field by offering the DDR dataset, which allows the evaluation of deep learning models and the investigation of clinical applications, especially in lesion recognition. They describe their plans, which include expanding and enhancing the dataset to include a wider variety of individuals and disease categories. In addition, they intend to design a framework that integrates DR grading and lesion segmentation, recognizing that precise lesion segmentation has the potential to improve classification accuracy given that DR grading is dependent on lesion appearance. To resolve the difficulties associated with the segmentation and detection of lesions in fundus images, the development of superior algorithms is still crucial.

In this paper [22], the authors offer a novel end-to-end system for enhancing images of the retinal fundus that are of poor quality. Inspiring themselves by the best possible transport theory, they suggest an unpaired image-to-image translation method for transporting images of poor quality to their high-quality counterparts. The authors establish a theoretical foundation for the sufficiency of a Generative Adversarial Network (GAN) model that includes a generator and distinction for this task. Their proposed method includes the following components: optimal transport-guided domain consistency, maximal information preservation consistency, and refined data resampling for lesion consistency. To deal with the information inconsistency between the images of poor quality and their improved versions, the authors present an information consistency mechanism to maintain structural consistency between the source and enhanced domains, such as optical discs, blood vessels, and lesions. They carry out perceptual and quantitative assessments on the EyeQ dataset using both no-reference as well as full-reference assessment metrics. Their proposed method outperforms two leading state-of-the-art rivals, showing an advantage in terms of frequently employed image-quality evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Additionally, the authors demonstrate the efficacy of their method for specific task assessment of diabetic retinopathy, demonstrating its potential for supporting clinical diagnoses in real-world settings. The work of the researchers provides a simple yet effective method for enhancing images of the retinal fundus that are of low quality. Utilizing the most effective transport theory and incorporating modules for domain consistency, data preservation, and lesion consistency, their suggested approach outperforms current approaches. The demonstrated gains in the image quality evaluation along with task-driven diabetic retinopathy assessment demonstrate the importance of their work in helping clinicians with real-world clinical diagnoses.

In their research [16], the researchers intend to address the main deficiency of existing models by concentrating on the categorization of every phase of Diabetic Retinopathy (DR), especially the early stages. They underline the growing incidence of diabetes in recent times and emphasize that about 30 percent of patients with diabetes will develop DR. Proliferative Diabetic Retinopathy (PDR) is the final stage of diabetic retinopathy (DR), which encompasses a range of stages from moderate to severe. If the disease is not detected in its early stages, it can cause floaters, impaired vision, and eventually blindness. Manually analysis of DR images is an exhausting and time-consuming endeavor that requires experts with extensive training. To surmount these limitations, the literature has proposed computer vision-based methods for the automatic identification and categorization of DR and its various stages. The researchers suggest a CNN ensemble-based structure for detecting and classifying various phases of DR in color fundus images. They train a combination of five deep CNN models, which are Xception, Dense121, Resnet50, Inceptionv3, and Dense169, using a publicly available Kaggle dataset of retina images. These models are used to represent rich features and improve classification performance at various phases of DR. The findings from experiments show that the suggested approach detects all phases of DR more effectively than existing methods. It also surpasses cutting-edge methods on the same Kaggle dataset. To further improve the accuracy of the detection of the initial stages, the researchers describe their plan to train distinct models dedicated to distinct phases of DR and then merge the results using ensemble techniques. This strategy seeks to enhance the precision of early-stage classification. In conclusion, the researchers present a CNN ensemble-based system for detecting and classifying various phases of DR in color fundus images. Their method outperforms existing techniques, demonstrating its potential to aid in a precise evaluation of DR at different stages. Future research will concentrate on training stage-specific algorithms to improve the accuracy of the initial stage detection.

Ophthalmology has been relying on fundus images for monitoring and diagnosing various eye diseases for a long time. However, the quality of these images can vary considerably because of variations in technology and ophthalmologists' expertise, leading to potential uncertainties and increased misdiagnosis risks when confronted with poor-quality (LQ) degraded fundus images. Thus, the restoration of actual fundus images has become a significant area of research. Despite the importance of this issue, there has been insufficient research into establishing a clinical standard for analyzing restorative methods in actual clinical settings. This paper [23] addresses the issue through the introduction of the Real Fundus (RF) dataset, which consists of 120 pairs of low- alongside high-quality (HQ) fundus images. This dataset seeks to address the problem of scarce data accessibility in the field. Using this dataset as a basis, the authors suggest RFormer, a Transformer-based Generative Adversarial Network, to fix degraded clinical fundus images. The most innovative component of the network's design is the Window-based Self-Attention Block (WSAB), which successfully encapsulates non-local self-similarity and long-range image dependencies. To improve the overall appeal of the recovered images, an algorithm based on a Transformer is implemented. Numerous tests performed on the newly developed clinical benchmark reveal that RFormer outperforms state-of-the-art (SOTA) techniques. In addition, the authors also expand their assessment to subsequent tasks that involve vessel segmentation and optic disc/cup identification, indicating that the proposed RFormer can effectively improve the performance of a variety of medical fundus image evaluation and use tasks. The exhaustive qualitative and quantitative outcomes acquired further substantiate RFormer's superiority over existing methods. Overall, this work not only establishes a baseline for the restoration of actual clinical fundus images but also provides useful insights into the possible application of Transformer-based techniques in this area, which is beneficial to the medical imaging community as a whole.

Fundus image enhancement serves an essential part in the detection and management of ocular diseases, but the presence of multiple complicated factors that diminish image quality presents a formidable obstacle to its accuracy. In this context [30], the authors present Learning Enhancement from Degradation (LED), an innovative diffusion model-based framework for enhancing fundus images. The framework uses a data-driven strategy to degradation to discover links between mismatched high-quality and poor-quality images. By employing the conditional diffusion model, the inverse enhancement procedure is learned in a paired fashion, leading to enhanced fundus images that preserve clinically significant features with enhanced clarity. During the inference phase, the proposed LED framework not only accomplishes impressive enhancement results but also integrates easily and effectively into current fundus image enhancement frameworks. To assess the effectiveness of LED, the researchers conduct experiments on a variety of downstream tasks using a variety of clinically pertinent metrics. Quantitatively and qualitatively, the findings illustrate the advantages of LED across state-of-the-art (SOTA) methodologies. In conclusion, the authors highlight the importance of their suggested diffusion model, the LED, for fundus image enhancement. LED effectively models the distribution of high-quality images through learning enhancing from accessible degradation, allowing for reliable enhancement results. In addition, the authors offer a coarse-to-fine performance enhancement framework that improves the performance of existing SOTA methods. By doing extensive testing on numerous fundus images and evaluation using multiple clinically relevant metrics, the LED method's efficacy is convincingly demonstrated. This work represents an important advancement in the field of fundus image processing and has the potential to improve the accuracy of diagnostics and quality in ophthalmology.

The authors of this paper [11] address the problem of retinal image quality assessment (RIQA) and suggest an approach termed Multiple Color-space Fusion Network (MCF-Net) to enhance the performance of RIQA. They begin by emphasizing the significance of RIQA in ensuring the accuracy of evaluations made by eye specialists or automated evaluation systems. Existing RIQA methods concentrate predominantly on the RGB color space and rely on small amounts of data with binary grade labeling (i.e., 'Accept' and 'Reject'). In order to surmount these limitations, the authors create an enormous Eye-Quality (EyeQ) dataset alongside 28,792 images of the retina from the EyePACS dataset. To evaluate RIQA methods, the EyeQ dataset provides a three-level quality assessment system (i.e., 'Good,' 'Usable,' and 'Reject'). This dataset is distinguished by its vast size, multiple levels of grading, and multi-modality. The authors examine the effects of various color spaces on RIQA and propose the MCF-Net, which combines a representation of various color spaces at the two levels: and prediction level. The MCF-Net is composed of numerous base networks for distinct color spaces (RGB, HSV, and LAB) as well as a fusion block that combines their outputs. The fusion block guarantees the incorporation of color-space depictions while preserving the autonomy and validity of the fundamental networks. In order to effectively train the MCF-Net, the authors maintain all the loss functions of the initial networks, including the fusion loss. To evaluate the efficacy of MCF-Net, tests have been done on the EyeQ dataset. The results demonstrate that the MCF-Net outperforms other methods for deep learning, demonstrating its cutting-edge performance in RIQA. In addition, the authors evaluate the performance of diabetic retinopathy (DR) identification techniques on images of varying quality grades and emphasize the strong dependence of automatic diagnostic systems on image quality. Overall, the paper provides insights into the significance of RIQA, presents an extensive EyeQ dataset using multi-level grading, suggests the MCF-Net for enhanced RIQA performance, and demonstrates the influence of the quality of images on automated diagnostic systems.

The authors of this paper [27] employed DenseNet-169, a densely connected CNN network, for DR detection [4]. The research utilized datasets obtained from the Diabetic Retinopathy Detection 2015 and APTOS 2019 Blindness Detection Kaggle competitions. To improve the image quality of the fundus images, preprocessing included eliminating noise, resizing, and adding Gaussian blur. Using data augmentation techniques, the disparity between the severity classes of DR was addressed. Utilizing the DenseNet-169 architecture, the proposed model obtained a promising 90% accuracy. In addition, a model based on regression was utilized, yielding an accuracy of 78%. Nonetheless, the researchers acknowledged specific constraints and suggested future avenues for advancement. These included the investigation of advanced aggregating methods, collaborative learning with multiple models, the incorporation of additional datasets, and the deployment of the developed system via mobile applications. In conclusion, the study focused on tackling the need for automated early detection of DR. Utilising the latest deep learning techniques and the DenseNet-169 architecture, the researchers have made significant strides in the accuracy of DR detection. The research underlined the potential of CNNs to automate the process of diagnosis and emphasized the significance of preprocessing data as well as enhancement techniques. The paper's discussion of limitations and future potential provides valuable insights for future studies and advancements in this field.

Using image processing techniques, the authors of this paper [17] propose a method for detecting Diabetic Retinopathy (DR), an eye disease associated with diabetes. The primary objective is to develop an effective and user-friendly instrument for eye specialists to detect DR in images of the retinal fundus. The authors use MATLAB (R2010a), a popular image-processing software utility, and an OpenCV (Computer Vision) framework to carry out the proposed method. The paper begins by defining image processing, which entails executing multiple processes on images to improve their appearance and retrieve useful information. Several techniques have been devised in the context of eye disease diagnosis for the early detection of DR based on characteristics such as blood vessels, hemorrhages, and exudates. The authors concentrate on the detection of DR by means of image enhancement techniques, specifically histogram equalization and adaptive histogram equalization. DR is a condition characterized by progressive retinal damage induced by elevated blood sugar levels in diabetic patients. It causes the retinal blood vessels to enlarge and leak, which eventually results in vision loss. The authors emphasize the severity of DR as a sight-threatening disease and offer statistics on the global prevalence of diabetes and DR, with a focus on India. Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR) are included in the classification of DR. NPDR is characterized by the enlargement of retinal blood vessels, whereas PDR is characterized by the formation of aberrant and fragile blood vessels that are novel. DR manifests trouble seeing at night, blurred vision, and black spots within the center of vision. Due to the risk of loss of vision, the authors emphasize the significance of early identification and management of DR. The authors then present their proposed DR detection methodology. They explain the Adaptive Histogram Equalisation (AHE) method, which is used to enhance image contrast. To resolve the drawbacks of AHE, they propose Contrast Limited Adaptive Histogram Equalisation (CLAHE), which prevents excessive noise amplification by limiting contrast enhancement. The proposed method uses CLAHE in conjunction with the OpenCV framework to extract blood vessels from images of the retinal fundus. To demonstrate their findings, they present a testing table containing the accuracy, specificity, and sensitivity outcomes of the proposed procedure on both healthy and diseased fundus images. The results demonstrate a high degree of precision for identifying DR and differentiating healthy fundus images from those with the disease. In conclusion, this work proposes an image processing-based method for DR detection. The authors emphasize the significance of early identification and management of DR and present their method using Adaptive Histogram Equalisation with CLAHE. The results indicate that the proposed procedure for identifying DR in retinal fundus images is effective. This study improves the field of ophthalmology by offering a potential diagnostic instrument for DR that is both efficient and accurate.

In this study [18], the authors examine diabetic retinopathy (DR), the leading cause of blindness in adults around the globe. They emphasize that an extended history of diabetes is an important predictor for the development of DR, accounting for roughly four out of five cases. However, they emphasize the critical fact that the early identification of DR can prevent more than 90 percent of cases that develop from progressing to blindness if the proper treatment is administered. Despite the availability of multiple effective treatments for DR, the authors note that neglect and failure to detect the condition promptly contribute substantially to the blindness of many DR patients. To surmount these obstacles, they investigate the applicability of Digital Image Processing (DIP) and Machine Learning (ML) techniques in the discipline of DR diagnosis. Recent improvements in these fields have paved the path for the creation of machines able to automatically identify the existence of DR using retinal images, the authors note. Yet, the authors admit that several factors degrade the standard of the obtained retinal images, thereby affecting the accuracy of DR results for detection. To address this problem, they suggest an innovative approach for early identification of blindness based on an algorithm for ensemble learning that utilizes spectral data obtained from retinal images. Using a dataset consisting of retinal images collected from people living in remote regions of South Asia, the efficacy of their suggested approach is evaluated. The results of the experiment indicate that the proposed method obtains an average classification accuracy of 91%, indicating its potential utility in detecting the severity of DR-related blindness. The authors observe, however, that the model's efficacy in identifying certain types of blindness remains inadequate, suggesting room for enhancement in future research. Despite its limitations, such as chaotic and mislabeled samples, they emphasize the uniqueness of the dataset used in the present study, as it presents intriguing opportunities for future research. The authors convey their desire to incorporate the images in poor condition that were omitted from this study in future research endeavors. This paper presents a method for detecting blindness based on an ensemble ML algorithm that analyses images of the retina to identify the presence and severity of

DR. The authors emphasize the capability of this method to extract color data from images of the retina and achieve high classification accuracy. They acknowledge the constraints of the dataset utilized and the room for improvement, highlighting the significance of future research in this area.

In this study [33], the authors concentrated on the application of the K-Nearest Neighbors (KNN) algorithm to the diagnosis of Diabetic Retinopathy (DR) using pictures of the retinal fundus. The proposed framework has two major components: attribute selection among selected features followed by normalization as the first component, and iterations with instance-based nearest neighborhood models as the following component. The study employs techniques such as normalization, tuning of parameters, and optimal feature selection to enhance the accuracy of classification of particular methods, like as the algorithm for decision trees and KNN classifiers. The results demonstrate that the KNN classifier obtains the highest accuracy of 81.99%, with an average weighted Receiver Operating Characteristics (ROC) value of 0.907%. The effectiveness of the KNN method contributes to its selection as the most effective classifier in this study. The conclusion emphasizes the efficacy of the suggested framework, which includes filtering and categorization. As a result of attribute selection among filtered features and normalization, the first component obtains an accuracy of 74.59%. In the subsequent component, which employs the KNN technique, the precision is increased to 81.99%. The proposed framework employs straightforward colored histogram filtering for processing images in the context of DR, which represents an innovative addition to the field. Using retinal fundus images, this study indicates the successful implementation of the KNN method for the diagnosis of DR. Using feature selection, normalization, and KNN classification, the proposed framework detects DRs with high precision. The findings of this study offer useful insights into possible future research and contribute to the expanding corpus of literature in the field of DR diagnosis.

In this paper [21], the authors examine the issue of automatic diabetic retinopathy detection, an important cause of blindness in the 20-to-65-year-old population. To solve this issue, they suggest a novel deep-learning hybrid strategy that employs transfer learning to train on the pre-trained Inception-ResNet-v2 framework. The hybrid model is created by adding a block of layers of CNN on top of Inception-ResNet-v2 to improve the model. Two datasets are used to assess the effectiveness of the suggested model: the Messidor-1 diabetic retinopathy dataset and the APTOS 2019 blindness detection dataset from Kaggle. The suggested framework achieves a test accuracy of 72.33 percent on the Messidor-1 dataset and 82.18 percent on the APTOS dataset, which surpasses other previously published results. The authors associate the higher accuracy of their algorithm on the Messidor-1 dataset to the Inception-ResNet-v2 architecture's inception block and additional connections. They emphasize that the positive features of the inception block, in conjunction with the improvements of remaining connections, permit an increase in network breadth while maintaining performance. In addition, the authors propose an additional CNN block motivated by the inception block, that further improves the efficacy of the model. This custom block is comprised of four separate CNN layers with varying filter sizes, and their results are concatenated. The ensemble generated by this method aids in the improved efficiency of the custom block in comparison to standard CNN layers. The paper concludes by introducing the combination of the Inception-ResNet-v2 deep neural network framework for the identification of diabetic retinopathy. The suggested approach integrates transfer learning from Inception-ResNet-v2 via a custom CNN block, resulting in greater accuracy than traditional approaches, especially on the Messidor-1 dataset. In future studies, the authors intend to examine the application of generative adversarial networks, also known as GANs, for data over-sampling as opposed to straightforward augmentation, and they plan to combine collaborative methods with various transfer learning models to enhance performance. This study contributes to the area of diabetic retinopathy detection by suggesting a new deep-learning approach and obtaining better results on two datasets. The authors discuss the benefits of their approach, highlight the importance of a hybrid structure, and propose potential directions for further study.

Chapter 3

Dataset

There are two datasets that we have worked on, one is the APTOS dataset, and another one is DDR_grading.

APTOS:
This is an illustration of the dataset's classification into training, validation, and testing groups for the APTOS 2019 Blindness Detection session. A total of 3662 samples were gathered from numerous individuals in rural India for the 2019 Asia Pacific Tele-Ophthalmology Society Blindness Detection (APTOS 2019 BD) dataset. The Aravind Eye Hospital in India has organized the dataset. Over a lengthy period, a variety of settings and situations were used to take the fundus pictures. The collected samples were then evaluated and labeled following the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRSS) by a team of skilled physicians. The APTOS 2019 BD samples are broken down into five categories according to the scaling system: No Diabetic Retinopathy (DR), Mild DR, Moderate DR, Severe DR, and Proliferative DR. [34]

DDR_grading:
This publicly accessible dataset includes 13,673 fundus images that were shot at a 45-degree FOV and identified to five DR phases. From the collection, 757 photos have DR lesions annotated. [15]

3.1 Data Sample

Class - 0: No Diabetic Retinopathy (DR)

Class - 1: Mild Diabetic Retinopathy (DR)

Class - 2: Moderate Diabetic Retinopathy (DR)

Figure 3.1: Class - 0: No Diabetic Retinopathy (DR) image

Figure 3.2: Class - 1: Mild Diabetic Retinopathy (DR) image

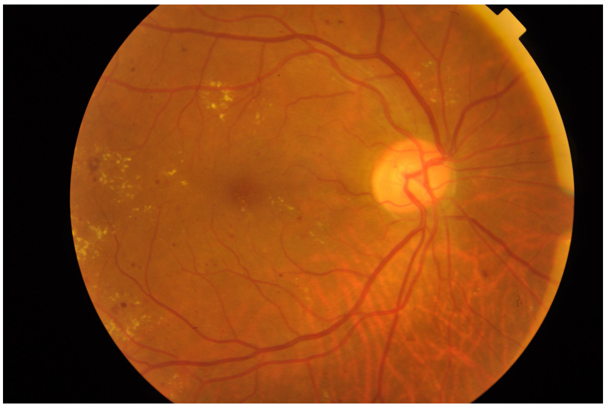


Figure 3.3: Class - 2: Moderate Diabetic Retinopathy (DR) image

Class - 3: Severe Diabetic Retinopathy (DR)

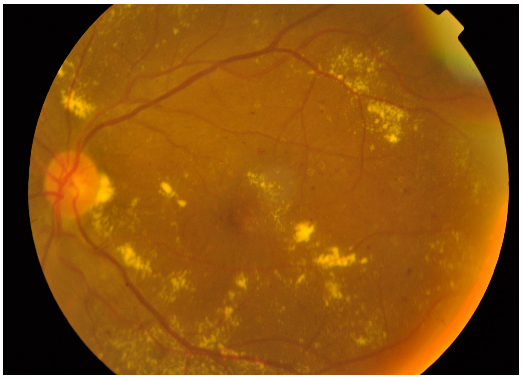


Figure 3.4: Class - 3: Severe Diabetic Retinopathy (DR) image

Class - 4: Proliferative Diabetic Retinopathy (DR)

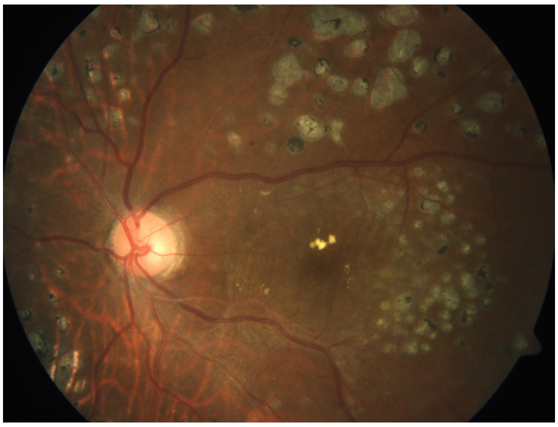


Figure 3.5: Class - 4: Proliferative Diabetic Retinopathy (DR) image

3.2 Data Spliting

In both datasets, we split the raw RGB images into three sectors such as training, testing, and validation. Here, we maintain the ratio consequently 70%, 15%, and 15%.

Training Set: A subset of a data collection that is used to fit (train) a model for the purpose of predicting or classifying variables that are available in the training set but uncertain in other data is referred to as a training set.

Testing Set: In the field of data mining, a testing set is a subset of a data set that is utilized to evaluate the expected future performance of a specific classification or prediction method. This model has been chosen from a pool of competing models on the basis of how well it performed with the validation data.

Validation Set: A data collection of examples that are used to tweak the hyperparameters (also known as the architecture) of a classifier is referred to as a validation data set. In certain circles, it is also referred to as the development set or simply the "dev set."

APTOS Dataset:

Table 3.1: Classification of APTOS Dataset

Class	Train	Test	Validation
0	1127	241	241
1	237	51	51
2	618	132	132
3	118	25	25
4	185	40	40

DDR_grading Dataset:

Table 3.2: Classification of DDR_grading Dataset

Class	Train	Test	Validation
0	3887	832	832
1	348	74	74
2	2729	584	584
3	161	34	34
4	599	128	128

3.3 Data Preprocessing

3.3.1 CLAHE

Images are equalized using Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE is an adaptation of Adaptive Histogram Equalization (AHE) that addresses the issue of contrast over-amplification. Instead of processing the entire image, CLAHE works with discrete sections called tiles. The false borders are then eliminated by combining the adjacent tiles using bilinear interpolation. We can use this algorithm to make photographs' contrast better. We may also use CLAHE on color photos; typically, this executes on the luminance channel, and the outcomes are considerably better for an HSV image after merely adjusting the luminance channel than they are for a BGR image after adjusting all the channels. [31] Block Size (BS) and Clip Limit (CL), two parameters that together constitute the CLAHE, are used. For increased image quality, these two factors are the most important. Since the input image has a very low intensity and a greater CL flattens its histogram, it gets brighter when CL is increased. The dynamic range widens, the visual contrast rises, and the BS rises. When employing image entropy, the two parameters found at the location of the largest curvature of the entropy result in a subjectively high-quality image. [25]

3.3.2 Normalization

The term 'Normalization' in statistics refers to the process of reducing the size of the data collection so that the normalized data ranges from -1 to 1. Comparing matching normalized values from two or more data sets using this normalizing technique is helpful. [35]

3.3.3 Green Channel Extraction

The retinal pictures typically have little contrast. The green channel of the image is initially extracted during preprocessing. Because of the extreme contrast in the green channel, microaneurysms are easily visible. To improve the green channel's contrast, preprocessing is used. The RGB and HSV channels are then independently lit equally and then concatenated to produce the illumination-equalized image. The green channel is isolated from the concatenated RGB, and CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to it to enhance the contrast at the tile level. The contrast transform function is calculated for each tile separately. [12]

3.3.4 Resize

We get various kinds of pixels and measure image data through both of the datasets. Moreover, their dimensions differ. To train the models we developed, we resized all image datasets in 512x512. Because in this measurement, the machine can work on the factors of detecting Diabetic Retinopathy stages. If we use larger than this, it gets complex and shows errors. And if we give smaller than this, it can not properly detect the factors because it cuts the important area of its factors in an image.

3.3.5 Augmentation

Image augmentation is typically performed as a preliminary step prior to model training. During training, augmented images are fed to the model alongside the original images, effectively expanding the training dataset. This augmented dataset assists the model in learning robust and invariant features, thereby enhancing its ability to generalize well to untested or undeployed data. Image augmentation entails a variety of transformation techniques that alter an image's visual appearance while preserving its semantic content. We use 10-degree rotation, vertical and horizontal flipping, and altering contrast or brightness transformation techniques here. For the APTOS dataset, in the augmentation part, upsampling was done to balance the data for each class. As class 0 contains enough data so there were made no augmentations in this class. In classes 1, 2, 3, and 4, data was augmented from 226 to 1000, 567 to 1000, 102 to 900, and 176 to 842 respectively for training. For test and validation, in classes 1, 2, 3, and 4, data was augmented from 50 to 210, 122 to 220, 21 to 190, and 38 to 179. For the DDR dataset, here also upsampling is done to balance the data for each class in the augmentation part. Since class 0 contains enough data so there were no changes as well in this class. In classes 1, 2, 3, and 4, data was augmented from 348 to 3885, 2729 to 3885, 161 to 3877, and 600 to 3885 respectively for training. For test and validation, in classes 1, 2, 3, and 4, data was augmented from 74 to 832, 584 to 834, 37 to 825, and 128 to 832. As the test and validation ratio was 15:15, so the number of images in the test and validation after augmentation are more or less the same.

Preprocessing steps for our model

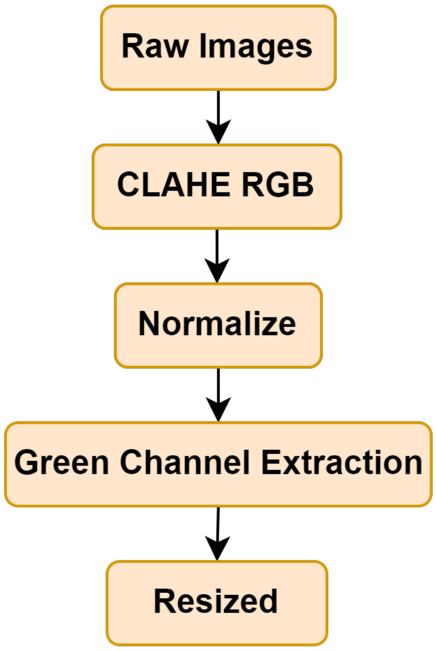


Figure 3.6: Preprocessing steps for our model

After preprocessing we get:

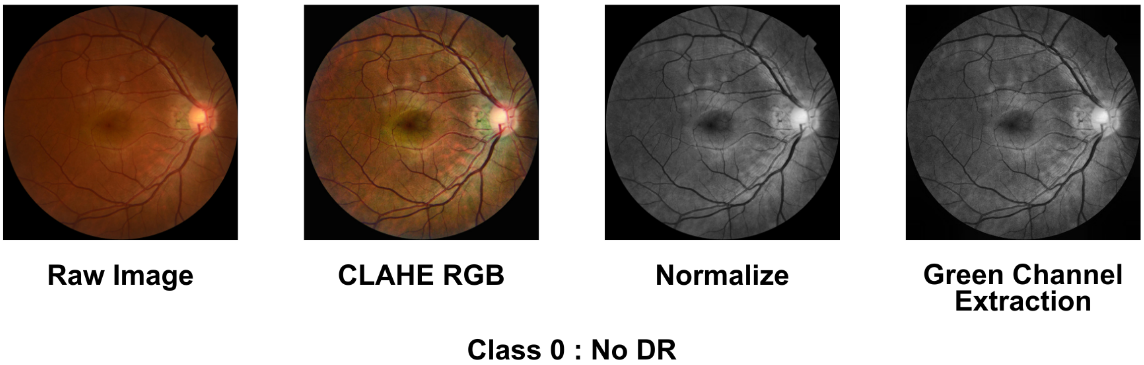


Figure 3.7: Preprocessed an image Class-0

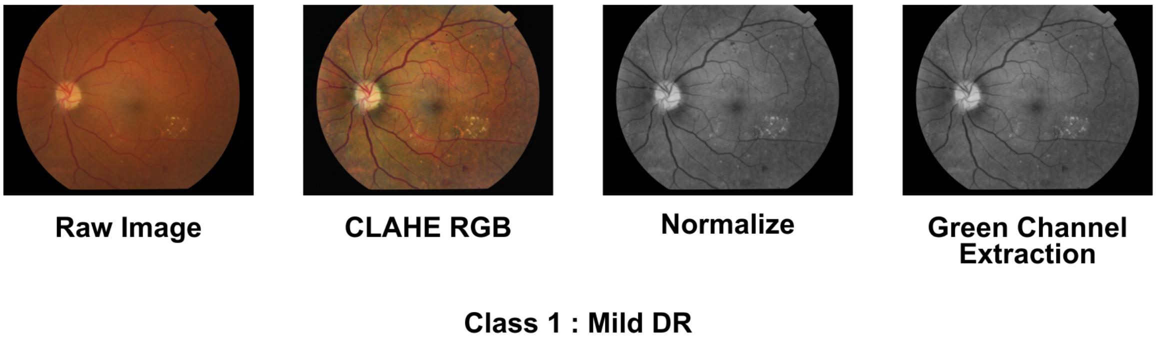


Figure 3.8: Preprocessed an image Class-1

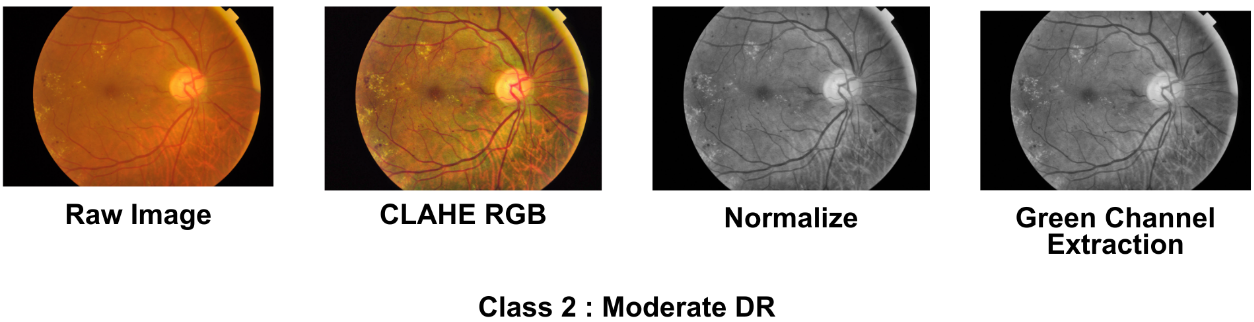


Figure 3.9: Preprocessed an image Class-2

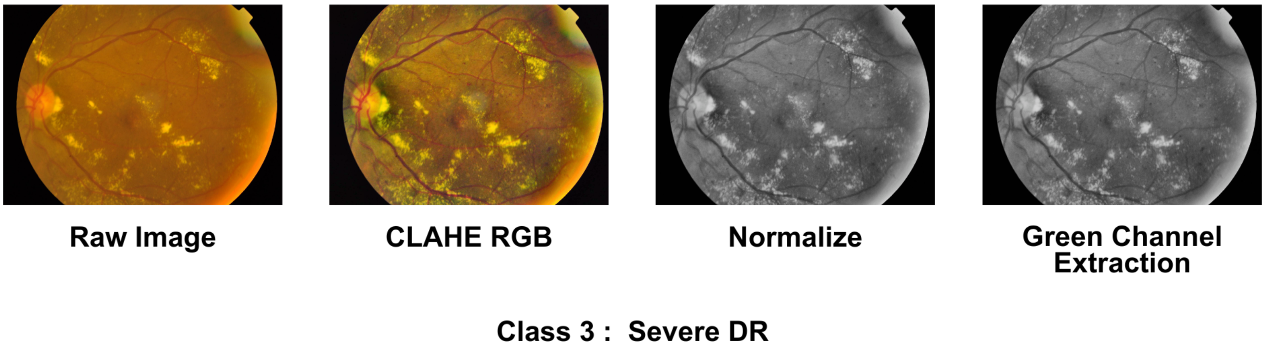


Figure 3.10: Preprocessed an image Class-3

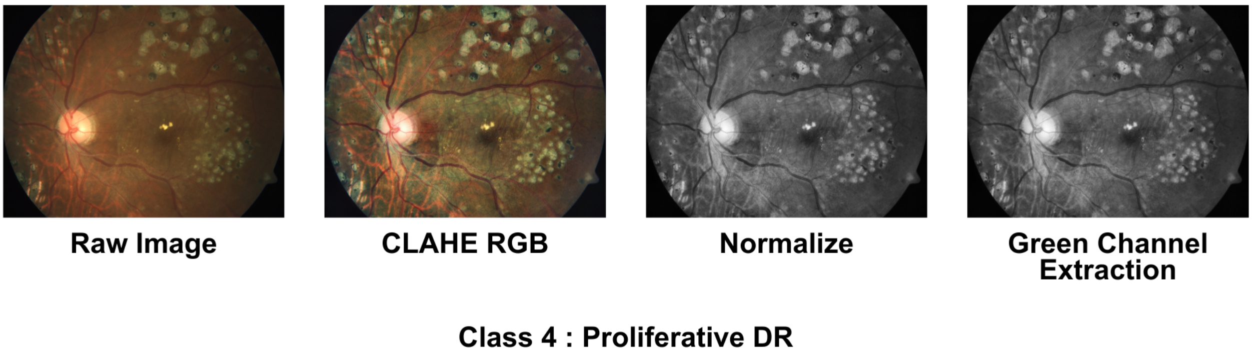


Figure 3.11: Preprocessed an image Class-4

Chapter 4

Research Methodology

4.1 Working Process

We collect a good number of datasets of fundus images from an open-source platform. These are RGB images. The inner, posterior surface of the eye is known as the fundus. It is made up of blood vessels, the macula, the retina, the optic disc, and the fovea. Fundus photography involves capturing images of the back of the eye using a specialized fundus camera that is pointed through the pupil. These images assist eye doctors in identifying, monitoring, and treating disease. [19]

Our research paper examines two distinct datasets which are APTOS and DDR Grading dataset. We begin by acquiring the unprocessed RGB images. Then, we utilize several preprocessing techniques, such as CLAHE in RGB images, then to normalize the data from -1 to 1, after that extract the green channel, and finally resize the image into 512x512 measurements. It is important to note that as both of our datasets are perfectly imbalanced, we augment the data to balance the five classes. In the augmentation part, we up sample the imbalanced classes to increase the number of images of those class. After conducting these experiments, we divided both datasets into 70% train, 15% validation, and 15% test sets, using the corresponding percentages. Here, four distinct CNN-Based model architectures are implemented, namely EfficientNetV2B3, EfficientNetV2S, Inception-RestnetV2, and MobileNetV2. We use training and validation data from both datasets separately to train these model architectures and test data to monitor the performance evolution of each model. Moreover, features are extracted from test datasets in the same way as train and validation data. In addition, we create a fusion model by removing the lowest layers up to the Global Average Pooling layer from the architectures of the mentioned model. It is because in the global average pooling layer all the features are exposed. After that, for subsequent work, we concatenate these features according to class wise from each model and split them into train, validation, and test sections. We train the KNN classifier using training and validation features and then evaluate its performance using test portion features.

Consequently, this fusion model will provide the most precise forecast. Here, KNN operates on the features. It uses the function Euclidean distance from the query point to find the nearest neighbors to select the best pixels for defining the correct parts. By doing so, we can predict the outcome more accurately. We employ three case types. We are examining the raw RGB imbalance datasets for Case-1. In Case 2, all preprocessing techniques are applied to these raw rgb imbalance datasets but here data remains imbalanced. Case 3 includes pre-processing techniques and the application of data augmentation to balance all classes in both datasets. For each scenario, we present the individual performance of each model and analyze their results for the two datasets mentioned and finally analyzed the result with performance of the fusion model with KNN classifier that is mentioned.

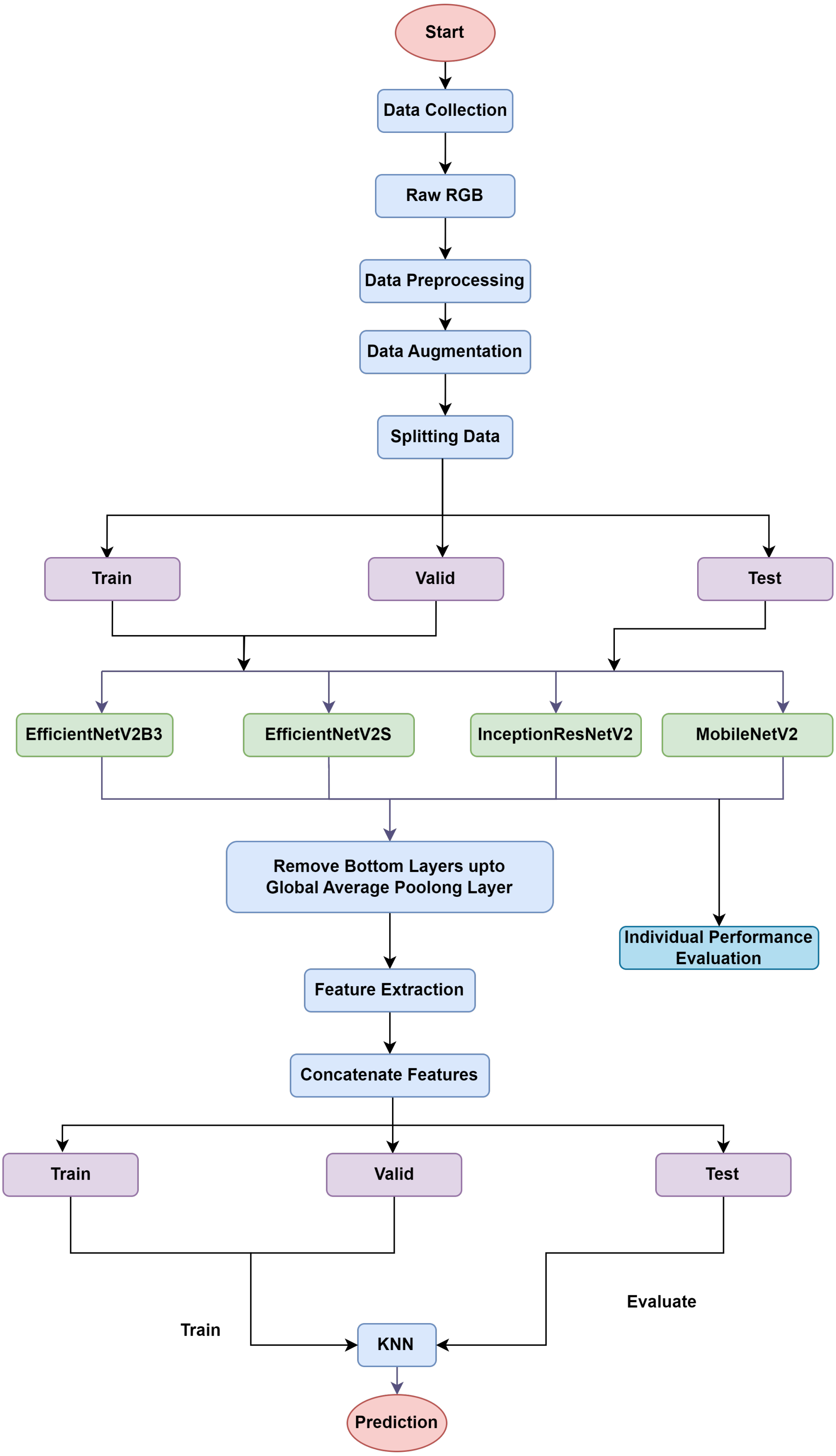


Figure 4.1: Model-architecture diagram

4.2 Used Architectures

For our research, we have used EfficientNetVB3, EfficientNetV2S, Inception-ResnetV2, and MobileNetV2. Besides, we have used KNN and ExplainableAI(LIME). These models are specially for the classification of images. The EfficientNet model's EfficientNetV2 variant has replaced the original version. Its goal is to increase the efficiency and precision of computer vision activities.[32]

4.2.1 EfficientNetV2S

In EfficientnetV2, With the help of the new 'Compound scaling' scaling technique introduced, the model's depth, width, and resolution had been optimized. The EfficientNetV2S is an EfficientNetV2 that is smaller and more productive than the EfficientNetV2. It has 224 million features, which is a significant number for a computer vision system. Despite being 6.8 times more compact than the original EfficientNetV2 model, the EfficientNetV2S model can be trained 11 times faster. To do this, the search space is constrained, and also model optimizations like the Fused-MBConv operation are applied. EfficientNetV2S performs admirably in mobile and embedded systems, which need both speed and efficiency. EfficientNetV2S has been identified as an appropriate fit for execution in mobile and embedded devices, as well as other contexts where speed and efficiency are crucial factors. The utilization of this approach may serve as a foundation for additional investigation into the development of efficient deep-learning models. EfficientNetV2S is an innovative iteration of EfficientNetV2 that offers enhanced compactness and speed compared to its predecessor. [36]

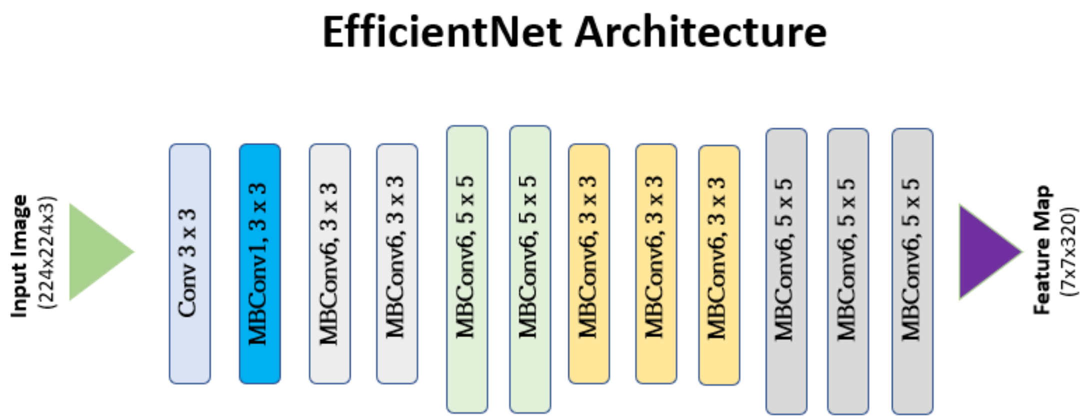


Figure 4.2: EfficientNet-architecture diagram

4.2.2 EfficientNetV2B3

On the other hand, EfficientNetV2B3 is a deep neural network design that was created especially for applications like picture categorization and is quite effective. By incorporating more complex design modifications and novel educational strategies, it surpasses the accomplishments of the EfficientNet and EfficientNetV2 models.

EfficientNetV2B3 has been significantly altered by the switch from the ReLU activation function to the Swish activation function. Swish outperforms ReLU in terms of precision and convergence speed. As a result, EfficientNetV2B3 can now discern intricate patterns in photos considerably more easily. EfficientNetV2B3 was created using both supervised and unsupervised learning techniques. As a result, the model can learn from high-level features and unlabeled data without the need for human annotations. The model also uses cutting-edge data augmentation techniques like RandAugment and CutMix, which boost its stability and broaden its application in a variety of situations. These techniques change the training data, which facilitates the model's ability to adapt to changes in position, illumination, and other factors that occur in the real world.

Since it is small and makes optimal use of those resources The model is suited for deployment on devices with limited access to such resources. For the categorization of images, EfficientNetV2B3 has undergone rigorous testing on a range of datasets and has consistently produced the best results while retaining a small model size. It outperforms its predecessors on benchmarks like ImageNet, CIFAR-10, and CIFAR-100, showing that it is more accurate and able to distinguish a wider range of objects and settings than its forerunners were. The model is suited for deployment on devices with limited access to such resources since it is small and makes optimal use of those resources. As a result, real-time photo categorization is now possible in situations when latency is a problem.

There are many real-world uses for the improvements made by EfficientNetV2B3, including self-driving automobiles, medical imaging, surveillance systems, and industrial quality control. It is a great option because it can combine precision and speed, especially for edge devices with limited resources. EfficientNetV2B3 provides a substantial advancement in photo categorization. To strengthen computer vision applications for academics and experts working in the field it is a special tool.[29]

4.2.3 Inception- ResnetV2

The Inception-ResNet-V2 is a convolutional neural network (CNN) architecture that fits the best parts of the Inception and ResNet models. In deep neural networks, dissipating gradients are a well-known problem. As gradients move through the network, they tend to get smaller and smaller, making it harder for earlier layers to learn. The Inception-ResNet-v2 model is made to deal with the problems that come with training very deep neural networks. In the ResNet design, the problem of gradient degradation is fixed by adding residual connections, which allow gradients to flow freely throughout the network

The Inception-ResNet-V2 model is an extension of the Inception architecture. It uses different filter sizes to record both local and global features in an image. Using residual links within the Inception modules makes it easy for gradients to spread smoothly. For effective training of networks with a lot of depth, it is important to use leftover connections, which are affected by skip connections in ResNet.

The Inception-ResNet-V2 architecture is made up of a stem block and many Inception-ResNet sections that come after it. The central block is in charge of extracting the first set of features, while the Inception-ResNet parts help the network learn more complex and abstract representations. The modules are made up of a mix of Inception blocks and ResNet blocks that are connected by residual links.

Inception architecture uses Parallel convolutional operations with different filter sizes. This lets the model catch a wide range of features. When this alternate structure is put into place, it makes it easier for the network to learn about both local and global representations. The ResNet architecture has skip connections that make it easier for the network to take over residual tasks. This trait makes it easier for the network to learn how to map identities. Through the synergistic combination of the Inception and ResNet architectures, the Inception-ResNet-V2 model has done very well on many test datasets, including ImageNet. The architecture has shown greater precision and better gradient flow within deep networks, making it a powerful framework for picture classification and other computer vision tasks. [7]

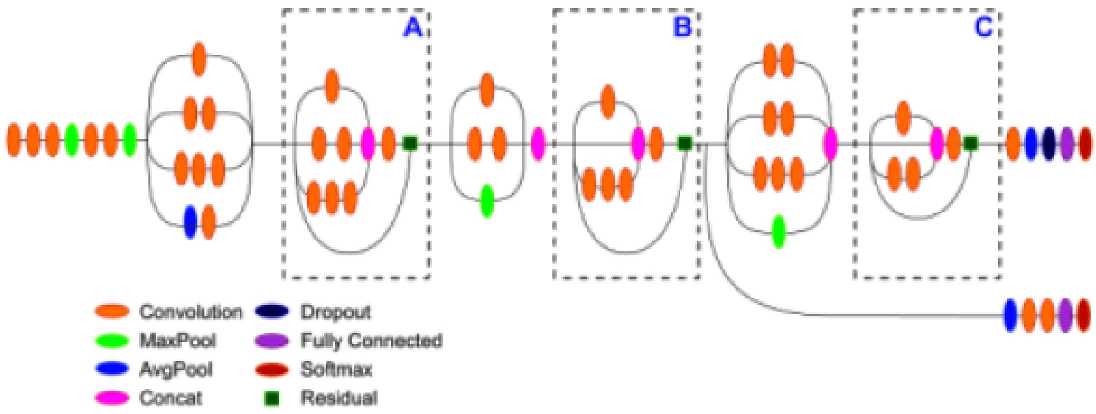


Figure 4.3: Inception-Resnet-V2-Architecture images

4.2.4 MobileNetV2

MobileNetV2 is a special convolutional neural network (CNN) software. The idea behind the design was to make it easier for smartphones and other devices to recognize and sort images more accurately. The goal of designing MobileNetV2's system was to find a balance between accuracy and efficiency. The goal of this study is to

come up with a system that can correctly identify objects while using as little memory and processing power as possible on different devices. The main ideas behind MobileNetV2 are the following:
Putting inverted residuals and linear bottlenecks into MobileNetV2 is a smart way to improve its performance. The technical language makes it sound like the software uses smart ways to encode pictures with fewer numbers and computations. The process includes making a puzzle less complicated so that the device can do its work as quickly as possible. By using this method, MobileNetV2 is able to do image recognition jobs with high accuracy while still being able to process information quickly. Tensor Processing Units (TPUs) have been tested with MobileNetV2 to see how well it works with specialized hardware. The chips were made to help with tasks that have to do with machine learning. MobileNetV2 works better because it makes changes that only affect the TPU. In this work, a method that uses both depthwise separable convolutions and 1x1 grouped convolutions was used. You can get better results if you use TPUs to their fullest limit.[10]

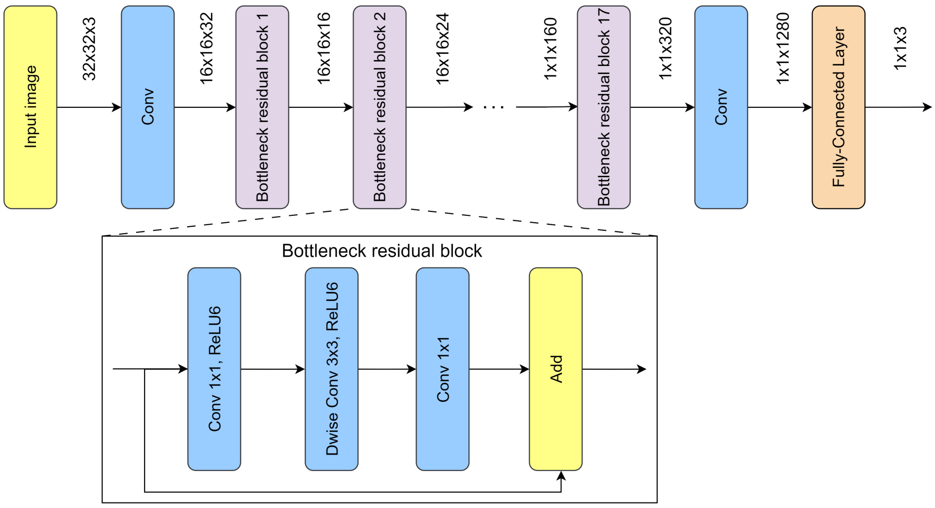


Figure 4.4: MobileNetV2

4.2.5 KNN

The k-nearest neighbor (KNN) algorithm is an algorithm for supervised machine learning that is used predominantly for tasks involving classification. Due to its versatility and simplicity, it has gained popularity in disease prediction. Based on the characteristics and labels of the training data, the algorithm predicts the categorization of unlabeled information. In KNN, the method classifies the query instance by taking into account the k proximal training data points, that are chosen by considering how close they are to the query. The number k indicates the 'nearest neighbors' that must be evaluated. After identifying the closest neighbors, the algorithm uses a simple majority selection rule to decide the query's ultimate classification. The simple design and understandability of the KNN algorithm are among its assets. It can manage datasets of different sizes, levels of noise, label counts, categories, and contexts, which makes it applicable to a vast array of classification problems. Moreover, the algorithm permits adaptations and changes to surmount its constraints and enhance precision. The traditional KNN algorithm has limitations that can affect its classification efficiency. It may, for instance, be impartial towards all classification-dependent neighbors, miss characteristics for estimating distances between points of data, and consider superfluous dataset features. However, due to its adaptability, the KNN algorithm has undergone numerous modifications to address these challenges. Such KNN variants vary in computational aspects, including the optimization of the k parameter, the improvement of distance calculations, the assignment of weights to distinct data points, and the truncation of the training datasets. These alterations seek to enhance the method's classification abilities and address its limitations. In conclusion, the KNN classifier is a supervised artificial intelligence algorithm utilized for tasks such as classification, specifically in predicting diseases. It finds the k closest neighbors from the training data to feed the search instance and uses the vote of the majority to figure out the final category. The method's accessibility and adaptability have made it a popular method, along with its drawbacks, which are capable of being mitigated through a variety of alterations and variations.

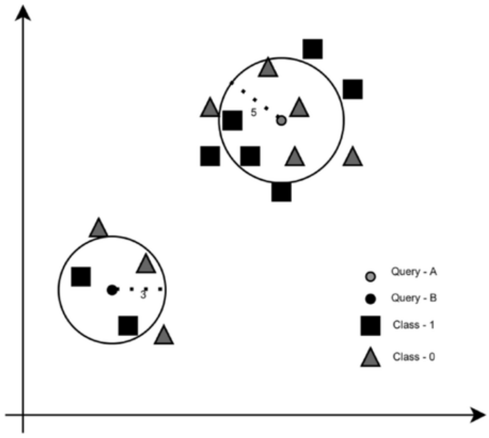


Figure 4.5: Visual illustration of the KNN algorithm

The example in Figure 1 shows how the KNN algorithm operates. Consider Query B with k = 3. The algorithm determines the class designations of Query B's three closest neighbors after identifying them. In this instance, two of the closest residents are members of Class 1, while one is a member of Class 0. Using the majority selection criterion, the algorithm assigns Query B to Class 1 because it is the class that occurs most frequently among its closest neighbors. Similarly, the value of k for Query A is set to 5. After locating the five closest neighbors, the algorithm finds that more than half of them belong to Class 0. Based on the majority selection criterion, Query A is classified as Class 0 as a result. [28]

4.2.6 ExplainableAI

ExplainableAI is characterized as artificial intelligence systems that provide a rationale for their predictions. ExplainableAI is a subset of the broader general designation for AI known as 'Interpretability'. Interpretability enables us to comprehend what a model is acquiring, the other facts it provides, and the rationale behind its recommendations in the context of the real-world problem we're attempting to solve. When model metrics are inadequate, interpretability is necessary. Model interpretability permits us to predict how a model will perform under various test conditions by comparing the training environment to the test environment. ExplainableAI Systems can be useful when attempting to comprehend the justification behind a particular machine learning model's prediction or decision. Shapley Additive Explanations (SHAP) and LIME are tools for explaining the decisions made by a machine learning model using local interpretations. Here, we are constructing our models with Lime.

LIME:
Regional Interpretable Model-Agnostic Explanations is the abbreviation for LIME.[24] LIME is an open-source guide created by Carnegie Mellon University researchers. It can be utilized to clarify the anticipated outcomes of any machine learning model, including credit scoring models and autonomous vehicles. LIME operates by disrupting the input data and observing the effect on the model's output. LIME attempts to fit a locally interpretable model that is compatible with the output of the original model, f(x), in the neighborhood surrounding the given example x.

Chapter 5

Result Analysis and Discussion

5.1 Experimental Setup

Our experiments are done on the thesis lab PC. Here are the hardware configurations.

Table 5.1: Experimental Setup

CPU	Intel core i7 12th Gen
GPU	Nvidia GeForce RTX 3080ti
Storage	1 TB SSD
RAM	64 GB

5.2 Result Analysis

Two datasets were utilized in this study. The APTOS-2019 dataset and DDR_grading are displayed here. These two datasets are perfectly balanced across five categories. Consequently, we utilize these datasets in three specific ways. In case one, we train and validate the imbalanced raw datasets without employing any preprocessing techniques. In the second case, we experimented with various preprocessing techniques to determine the performance of these models. The final step involved balancing the dataset in each classification using augmentation, followed by the application of preprocessing techniques to obtain the final results.

5.2.1 APTOS-2019 Dataset

Case-1: We tried with imbalanced raw data without doing any kind of pre-processing techniques. Here, we get -

The accuracy of the predictions is high, but not because of accurate results. Instead, it is mainly due to a high number of false predictions. Therefore, while the overall

Table 5.2: Case-1

Models	Accuracy
EfficientNetV2B3	80%
EfficientNetV2S	81%
Inception-ResnetV2	78%
MobileNetV2	73%
Fusion model	82%

predicted outcome may seem average, it is not in a positive sense.

Case-2: We tried different methods of preprocessing in these imbalanced datasets. Here we get-

Table 5.3: Case-2

Models	Accuracy
EfficientNetV2B3	79%
EfficientNetV2S	82%
Inception-ResnetV2	82%
MobileNetV2	78%
Fusion model	83%

In this section, this structure produces results that are equivalent to the first case. However, these predictions are incorrect, leading to a sort of skewed outcome in stages 0 and 2.

Case-3: We balanced these datasets with augmentations in 4 classifications except 0 classifications. Then apply preprocessing techniques and train these model architectures to get the output results.

Table 5.4: Case-3

Models	Accuracy
EfficientNetV2B3	84%
EfficientNetV2S	86%
Inception-ResnetV2	83%
MobileNetV2	85%
Fusion model	87%

After augmentations, we trained our model and balanced all categories. Utilizing preprocessing techniques on well-balanced data yields superior results.

Comparative Better Result on EfficientNetV2S architecture in Individual models

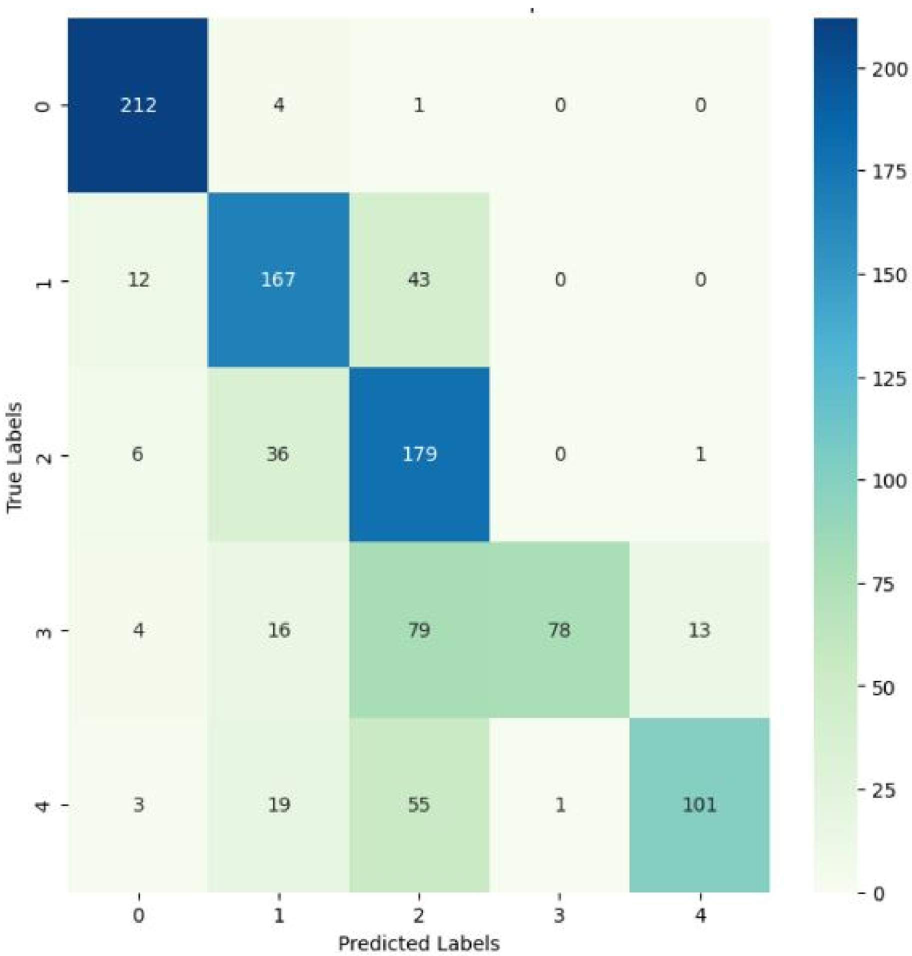


Figure 5.1: Result on EfficientNetV2S architecture

Result of Fusion KNN Model on APTOS

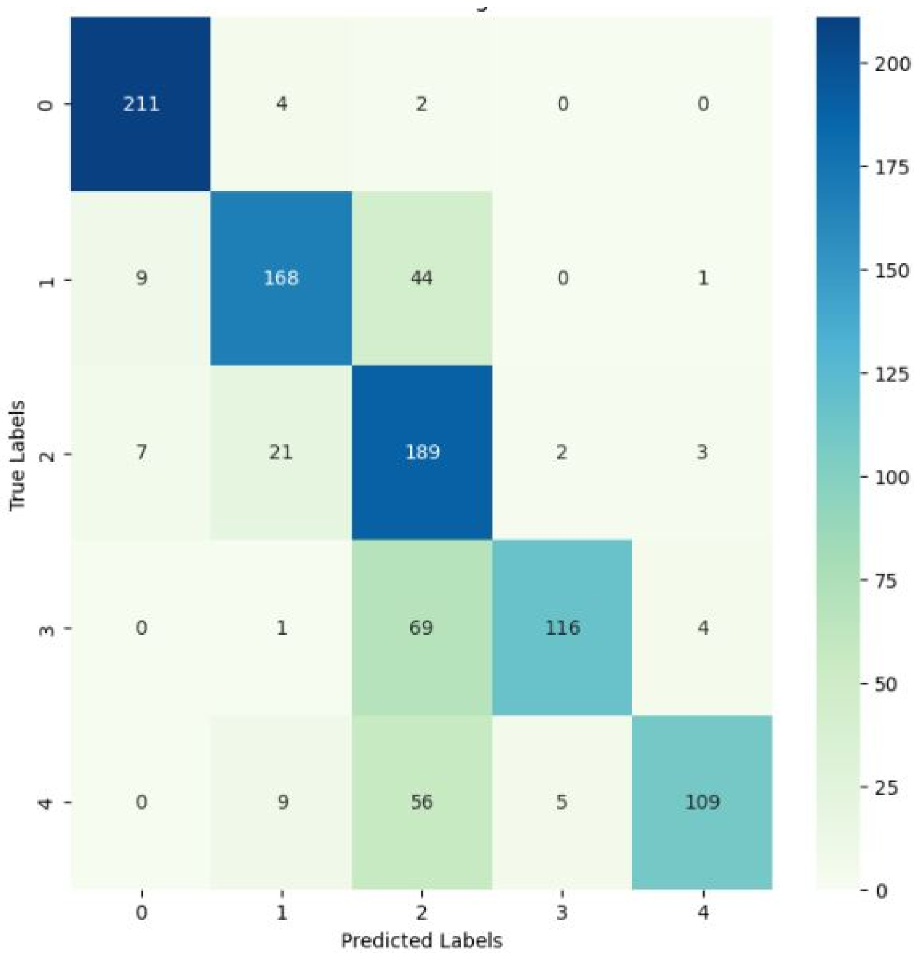


Figure 5.2: Result of Fusion KNN Model on APTOS

5.2.2 DDR_grading Dataset

Case-1: We tried with imbalanced raw data without doing any kind of pre-processing techniques. Here, we get -

Table 5.5: Case-1

Models	Accuracy
EfficientNetV2B3	60%
EfficientNetV2S	68%
Inception-ResnetV2	67%
MobileNetV2	66%
Fusion model	71%

In this section, it is apparent that even with this imbalanced raw dataset, the aforementioned architectures produce not-so-good results. It exhibits a good number of false predictions. This can accurately identify only two of the five classifications. The remainder of the classification is filled with unfavorable findings.

Case-2: We tried different methods of preprocessing in these imbalanced datasets. Here we get-

Table 5.6: Case-2

Models	Accuracy
EfficientNetV2B3	67%
EfficientNetV2S	84%
Inception-ResnetV2	79%
MobileNetV2	64%
Fusion model	73%

In this section, this framework gives equivalent results rather than the first case. These predictions, however, are erroneous, resulting in a sort of skewed outcome in stages 0 and 2.

Case-3: We balanced these datasets with augmentations in 4 classifications except 0 classifications. Then apply preprocessing techniques and train these model architectures to get the output results.

After augmenting our model, we trained it and balanced each category. Using preprocessing techniques on well-balanced data produces superior outcomes.

Table 5.7: Case-3

Models	Accuracy
EfficientNetV2B3	79%
EfficientNetV2S	81%
Inception-ResnetV2	79%
MobileNetV2	78%
Fusion model	85%

Comparative Better Result on EfficientNetV2S architecture in Individual models

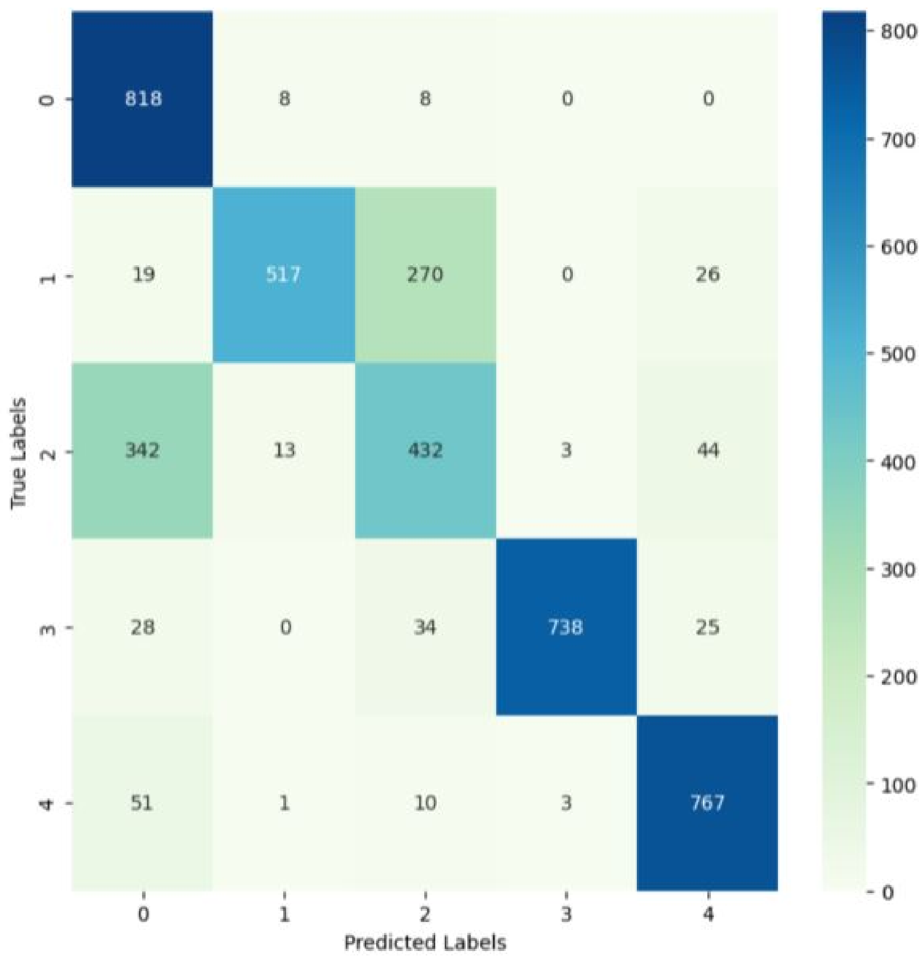


Figure 5.3: Result on EfficientNetV2S architecture

Result of Fusion KNN Model on DDR_grading

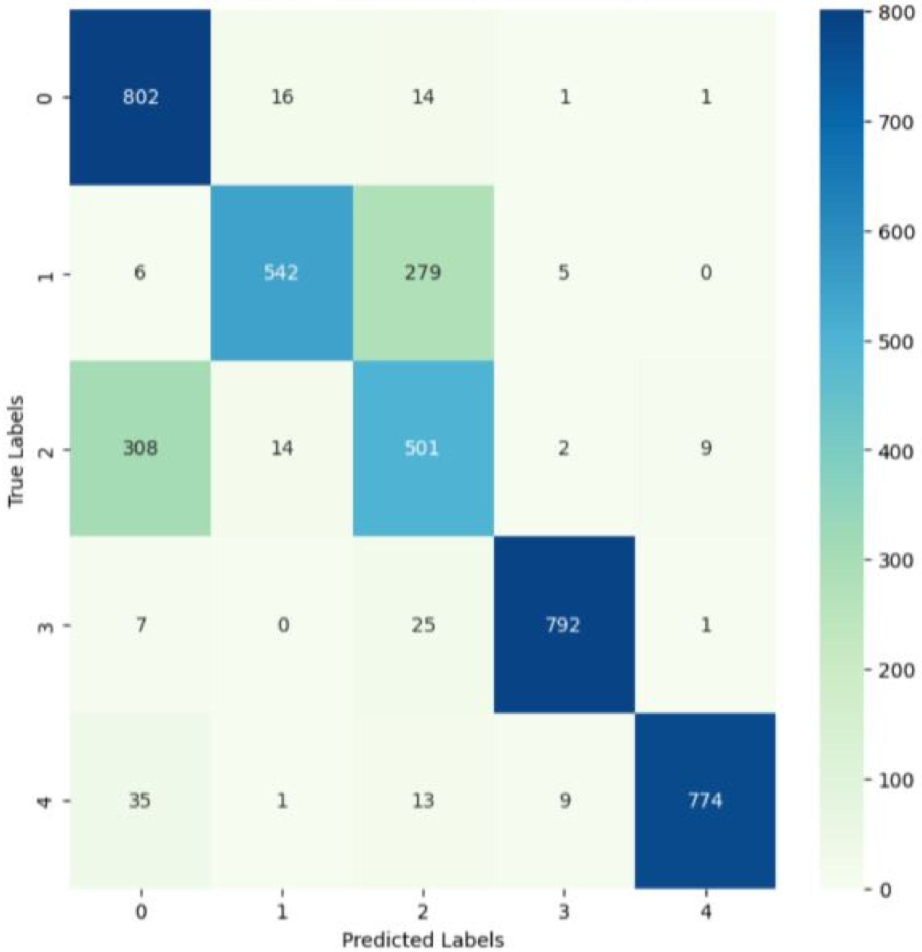


Figure 5.4: Result of Fusion KNN Model on DDR_grading

5.3 Discussion

From the above result analysis, we can see how the individual models and the fusion models are performing in each case. After evaluating these models using the APTOS-2019 and DDR grading datasets, we have determined that case 3 which is balanced and preprocessed data provides better performance than the other two cases. In each of the three instances, our fusion model, from which all features were extracted, yielded better outcomes than the individual models. However, this is since, in case 3, we must first balance all five stages of Diabetic Retinopathy (DR) as both datasets that we have mentioned are perfectly imbalanced data. Then, we employ a variety of preprocessing techniques to make the factors more visible and identify the stages. Due to the severity of this disease, it is essential to determine whether or not a patient is currently experiencing symptoms. In the interim, it will be simple to obtain the appropriate treatment if the disease is detected early. Here, we are working with distinct hybrid model variants as well as a fusion model that combines the features that were extracted from each model. After the training and validation phases, we can evaluate its performance based on its outputs. Our two datasets are completely imbalanced and somewhat convoluted. In some images, it is extremely challenging to distinguish between each class. Therefore, we begin by examining this raw imbalance data, where accuracy is high but predictions are inaccurate which is based on the false predictions as the data were highly imbalanced. Then, we apply a variety of preprocessing techniques, including CLAHE, Green channel extraction, and normalization from -1 to 1. After preparing this data, we train our models to determine the performance characteristics of these architectures. Comparatively superior to the previous model, it still predicts inaccurate outcomes as data were imbalanced in each class. Finally, we apply augmentation to each dataset stage and balance the classes. In this instance, we train and validate these models with this data to determine their performance. From three instances, case 3 is showing better performance as here both datasets are balanced along with preprocessed. Moreover, if we see at the individual model performances, the EfficientNetV2S model is showing better performance among all the individual models in all three instances. However, now if we come to the fusion model of all four models, in every instance, the Fusion model with KNN classifier provides better performance (in terms of accuracy) than the individual hybrid model architectures. After examining each step, we can conclude that EfficientNetV2S performs well in individual model tests, whereas the KNN fusion model performs better overall. After applying augmentation techniques to the APTOS dataset, there is a clear enhancement in accurately predicting class 3 instances. This improvement is accompanied by a notable decrease in false predictions. Likewise, there is an increase in accurately predicting class 4 instances, albeit to a lesser extent, and a slight reduction in false predictions in the fusion model in comparison to the better-performing of four models. The limited availability of data in class 4 poses a challenge for achieving higher accuracy. Nevertheless, the utilization of augmentation techniques has proven to be advantageous in improving predictions for both classes. It is worth noting for the DDR dataset that the high number of false predictions in class 2 could be attributed to the presence of a considerable amount of flawed images from other classes. Despite attempts to augment the dataset, increasing the number of flawed images might be inevitable. Additionally, distinguishing the precise correct class for labeled images poses a challenge, particularly when there are subtle differences between class 1 and class 2. Consequently, this situation could contribute to a higher occurrence of false predictions.

5.4 Findings

When working with these two datasets, we encounter a variety of challenges. Those challenges are described below.

5.4.1 Image Quality:

The image dataset we used includes low-quality images with various issues. These images suffer from low contrast, with areas that are excessively dark or bright, causing color imbalances. Some images also exhibit a light green tint. Moreover, the presence of black shadows further hampers the visibility of crucial blood vessels, which are essential for detection purposes. Furthermore, a significant portion of the dataset consists of flawed images due to the capturing process. Certain images are affected by reflective lights, making it challenging to identify the factors associated with diabetic retinopathy. Additionally, the overall poor quality of the images presents difficulties in visually distinguishing the specific factors that can be detected by the relevant class.

5.4.2 Imbalanced Dataset:

The dataset we have used APTOS-2019 and DDR grading both datasets are completely imbalanced in each class. From these diabetic retinopathy datasets, in both datasets class 0 contains a huge amount of images whereas the other classes contain very less comparatively. Therefore, it is preferable to provide biased results in the first instance which is imbalanced raw data. In APTOS and DDR grading dataset, we see class 1, 3 and 4 contains very less amount of images which are resulting in false predictions. Moreover, the model can detect mostly 0 and 2 classes with relative ease, but 1, 3, and 4 classes are predicted incorrectly. As these two classes contain a substantial number of images in raw format, the majority of the model's performance is dependent on these components. Therefore, the first case demonstrates better performance resulting in biased predictions. However, the perfectly imbalanced data is one of the core reasons why the new models EfficidentNetV2B3 and EfficientNetV2S are not showing optimal high accuracy individually.

5.4.3 Focused Area With ExplainableAI:

Lastly, as we already mentioned about Explainable AI, we use this to visualize the area that the model is learning from the images. However, when ExplainableAI is used to describe the actual process of identifying stages through the retina's eye factors, it is a noticeable fact that the model we used, is incapable of performing some of the additional required tasks. Even though we have expert reviews on how to work on collective factors to detect DR stages, these models were incapable of learning from the additional factors of the images and producing accurate results in these instances. In our case, we visualize the model's mechanism using lime. In this image, we centered the machine's attention on the factors necessary to identify DR classes in the left image and the right image illustrates how our model's architecture prioritizes DR class detection factors. To ensure its operational quality, we consulted a physician about how to detect DR classes practically and on which significant factors they focus to detect DR. According to Dr. Tamal Kanti Roysarkar, a senior consultant in the Ophthalmology department at the ASGAR ALI Hospital, they are also inquiring about additional factors that the machine was unable to detect the additional factors that they mentioned. As in the right-left picture, according to the senior consultant, they focus on the blood vessels of the upper side and the lower side horizontally that is marked, the left side of the optical disc and the near about area around the optical disc, and the macula of the retina. In the right picture of Figure 5.5, we showed the visualization of the learning area that our used model focuses on and we noticed that some additional factors like the left side of the optical disc area and macula are not focused by our used model architectures. The possible reasons we found out from our research that is there may be some kind of similarity in the images in those areas for which the models are not focusing on those mentioned areas and the models are focusing on the different areas. Which, consequently, the models are not performing and gaining an optimal outcome in this sector. These images were extracted from the efficientNetV2S training mechanism.

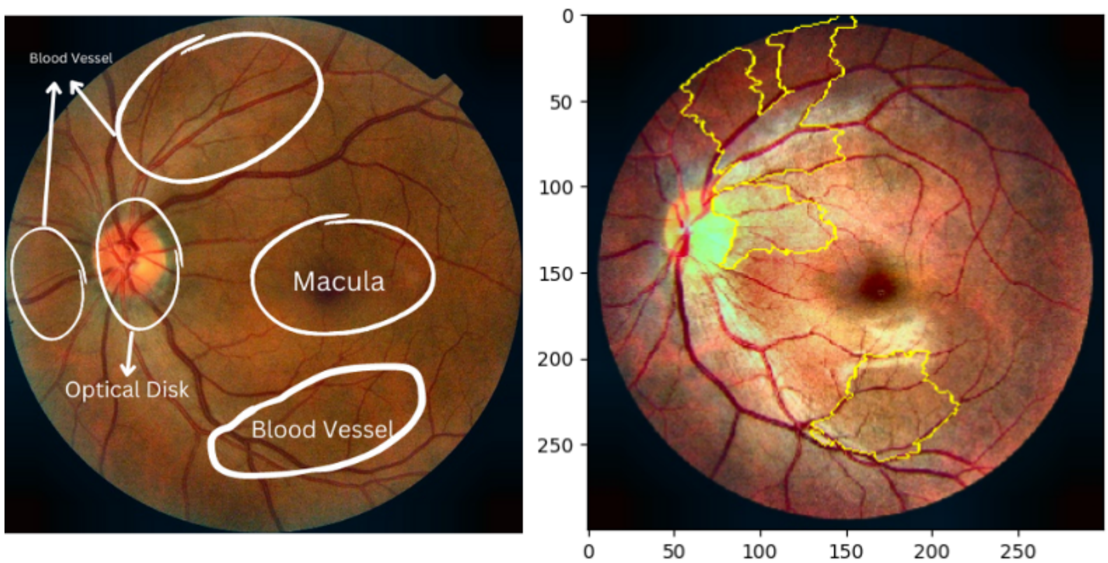


Figure 5.5: Before & After Apply Lime in an image

Chapter 6

Conclusion

In this study, we evaluate the ability of CNN-based models to detect diabetic retinopathy (DR), a serious eye disease that, if not diagnosed early, can cause vision loss and eventual blindness. In addition, traditional methods of specialist diagnosis are time-consuming and error-prone. To address this issue, we have developed computer vision-based methods for automatically detecting and classifying DR stages from retinal imaging scans. In addition, we have utilized numerous CNN architectures, including EfficientNetV2B3, EfficientNetV2S, Inception-ResNetV2, and MobileNetV2, as well as a KNN classifier. These models are trained using a dataset of retinal images with varying DR stages. In addition, accuracy and confusion matrices were used to evaluate our results. Nonetheless, poor image quality and imbalanced class variations in the datasets present obstacles. Despite these obstacles, our models demonstrate promising results in detecting specific DR stages after preprocessing. Correctly predicting classes for a substantial number of raw images remains a challenge. It is crucial to note that the efficacy of our models is highly dependent on image quality and specific class components. We have also implemented an Explainable AI to comprehend the factors that contribute to DR stage detection and the causes of incorrect outcomes. Even though CNN-based models have the potential to automate DR detection, additional research and development are necessary to overcome the obstacles identified in this study. Our work contributes to ongoing efforts to identify and classify the earliest stages of DR using computer vision and deep learning. Such models can aid medical professionals in making accurate diagnoses, resulting in prompt treatment and better outcomes for diabetic retinopathy patients, thanks to technological advancements and refined methodologies.

Bibliography

[1] E. T. D. R. S. R. GROUP, "Grading diabetic retinopathy thy from stereoscopic color fundus photographs- an extension of the modified airlie house classification," *Ophthalmology*, 1991. DOI: https://doi.org/10.1016/S0161-6420(13)38012-9.

[2] P. Vashist, S. Singh, N. Gupta, and R. Saxena, "Role of early screening for diabetic retinopathy in patients with diabetes mellitus: An overview," en, *Indian J. Community Med.*, vol. 36, no. 4, pp. 247–252, Oct. 2011.

[3] J. K. C.A. Harper, "Diabetic retinopathy management guidelines," *Expert Review of Ophthalmology*, 2012. DOI: https://doi.org/10.1586/eop.12.52.

[4] D. B. R. Taylor, "Handbook of retinal screening in diabetes: Diagnosis and management," 2012.

[5] L. deng, "Apsipa trans. signal inf process," *A tutorial survey of architectures, algorithms, and applications for deep learning*, 2014.

[6] X. L. X.W. Chen, "Big data deep learning: Challenges and perspectives," *IEEE Access*, 2014. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6817512.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, *Inception-v4, inception-resnet and the impact of residual connections on learning*, 2016. arXiv: 1602.07261 `[cs.CV]`.

[8] P. H. Scanlon, "Screening for diabetic retinopathy," in *Practical Manual of Diabetic Retinopathy Management*. John Wiley Sons, Ltd, 2017, ch. 4, pp. 40–53, ISBN: 9781119058984. DOI: https://doi.org/10.1002/9781119058984.ch4. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119058984.ch4. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119058984.ch4.

[9] R. K. M.T. Esfahani M. Ghaderi, "Leonardo electron j pract technol.," *Classification of diabetic and normal fundus images using new deep learning method*, 2018.

[10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," Jun. 2018, pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

[11] H. Fu, B. Wang, J. Shen, *et al.*, "Evaluation of retinal image quality assessment networks in different color-spaces," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, *et al.*, Eds., Cham: Springer International Publishing, 2019, pp. 48–56, ISBN: 978-3-030-32239-7.

[12] G. Indumathi and V. Sathananthavathi, "Chapter 5 - microaneurysms detection for early diagnosis of diabetic retinopathy using shape and steerable gaussian features," in *Telemedicine Technologies*, H. D. Jude and V. E. Balas, Eds., Academic Press, 2019, pp. 57–69, ISBN: 978-0-12-816948-3. DOI: https://doi.org/10.1016/B978-0-12-816948-3.00005-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128169483000052.

[13] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2019.06.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519305377.

[14] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2019.06.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519305377.

[15] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2019.06.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519305377.

[16] S. Qummar, F. G. Khan, S. Shah, *et al.*, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150 530–150 539, 2019. DOI: 10.1109/ACCESS.2019.2947484.

[17] S. S.A, M. S. B., and R. Naveen, "Diabetic retinopathy detection using image processing," *International Journal of Engineering and Advanced Technology*, vol. 8, pp. 937–941, Aug. 2019. DOI: 10.35940/ijeat.F1179.0886S19.

[18] N. Sikder, M. S. Chowdhury, A. Shamim Mohammad Arif, and A.-A. Nahid, "Early blindness detection based on retinal images using ensemble learning," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1–6. DOI: 10.1109/ICCIT48885.2019.9038439.

[19] May 2020. [Online]. Available: https://www.aao.org/eye-health/anatomy/fundus.

[20] U. Ishtiaq, S. Abdul Kareem, E. R. M. F. Abdullah, G. Mujtaba, R. Jahangir, and H. Y. Ghafoor, "Diabetic retinopathy detection through artificial intelligent techniques: A review and open issues," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15 209–15 252, Jun. 2020, ISSN: 1573-7721. DOI: 10.1007/s11042-018-7044-8. [Online]. Available: https://doi.org/10.1007/s11042-018-7044-8.

[21] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence*, V. Bhateja, S.-L. Peng, S. C. Satapathy, and Y.-D. Zhang, Eds., Singapore: Springer Singapore, 2021, pp. 679–689, ISBN: 978-981-15-5788-0.

[22] W. Wang, F. Wen, Z. Yan, and P. Liu, "Optimal transport for unsupervised denoising learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 2104–2118, 2021.

[23] Z. Deng, Y. Cai, L. Chen, *et al.*, "Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4645–4655, 2022. DOI: 10.1109/JBHI.2022.3187103.

[24] A. Kumar, *What is explainable ai? concepts amp; examples*, Mar. 2022. [Online]. Available: https://vitalflux.com/what-is-explainable-ai-concepts-examples/.

[25] P. Marimuthu, "Image contrast enhancement using clahe," en-US, *Analytics Vidhya*, Aug. 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/08/image-contrast-enhancement-using-clahe/.

[26] "Prevalence of diabetic retinopathy in india stratified by known and undiagnosed diabetes, urban–rural locations, and socioeconomic indices: Results from the smart india population-based cross-sectional screening study," *The Lancet Global Health*, vol. 10, no. 12, e1764–e1773, 2022, ISSN: 2214-109X. DOI: https://doi.org/10.1016/S2214-109X(22)00411-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214109X22004119.

[27] N. Shaikh, H. Tawde, S. Vishwakarma, and D. K. K. Tripathi, *Blindness avoid (diabetic) retinopathy detection*, Mar. 2022. DOI: 10.22214/ijraset.2022.41017. [Online]. Available: http://dx.doi.org/10.22214/ijraset.2022.41017.

[28] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of k-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, Apr. 2022. DOI: 10.1038/s41598-022-10358-x. [Online]. Available: https://doi.org/10.1038/s41598-022-10358-x.

[29] K. Azam, M. A. Azam, M. A. Qureshi, K. B. Khan, and M. A. Azam, "Efficientnet aspp deep network for malignant ultrasound breast cancer segmentation," in *2023 IEEE International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEST)*, 2023, pp. 1–6. DOI: 10.1109/ICEST56843.2023.10138837.

[30] P. Cheng, L. Lin, Y. Huang, H. He, W. Luo, and X. Tang, "Learning enhancement from degradation: A diffusion model for fundus image enhancement," 2023. DOI: 10.48550/ARXIV.2303.04603. [Online]. Available: https://arxiv.org/abs/2303.04603.

[31] GeeksforGeeks, "Clahe histogram equalization opencv," en-us, *GeeksforGeeks*, May 2023. [Online]. Available: https://www.geeksforgeeks.org/clahe-histogram-eqalization-opencv/.

[32] A. Sarkar, *Efficientnetv2: Faster, smaller, and higher accuracy than vision transformers*, May 2023. [Online]. Available: https://towardsdatascience.com/efficientnetv2-faster-smaller-and-higher-accuracy-than-vision-transformers-98e23587bf04.

[33] T. Vijayan, M. Sangeetha, A. Kumaravel, and B. Karthik, “Feature selection for simple color histogram filter based on retinal fundus images for diabetic retinopathy recognition,” *IETE Journal of Research*, vol. 69, no. 2, pp. 987–994, 2023. DOI: 10.1080/03772063.2020.1844082. eprint: https://doi.org/10.1080/03772063.2020.1844082. [Online]. Available: https://doi.org/10.1080/03772063.2020.1844082.

[34] en. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection/data.

[35] A. K. Srivastav, “Normalization formula,” [Online]. Available: https://www.wallstreetmojo.com/normalization-formula/#:~:text=In%5C%20statistics%5C%2C%5C%20the%5C%20term%5C%20%5C%E2%5C%80%5C%9Cnormalization,two%5C%20or%5C%20more%5C%20data%5C%20sets.

[36] K. Team, *Keras documentation: Efficientnetv2 b0 to b3 and s, m, l*, en. [Online]. Available: https://keras.io/api/applications/efficientnet_v2/.