Automated Detection and Classification of Diabetic Retinopathy using Deep Learning Approach

Master of Science

In

Machine Learning and Artificial Intelligence

By

Kapil Madan

Thesis Supervisor – Karthick Kaliannan Neelamohan

ACKNOWLEDGEMENTS

The thesis presented here is a culmination of the support, guidance and generosity of those around me without which this work would not have been possible. When I first started this research, I did not know where to start and where it was going to take me by the time it ends.

I would like to express my special appreciation and thanks to my thesis supervisor, Karthick Kaliannan Neelamohan for encouraging and supporting me throughout my research. Your advice has been a critical part from the very beginning to the end of this work in different aspects of analysis and presentation of this work.

I would also like to extend my gratitude to the faculty of Liverpool John Moores University – Dr. Manoj Jayabalan, Dr. Rik Das and Dr. Ahmed and to Upgrad for their dedicated and invaluable guidance and expert academic support.

I am also grateful to my organization, VMware for supporting and funding my higher education for this research work and for providing me an opportunity to grow as a research scientist.

A special thanks to my mother and father. My words cannot express how grateful I am for all the sacrifices that they have made on my behalf and for motivating me to strive towards my goal. I wouldn't have followed this academic path without the values and examples set for me by my parents.

My gratitude for all the open-source community. Culture of collaboration is key for our success as a civilization.

Without all of them this research thesis would not be possible.

ABSTRACT

Diabetic Retinopathy (DR) is an eye complication and is the major cause of blurred vision and blindness worldwide. Early detection of DR is imperative to prevent vision loss in population suffering from diabetes. The detection of DR is done through retinal fundus images and is usually performed by ophthalmologists at clinical facilities who identify the presence and significance of many subtle features, a process which is cumbersome, time consuming and prone to human errors thus resulting in longer waiting time between early diagnosis and treatment. Additionally, the detection process requires advanced technologies which might not be available in villages or remote rural areas. Automated detection of diabetic retinopathy has potential benefits such as increased efficiency and reproducibility, wide coverage of screening programs, reducing barriers to clinical access, and improving patient outcomes by providing early detection and treatment. Artificial Intelligence (AI) based deep learning techniques are one of the powerful methods and has become a state-of-art providing much more faster and efficient results than manual system for automated analysis of images in various sectors. It is also providing promising results in healthcare domain on medical images for tasks like classification, segmentation and object detection. The primary objective of this research will be to develop, compare, evaluate robust and computationally efficient deep learning (DL) algorithms for early detection and classification of DR stages. This research aims to explore DL methods based on convolutional neural network (CNN) with and without transfer learning. For transfer learning, the pre-trained model developed are - VGG-16, ResNet50 and DenseNet121. Additionally, the research also evaluates and compares performance of Capsule Networks (CapsNet) with CNN's.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DR Diabetic Retinopathy.

AI Artificial Intelligence.

DL Deep Learning.

CNN Convolutional Neural Networks.

CapsNet Capsule Network.

UK United Kingdom.

NPDR Non proliferative diabetic retinopathy.

PDR Proliferative diabetic retinopathy.

MA Microaneurysms.

LDA Linear Discriminant Analysis

SVM Support Vector Machine

KNN K Nearest Neighbor

GPU Graphical Processing Unit

ANN Artificial Neural Network

ILSVRC ImageNet Large Scale Visual Recognition Challenge

TP True Positive

TN True Negative

FP False Positive

FN False Negative

KD Kaggle Dataset

WHO World Health Organization

CAD Computer Aided Detection

RFI Retinal Fundus Images

LRF Low Resolution Fundus

HRF High Resolution Fundus

RGB Red Green Blue

CLAHE Contrast Limited Adaptive Histogram Equalization

ROC Retinopathy Online Challenge

CAM Class Activation Map

ReLU Rectified Linear Unit

OCT Optical Coherence Tomography

1. INTRODUCTION

1.1 Background of the Study

Diabetic retinopathy (DR) is one of the major causes of blindness in the world for people in working age who are suffering from diabetes (Pratt, 2019). Approximately 420 million people worldwide have been diagnosed with diabetes. The prevalence of this disease has doubled in the past 30 years. Of those with diabetes, approximately one-third are expected to be diagnosed with DR, a chronic eye disease that can progress to irreversible vision loss (Lam et al., 2018). DR occurs due to the existence of microvascular damage to blood vessels of the light sensitive tissue (retina) at the back of the eye. There are two primary stages of DR – Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). In NDPR, narrow bulges called microaneurysms (MA) and other abnormalities including, hemorrhages, exudates, venous beading etc. can be identified. Large retinal blood vessels also start dilating and become irregular in diameter and as more vessels become blocked, NPDR progresses from mild to moderate and then to severe with presence of more abnormalities. Figure 1 depicts retinal features of DR. In PDR, the damaged blood vessels leak a transparent jelly-like fluid that fills the center of the eye causing the development of abnormal blood vessels in the retina. Pressure builds up in the eyeball because of the newly grown blood vessels that interrupt the normal flow of the fluid which damages the optic nerve and leads to blindness. The classification rule of DR is given in Table 1.

It is imperative that people suffering from diabetes are screened for earlier signs of DR so that more severe cases could be prevented . However, detection of DR is a challenging task as it has minor symptoms at its nascent stages those symptoms are very hard to detect. Additionally, the detection process is time consuming and costly that needs to take place at skilled clinical facilities under the supervision of trained experts. Due to these challenges, there could be variations in classification across clinics depending upon the grader which might lead to delayed and inefficient diagnostic mechanism. Automated detection of DR through machine learning and artificial intelligence-based methods could overcome all these challenges.

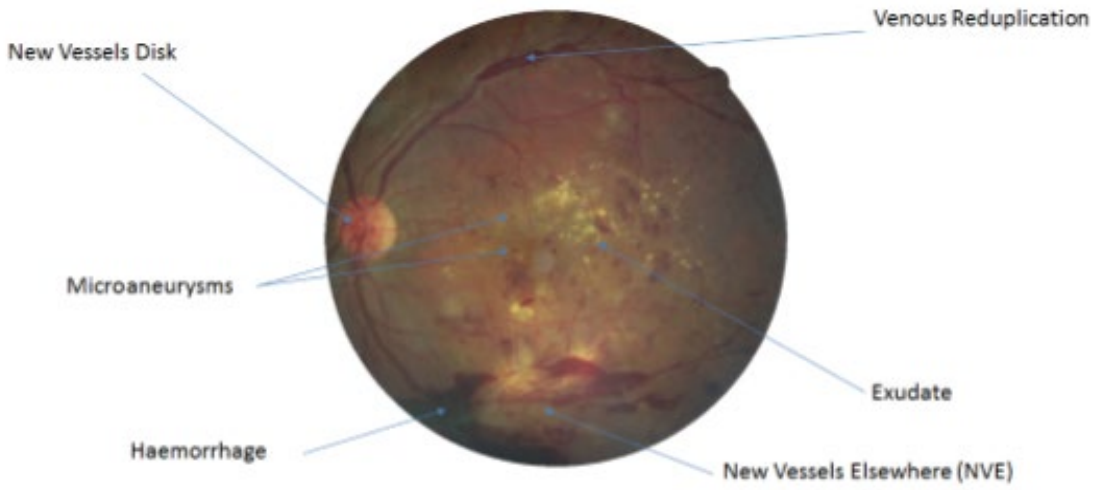


Figure 1.1: Example of retinal features in eye fundus images.

Disease Grades	Finding
Grade 0: No Diabetic retinopathy	No visible signs of abnormalities
Grade 1: Mild NPDR	Micro-aneurysms (MA) only
Grade 2: Moderate NPDR	At least one hemorrhage or MA and/or at least one of the following: • Retinal hemorrhages. • Hard/soft exudates • Venous beading
Grade 3: Severe NPDR	Any of the below mentioned symptoms but no symptoms of PDR. This is generally called 4-2-1 rule. • More than 20 intraretinal hemorrhages in each of the four quadrants

	• Venous beading in two or more quadrants • Intraretinal microvascular abnormalities in one or more quadrants
Grade 4: PDR	One of either: • Neo-vascularization • Vitreous/preretinal hemorrhages.

Table 1.1: Classification of DR stages

Previous efforts have been made in this area using feature engineering i.e., image feature extraction and traditional machine learning method. The retinal features are extracted using image processing techniques and passed to a classifier viz. artificial neural networks (ANN's), sparse representation, linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbors (KNN) and so on. However, there lies a complex combination of features that leads to different stages of DR. Additionally, feature engineering process involved is time consuming and require domain expertise. Deep learning (DL) belongs to the wide family of machine learning methods that have been known to overcome all these limitations and aims to identify the salient low-level features like edges, textures etc. of an image without explicit feature engineering. In 2012, (Krizhevsky et al., 2012) introduced convolutional neural networks (CNN), a deep learning technique known as became popular for solving the image classification problem . The model recorded state-of-the-art performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition (Jia Deng et al., 2009) outperforming other commonly used machine learning techniques. CNN's are most applicable to image classification, segmentation and object detection tasks and requires large amount of data for training. The most popular method initially, when applying CNNs is transfer learning. Transfer learning involves using pre-trained models on natural (non-medical) images to reduce the training time that is required to undertake the computationally heavy CNN training over large data sets. Capsule Networks also known as CapsNet presented in (Sabour et al., 2017) is another widely used deep learning technique that solves some of the disadvantages of CNN's. In work presented

by (Gritsevskiy and Korablyov, 2018), CapsNet was able to generalize information over 25 times faster than a CNN's.

1.2 Problem Statement

The problem statement that the research is trying to solve can be defined as-

To design, build, test, evaluate and compare automated screening models for classifying diabetic retinopathy disease into 5 classes, using suitable deep learning techniques

The research is aimed to solve the following problem:

1. ***Data Imbalance***

   More than 70% of the images in Kaggle dataset (KD) are from normal any complications related to DR (Grade 0). Moreover, images related to severe cases of DR are rarer than images related to moderate and mild stages of DR. Therefore, classification algorithm needs to learn from sparse information and hence appropriate data augmentation methods must be applied to handle class imbalance problem appropriately.

2. ***Complexity of diagnosis***

   There are numerous and complex features that are involved in classifying DR into several stages and the different combinations of these features result in different classification results. Therefore, any proposed algorithm must take combination of these features into account and hence image-preprocessing techniques must be applied so that complex features are easier to detect and improves classification accuracy.

3. ***Choosing appropriate method for high level performance***

   There is a wide range of DL based methods used for data and image analysis classification task. Hence, the proposed model should be robust and computationally efficient to be used for further research and clinical processes. Ideally, the outcome should match or beat the agreement between graders.

1.3 Aims and Objectives

The aims and objectives of the research is given below:

1. ***To identify the most suitable image pre-processing technique that can be effectively utilized making retinal features more visible***

   The dataset that will used in the research comes from different models and variants of cameras so the images would be affected by color and . These variations might not let the features be visible and hinder with classifier performance. Therefore. it is imperative to remove any variation and noise from images so that the model could identify DR features more accurately. Several experiments need to be performed on eye images to identify which pre-processing technique gives better results.

2. ***To determine the most effective method that addresses the class imbalance issue appropriately***

   Any medical image analysis task generally needs to take account of the fact that it will see more negative cases than the positive ones and hence the tasks should be able to learn from minimal amounts of data with respect to the class which has less images. From a medical point of view, clinicians would expect model that have a high sensitivity meaning a smaller number of false negatives. A classification model for DR that has a high sensitivity indicates that the technique is robust and will not miss any positive case during screening. Hence, the deep learning model to be implemented for the classification of diabetic retinopathy screening must provide high sensitivity.

3. ***To compare and evaluate the most robust and efficient DL based method that can be utilized for multi class classification for DR***

   The research will aim to explore three DL based techniques. The most appropriate deep learning method to extract the features of DR from image data and to accurately classify into 5 stages needs to be identified.

4. ***To suggest a method to interpret and explain the prediction***

   The application of DL based models is widely perceived to be black box, i.e., non-interpretable, in the clinical community. It is very difficult to identify that if the model built for DR severity classification is adhering to grading framework as shown in Table 1 or if the model has learned its own framework for classification task. Hence, we need to answer if DL based models can provide regional and pixel-based prediction interpretations. We need to explore and explain that what are the predictors how the model has reached to is prediction.

1.4 Research Question

The following questions needs to be explored from the research-

1. ***Can we suggest that transferring features from ImageNet model trained on natural images are suitable for eye images?***

   ImageNet pre-trained model is mainly trained using natural images which contains more information than medical images. Because of the big difference between natural images and medical images, we need to fine-tune our networks. Additionally, the models based on ImageNet have millions of parameters involved. Transfer learning may have a very limited effect when we switch the domain from one type to another. Hence, this case may be no better than training from scratch, as the networks learn very different high-level features in the two tasks. Certainly, we know if we have enough data, training from scratch is a feasible approach. Hence, we need to explore if competitive results could be achieved by training CNN's from scratch rather than using ImageNet based pre-trained model. Additionally, we would also want to compare the parameters involved in both type of learning.

2. ***Is CapsNet more suitable than CNN's for classifying retinal fundus images?***

   The aim of the research is to perform a 5-class classification for DR and classification rules mentioned in Table 1 are complex and dependent on the presence of features in accordance with the quadrants of eye image. The closer a feature is to the center of the eye i.e., macula, the area in the center of the retina, the more effect it has

towards the severity of DR classification. This makes our research more challenging as CNN's are not spatially invariant. Instead, they rely on pooling layers to achieve translation invariance, and on data-augmentation to handle rotation invariance. Additionally, CNN requires lot of data for training process other it suffers with model overfitting issues. Recently, CapsNet (Sabour et al., 2017) was introduced as an alternative DL based architecture and training approach to overcome the disadvantages on CNN's and model the spatial variance of a feature or an object in the image. This property is known as equivariance and this type of learning is also called as one-shot learning type of vision. The research conducted will answer if CapsNet are more suited to perform a 5-class classification on retinal fundus images than CNN's.

1.5 Scope of the Study

Due to time limitations, the following points were not considered in scope of the research-

1. The data in the research is taken from Kaggle dataset provide by Eye-PACS, an organization which provides end-to-end services to implement a successful blindness prevention program. Hence validation on images and their grades will not be a part of this research.

2. As AI based deep learning methods are a part of ongoing research, there are wide number of techniques and models available in deep learning. We will limit our techniques to only those that will answer our questions given in section 3 and will help us to achieve objectives given in section 4. The research will be limited to the application of CNN's models, pre-trained CNN models that were a part of ILSVRC challenge and Capsule networks.

1.6 Significance of the Study

The increasing rate of diabetes is well known to everyone around the world. 422 million people are now living with some type of diabetes, and this number is projected to increase rapidly in future (WHO, 2018).

The research will be significant in below mentioned cases-

1. ***Screening programs covered by public and private authorities***

   With the advent of high-resolution retinal imaging system many countries in the world have implemented screening programs to cover their citizens for DR detection at an early stage to minimize the risk of vision loss. The impact of this screening program is clear that from the fact that DR is no longer the leading cause of vision impairment in the UK(Liew et al., 2014). However, conducting screening programs are complex and costly and required highly trained graders. These programs require at least annual screening of people with diabetes and as a result produce large quantity of digital images that need to be studied. The screening process is also costly and repetitious which puts pressure on graders to produce high quality results over a longer period. Additionally, given the increasing rise of diabetes, country-wide screening programs might struggle to meet the ever-rising demand using manual grading approach. Hence DR diagnosis through automated screening will be helpful to cover large mass in screening program with cost-effective and standardized way.

2. ***Availability in areas which lack clinical resources***

   Advanced methods and technology could not be available in rural areas and hence delayed process in detection of DR can lead to propagate the disease to higher stages and introducing more complications and cost in treatment. Rural areas can benefit from automated detection is several ways. With application of an AI system in those areas, the photos could be uploaded to an online server and the server will process the images and provide results in less time. Moreover, DR could be detected at early stages and necessary steps could be taken to prevent its advancement to higher stages. Hence, automated detection can be identified as a cost-effective, less time-

consuming use of health service resource in remote areas which lack resources and facilities.

1.7 Structure of the Study

This thesis is organized as follows-

In chapter 1, a brief overview and background of the work is presented. This chapter discusses aims and objectives of the research and all the research related questions that are to be answered through this study. It also discusses scope and significance of the study that is conducted and usefulness of the research in health sector.

Chapter 2 starts with providing a survey of different retinal image datasets available for conducting DR classification study and some image-preprocessing techniques used in previous studies. The chapter then discusses different studies conducted in the past with respect to tradition and deep learning approach and compares their result. This chapter helps us in identifying the right direction in which the research needs to be carried out so that proper analysis could be performed and a significant contribution could be made to the field of study.

In chapter 3, a brief discussion on research methodology is discussed. This chapter consist of detailed discussion on data selection, pre-processing and transformation steps and how several challenges with dataset like class imbalance , detection complex retail features etc. are handled . It also provides some data visualization of all five stages of DR and provides justification on how pre-processing techniques helps to increase accuracy of the model. This is then followed by proposed DL methods applied with a detailed analysis on each method and how the research is conducted for each of these. All the results and comparison of different DL methods is done and justification is provided for selecting the best models out of this.

Chapter 4 provides a summary and evaluation of the result and conclusions from the study. It summarizes the answers to research questions discussed in chapter 1 and also the objective achieved by it.

2. LITERATURE REVIEW

2.1 Introduction

Various researchers across the world have taken a shot at the point of detecting diabetic retinopathy and its classification. Different methodologies are applied by them using computer aided technologies (CAD) to detect and classify DR. This chapter provides a brief insight to different datasets available for research, medical image pre-processing, past work done for detecting and classifying DR.

2.2 Datasets for classifying DR

Eye images known as retinal fundus images (RFI) and the camera which is used to take these images is called fundus camera. This fundus camera is able to take images of internal surface of eyes like retina, posterior pole, macula, optic disc, and blood vessels. Hence, image acquisition is a leading step for medical diagnosis. Most of the researchers used publicly available datasets to conduct their research while a limited number of research's was conducted using private dataset. We review some of the publicly available dataset in this section.

1. ***Kaggle Dataset***

   Kaggle dataset (KD) is widely used dataset which has been used by researchers across the globe. The dataset was produced by EyePACS and contains 88,702 retinal fundus images. Out of these, 35,126 eye images were assigned for training and 53.576 for testing. (Gulshan et al., 2016) classified DR using this dataset.

2. ***E-Ophtha***

   E-Ophtha dataset was created by tele medical network and divided it into two parts – E-Ophtha EX and E-ophtha MA having 381 and 82 images respectively for identifying exudates and micro-aneurysm respectively. (Kusakunniran et al., 2018) used this dataset for DR study.

3. ***Retinopathy Online Challenge (ROC)***

It is made available by The University of Iowa and fundus images were taken using the Canon CR5-45 camera. It contains 100 eye images. Out of this, 50 images are used for training purpose and the rest 50 for testing purpose. (Chudzik et al., 2018) used this dataset for detection of MA's

4. ***DIARETDB0***

In the DIARETDB0 dataset, there are 130 eye images out of which 110 are considered for training and rest for testing purpose. (Nijalingappa and Sandeep, 2015) used this dataset for the identification of DR

5. ***DIARETDB1***

In this dataset there are 89 images out of which 84 are used for train purpose and rest for test purpose. (Bui et al., 2017) used the DIARETDB1 dataset to find out the cotton wool spots.

6. ***STARE***

This dataset is a publicly available dataset made available by the University of California, San Diego and it contains 400 eye fundus images taken by the Topcon camera. (Mo and Zhang, 2017) used STARE dataset for the segmentation of retinal vessels.

7. ***DRIVE***

This dataset is was made available by Holland for educational and research purposes. It contains 40 eye fundus images taken by a non-mydriatic Canon CR5 camera out of which 20 was used for training purpose and rest for testing purpose. (Wang et al., 2015) used this dataset to identify the retinal blood vessels.

8. ***MESSIDOR***

This dataset was produced by Department of ophthalmology, to facilitate the researchers for educational and research purposes. There are total 1200 eye images. (Nazir et al., 2020) used the Messidor dataset to screen out the DR.

9. ***ARIA***

ARIA dataset contains 143 retinal digital fundus images captured using a Zeiss FF450 fundus camera with fifty degrees of FOV.(Arunkumar and Karthigaikumar, 2017) used this dataset for the multi class disease classification using deep learning.

For research on medical fundus images, widely accessible resources are available as shown in Table 2.. The purpose of these databases is to check the strength of automatic screening of DR and then compare the results with current techniques.

Name	Number of images	Image Resolution
DRIVE	20 color fundus testing images. 20 color fundus training images.	768 × 584
Image-Ret (DIARETB0, DIARETB1)	DIARETB0: total 130 images in which 20 images are normal and 110 with DR. DIARETB1: total 89 images: five images are normal and 84 images with DR.	1500 × 1152
Messidor	1200 images.	1440 × 960, 2240 × 1488, and 2304 × 1536
Retinopathy Online Challenge	100 digital fundus images.	768 × 576, 1058 × 1061, and 1389 × 1383
E-ophtha EX	It contains 47 images with exudates and 35 images with no lesion.	2048 × 1360

E-ophtha MA	It contains 148 images with microaneurysms or small hemorrhages and 233 images with no lesion.	2048 × 1360
Eye-PACS Kaggle Dataset	88,702 images out of which 35,126 are used or training purpose	3500 × 2500
STARE	400 images.	605 × 700
ARIA	143 images in which 115 images are abnormal and 54 images are healthy.	2196 × 958

Table 2.1: Publicly available fundus image dataset

2.3 Medical Image Pre-processing

Fundus eye images are used to image the retina, optic disc, macular regions and the posterior surface of an eye. These regions are used by ophthalmologists for diabetics screening for grading of diabetic retinopathy (DR). There are complex features in fundus images which are used in characterization of DR such as exudates, micro aneurysms, hemorrhages and blood vessels which are not clearly visible as images suffer from noise and quality degradation after they have been taken from cameras with different resolution and adjustments. To clearly identify features and improve accuracy of classification technique, removal of noise and image enhancement techniques are used. They are mentioned below:

1. ***DENOISING***

   Speckle noise is one of the major problems in medical images. It arises due to the multiplication of certain unwanted signal with the original signal resulting in reduction

of image information and quality. In fundus images only, Low Resolution Fundus (LRF) images suffer from noise. High Resolution Fundus (HRF) images have zero noise as they are captured through highly specialized cameras. Noise removal may be performed either in spatial or frequency domain. Though most of the medical images require denoising in frequency domain, filtering in spatial domain is suitable for fundus images because they require the need of sharp detection of edges. Non-linear filtering may be used for fundus images. Though this may require consistent processing time, it achieves better results as compared to linear filtering and preserves image information Therefore spatial non-linear filtering techniques like mean filter, wiener filter, Gaussian filter, median filter may be used for effective noise removal while preserving the edges in fundus images.

2. ***IMAGE ENHANCEMENT***

Image enhancement is used to improve the visual effects and the clarity of image or to make the original image more conducive for computer to process. It involves collection of techniques to improve the visual appearance of an image for analysis by a human or machine. It can improve the image appearance by increasing dominance of certain features and decreasing ambiguity between different regions of the image. Each of these features in eye image have different color identifications and different contrast features which makes its detection possible in various channels. The most common technique for the enhancement of contrast of an image is Contrast Limited Adaptive Histogram Equalization (CLAHE). It partitions the green channel of the retinal image into contextual regions and then histogram equalization is applied to each region. This equalizes the distribution of used gray levels and makes the hidden features more visible. CLAHE process on small region of image rather than the entire image itself.

2.4 DR Screening Using Traditional ML Algorithm's

Before the development of deep learning algorithms, especially deep neural networks, a feature extraction step was needed for general computer vision tasks. Features, in this case, are distinguishing and significative small image patches. Following the feature extraction step, traditional machine learning classification algorithms, such as support vector machines, logistic regression or decision trees, are trained using extracted features to classify the image. This pipeline was applied to almost all traditional computer vision tasks including

image classification, image segmentation, etc. Automatic DR detection based on traditional computer vision techniques follows the same pipeline. First, specific features are extracted from the fundus images using single or combinations of manually designed feature extraction algorithms. Classifiers are then trained using the extracted feature to classify the DR stages.

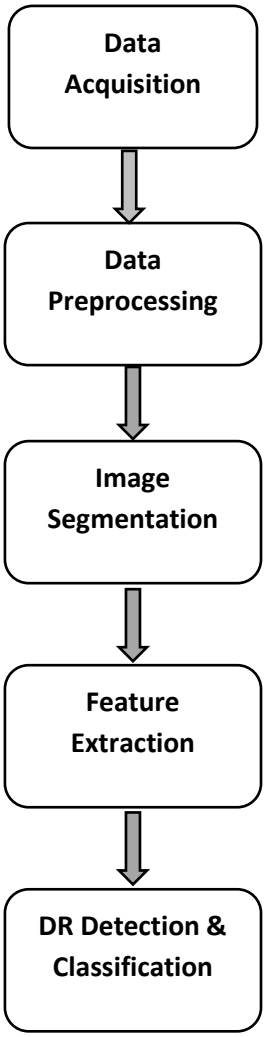


FIGURE 2.1: Traditional Image Classification Approach.

Several such works are briefly reviewed below-

(Williamson et al., 1996) used feature engineering approach to build an artificial neural network (ANN). The network was trained on 83 labeled images to detect features like

vessels, exudates, and hemorrhages. The model was then tested on 100 images to determine if a fundus image contains those features of disease. The reported accuracy for the detection of vessels, exudate and hemorrhages were 91%, 93%, and 74% respectively. The network achieved a sensitivity and specificity of 88.5% and 83.5% respectively for the detection of DR. The test set comprised small number of images, and thus it can be argued is not representative of an ideal screening dataset. However, the work showed that ANN's were able to detect features of DR although more robust models are required

(Acharya et al., 2009) created a five-class classification model by detecting retinal features as hemorrhages, micro-aneurysms, exudate and blood vessels from eye images using image segmentation technique. These retinal features were then fed into a support vector machine (SVM) classifier for multiclass classification. This model produced a sensitivity and specificity of 82% and 86% respectively for PDR and NDPR stages and an average accuracy of 85.9% on the five-class classification problem. This method was tested on a small dataset containing approximately 20 images per class.

(Adarsh and Jeyakumari, 2013) used feature engineering to identify retinal feature and fed them to (SVM) classifier to diagnose DR in to five classes. Different image pre-processing methods were applied to detect retinal blood vessels, exudate, micro-aneurysms, and texture features. The lesion area and texture features were used to build a feature vector for a multi-class SVM task. This model reported accuracies of 96% and 94% for DR classification on the image dataset DIARETDB0 and DIARETDB1 containing 89 and 130 images respectively.

Recursive region growing segmentation algorithms was also adopted in research by (Singalavanija et al., 2006) to extract DR visual features such as exudates, hemorrhages, and microaneurysms. Diabetic patients (182 patients, 336 eyes) were examined by retinal specialists; 221 eyes had a normal fundus and 115 eyes had non-proliferative diabetic retinopathy. Digital retinal images were taken of these 336 eyes and interpreted by the automated screening program . The study was carried out in three steps. Step 1 was to collect baseline retinal image data of 600 eyes of normal subjects with normal fundi and data of 300 eyes of diabetic patients with diabetic retinopathy. All data were recorded by digital fundus camera. Step 2 was to analyse all retinal images for normal and abnormal features. By this method, the automated computerized screening program was developed. The program preprocesses colour retinal images and recognizes the main retinal

components (optic disc, fovea, and blood vessels) and diabetic features such as exudates, haemorrhages, and microaneurysms. All of the accumulated information is interpreted as normal, abnormal, or unknown. Step 3 was to evaluate the sensitivity and specificity of the computerized screening program by testing the program on diabetic patients and comparing the program's results with the results of screening by retinal specialists. The features were then used for DR binary classification. The sensitivity and specificity of the proposed system are 74.8% and 82.7% respectively.

(Kahai et al., 2006) created a decision support system for early detection of DR(presence of microaneurysms) using Bayes optimality criteria. The study made use of 143 retinal images provided by the Louisiana State University Eye Center. Supervised learning was performed for training, whereas unsupervised learning was used to test the system. They compared the retinal images of the diabetic patients which do not manifest microaneurysms with those which do. Moderate-to-severe cases were considered for the case where in microaneurysms are present .AYES decision (abnormal) corresponds to the presence of microaneurysms for the moderate and severe cases of NPDR and a NO decision (normal) relates to the absence of microaneurysms. Each decision has an associated cost that is represented by the Bayes risk. The method was able to identify the early stage of DR with a sensitivity of 100% and specificity of 67%.

(Costa et al., 2018), presented an adversarial auto encoder for the synthesis of the retinal vessel network. In this technique, both structures achieve the optimal solution of differentiable loss functions. Finally, the resultant structure provides end to end fundus image synthesis model to generate the fundus images according to the requirement of users. The proposed technique generates as many synthetic retinal images as the user requires. To simplify the evaluation of our results, we generated a fixed dataset containing the same amount of image pairs (vessel network/retinal image) as in our initial training dataset (614 pairs) and performed experimental qualitative and quantitative comparisons on them. This dataset will be denoted as Synthetic Dataset (SD). As the model outputs pairs of images with 256 × 256 resolution, every real image was downscaled to the same size in order to perform a meaningful comparison. The models trained with real images obtained an average AUC of 0.887 ± 0.004, while when using only synthetic images, the average AUC was 0.841 ± 0.009.

(Villalobos-Castaldi et al., 2010) presented an automatic, efficient and fast approach for vessel extraction using retinal fundus images. The proposed method described in this paper uses one of the basic approaches to edge detection: the enhancement/thresholding method to achieve a fast algorithm for automated detection of blood vessels in retinal images. Their model reported a sensitivity of 0.96 , specificity of 0.94 and accuracy of 0.98.

A cataract classification technique was developed by (Almotiri et al., 2018). The experiments were performed on fundus images to classify and grade cataracts. The radial basis function network was applied for the cataract grading into mild and severe. The implementation of the presented method was performed on MATLAB and SVM applied for classification. The proposed method reported an accuracy of 0.93

(Roychowdhury, 2016) introduced a fundus image classification approach to finding the retinal image features to decrease the complexity of computational time as well as increase the accuracy of DR detection. First of all, feature extraction was performed on the basis of pixels and region of the retinal images. In the next step authors performed features ranking technique for optimal classification. Furthermore, the decision forest and decision tree classifier applied to classify the DR lesion and vessels. The study reported an accuracy of 0.83.

2.5 DR Screening Using Deep Learning Algorithms

Machine learning algorithms are only used for numerical optimization based on the features designed by humans. Thus, domain knowledge of the DR disease plays a crucial role in building effective DR classification models. The limitation of applying conventional computer vision algorithms for DR lesion detection or DR classification is that the manually designed features often require a long time and experience to design and validate. DL based algorithms be an alternative to the above-mentioned methods for image classification tasks as they have the capacity to automatically detect features of images and are new state-of-art.

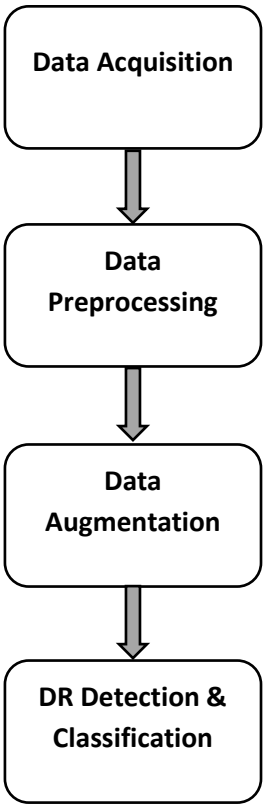


FIGURE 2.2: Deep Learning based approach for Image Classification

Several works which use deep neural network models to detect DR automatically are briefly reviewed below.

(Xu et al., 2017) classifies images into DR and non-DR grades and compares traditional feature extraction-based approach with CNN model approach on 800 labelled images of Kaggle dataset. Four different feature engineering approaches were used to identify hard exudates, retinal lesions, MA's, and blood vessels respectively. The images were resized to 224*224 pixel and were passed as an input to a 13-layer CNN model. The result shows that CNN based methods are superior to feature extraction approach and reported accuracy of CNN model is 94.5 % while those of feature extraction approaches are below 90%. However, the paper does not present any evaluation of sensitivity and specificity scores. Therefore, it is not possible to determine how well the model has learned complex features of DR.

A similar work to this on developing CNN model with dropout techniques but with more layers was presented by (Chandore, 2017) in which they implemented a deep CNN network which contained 24 layers. The model used images of 448 × 448 pixels and was trained on a KD of 61,000 eye images for DR classification. The model was validated on 14,000 eye images and achieved and accuracy of 85% with reasonable sensitivity and specificity values of 81% and 88% respectively.

(Mahapatra, 2016) introduced a convolutional neural network-based retinal image quality assessment technique using various handcrafted features. The reported approach was also based on saliency maps to collect the unsupervised information used for the decision making of retinal quality images. The saliency maps collected multiple scales of every pixel to achieve local and global information of retinal images.

(Gulshan et al., 2016) used transfer learning methodology using Inceptionv3 model pre-trained on ImageNet dataset to classify images into normal and DR images. An ensemble of 10 networks was used for classification the final prediction was calculated by taking a linear average over all predictions given by the ensemble. The dataset that was used in training contained 128,175 eye images. The observed results were validated on two test datasets, Eye-PACS-1 with sizes 9963 images and Messidor-2 with size of 1748 images, had sensitivities of 90.3 % and 87%, respectively while specificity was 98% and 98.5% respectively. The area under receiver operating curve is0.991 for Eye-PACS-1 and 0.990 for Messidor-2. There were 22 million parameters in the model and hence the model was heavy weight requiring high training time and resources.

A comparison of transfer learning models was performed by (Lam et al., 2018) in which they compared five pre-trained model that were a part of ILSVRC challenge – VGG-16, AlexNet, Resnet, GoogleNet and InceptionV3 for binary and multi-class classification. Images were resized to 2048*2048 pixels and the approach was to use the Kaggle dataset to train the model and tested the proposed approach on 195 images of E-Optha dataset. The authors considered a multi-class classification task with five DR grades. The model which outperformed other models was InceptionV3 with a multi-class accuracy of 96% and a binary-class accuracy of 98%.

(Khojasteh et al., 2019)applied residual networks (ResNet-50) with SVM to get better results for the detection of retinal exudates. The author investigated different convolutional

neural networks techniques and then achieved a better technique with the high performance of exudates identification. The model was able to obtain an accuracy of 98%.

A transfer learning method was used by (Ashikur et al., 2020) in which involves using pre-trained models on natural image datasets and retraining the weights towards a different medical image dataset. Pre-trained models that were user are VGG-16, AlexNet and GoogleNet . The results showed that GoogleNet outperformed other models and reported sensitivity and specificity of 95% and 96% respectively. The paper also showed that the results for classifying the earlier stages of DR were still less with 29% sensitivity reported for the mild stage of DR.

(Pérez et al., 2020) argued that transfer learning methods based on pre-trained models such as DenseNet, AlexNet and Inception that were a part of ILSVRC Image Net challenge contains huge number of parameters and thus are not suitable for quality assessment on light weight devices. They presented a lightweight CNN model named MFQ-NET trained on Kaggle dataset suitable to run on mobile devices. The main idea is to train an initial smaller model on image patches and later extend it to full images. The network consists of a fifteen layers CNN, which is structured on two main blocks: patch feature extraction (PFE) block and the image classification (IC) block. The first block is pretrained with 224 × 224 × 3 patches extracted from the original images. The second block takes the output of the first block, which is extended from patches to full images, and makes the prediction for a full 896×896×3 image. The results in the paper show that the MFQ-Net was as effective as other pre-trained deep models but with a size which is one to two orders of magnitude less than these models and achieved an accuracy of 92% and 85% for binary and three class classifiers. The precision for binary and three class classifiers was 0.87 and 0.85 respectively and recall was 0.94 and 0.85 respectively

2.6 Discussion

Many past works have employed feature engineering and traditional methods for its classification. But manually diagnosing DR from retinal images is time-consuming and challenging. To end this, several works have applied deep CNN models to diagnose DR automatically. Overall, on comparing the performance and analyzing the results of both

traditional and Deep Learning-based methods, the DL-based methods outperform the conventional methods, as discussed in the literature study. Additionally, all the traditional methods described above were trained and tested on smaller dataset of approximately 100 images or less and the reported results do not consider any aspects of time spent in diagnosis. On individually reviewing the DL-based methods, Conventional Neural Network (CNN) and its pre-trained architectures have been used by most of the researchers, and it has produced more potential results.

However, CNN suffers from different issues. One of them is data annotation, where it requires the ophthalmologists' services to label the retinal fundus images. Class imbalance and overfitting are the other issues which may result in biased prediction. An increase in data increases the performance of the DL-based systems, which may not be possible in all kinds of problems.

Another thing worth noting is that the proposed methods do not give any clarity on the complexity of the network required. Similar results were obtained when using a CNN which is 13-layer deep and InceptionNet which was 48-layer deep. As the depth of the network increases, the number of parameters also increases and thus increases the amount of training time and computational resources.

Hence, more research work must be performed on this area to encounter all the above drawbacks and to increase the robustness and performance of the system. This thesis is therefore focused on applying end-to-end, accurate and computationally efficient CNN models for automatic DR classification so that those could be compared and it could be suggested that how much deep a network is needed to be for DR classification.

Year	Author	Methodology	Outcome
1996	(Williamson et al., 1996)	Feature Extraction	Sensitivity – 88.40% Specificity – 83.50%
2006	(Singalavanija et al., 2006)	Image Segmentation	Sensitivity – 74.8% Specificity– 82.7%

2006	(Kahai et al., 2006)	Decision Support System	Sensitivity – 100% Specificity – 67%
2009	(Acharya et al., 2009)	Feature Extraction	Sensitivity – 82% Specificity – 86%
2010	(Villalobos-Castaldi et al., 2010)	Feature Extraction	Sensitivity – 96% Specificity – 94% Accuracy- 98%
2013	(Adarsh and Jeyakumari, 2013)	Feature Extraction	Accuracy -96%. No sensitivity and specificity reported
2016	(Roychowdhury, 2016)	Feature Extraction and ranking with decision trees	Accuracy- 83%
2017	(Xu et al., 2017)	CNN	Accuracy -94.5%. No sensitivity and specificity reported
2018	(Costa et al., 2018)	Auto encoder-decoder	Average AUC- 0.88 No other data
2018	(Almotiri et al., 2018)	Feature Extraction with SVM	Accuracy- 93%
2017	(Chandore, 2017)	CNN	Sensitivity – 81% Specificity – 88%
2016	(Gulshan et al., 2016)	Inceptionv3	Sensitivity – 90.3% Specificity – 87%

2018	(Lam et al., 2018)	Comparison of following models – AlexNet VGG16 GoogleNet Resnet Inceptionv3	Best model was Inceptionv3 with- Accuracy – 98%
2020	(Ashikur et al., 2020)	Alexnet VGGNet GoogleNet	GoogleNet performed best with- Sensitivity – 95% Specificity – 96%
2020	(Pérez et al., 2020)	CNN (MFQ-Net)	Sensitivity –87% Specificity – 94%

Table 2.2: Literature Review

3. RESEARCH METHODOLOGY

3.1 Introduction

This research aims to compare different deep learning-based CNN models to identify which one is the most efficient and robust model to classify DR from fundus images. In this chapter we discuss various approaches for image-preprocessing and proposed modelling techniques to answer fulfill our aims and objectives discussed in chapter 1. We will also discuss different strategies to overcome the challenges and limitations discussed in chapter 2. We will then build different models varying from 5 convolutional layers to 121 convolutional layer using vanilla CNN's, transfer learning and CapsNET approach to understand the impact of depth and pre-trained weights on prediction and classification.

3.2 Research Methodology

Figure 4 depicts the basic training workflow that has been carried out in this thesis. We divided the image into 3 RGB channels and performed image enhancement using CLAHE on the green channel. After this channel merging was performed and enhanced RGB image of eye is obtained. We then applied several image augmentation techniques like rotation, flipping etc. to handle class imbalance problem. After this we developed five different proposed models and compared their result to identify the best suited model for DR classification.

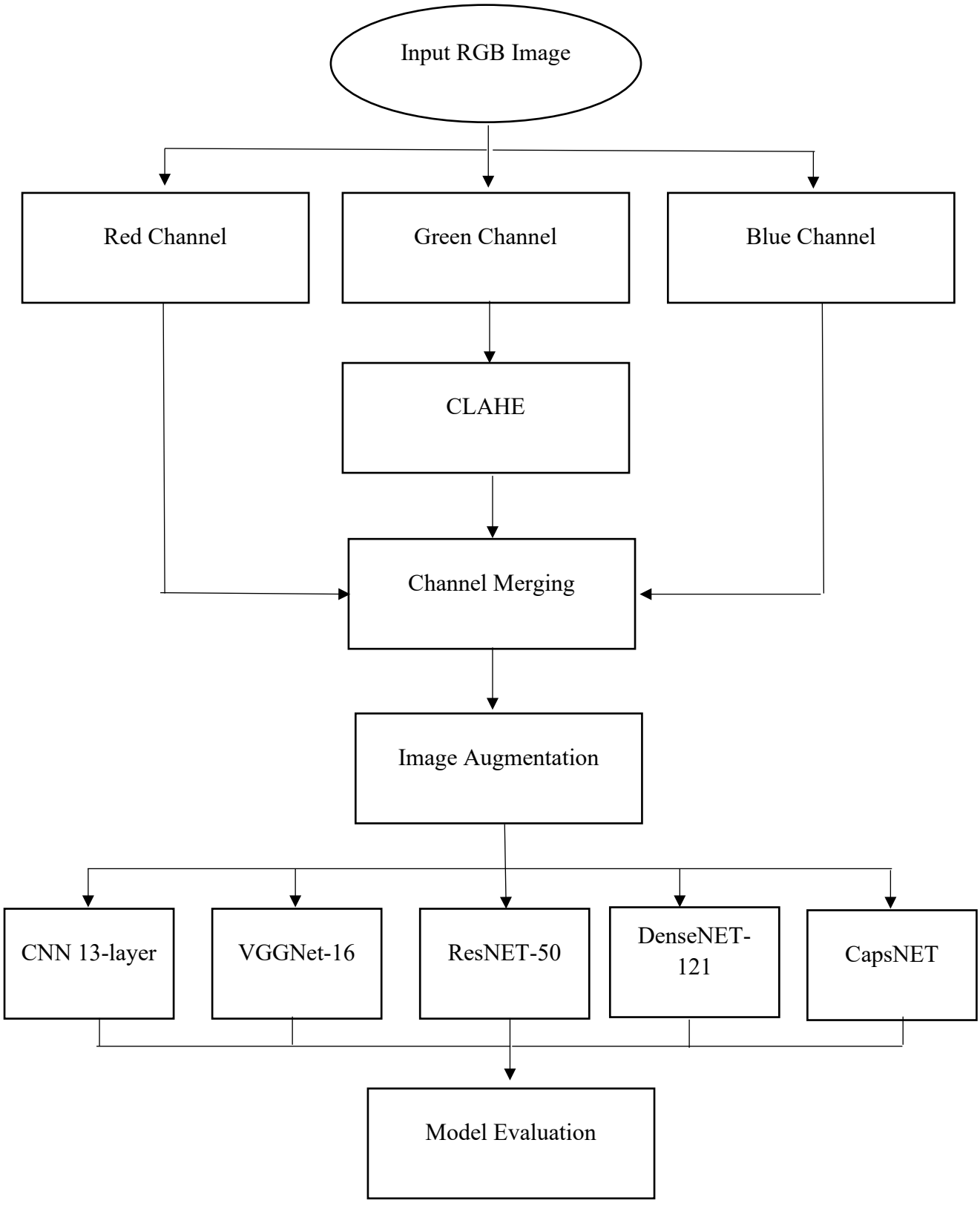


FIGURE 3.1: Model Training Workflow

3.2.1 Dataset Description

Kaggle dataset (KD) is a well-known and widely used for the detection of diabetic retinopathy. This dataset contains the total number of 88,702 retinal fundus photographs. This dataset was produced by Eye-PACS to facilitate the researchers without any cost. In this dataset, 35,126 fundus photographs were assigned for training purposes and 53,576 for the testing. A grader has graded the level of DR in each patient's eye using the fundus images and according to the five point scale presented earlier in Table 1. The total number of images in each class is given in Table 3.

Grades	Train Images	Test Images
0	25,810	39,534
1	2443	3,762
2	5292	7,861
3	873	1,214
4	708	1206

Table 3.1: Dataset overview

3.2.2 Exploratory Data Analysis

The dataset is divided into 5 severity levels as given in table 1. The sizes of the training images vary from 400×315 to 5184×3456. A general visualization of retinal images is given in Figure 5

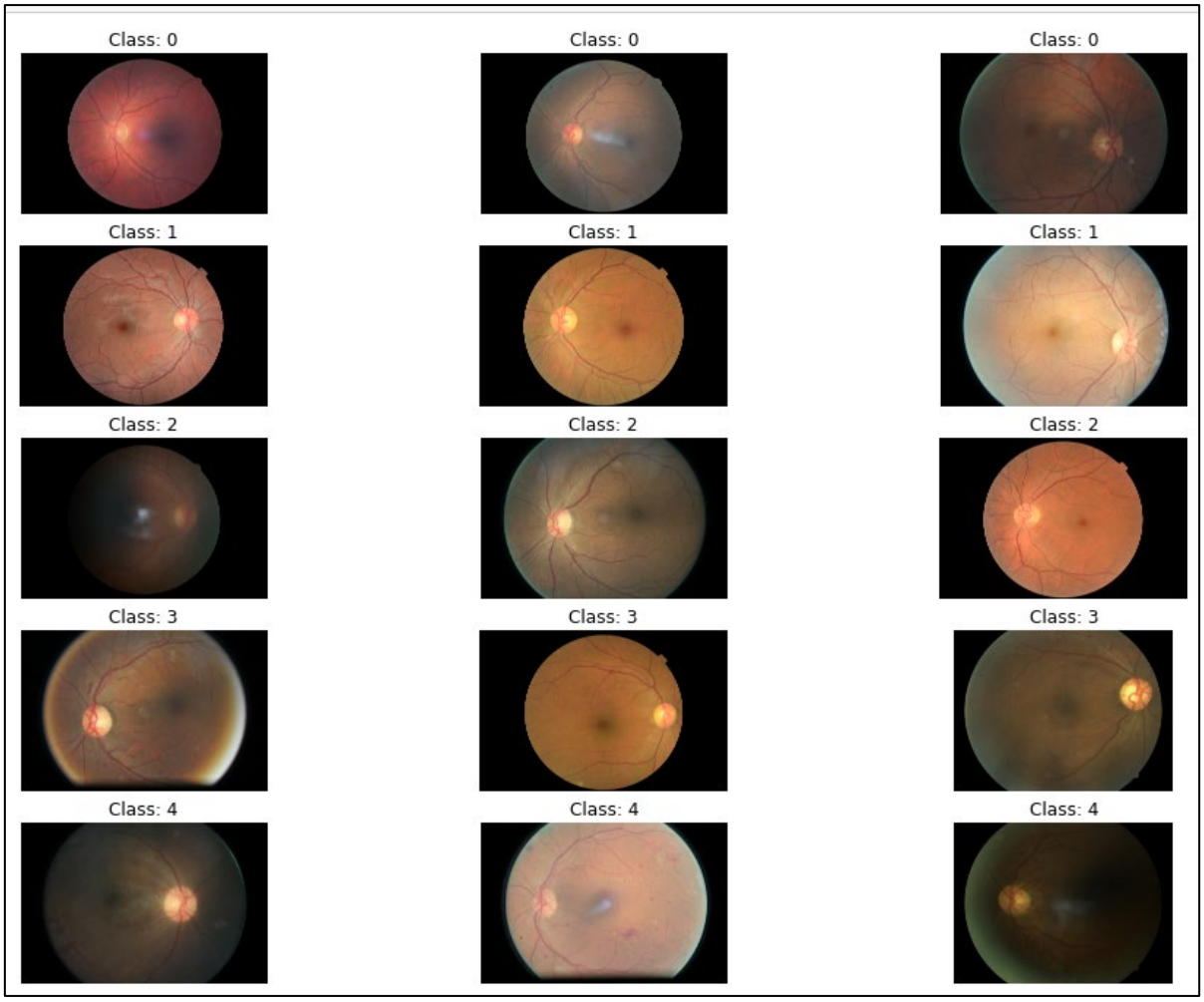


FIGURE 3.2: Image Visualization from Eye-PACS dataset

Figure 6 depicts three RGB (red, green and blue) channels of the image. It can be observed that red channel has extra illumination in between thus hiding the blood vessels while the blue channel hides the hemorrhages as compared to original image. However, the green channel is crisper in displaying retinal features of eye as compared to blue and red channel.

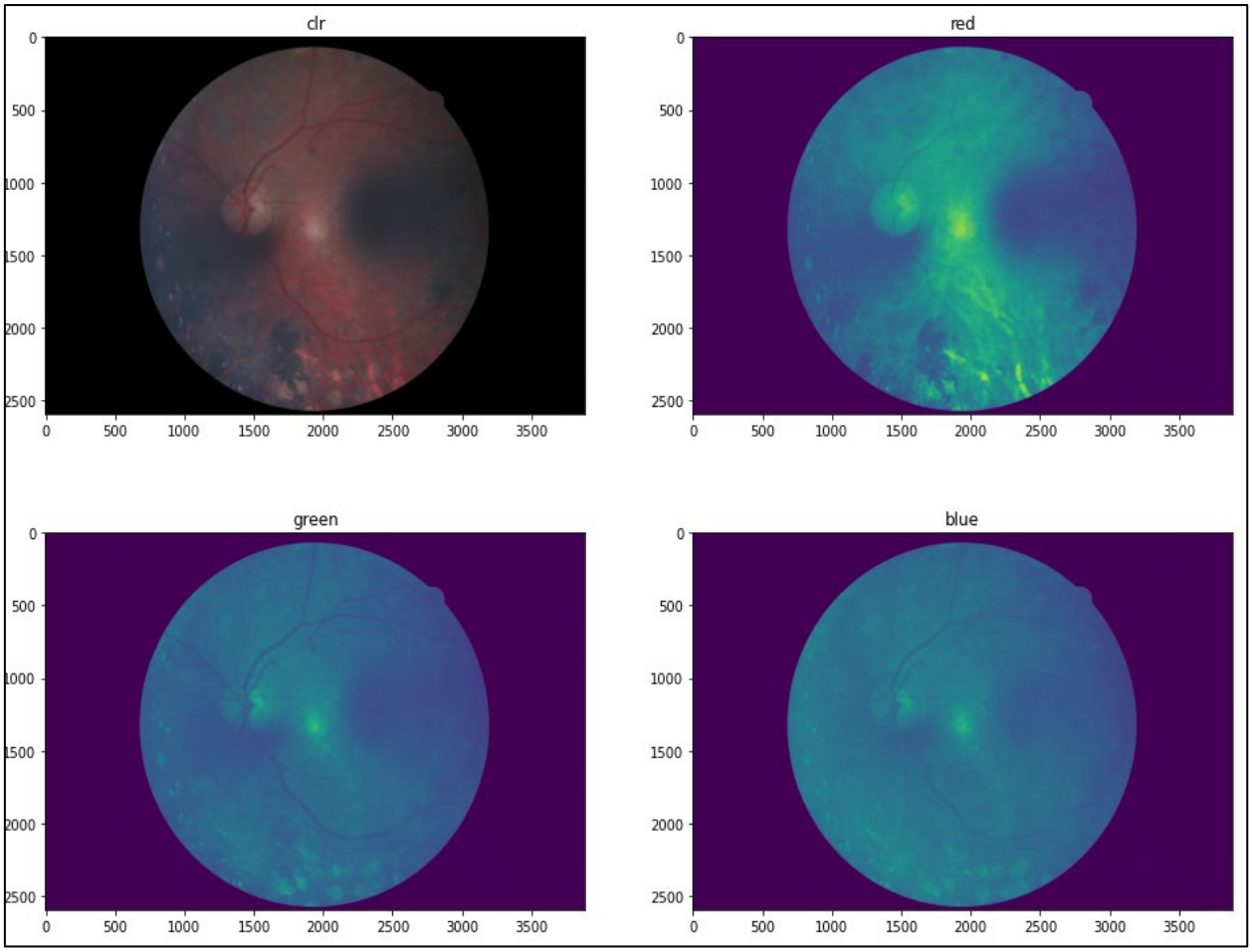


FIGURE 3.3: RGB channels of eye image

3.2.3 Image Preprocessing

Because of variability in the acquisition of fundus images, they are sometimes poorly contrasted and contains noise. Hence, noise removal and contrast enhancement is essential to improve the contrast of the image. In this work, image pre-processing for image improvement is done using OpenCV library and 6 different types of images were used to identify the best pre-processing technique. These were –

1. Raw RGB image cropped and resized.

2. Green channel image from the cropped and resized raw RGB image.

3. Contrast Limited Adaptive Histogram Equalization (CLAHE) of the cropped raw RGB image.
4. Green channel of the CLAHE image.
5. Green channel of the RGB image after CLAHE of luminance channel.
6. Hue Saturation Value (HSV) colour space conversion from RGB.

First, the images were cropped to remove the black background so that unnecessary convolution can be avoided. The images were than downsized to 256*256 for transfer learning models and 192*192 pixels for CapsNET model respectively and noise was removed using gaussian filter.

In order to confirm that the features were more visually apparent, preliminary networks were trained for a maximum of 10 epochs with various learning rates and the value of the loss function was recorded. In all scenarios, CLAHE image processing on green channel has smallest loss after arbitrary 10 epochs of training.

Figure 7 depicts gaussian filter applied on green channel. It can be clearly observed that image retinal features are more clearly visible after removing noise using gaussian filter.

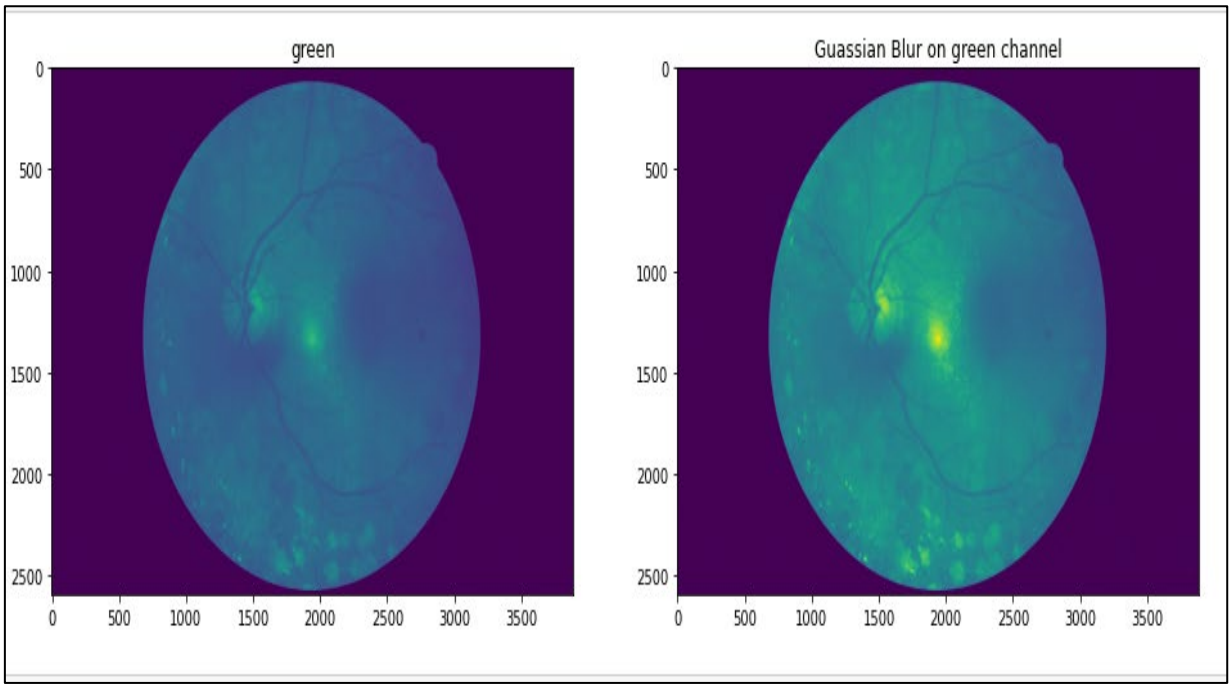


FIGURE 3.4: Gaussian filter on Green channel of eye image

CLAHE is designed to operate on small tiles in the image, and not on the whole image. For each tile, the contrast transform function is calculated with adaptive histogram equalization. The contrast of each tile is improved in such a way that the histogram of the output region will match that specified region by the distribution value. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that may be presented in the image. Figure 8 depicts eye image of green channel with gaussian filter and image after applying CLAHE to it. It could be clearly observed that image has been enhanced after applying CLAHE.

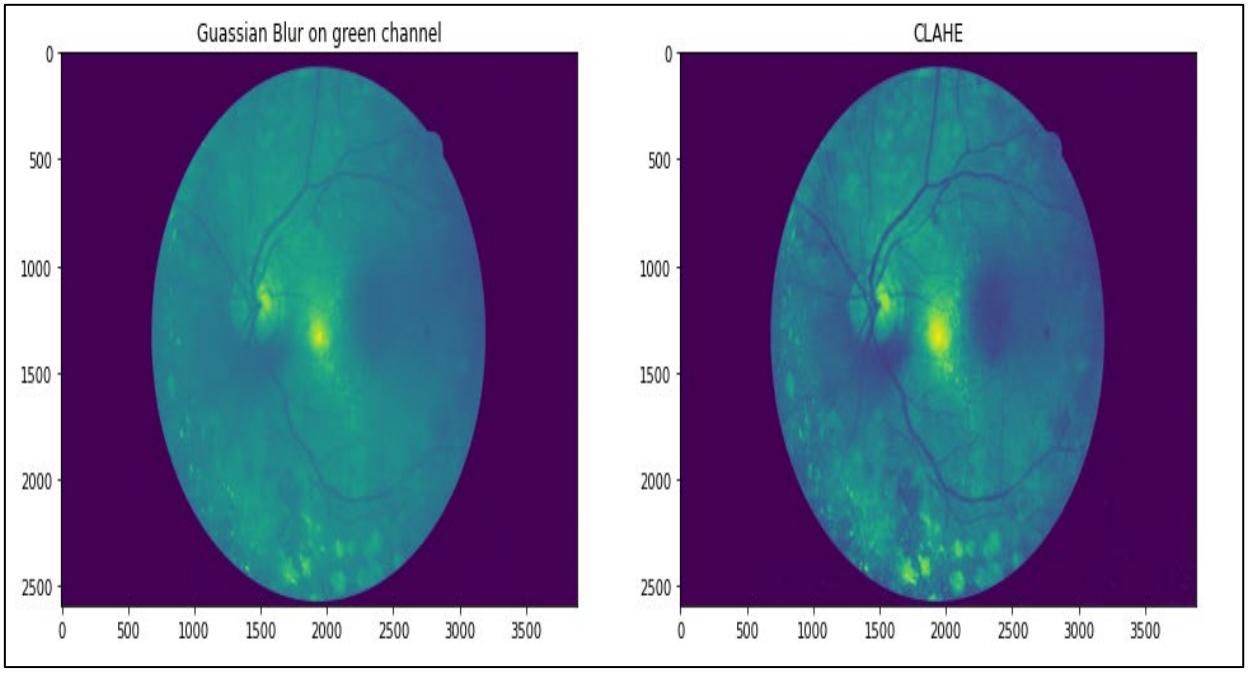


FIGURE 3.5: CLAHE on green channel of eye image

The last pre-processing image is achieved by merging and concatenating the pre-processed green component with the original red and blue components. Figure 9 illustrates visual comparison between original image and the result of pre-processing image.

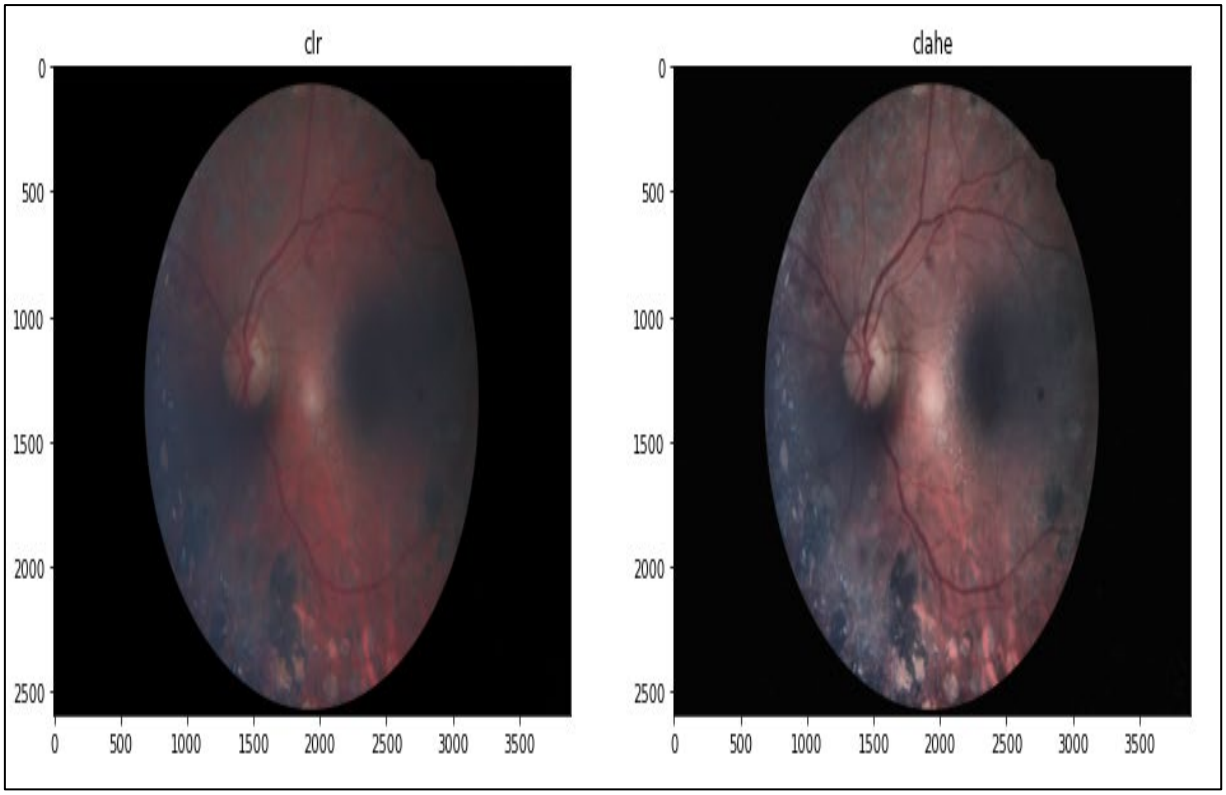


FIGURE 3.6: RGB image after applying CLAHE

3.2.4 Image Augmentation

Figure 10 depicts training samples in each class. To deal with the class imbalance problem, the state-of-the-art solutions for learning from imbalanced data include sampling methods –

1. ***Undersampling*** - The majority class is reduced in size to meet the minority classes.
2. ***Oversampling*** - The minority class is oversampled to balance the dataset.
3. ***Cost-sensitive learning*** -Assigning different weights to samples from different classes.

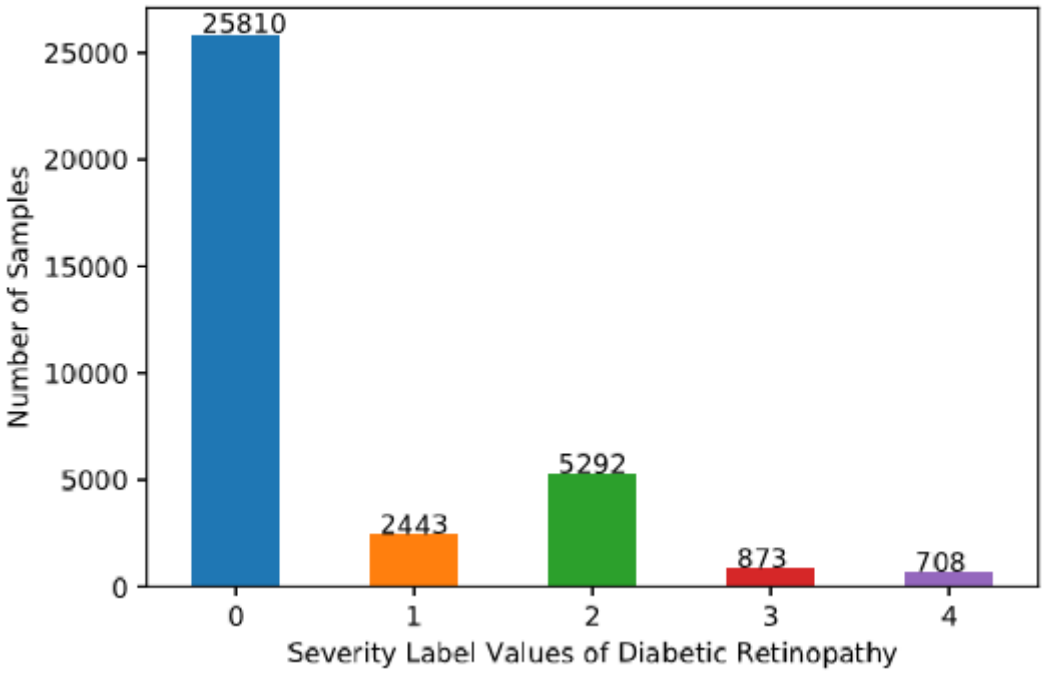
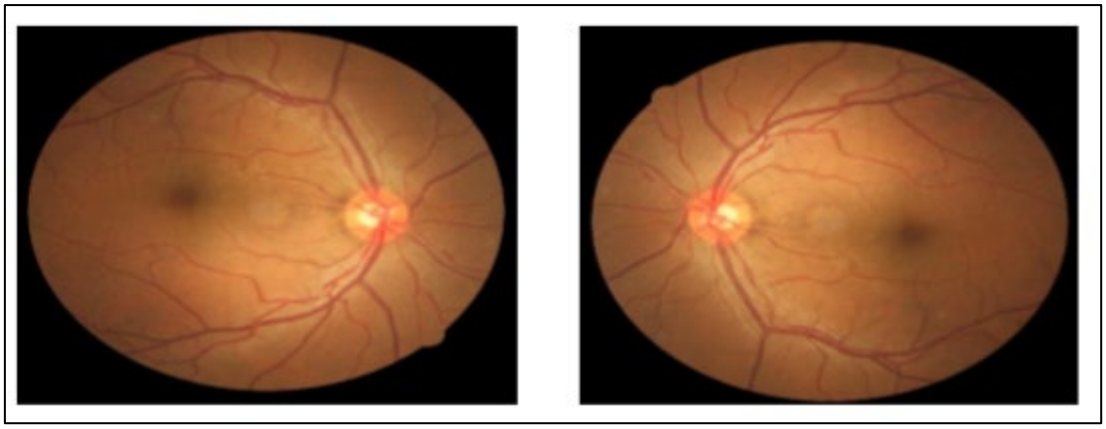


FIGURE 3.7: Image samples in each class

As deep learning models usually have millions of parameters, a large amount of data is required for training such models. Undersampling will result in training models with limited data (i.e., some training samples from majority class are dropped) and cost-sensitive learning needs us to carefully pre-define the weights assigned to each class which may be time-consuming. Thus, we choose to augment the minority classes using several image augmentation techniques to balance the training set. We applied following types of image augmentation techniques-

1. Flipping Horizontally
2. Flipping Vertically
3. Rotation – 90 degrees and 180 degrees.

Figure 3.8 depicts the eye images after applying the augmentations.



a. Flipping Horizontally b. Flipping Vertically

c. 90 Degree Rotation d. 180 Degree Rotation

FIGURE 3.8: Image Augmentation Techniques

3.3 Evaluation Metrics

A one versus all mechanism will be adopted to find the confusion matrix i.e. we need to compute a confusion matrix for every class and consider that class as the positive class and all other negative class. The confusion matrix is shown in Figure 5. We will use sensitivity, specificity, accuracy, and AUC metrics as evaluation benchmark. If P' represents the positive predicted binary class and N' represents the negative predicted binary class.

Similarly, for the ground truth, P represents the positive case and N represents the negative case.

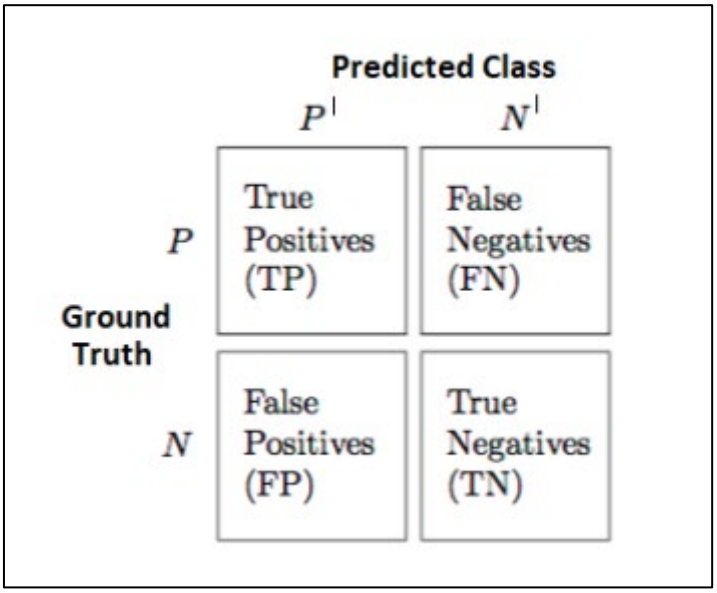


Figure 3.9: Confusion Matrix

The formulas for the same are given below:

Accuracy:

$$Correctly\ Classified\ Images/All\ Images$$

$$\frac{(\mathrm{TP}+\mathrm{TN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FN}+\mathrm{FP})}$$

Sensitivity:

$$Correctly\ classified\ positive\ images/All\ positive\ images$$

$$\frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})}$$

Specificity:

$$Correctly\ classified\ negative\ images/All\ negative\ images$$

$$\frac{\mathrm{TN}}{(\mathrm{TN}+\mathrm{FP})}$$

4: DEVELOPMENT OF DEEP LEARNING MODELS

4.1 Introduction

In this chapter different CNN models are developed. This chapter contributes to training, validation and evaluation of different models. We will discuss and compare architecture and see outcomes of applying different CNN based DL modelling techniques to solve DR classification problem. This will fulfill three of our research objective research objectives discussed in section 1.3 about developing and evaluation different CNN models. We start by developing a vanilla CNN of 5 convolution layers. We then increase the convolution layers to 16, 50, 121 gradually and implement transfer learning. This will fulfill one of our research questions discussed in section 1.4 about finding how deep we need to go while building deep networks for DR classification.

4.2 Proposed Models

4.2.1 CNN

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, CNN's can learn these filters/characteristics.

A CNN can successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

4.2.1.1 Architecture

The proposed CNN architecture contains 5 convolutional layers with a kernel size of (3,3). There are 5 max-pooling layers with size of (2,2). Each convolutional layer is followed by Batch Normalization layer to control model overfitting. After the 5th convolutional layer, there is a dropout of 30% followed by dense layer of 2048 neurons. Table 4.1 shows the architecture of proposed CNN model

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 64)	1792
batch_normalization (BatchNo	(None, 256, 256, 64)	256
activation (Activation)	(None, 256, 256, 64)	0
max_pooling2d (MaxPooling2D)	(None, 128, 128, 64)	0
conv2d_1 (Conv2D)	(None, 128, 128, 128)	73856
batch_normalization_1 (Batch	(None, 128, 128, 128)	512
activation_1 (Activation)	(None, 128, 128, 128)	0
max_pooling2d_1 (MaxPooling2	(None, 64, 64, 128)	0
conv2d_2 (Conv2D)	(None, 64, 64, 256)	295168
batch_normalization_2 (Batch	(None, 64, 64, 256)	1024
activation_2 (Activation)	(None, 64, 64, 256)	0
max_pooling2d_2 (MaxPooling2	(None, 32, 32, 256)	0
conv2d_3 (Conv2D)	(None, 32, 32, 512)	1180160
batch_normalization_3 (Batch	(None, 32, 32, 512)	2048
activation_3 (Activation)	(None, 32, 32, 512)	0
max_pooling2d_3 (MaxPooling2	(None, 16, 16, 512)	0
conv2d_4 (Conv2D)	(None, 16, 16, 1024)	4719616
batch_normalization_4 (Batch	(None, 16, 16, 1024)	4096
activation_4 (Activation)	(None, 16, 16, 1024)	0
max_pooling2d_4 (MaxPooling2	(None, 8, 8, 1024)	0
dropout (Dropout)	(None, 8, 8, 1024)	0
flatten (Flatten)	(None, 65536)	0
dense (Dense)	(None, 2048)	134219776
dense_1 (Dense)	(None, 5)	10245

Total params: 140,508,549
Trainable params: 140,504,581
Non-trainable params: 3,968

Table 4.1: Proposed CNN Architecture

4.2.1.2 Model Training

The training took place for around 240 minutes for image size of 256*256 with a batch size of 128. Figure 4.1 denotes graph of accuracy with respect to number of epochs.

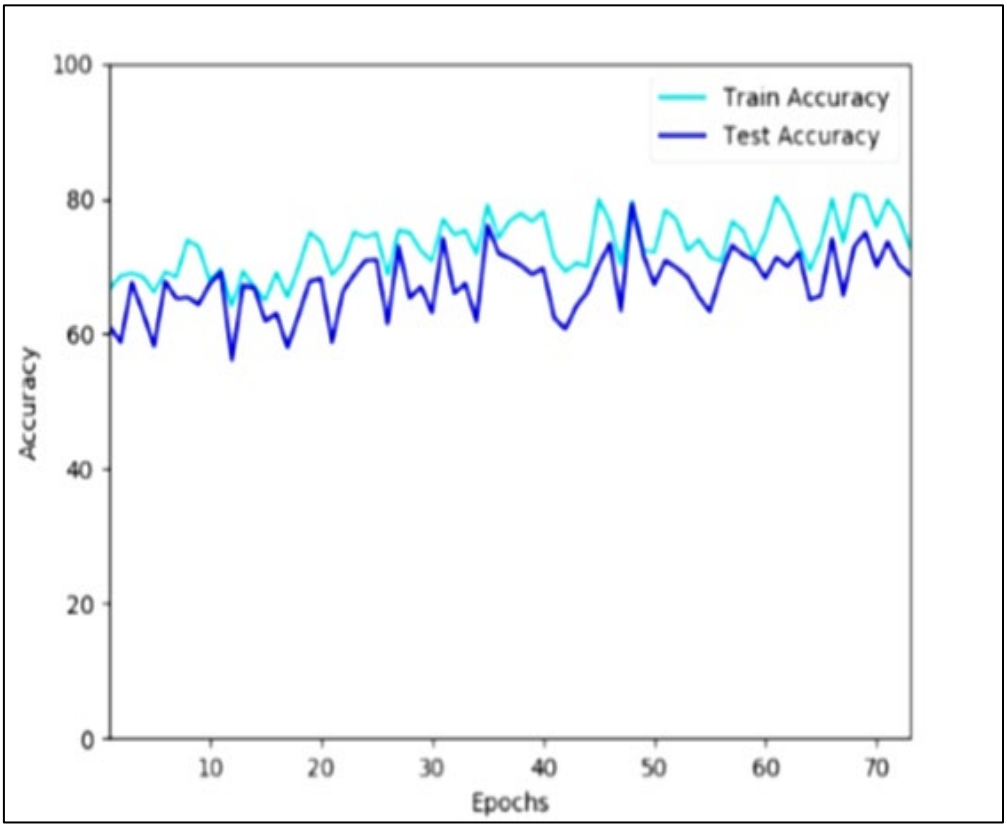


Figure 4.1: Accuracy v/s Epochs for modified CNN model

A learning rate of 0.1 is used in initial epochs and then was decreased by a factor of 10 if the training loss is not decreased in 10 epochs. Early stopping criteria was used on validation loss with a patience of 12 epochs. If there is no further decrease in validation loss until 12 epochs, the training will be terminated. Figure 4.2 depicts training and test log loss plotted against epochs.

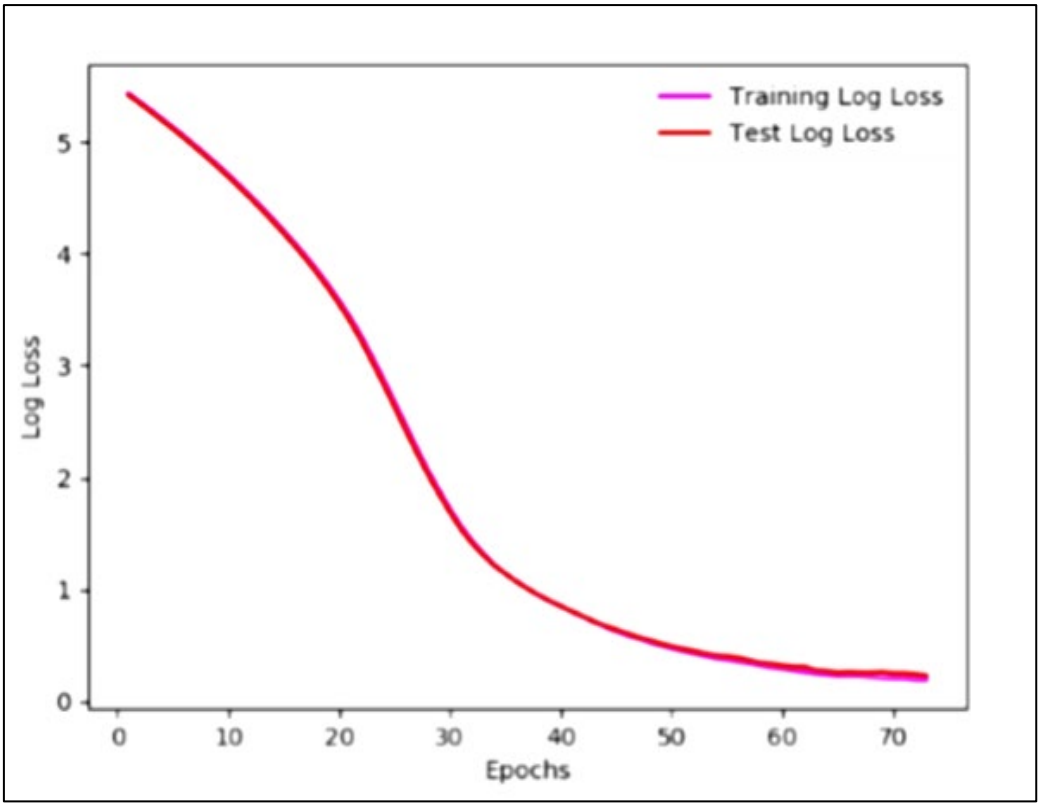


Figure 4.2: Log loss v/s Epochs for Proposed CNN Model

4.2.1.3 Model Evaluation

The results for class-wise specificity, sensitivity and accuracy for CNN model is given in Table 4.2.

Results for Proposed CNN Model			
Class Label	***Specificity***	***Sensitivity***	***Accuracy***
Class 0	0.76	0.85	0.81
Class 1	0.55	0.50	0.52
Class 2	0.67	0.68	0.73

Class 3	0.52	0.28	0.44
Class 4	0.73	0.54	0.64

Table 4.2: Proposed CNN Model Results

The class-wise ROC curve (AUC score) for CNN model is given in Figure 4.3. It also depicts macro average ROC area and micro average ROC area.

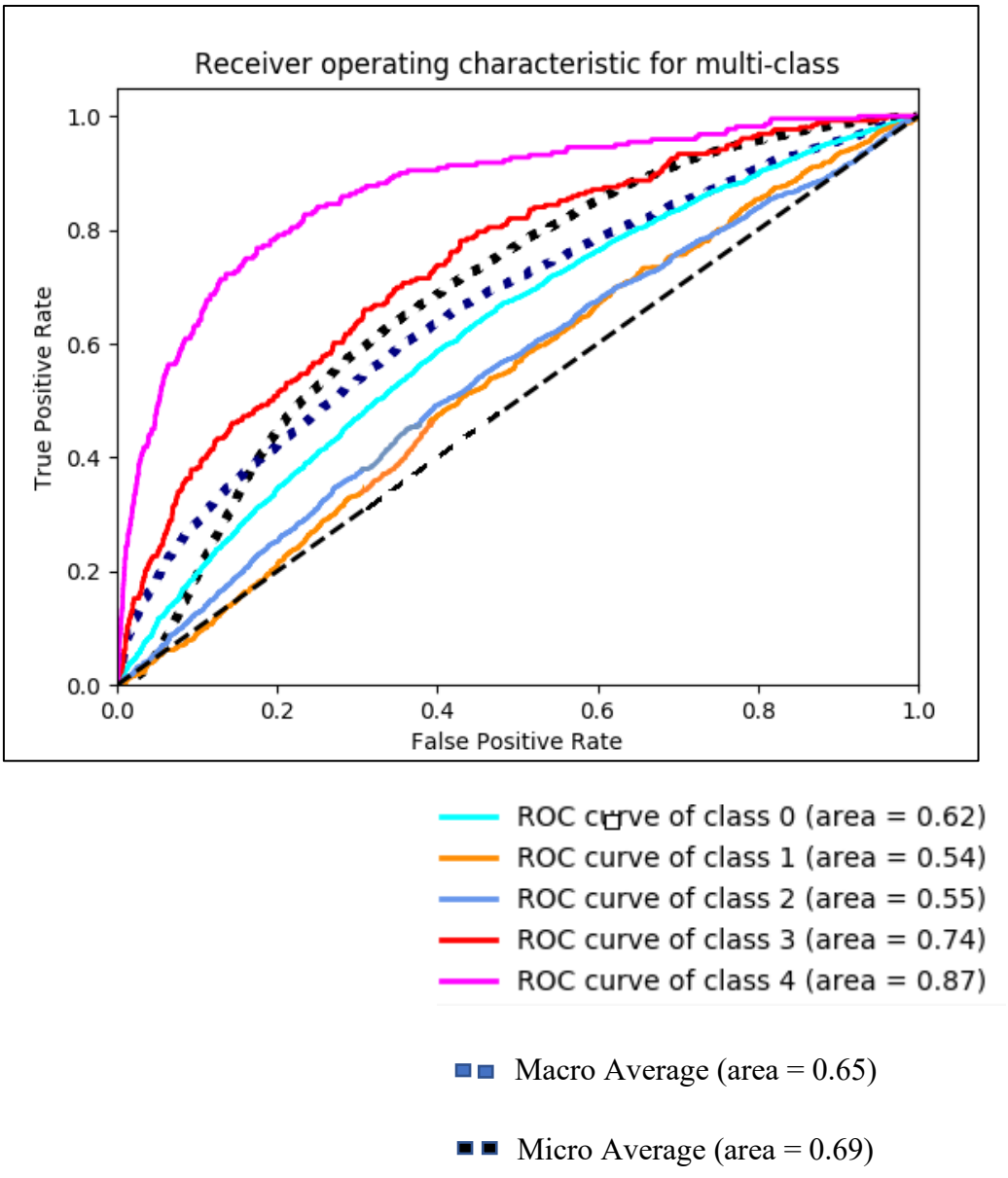


Figure 4.3: ROC Curve (AUC score) for Proposed CNN Model

With a micro average ROC score of 0.69, the evaluation results of vanilla CNN shows that the model was able to correctly extract features from fundus images. The ROC score of 0.87 for class 4 also shows that the model was able to handle class-imbalance problem efficiently and class 4 contains lesser number of images when compared to other classes. The sensitivity, specificity and accuracy score show that the 5-layer CNN performed well on the dataset and was able to classify and predict classes of DR. However, the sensitivity score for class 3 was very low which shows that vanilla CNN's were not able to correctly predict the positive cases for moderate DR.

4.2.2 Transfer Learning using VGG-16

VGG16 is a convolutional neural network model that achieved 92.7% top-5 test accuracy in ImageNet challenge, which is a dataset of over 14 million images belonging to 1000 classes.

4.2.2.1 Architecture

The original VGG16 network architecture contains 5 groups of convolutional layers that in total include 13 convolutional layers, each with a kernel size of (3,3), 5 max-pooling layers, each with a pooling size of (2,2). The network accepts 3-channel image of resolution 224*224. Table 4.3 shows the architecture of VGG-16.

We have tweaked the VGG-16 architecture a little. For this model we use image size of (256*256). After getting the output of block5 of VGG16 model we insert a Flatten layer then we insert a Dropout (Nitish Srivastava Geoffrey Hinton Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov, 2018) layer to control the overfitting and to reduce the number of parameters so that model can be more robust to test dataset. Then we insert five more blocks. Each block consists Dense layer, followed by Dropouts, Batch Normalization layer , which is followed by LeakyReLU layer (Xu et al., 2015) to control overfitting.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 224, 224, 64)	1792
conv2d_2 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d_1 (MaxPooling2	(None, 112, 112, 64)	0
conv2d_3 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_4 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 56, 56, 128)	0
conv2d_5 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_7 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_3 (MaxPooling2	(None, 28, 28, 256)	0
conv2d_8 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_10 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_4 (MaxPooling2	(None, 14, 14, 512)	0
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_13 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_5 (MaxPooling2	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_1 (Dense)	(None, 4096)	102764544
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 4096)	16781312
dropout_2 (Dropout)	(None, 4096)	0
dense_3 (Dense)	(None, 2)	8194

Total params: 134,268,738
Trainable params: 134,268,738
Non-trainable params: 0

Table 4.3: VGG-16 Architecture

We use LeakyReLU activation function because of dying ReLU problem in neural network. Then after loading pre-trained weights, fully connected layers are removed

from the network up to last densely connected layer of size 4096 hidden units. Table 4.4 depicts modified VGG-16 network.

VGG16 Architecture		
Layer (type)	Output Shape	Number of Parameter
InputLayer	(None, 256, 256, 3)	0
VGG16 (model)	multiple	14714688
Flatten	(None, 32768)	0
Dropout	(None, 32768)	0
Dense	(None, 4096)	134221824
BatchNormalization	(None, 4096)	16384
LeakyReLU	(None, 4096)	0
Dropout	(None, 4096)	0
Dense	(None, 2048)	8390656
BatchNormalization	(None, 2048)	8192
LeakyReLU	(None, 2048)	0
Dense	(None, 1024)	2098176
BatchNormalization	(None, 1024)	4096
LeakyReLU	(None, 1024)	0
Dense	(None, 512)	524800
BatchNormalization	(None, 512)	2048
LeakyReLU	(None, 512)	0
Dense	(None, 10)	5130
BatchNormalization	(None, 10)	40
LeakyReLU	(None, 10)	0
Dense(Predictions)	(None, 5)	55

Table 4.4: Modified VGG-16 Architecture

4.2.2.2 Model Training

The training took place for around 432 minutes for image size of 256*256 with a batch size of 64. Figure 4.4 denotes graph of accuracy with respect to number of epochs.

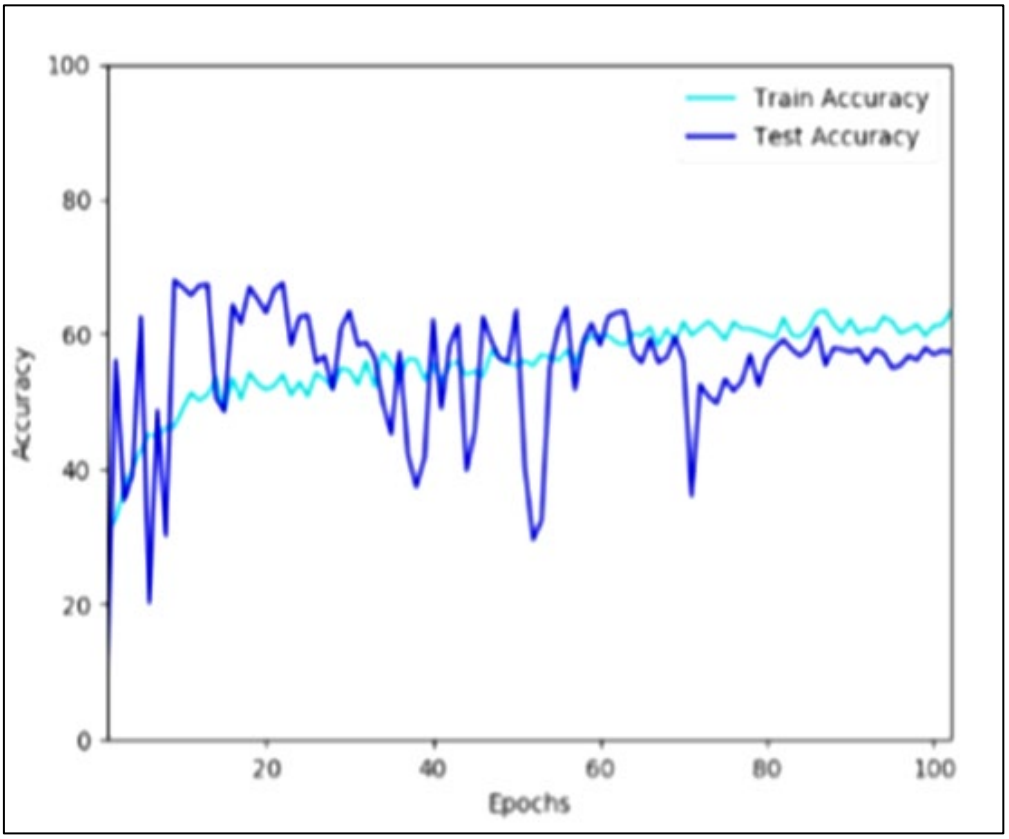


Figure 4.4: Accuracy v/s Epochs for Modified VGG-16

A learning rate of 0.1 is used in initial epochs and then was decreased by a factor of 10 if the training loss is not decreased in 10 epochs. Early stopping criteria was used on validation loss with a patience of 12 epochs. If there is no further decrease in validation loss until 12 epochs, the training will be terminated. Figure 4.5 depicts training and test log loss plotted against epochs.

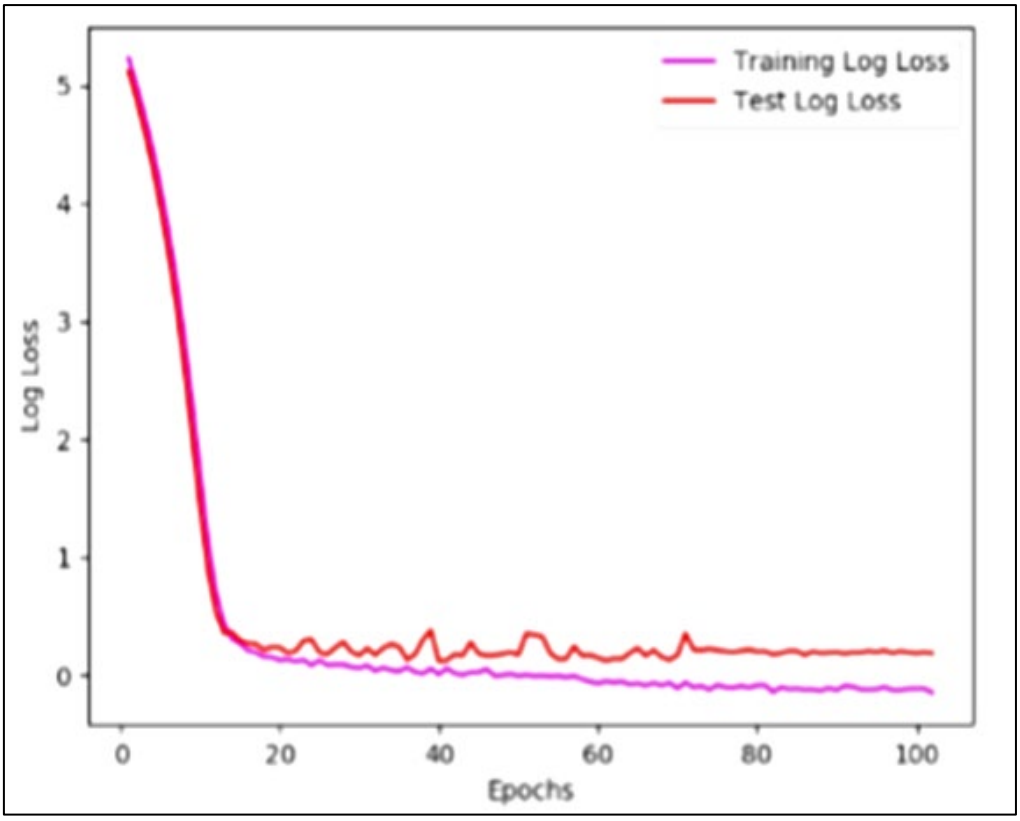


Figure 4.5: Log loss v/s Epochs for modified VGG-16

4.2.2.3 Model Evaluation

The results for class-wise specificity, sensitivity and accuracy for modified VGG-16 model is given in Table 4.5.

Results for Modified VGG-16 Model			
Class Label	***Specificity***	***Sensitivity***	***Accuracy***
Class 0	0.80	0.91	0.86
Class 1	0.58	0.46	0.53

Class 2	0.66	0.71	0.70
Class 3	0.56	0.31	0.42
Class 4	0.76	0.58	0.66

Table 4.5: Modified VGG-16 Model Results

The class-wise ROC curve (AUC score) for modified VGG-16 model is given in Figure 4.6. It also depicts macro average ROC area and micro average ROC area.

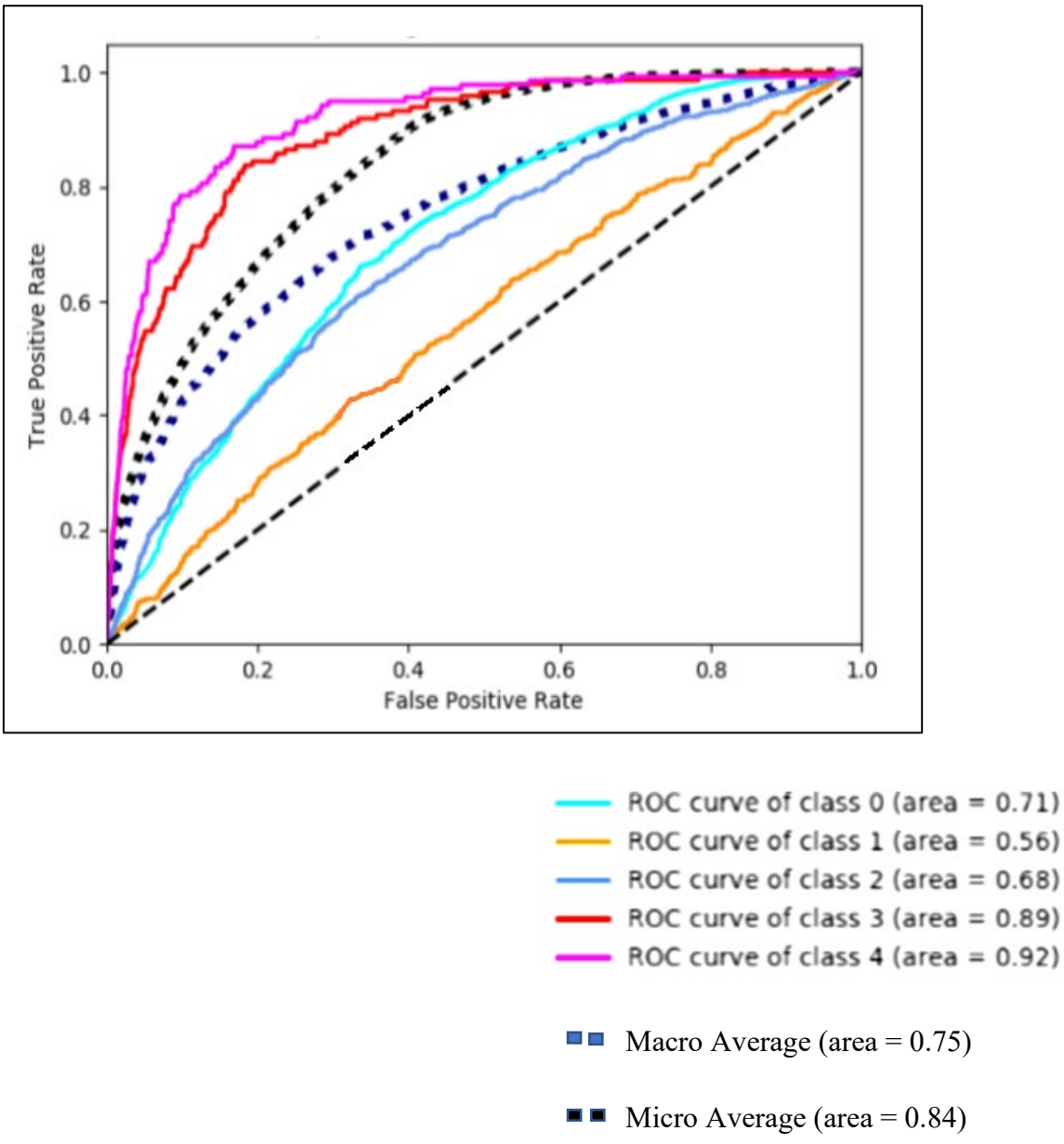


Figure 4.6: ROC Curve (AUC score) for Modified VGG-16

With a micro average ROC score of 0.84, the evaluation results of VGG-16 with LeakyRELU shows that the model was able to correctly extract features from fundus images. The ROC score of 0.92 for class 4 also shows that the model was able to handle class-imbalance problem efficiently and class 4 contains lesser number of images when compared to other classes. The sensitivity, specificity and accuracy score show that the 16-layer VGG model performed well on the dataset and was able to classify and predict classes of DR. However, the sensitivity score for class 1 and class 3 was very low which shows that proposed VGG-16 model was not able to correctly predict the positive cases for mild and moderate DR.

4.2.3 Transfer Learning using ResNET-50

Deep residual networks were a breakthrough idea which enabled the development of much deeper networks (hundreds of layers as opposed to tens of layers).

It's a generally accepted principle that deeper networks are capable of learning more complex functions and representations of the input which should lead to better performance. However, many researchers observed that adding more layers eventually had a negative effect on the final performance. This behavior was not intuitively expected, as explained by the authors below.

Let us consider a shallower architecture and its deeper counterpart that adds more layers onto it. There exists a solution by construction to the deeper model: the added layers are identity mapping, and the other layers are copied from the learned shallower model. The existence of this constructed solution indicates that a deeper model should produce no higher training error than its shallower counterpart. But experiments show that our current solvers on hand are unable to find solutions that are comparably good or better than the constructed solution (or unable to do so in feasible time).

This phenomenon is referred to by the authors as the degradation problem - alluding to the fact that although better parameter initialization techniques and batch normalization allow for deeper networks to converge, they often converge at a higher error rate than their shallower counterparts. In the limit, simply stacking more layers degrades the model's ultimate performance.

The authors propose a remedy to this degradation problem by introducing residual blocks in which intermediate layers of a block learn a residual function with reference to the block input. You can think of this residual function as a refinement step in which we learn how to adjust the input feature map for higher quality features. This compares with a "plain" network in which each layer is expected to learn new and distinct feature maps. In the event that no refinement is needed, the intermediate layers can learn to gradually adjust their weights toward zero such that the residual block represents an identity function.

4.2.3.1 Architecture

The ResNet-50 model consists of 5 stages each with a convolution and Identity block. Each convolution block has 3 convolution layers and each identity block also has 3 convolution layers. The ResNet-50 has over 23 million trainable parameters. Table 4.7 depicts ResNET-50 architecture.

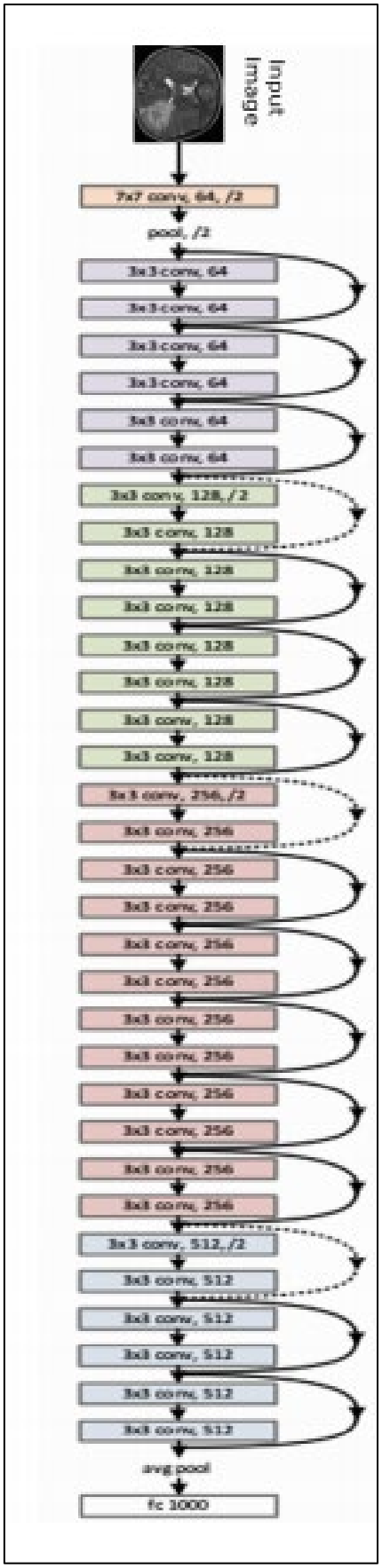


Figure 4.7: ResNET-50 Architecture

4.2.3.2 Model Training

The training took place for around 640 minutes for image size of 256*256 with a batch size of 64. Figure 4.8 denotes graph of accuracy with respect to number of epochs.

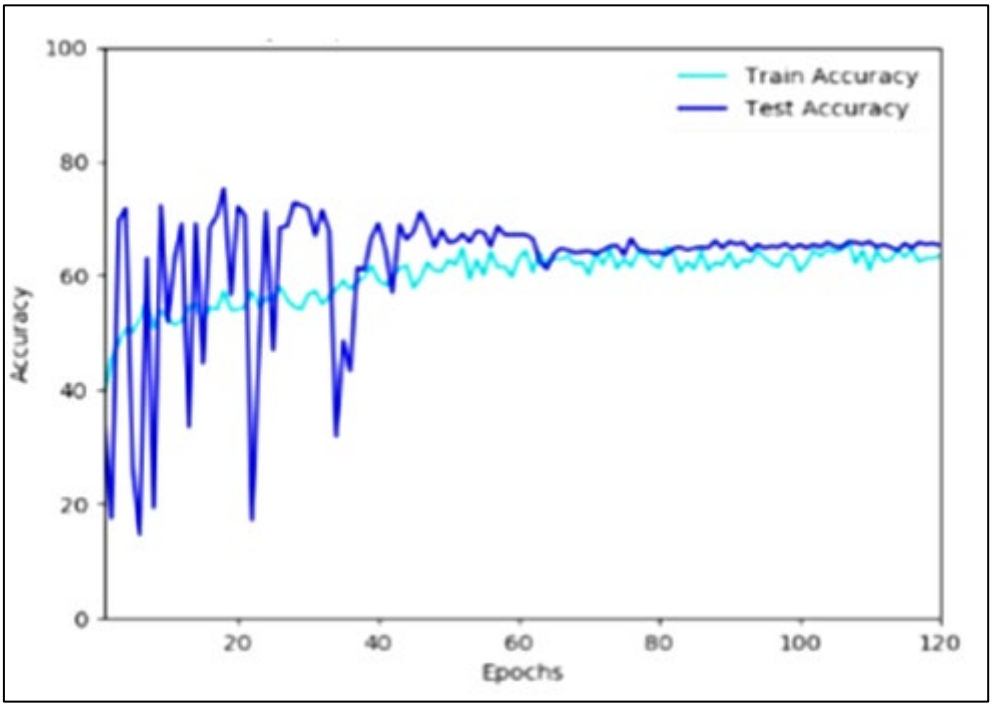


Figure 4.8: Accuracy v/s Epochs for ResNET-50

A learning rate of 0.1 is used in initial epochs and then was decreased by a factor of 10 if the training loss is not decreased in 10 epochs. Early stopping criteria was used on validation loss with a patience of 12 epochs. If there is no further decrease in validation loss until 12 epochs, the training will be terminated. Figure 4.9 depicts training and test log loss plotted against epochs.

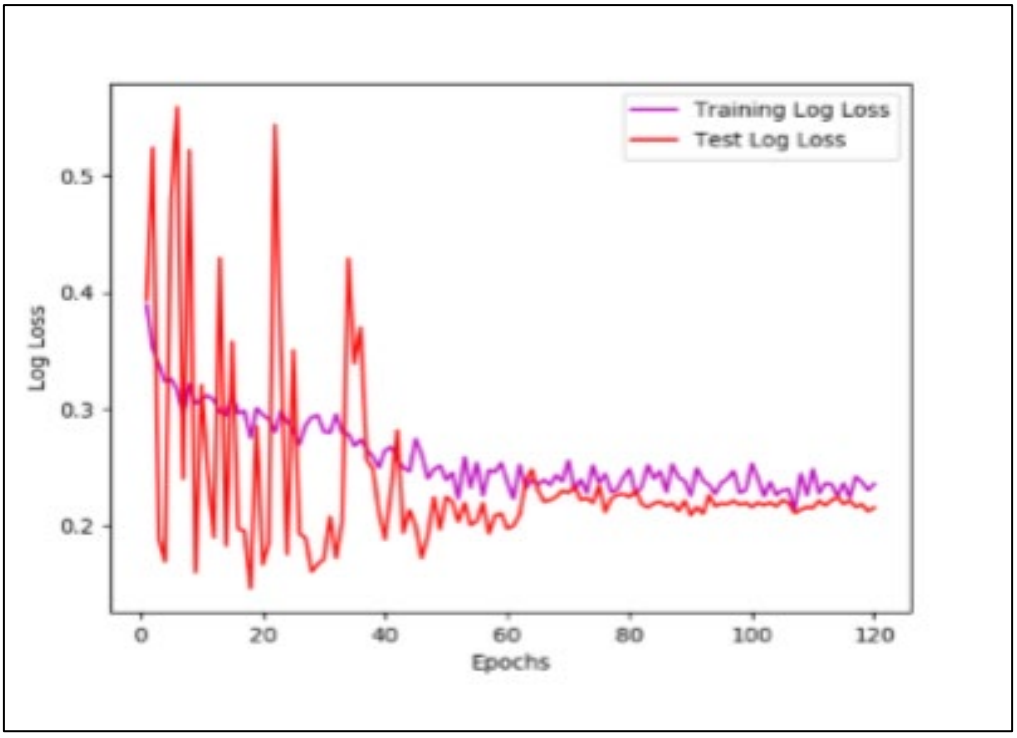


Figure 4.9: Log loss v/s Epochs for ResNET-50

4.2.3.3 Model Evaluation

The results for class-wise specificity, sensitivity and accuracy for modified VGG-16 model is given in Table 4.6.

Results for Modified ResNET model			
Class Label	***Specificity***	***Sensitivity***	***Accuracy***
Class 0	0.85	0.92	0.89
Class 1	0.63	0.55	0.61
Class 2	0.71	0.78	0.75
Class 3	0.64	0.48	0.58

Class 4	0.78	0.67	0.65

Table 4.6: ResNET-50 Results

The class-wise ROC curve (AUC score) for ResNET-50 model is given in Figure 3.18. It also depicts macro average ROC area and micro average ROC area.

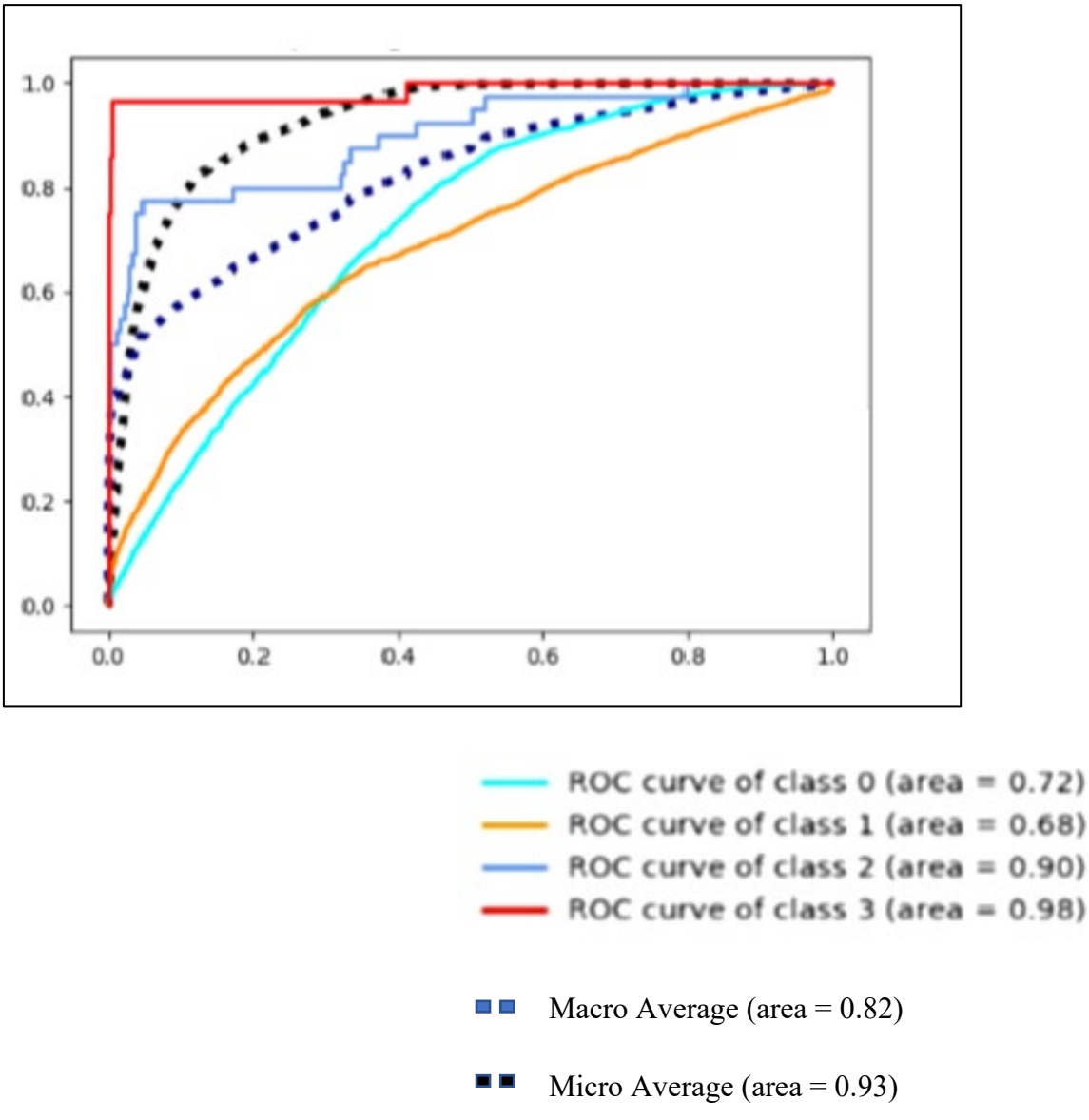


Figure 4.10: ResNET-50 Results

With a micro average ROC score of 0.93, the evaluation results of ResNET-50 shows that the model was able to correctly extract features from fundus images. The ROC score of 0.98 for class 3 also shows that the model was able to handle class-imbalance problem efficiently and class 3 contains lesser number of images when compared to other classes. The sensitivity, specificity and accuracy score show that the 50-layer ResNET model performed well on the dataset and was able to classify and predict classes of DR.

4.2.4 Transfer Learning using DenseNET-121

DenseNet (Dense Convolutional Network) is an architecture that focuses on making the deep learning networks go even deeper, but at the same time making them more efficient to train, by using shorter connections between the layers. DenseNet is a convolutional neural network where each layer is connected to all other layers that are deeper in the network, that is, the first layer is connected to the 2nd, 3rd, 4th and so on, the second layer is connected to the 3rd, 4th, 5th and so on. This is done to enable maximum information flow between the layers of the network.

To preserve the feed-forward nature, each layer obtains inputs from all the previous layers and passes on its own feature maps to all the layers which will come after it. Unlike Resnets it does not combine features through summation but combines the features by concatenating them.

4.2.4.1 Architecture

DenseNet consists of two important blocks other than the basic convolutional and pooling layers. they are the Dense Blocks and the Transition layers. DenseNet starts with a basic convolution and pooling layer. The first convolution block has 64 filters of size 7x7 and a stride of 2. It is followed by a MaxPooling layer with 3x3 max pooling and a stride of 2. Then there is a dense block followed by a transition layer, another dense block followed by a transition layer, another dense block followed by a transition layer, and finally a dense block followed pooling layer and a fully connected layer.

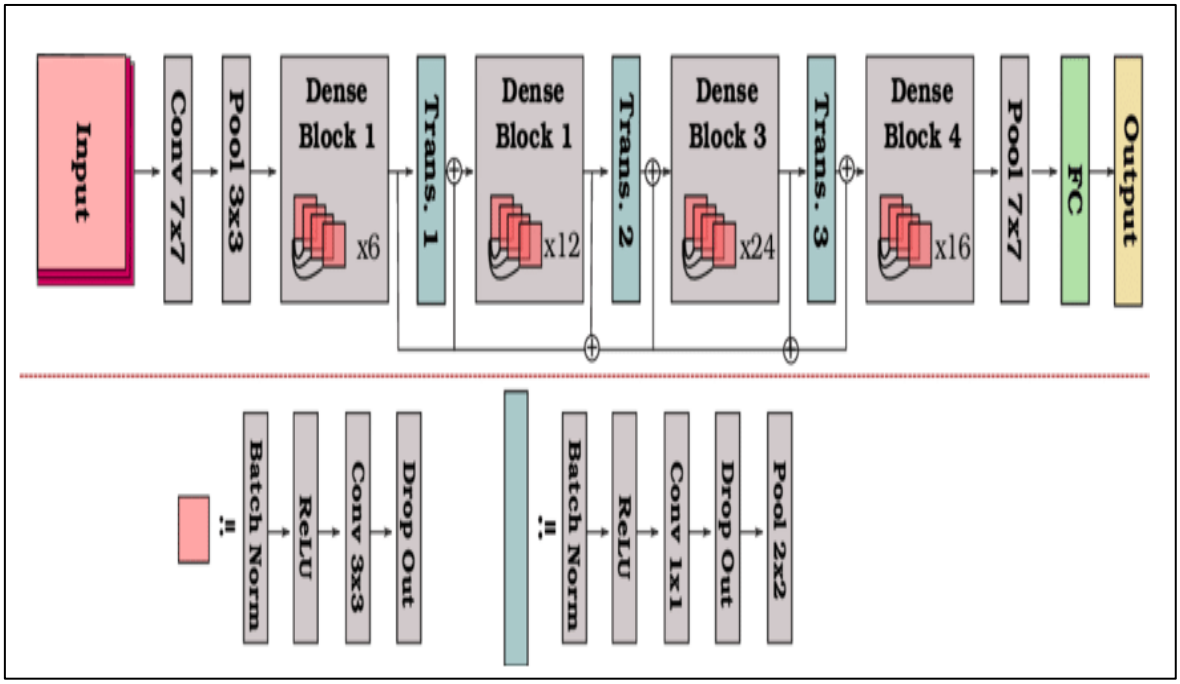


Figure 4.11: Architecture of DenseNet-121

4.2.4.2 Model Training

The training took place for around 860 minutes for image size of 256*256 with a batch size of 64. Figure 4.11 denotes graph of accuracy with respect to number of epochs.

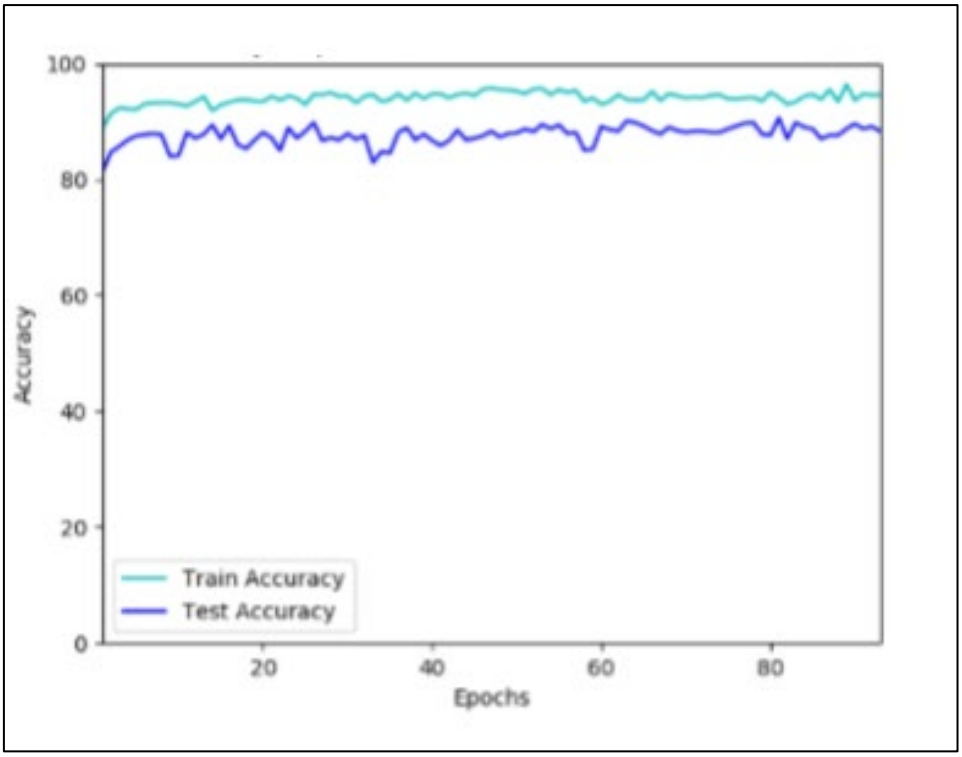


Figure 4.12: Accuracy v/s Epochs for DenseNET-121

A learning rate of 0.1 is used in initial epochs and then was decreased by a factor of 10 if the training loss is not decreased in 10 epochs. Early stopping criteria was used on validation loss with a patience of 12 epochs. If there is no further decrease in validation loss until 12 epochs, the training will be terminated. Figure 4.13 depicts training and test log loss plotted against epochs.

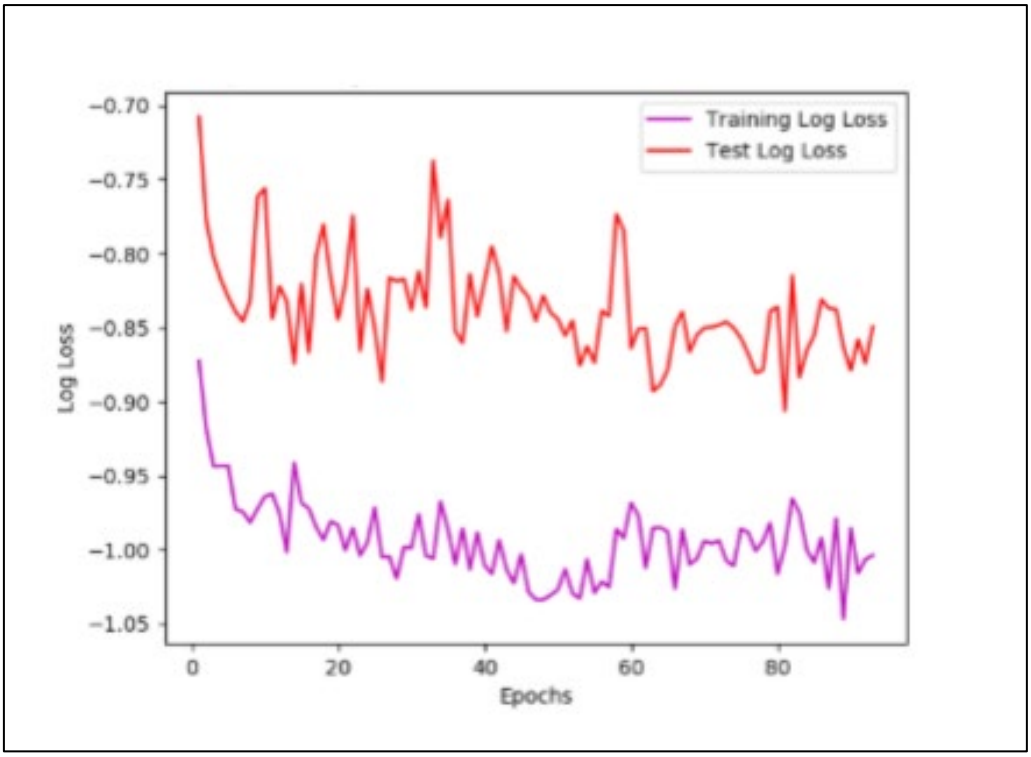


Figure 4.13: Log loss v/s Epochs for DenseNET-121

4.2.4.3 Model Evaluation

The results for class-wise specificity, sensitivity and accuracy for DenseNet-121 model is given in Table 4.7

Results for DenseNet-121			
Class Label	***Specificity***	***Sensitivity***	***Accuracy***
Class 0	0.86	0.93	0.86
Class 1	0.58	0.55	0.60
Class 2	0.61	0.68	0.75
Class 3	0.60	0.45	0.54

Class 4	0.82	0.55	0.72

Table 4.7: DenseNet-121 Results

The class-wise ROC curve (AUC score) for ResNET-50 model is given in Figure 4.13. It also depicts macro average ROC area and micro average ROC area.

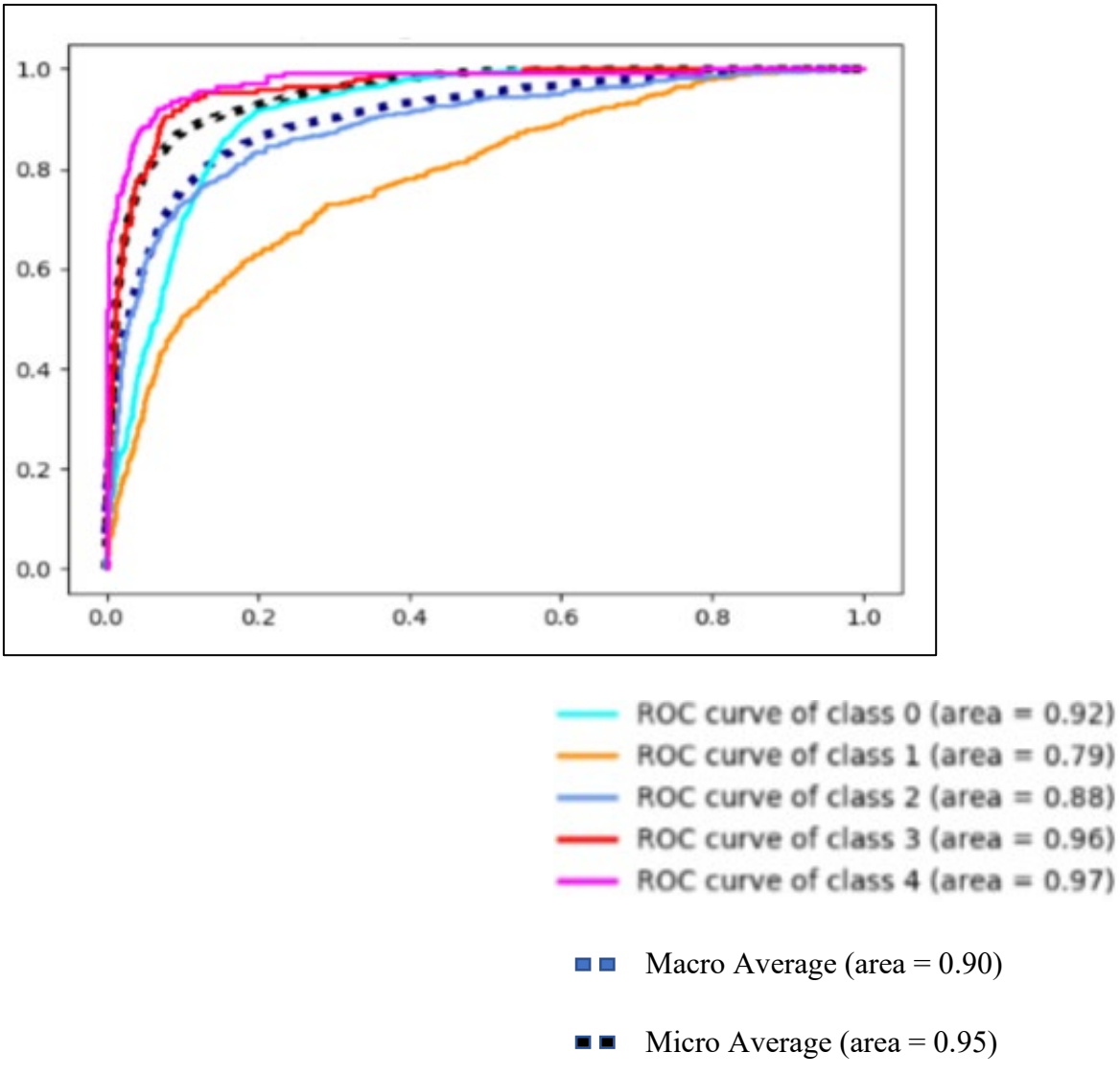


Figure 4.14: Logloss v/s Epochs for DenseNET-121

With a micro average ROC score of 0.95, the evaluation results of DenseNET-121 shows that the model was able to correctly extract features from fundus images. The ROC score of 0.97 for class 4 also shows that the model was able to handle class-imbalance problem efficiently and class 4 contains lesser number of images when compared to other classes. The sensitivity, specificity and accuracy score show that the 121-layer DenseNET model performed well on the dataset and was able to classify and predict classes of DR. However, the sensitivity score for class 3 was very low which shows that proposed DenseNET-121 model was not able to correctly predict the positive cases for moderate DR.

4.2.5 Capsule Network (CapsNet)

4.2.5.1 Architecture

The Capsule Network (Gritsevskiy and Korablyov, 2018) which performs well on MNIST dataset and also required less number of epochs during training but due to large number of kernels at the first and second layer the number of parameters are very high so it increase the time complexity of the model.

4.2.5.2 Proposed Model

We did not make any changes in the CapsNet and just use it for DR classification. We have used an image size of 192*192 in order to reduce training time. Table 4.8 depicts the architecture for capsule networks.

CapsNet Architecture		
Layer (type)	Output Shape	Number of Parameter
InputLayer	(None, 192, 192, 3)	0
Conv2D	(None, 185, 185, 256)	49408
LeakyReLU	(None, 185, 185, 256)	0
primarycap_Conv2D	(None, 89, 89, 256)	5308672
primarycap_Reshape	(None, 253472, 8)	0
primarycap_squash	(None, 253472, 8)	0
digitcaps	(None, 5, 16)	163489440
InputLayer	(None, 5)	0
Mask	(None, 16)	0
Dense	(None, 512)	8704
LeakyReLU	(None, 512)	0
Dense	(None, 1024)	525312
LeakyReLU	(None, 1024)	0
Dense	(None, 110592)	113356800
output	(None, 5)	0
out_recon	(None, 192, 192, 3)	0

Table 4.8: CapsNET Architecture

4.2.5.3 Model Training

The training took place for around 652 minutes for image size of 192*192 with a batch size of 32. Figure 4.14 denotes graph of accuracy with respect to number of epochs.

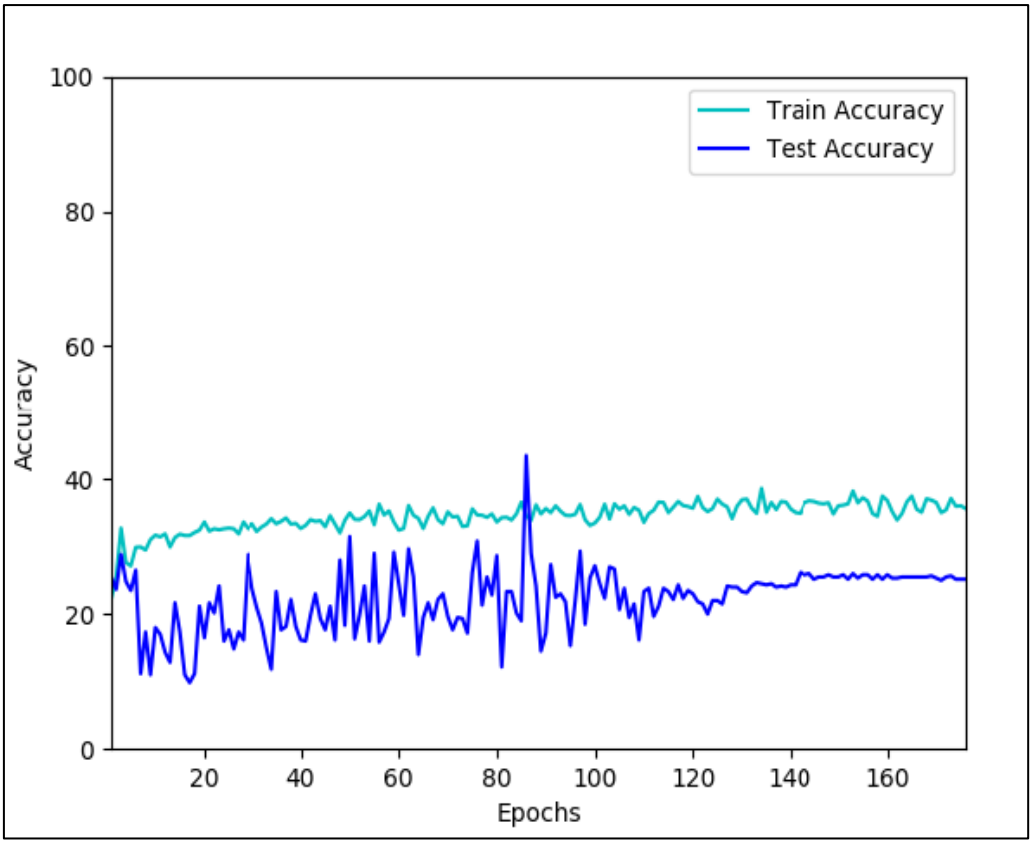


Figure 4.15: Accuracy v/s number of epochs for CapsNet Model

A learning rate of 0.1 is used in initial epochs and then was decreased by a factor of 10 if the training loss is not decreased in 15 epochs. Early stopping criteria was used on validation loss with a patience of 20 epochs. If there is no further decrease in validation loss until 20 epochs, the training will be terminated. Figure 4.15 depicts training and test log loss plotted against epochs.

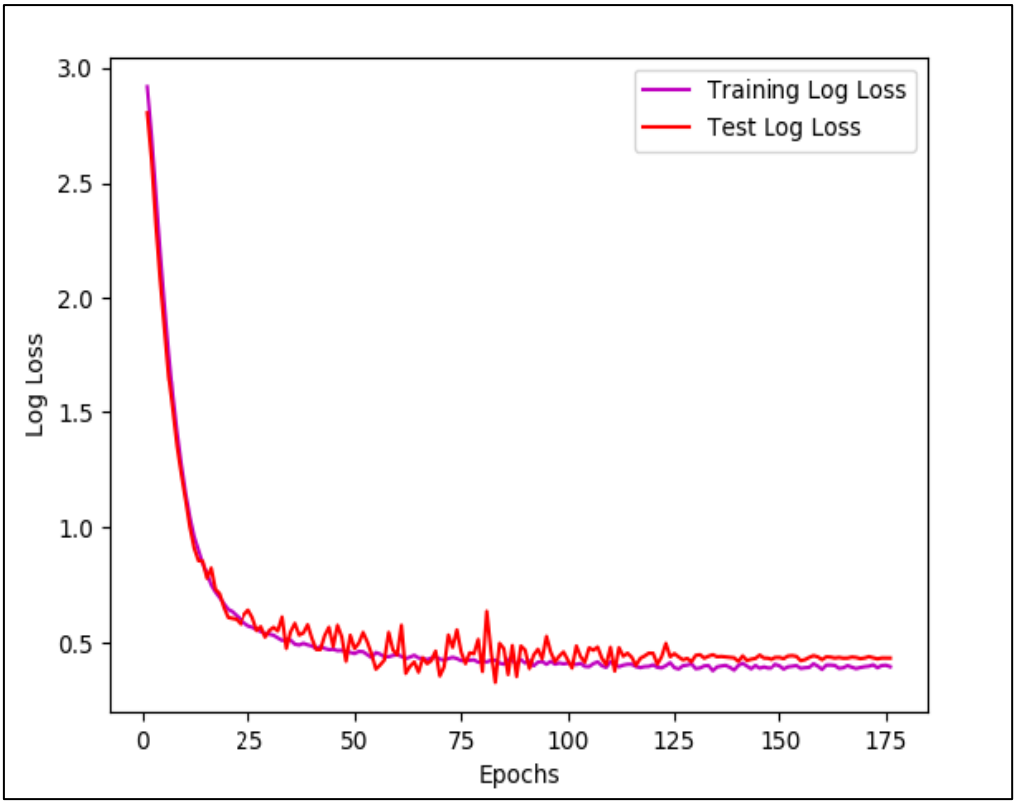


Figure 4.16: Log loss v/s number of epochs for CapsNet Model

4.2.5.5 Model Evaluation

The results for specificity, sensitivity and accuracy for CapsNet model is given in Table 4.9.

Results for CapsNet Model			
Class Label	***Specificity***	***Sensitivity***	***Accuracy***
Class 0	0.81	0.63	0.69
Class 1	0.11	0.34	0.46
Class 2	0.37	0.31	0.52

Class 3	0.29	0.39	0.32
Class 4	0.35	0.31	0.45

Table 4.9: CapsNET Model Results

The class-wise ROC curve (AUC score) for CapsNET model is given in Figure 4.16. It also depicts macro average ROC area and micro average ROC area.

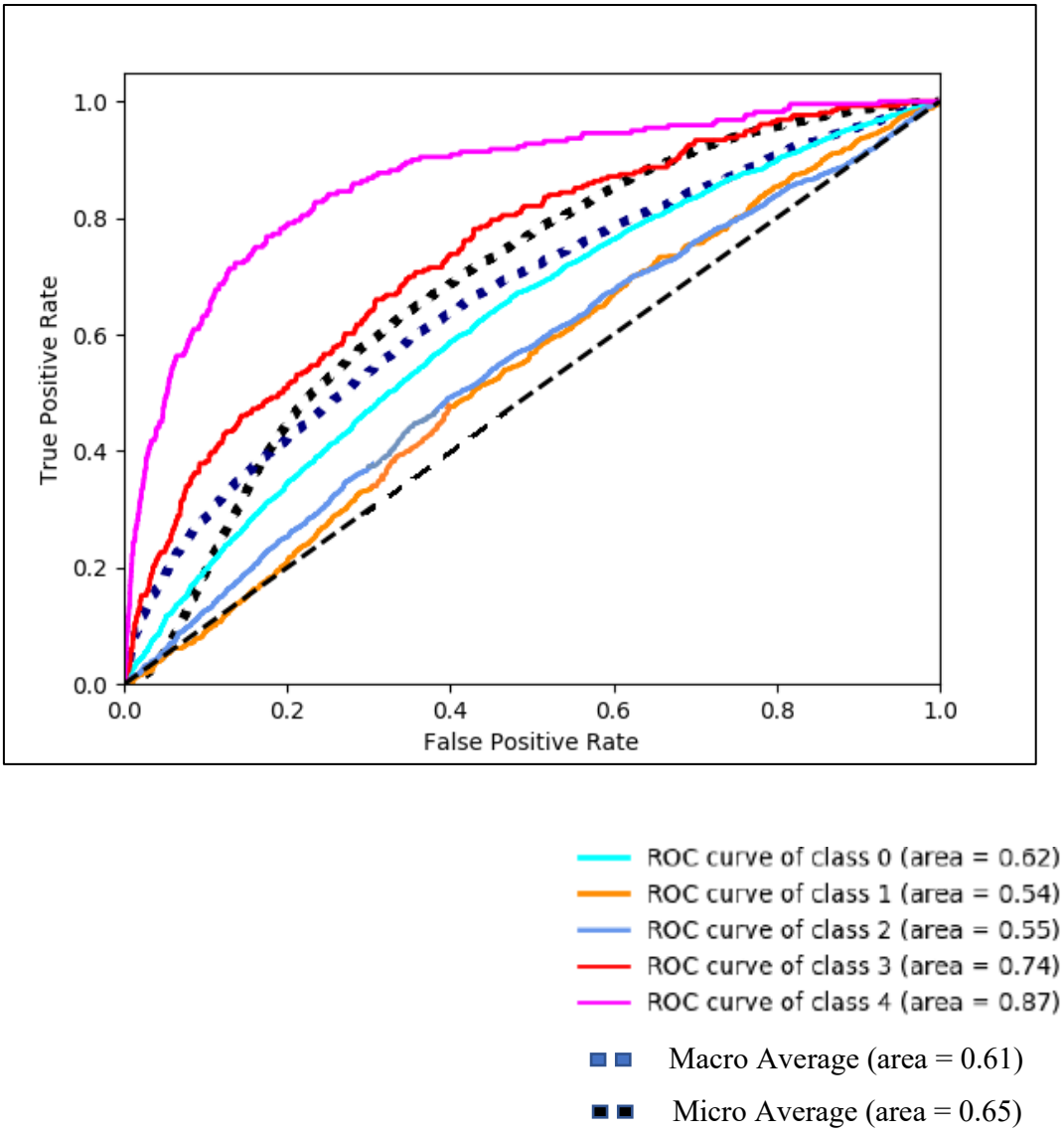


Figure 4.17: ROC curve (AUC score) for CapsNet Model

With a micro average ROC score of 0.65, the evaluation results of CapsNET shows that the model was able to correctly extract features from fundus images. The ROC score of 0.87 for class 4 also shows that the model was able to handle class-imbalance problem efficiently as class 4 contains lesser number of images when compared to other classes. The sensitivity, specificity and accuracy score show that the CapsNET model was not robust and accurate when compared to vanilla CNN and models based on transfer learning.

4.3 Discussion

We observed that DenseNET-121 outperformed all other models in AUC score. The AUC score of 0.95 for classification of DenseNET model suggest that classification accuracy has peaked. The limitation which is remaining is either around the subjectivity of the image being gradable or not. This observation also shows that deeper CNN's are more beneficial in classifying DR. The sensitivity and specificity data show that ResNET-50 was very effective in multiclass classification. Sensitivity is the ability of a test to correctly identify patients with a disease. Specificity is the ability of a test to correctly identify people without the disease. In problem, such as classification od DR, sensitivity is much more important metrics as compared to specificity as it is important to correctly identify people with a disease that without a disease. Hence, it can be concluded that transfer learning was more effective as compared to vanilla CNN's and CapsNET.

Table 4.10 fives a comparison of training time and number of epochs taken by each of the proposed model to achieve a minimal train and test loss.

Model	***Training Time (in minutes)***	***Number of epochs***
CNN	120	85
VGG-16	210	110
ResNET-50	430	135
DenseNET-121	680	95

CapsNet	940	175

Table 4.10: Training Time and Number of Epochs for proposed models

The training time for DenseNET is 5 times much more as compared to vanilla CNN. The proposed ResNET model took two-third of time of what DenseNET took. The VGG-16 took only one-fifth time of what DenseNET took. The model which leads in training time was CapsNET which took 10 times more. This is understood from the fact that parameters involved increases as we move from vanilla CNN to CapsNET which in turn increases the training time. The increasing order of the proposed model in terms of training time and parameters in found to be - *CNN<VGG-16<ResNET-50<DenseNET-121<CapsNET.*

Comparing the number of epochs, it was observed that CapsNET took the greatest number of epochs for getting a minimal train and test loss. The CapsNET was then followed by ResNET-50, VGG-16 and DenseNET-121 respectively. The lowest number of epochs was observed by vanilla CNN. The increasing order of proposed model based on number of epochs is found to be- *CNN<DenseNET-121<VGG-16<ResNET-50<CapsNET.*

The aim of the proposed models was to reduce overfitting and increase feature extraction. The testing accuracy and validation accuracy are corelated well which shows that overfitting was not an issue and models were able to learn the features correctly. However, in order for proposed models to be applicable widely there is a need to give justification for the prediction provided by the models. Hence, there is a need to provide some methods which could help us to interpret the results of the proposed models.

5: INTERPRETING CNN PREDICTIONS

5.1 Introduction

The CNN based DL models developed in chapter 4 does not give any insights or reasoning behind the predictions. The models do not provide any features that have been extracted in order to arrive to result and only work based on the ground truth are considered as black box. Hence, we also wish to determine which features could have led to the prediction. This is also one of our research questions discussed in section 1.4. This chapter introduces methods to understand the interpretation of CNN models. We will use DenseNet 121 model to interpret the predictions. These interpretations of predictions might also prove helpful for clinicians to understand the features and how the model has concluded the classification. Furthermore, after understanding the interpretation there is also a scope to adjust and improve our models accordingly and to encourage future studies.

5.2 Related Work

There has only been a limited number of methods which could explain the prediction process. These are discussed below:

(Simonyan et al., 2014) considered two visualization techniques which are based on computing the gradient of the class score with respect to the input image. The first one generates an image that maximizes the class score and in turn visualizing the notion of the class captured by a CNN. The second technique computes a class saliency map, specific to a given image and class. We show that such maps can be employed for weakly supervised object segmentation using classification CNN. Finally, we establish the connection between the gradient based CNN visualization methods and deconvolutional networks

(Zhou et al., 2016) showed how the global pooling layer explicitly enables the convolutional neural network to have remarkable localization ability despite being trained on image-level labels. While this technique was previously proposed as a means for regularizing training, we find that it builds a generic localizable deep representation that can be applied to a variety of tasks. The network was able to localize the discriminative image regions on a variety of tasks despite not being trained for them.

(Poplin et al., 2017) demonstrated a method called soft attention heat maps . This method was used to visualize which regions of a fundus image the CNN prediction model is using to try to predict age, gender and smoker/nonsmoker. This work demonstrated the use of feature visualization methods on fundus images which in turn serves to provide better understanding of the CNN model's prediction process. These feature visualization methods can be further extended to view the aspects of disease in a fundus image that have led to disease severity prediction.

5.3 Proposed Methods

The visualization demonstrated in this section are taken from the DenseNET 121 model. This model was evaluated as the best model in terms of AUC score out of all the models that has been developed. Once the model is trained, the parameters remain unchanged during the whole process of visualization technique. That model than can be used to produce Class Activation Maps (CAM's) and saliency maps

5.3.1 Class Activation Maps

The class activation technique is based on (Zhou et al., 2016). We inserted a global average pooling layer after the final convolutional layer. This helped in the localizing the region. An input image is then passed to the trained CNN and weights are activated. These weights are then projected to the feature maps to identify the important regions within the class. Figure 5.1 demonstrates the CAM produced from one of the fundus images from class 0.

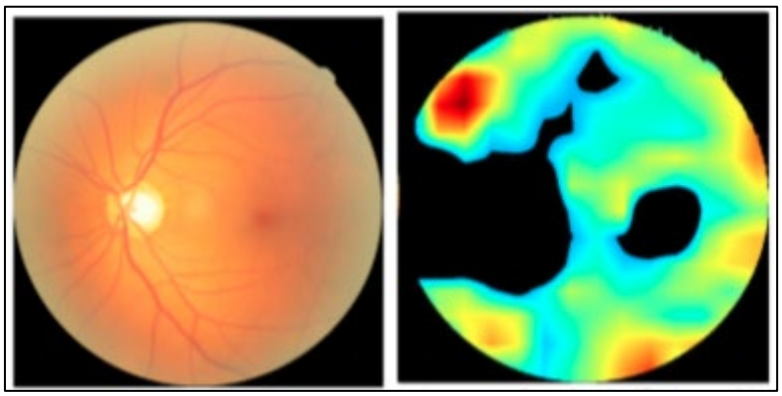


Figure 5.1: Class Activation Map from fundus image

5.3.2 Saliency Maps

The saliency map technique is based on (Simonyan et al., 2014) .They are calculated by taking the gradient of the output with respect to the input image pixels A positive value in gradient demonstrates that the pixel contributes more towards the output class value. Therefore, the larger the gradient of a pixel, the more the image is relied on this pixel for classification process. In this producing the gradient for each pixel produces a saliency map for the fundus image Figure 5.2 demonstrates the saliency map produced from one of the fundus images from class 0.

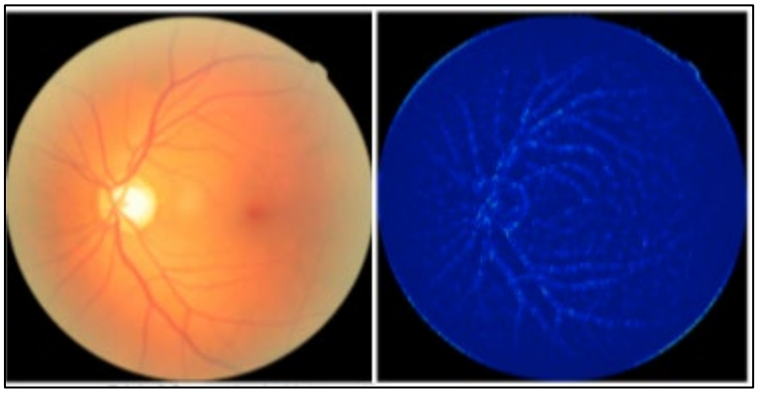


Figure 5.2: Saliency Maps from fundus image

5.4 Discussion

The conclusions for each of the DR grade are as follows:

1. Figure 5.3 demonstrates the CAM and saliency for grade 0. The CAM stretches to the majority part of the retina. It suggests that model correctly identifies it as no features for DR in whole retina. Pixels are considered around the vessel structure as important in classification. The pixels are very light and indicate no DR.

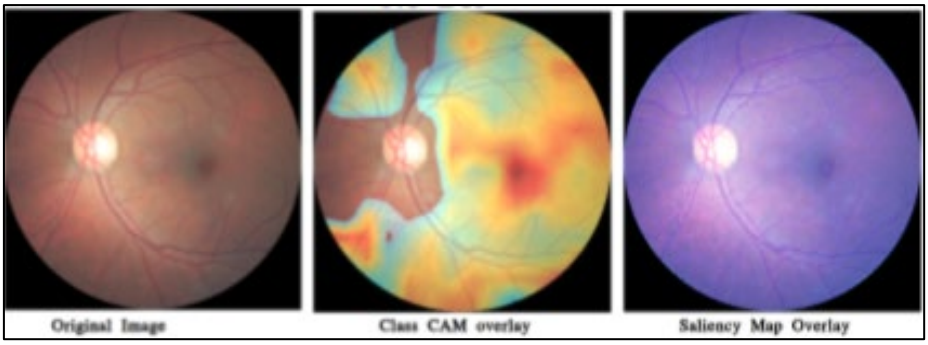


Figure 5.3: CAM and Saliency maps for grade 0

2. Figure 5.4 and 5.5 demonstrates the CAM and saliency for grade 1 and grade 2 respectively. These are the cases of mild and moderate DR. The CAM demonstrates that the important regions which leads to mild and moderate DR lies around the vessel or around the macular region which is the center of retina. This shows the initial sign of MA's and hemorrhages and corresponds to the grading system discussed in table 1.1. The pixels are very light for saliency maps for the mild and moderate cases of DR

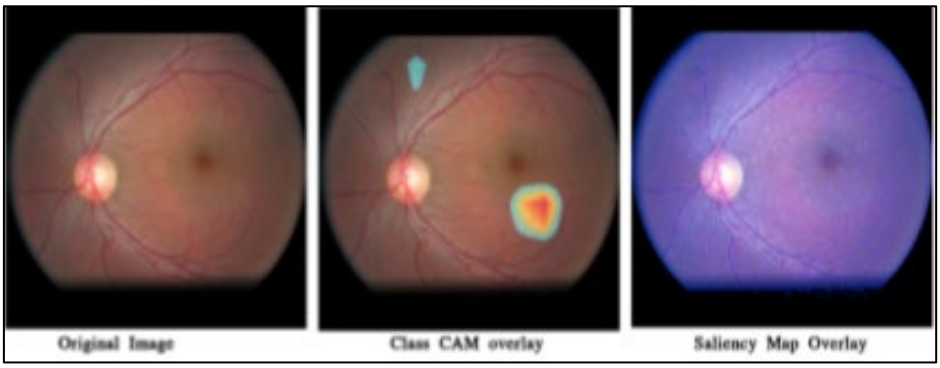


Figure 5.4: CAM and Saliency maps for grade 1

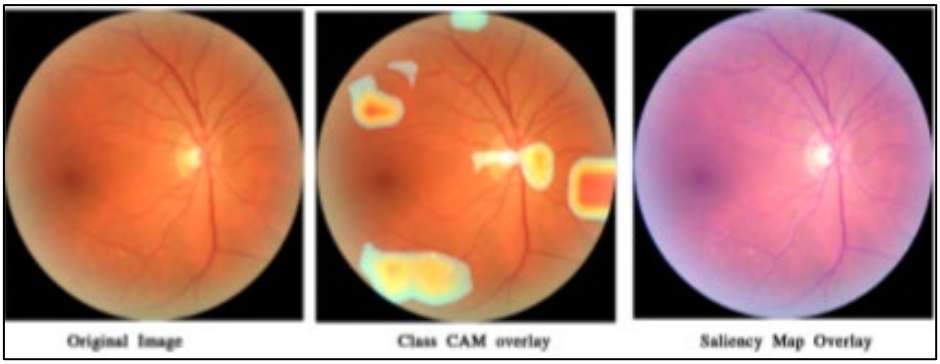


Figure 5.5: CAM and Saliency maps for grade 2

3. Figure 5.6 and 5.7 demonstrates the CAM and saliency for grade 3 and grade 4 respectively. These are the images which belong to severe and proliferative cases of DR. It shows the important regions in red close to the macula. Severe and proliferative cases of DR require the MA's and hemorrhages to appear all over the retina as compared to mild and moderate cases where this is localized to a quadrant. Hence if the model notices these features in other quadrant of retina, it allocates it with a grade of 3 or 4 otherwise 1 or 2. The saliency maps clearly shows the clinical features like MA's, hemorrhages and cotton wool spots which were not visible in case of no DR and were very light in case of mild or moderate DR.

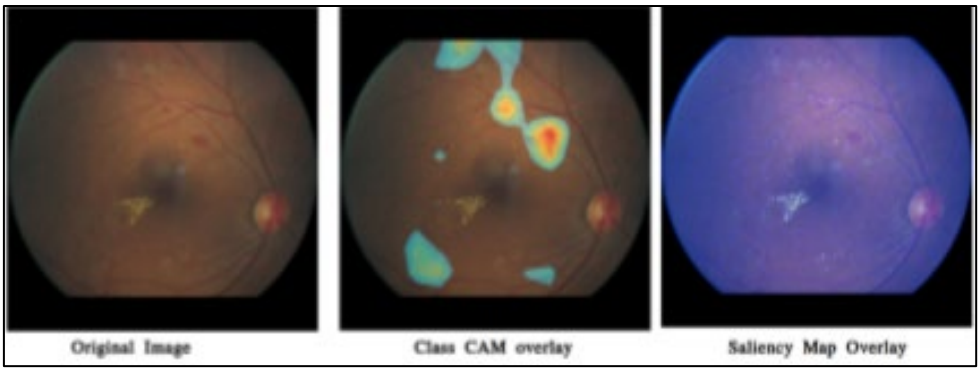


Figure 5.6: CAM and Saliency maps for grade 4

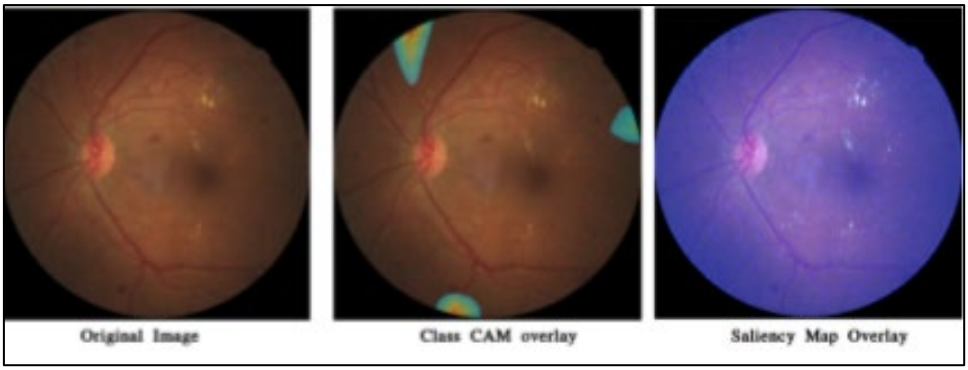


Figure 5.7: CAM and Saliency maps for grade 4

4. The figures 6.6 and 6.7 also indicate that for saliency maps, CNN model is not considering the neovascularization of the optic disc which is an important feature for proliferative case of DR

6: CONCLUSION

This chapter discusses the 5 DR classification techniques, main findings and future work that could be conducted based on these. Section 6.1 summarizes the findings. Section 6.2 discusses the main finding and contributions. Section 6.3 discusses possibilities of future work.

6.1 Summary

In this study, five different DL based techniques were used to classify DR, an eye disease which causes vision loss. A multi-class classification was performed on images that were pre-processed. The pre-processing process involves removal of noise using gaussian blur, image enhancement using CLAHE. The images were then augmented and fed to different CNN models. The parameters learned by the model was then validated on test set from the same dataset and a validation set.

The first approach was developed on basic concept of CNN's in which 5 convolutional layers were used to extract features from eye images. Each of the convolutional layer was followed by batch normalization to prevent overfitting of model. The dense layer in the end would then corelate the extracted features into a prediction. The reported results showed that CNN was effective in classifying grades of DR. However, it could not produce the accuracy and performance that could be significant in predicting all the classes.

In the second approach, we have used transfer learning process and developed a VGG-16 network. We modified the VGG-16 to use LeakyReLU instead of ReLU as activation function. A Rectified Linear Unit (A unit employing the rectifier is also called a rectified linear unit ReLU) has output 0 if the input is less than 0, and raw output otherwise. That is, if the input is greater than 0, the output is equal to the input. The operation of ReLU is closer to the way our biological neurons work. ReLU is non-linear and has the advantage of not having any backpropagation errors unlike the sigmoid function, also for larger Neural Networks, the speed of building models based off on ReLU is very fast. The problem we see in ReLU is the Dying ReLU problem where some ReLU Neurons essentially die for all inputs and remain inactive no matter what input is supplied, here no gradient flows and if large number of dead neurons are there in a Neural Network it's performance is affected,

this can be corrected by making use of what is called Leaky ReLU where slope is changed causing a leak and extending the range of ReLU.

In the third approach, we used another transfer learning approach and developed ResNET-50. The layers we increased from 5 in CNN, 16 in VGG-16 and then to 50 in ResNET-50. We observed that as we moved deep in the network, we were able to achieve better results.

In the fourth approach, we developed DenseNET-121 to confirm that we achieve better results as we move deeper in the network. We observe that this approach gives most appropriate results out of all the approaches that has been tried till now. The macro average and micro average ROC score was above 90 for DenseNET and below 90 for all the models build before this.

In the fifth approach, we developed a CapsNET model which outperformed CNN's in MNIST dataset challenge. CNN's has a major limitation of the inability to retain spatial relationship between learned features in deeper layers. Capsule network with dynamic routing was introduced in 2017 with a speculation that it can overcome this limitation. We observed that CapsNET gave similar results to our first approach of vanilla CNN's. The approaches which involved transfer learning outperformed the CapsNET approach.

Finally, the DenseNET approach was presented to demonstrate how to extract the learned features. Saliency maps proposed a per-pixel visualization of the classification prediction and CAMs proposed a regional based approach to interpreting the prediction. This would help the clinicians to understand the white box structure of CNN's and to interpret how it arrives to the results.

6.2 Main Findings and Contribution

This section discusses the main findings and contributions of this research in the context of aims and objectives and research questions discussed in section 1.3 and section 1.4 respectively.

Below each aims and objective is considered and explained how the research work addresses each of them

1. ***To identify the most suitable image pre-processing technique that can be effectively utilized making retinal features more visible***

Out of the three channels of eye images – Red, Blue and Green, the green channel was observed to be the most appropriate channel for feature extraction. Image noise was removed using gaussian blur and then CLAHE was applied on the green channel for contrast improvement. After this all the 3 channels were merged. There were 5 other different image pre-processing techniques tested on the images. This technique showed the lowest loss out of the various other techniques applied for image pre-processing in this research.

2. ***To determine the most effective method that addresses the class imbalance issue appropriately***

The ResNET-50 most the most effective model which handled the class imbalance problem and outperformed all other models in terms of sensitivity and specificity for all classes. Table 6.1 and Table 6.2 gives a comparison in terms of specificity and sensitivity for all the approaches.

Class Label	***CNN***	***VGG-16***	***ResNET-50***	***DenseNET-121***	***CapsNET***
Class 0	0.76	0.80	0.85	0.86	0.81
Class 1	0.55	0.58	0.63	0.58	0.11
Class 2	0.67	0.66	0.71	0.61	0.37
Class 3	0.52	0.56	0.64	0.60	0.29
Class 4	0.73	0.76	0.78	0.82	0.35

Table 6.1: Specificity comparison for all models

Class Label	***CNN***	***VGG-16***	***ResNET-50***	***DenseNET-121***	***CapsNET***
Class 0	0.85	0.91	0.92	0.93	0.63

Class 1	0.50	0.46	0.55	0.55	0.34
Class 2	0.68	0.71	0.78	0.68	0.31
Class 3	0.28	0.31	0.48	0.45	0.39
Class 4	0.54	0.58	0.67	0.55	0.31

Table 6.2: Sensitivity comparison for all models

3. ***To compare and evaluate the most robust and efficient DL based method that can be utilized for multi class classification for DR***

The DenseNET-121 model was able to produce the best classification accuracy on test and validation sets. Table 6.3 and Table 6.4 gives a comparison between the macro and micro average ROC score for all the approaches.

Class Label	***CNN***	***VGG-16***	***ResNET-50***	***DenseNET-121***	***CapsNET***
Class 0	0.65	0.75	0.82	0.90	0.61

Table 6.3: Macro Average ROC score comparison for all models

Class Label	***CNN***	***VGG-16***	***ResNET-50***	***DenseNET-121***	***CapsNET***
Class 0	0.69	0.84	0.93	0.95	0.65

Table 6.4: Micro Average ROC score comparison for all models

4. ***To suggest a method to interpret and explain the prediction***

   The saliency maps demonstrated a per-pixel visualization and CAMs demonstrated region-based visualizations and suggested that CNN's were able to identify MA's, hemorrhages and exudates. However, it misses neovascularization which is an important marker for proliferative DR.

Below each research question is considered and explained how the research work addresses each of them

1. ***Can we suggest that transferring features from ImageNet model trained on natural images are suitable for eye images?***

   The answer to this question is Yes. We have observed in the research that transfer learning methodology were more robust and accurate in classifying grades of DR when compared to vanilla CNN and CapsNET. Table 6.5 shows the comparison of accuracy for all the models.

Class Label	*CNN*	*VGG-16*	*ResNET-50*	*DenseNET-121*	*CapsNET*
Class 0	0.81	0.86	0.89	0.86	0.69
Class 1	0.52	0.53	0.61	0.60	0.46
Class 2	0.73	0.70	0.75	0.75	0.52
Class 3	0.44	0.42	0.58	0.54	0.32
Class 4	0.64	0.66	0.65	0.72	0.45

Table 6.5: Accuracy comparison for all models

2. ***Is CapsNet more suitable than CNN's for classifying retinal fundus images?***

The answer to this question is No. The accuracy in classifying each of the classes for CapsNET is less when compared to CNN's. Additionally, the training time and parameters involved were much for for CapsNET than CNN. Hence, we could conclude that CNN's were better than CapsNET in classifying DR

6.3 Future Work

In this section sone of the future research direction is discussed

1. ***More training images containing neovascularization***

As we observed in the research findings that CNN models were able to interpret MA's, hemorrhages and exudates but not neovascularization. The reason for this was lack of training data containing more images that shows neovascularization. Having more images for this would make it clear for the model that it should identify it and classify it as proliferative DR. In addition to this, object detection methodology could also be used to generate labelled regions of neovascularization.

2. ***Classification using optic coherence tomography imaging***

OCT is a non-invasive imaging technique relying on low coherence interferometry to generate in vivo, cross-sectional imagery of ocular tissues. Originally developed in 1991 as a tool for imaging the retina, OCT technology has continually evolved and expanded within ophthalmology as well as other medical specialties. The future research work could be directed to study the features using optic coherence tomography (OCT) images

3. ***Using encoder-decoder architecture for DR classification***

The architecture of the encoder network is topologically identical to the CNN's. The role of the decoder network is to map the low-resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. The novelty lies is in the

manner in which the decoder up samples its lower resolution input feature map's. Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear up sampling. Hence, comparing an encoder-decoder based architecture could be a possible future research direction

REFERENCES

4. Acharya, U.R., Lim, C.M., Ng, E.Y.K., Chee, C. and Tamura, T., (2009) Computer-based detection of diabetes retinopathy stages using digital fundus images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2235, pp.545–553.
5. Adarsh, P. and Jeyakumari, D., (2013) Multiclass SVM-based automated diagnosis of diabetic retinopathy. *International Conference on Communication and Signal Processing, ICCSP 2013 - Proceedings*, pp.206–210.
6. Almotiri, J., Elleithy, K. and Elleithy, A., (2018) Retinal vessels segmentation techniques and algorithms: A survey. *Applied Sciences (Switzerland)*, 82.
7. Arunkumar, R. and Karthigaikumar, P., (2017) Multi-retinal disease classification by reduced deep learning features. *Neural Computing and Applications*, [online] 282, pp.329–334. Available at: https://doi.org/10.1007/s00521-015-2059-9.
8. Ashikur, M., Arifur, M. and Ahmed, J., (2020) Automated Detection of Diabetic Retinopathy using Deep Residual Learning. *International Journal of Computer Applications*, 17742, pp.25–32.
9. Bui, T., Maneerat, N. and Watchareeruetai, U., (2017) Detection of cotton wool for diabetic retinopathy analysis using neural network. *2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCIA)*, pp.203–206.
10. Chandore, V., (2017) Automatic Detection of Diabetic Retinopathy using deep Convolutional Neural Network. 3, pp.633–641.
11. Chudzik, P., Majumdar, S., Calivá, F., Al-Diri, B. and Hunter, A., (2018) Microaneurysm detection using fully convolutional neural networks. *Computer methods and programs in biomedicine*, 158, pp.185–192.
12. Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abramoff, M., Mendonca, A.M. and Campilho, A., (2018) End-to-End Adversarial Retinal Image Synthesis. *IEEE transactions on medical imaging*, 373, pp.781–791.
13. Gritsevskiy, A. and Korablyov, M., (2018) Capsule networks for low-data transfer learning. *arXiv*, pp.1–11.
14. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L. and Webster, D.R., (2016) Development and validation of

a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA - Journal of the American Medical Association*, 31622, pp.2402–2410.

15. Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li and Li Fei-Fei, (2009) ImageNet: A large-scale hierarchical image database. pp.248–255.
16. Kahai, P., Namuduri, K.R. and Thompson, H., (2006) A decision support framework for automated screening of diabetic retinopathy. *International Journal of Biomedical Imaging*, 2006, pp.1–8.
17. Khojasteh, P., Passos Júnior, L.A., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J.P. and Kumar, D.K., (2019) Exudate detection in fundus images using deeply-learnable features. *Computers in biology and medicine*, 104, pp.62–69.
18. Krizhevsky, B.A., Sutskever, I. and Hinton, G.E., (2012) Cnn实际训练的. *Communications of the ACM*, 606, pp.84–90.
19. Kusakunniran, W., Wu, Q., Ritthipravat, P. and Zhang, J., (2018) Hard exudates segmentation based on learned initial seeds and iterative graph cut. *Computer methods and programs in biomedicine*, 158, pp.173–183.
20. Lam, C., Yu, C., Huang, L. and Rubin, D., (2018) Retinal Lesion Detection With Deep Learning Using Image Patches. *Investigative opthamology and visual science*, 59, pp.590–596.
21. Liew, G., Michaelides, M. and Bunce, C., (2014) A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open*, 42, pp.1–6.
22. Mahapatra, D., (2016) Iowa Research Online Retinal Image Quality Classification Using Neurobiological Models of the Human Visual System Retinal Image Quality Classification Using System.
23. Mo, J. and Zhang, L., (2017) Multi-level deep supervised networks for retinal vessel segmentation. *International journal of computer assisted radiology and surgery*, 1212, pp.2181–2193.
24. Nazir, T., Irtaza, A., Javed, A., Malik, H., Hussain, D. and Naqvi, R.A., (2020) Retinal Image Analysis for Diabetes-Based Eye Disease Detection Using Deep Learning. *Applied Sciences*, [online] 1018. Available at: https://www.mdpi.com/2076-3417/10/18/6185.
25. Nijalingappa, P. and Sandeep, B., (2015) Machine learning approach for the

identification of diabetes retinopathy and its stages. *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp.653–658.

26. Nitish Srivastava Geoffrey Hinton Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov, (2018) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 15, pp.7642–7651.
27. Pérez, A.D., Perdomo, O. and González, F.A., (2020) A lightweight deep learning model for mobile eye fundus image quality assessment. January 2020, p.67.
28. Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnel, M. V., Corrado, G.S., Peng, L. and Webster, D.R., (2017) Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. *arXiv*.
29. Pratt, H., (2019) Deep Learning for Diabetic Retinopathy Diagnostics. *Liverpool University*. [online] Available at: https://livrepository.liverpool.ac.uk/3046567/.
30. Ramasubramanian, B. and Selvaperumal, S., (2016) A comprehensive review on various preprocessing methods in detecting diabetic retinopathy. *International Conference on Communication and Signal Processing, ICCSP 2016*, pp.642–646.
31. Roychowdhury, S., (2016) Classification of large-scale fundus image data sets: a cloud-computing framework. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2016, pp.3256–3259.
32. Sabour, S., Frosst, N. and Hinton, G.E., (2017) Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 2017-DecemNips, pp.3857–3867.
33. Simonyan, K., Vedaldi, A. and Zisserman, A., (2014) Deep inside convolutional networks: Visualising image classification models and saliency maps. *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, pp.1–8.
34. Singalavanija, A., Supokavej, J., Bamroongsuk, P., Sinthanayothin, C., Phoojaruenchanachai, S. and Kongbunkiat, V., (2006) Feasibility Study on Computer-Aided Screening for Diabetic Retinopathy. *Japanese Journal of Ophthalmology*, [online] 504, pp.361–366. Available at: https://doi.org/10.1007/s10384-005-0328-3.
35. Ting, D.S.W., Pasquale, L.R., Peng, L., Campbell, J.P., Lee, A.Y., Raman, R., Tan,

G.S.W., Schmetterer, L., Keane, P.A. and Wong, T.Y., (2019) Artificial intelligence and deep learning in ophthalmology. *The British journal of ophthalmology*, 1032, pp.167–175.

36. Villalobos-Castaldi, F.M., Felipe-Riverón, E.M. and Sánchez-Fernández, L.P., (2010) A fast, efficient and automated method to extract vessels from fundus images. *Journal of Visualization*, [online] 133, pp.263–270. Available at: https://doi.org/10.1007/s12650-010-0037-y.
37. Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y. and Yang, G., (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, 149, pp.708–717.
38. WHO, W.H.O., (2018) *GLOBAL REPORT ON DIABETES*.
39. Williamson, T.H., Gardner, G.G., Keating, D., Kirkness, C.M. and Elliott, A.T., (1996) Automatic detection of diabetic retinopathy using neural networks. *Investigative Ophthalmology and Visual Science*, 373, pp.940–944.
40. Xu, B., Wang, N., Chen, T. and Li, M., (2015) Empirical Evaluation of Rectified Activations in Convolutional Network. [online] Available at: http://arxiv.org/abs/1505.00853.
41. Xu, K., Feng, D. and Mi, H., (2017) Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules*, 2212.
42. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A., (2016) Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, pp.2921–2929.

APPENDIX A: RESEARCH PROPOSAL

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DR Diabetic Retinopathy.

AI Artificial Intelligence.

DL Deep Learning.

CNN Convolutional Neural Networks.

Caps Net Capsule Network.

UK United Kingdom.

NPDR Non proliferative diabetic retinopathy.

PDR Proliferative diabetic retinopathy.

MA Microaneurysms.

LDA Linear Discriminant Analysis

SVM Support Vector Machine

KNN K Nearest Neighbor

GPU Graphical Processing Unit.

ANN Artificial Neural Network

ILSVRC ImageNet Large Scale Visual Recognition Challenge

TP True Positive

TN True Negative

FP False Positive

FN False Negative

1. Background

1.1 Overview

Diabetic retinopathy (DR) is one of the major causes of blindness in the world for people in working age who are suffering from diabetes (Pratt, 2019). Approximately 420 million people worldwide have been diagnosed with diabetes. The prevalence of this disease has doubled in the past 30 years. Of those with diabetes, approximately one-third are expected to be diagnosed with DR, a chronic eye disease that can progress to irreversible vision loss (Lam et al., 2018). DR occurs due to the existence of microvascular damage to blood vessels of the light sensitive tissue (retina) at the back of the eye. There are two primary stages of DR – Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). In NDPR, narrow bulges called microaneurysms (MA) and other abnormalities including, hemorrhages, exudates, venous beading etc. can be identified. Large retinal blood vessels also start dilating and become irregular in diameter and as more vessels become blocked, NPDR progresses from mild to moderate and then to severe with presence of more abnormalities. Figure 1 depicts retinal features of DR. In PDR, the damaged blood vessels leak a transparent jelly-like fluid that fills the center of the eye causing the development of abnormal blood vessels in the retina. Pressure build up in the eyeball because of the newly grown blood vessels that interrupt the normal flow of the fluid which damages the optic nerve and leads to blindness. The classification rule of DR is given in Table 1.

It is imperative that people suffering from diabetes are screened for earlier signs of DR so that more severe cases could be prevented . However, detection of DR is a challenging task as it has minor symptoms at its nascent stages those symptoms are very hard to detect. Additionally, the detection process is time consuming and costly that needs to take place at skilled clinical facilities under the supervision of trained experts. Due to these challenges there could be variations in classification across clinics depending upon the grader which might lead to delayed and inefficient diagnostic mechanism. Automated detection of DR through machine learning and artificial intelligence-based methods could overcome all these challenges.

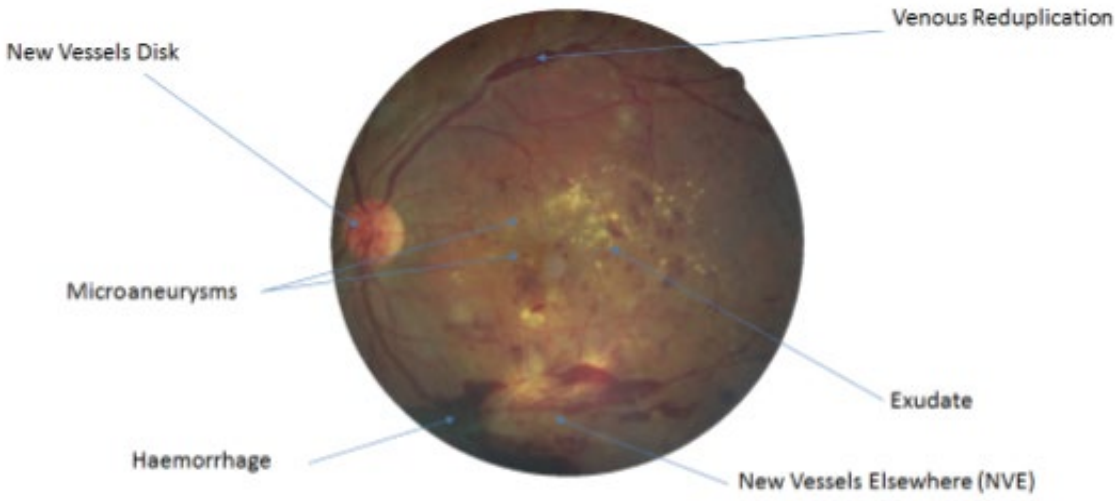


FIGURE 1.1: Example of retinal features in eye fundus images for detection of DR.

Disease Grades	Finding
Grade 0: No Diabetic retinopathy	No visible signs of abnormalities
Grade 1: Mild NPDR	Micro-aneurysms (MA) only
Grade 2: Moderate NPDR	At least one hemorrhage or MA and/or at least one of the following: • Retinal hemorrhages. • Hard/soft exudates • Venous beading
Grade 3: Severe NPDR	Any of the below mentioned symptoms but no symptoms of PDR. This is generally called 4-2-1 rule. • More than 20 intraretinal hemorrhages in each of the four quadrants

	• Venous beading in two or more quadrants • Intraretinal microvascular abnormalities in one or more quadrants
Grade 4: PDR	One of either: • Neo-vascularization • Vitreous/preretinal hemorrhages.

Table 1.1: Classification of DR stages

Previous efforts have been made in this area using feature engineering i.e. image feature extraction and traditional machine learning method. The retinal features are extracted using image processing techniques and passed to a classifier viz. artificial neural networks (ANN's), sparse representation, linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbors (KNN) and so on. However, there lies a complex combination of features that leads to different stages of DR. Additionally, feature engineering process involved is time consuming and require domain expertise.
Deep learning (DL) belongs to the wide family of machine learning methods that have been known to overcome all these limitations and aims to identify the salient low-level features like edges, textures etc. of an image without explicit feature engineering. In 2012, (Krizhevsky et al., 2012) introduced convolutional neural networks (CNN), a deep learning technique known as became popular for solving the image classification problem . The model recorded state-of-the-art performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition (Jia Deng et al., 2009) outperforming other commonly used machine learning techniques. CNN's are most applicable to image classification, segmentation and object detection tasks and requires large amount of data for training. The most popular method initially, when applying CNNs is transfer learning. Transfer learning involves using pre-trained models on natural (non-medical) images to reduce the training time that is required to undertake the computationally heavy CNN training over large data sets. Capsule Networks also known as CapsNet presented in (Sabour et al., 2017) is another widely used deep learning technique that solves some of

the disadvantages of CNN's. In work presented by (Gritsevskiy and Korablyov, 2018), CapsNet was able to generalize information over 25 times faster than a CNN's.

1.2 Purpose

Through recent advancements in DL techniques, and the innovations in GPU computing power, it is now possible to leverage DL methods to more complex tasks such as medical image analysis. The main purpose of the research will be-

To design, build, test, evaluate and compare automated screening models for classifying diabetic retinopathy disease into 5 classes, using suitable deep learning techniques

1.3 Research Problems & Challenges

1. **Data Imbalance**

   More than 70% of the images in Kaggle dataset (KD) are from normal any complications related to DR (Grade 0). Moreover, images related to severe cases of DR are rarer than images related to moderate and mild stages of DR. Therefore, classification algorithm needs to learn from sparse information and hence appropriate data augmentation methods must be applied to handle class imbalance problem appropriately.

2. **Complexity of diagnosis**

   There are numerous and complex features that are involved in classifying DR into several stages and the different combinations of these features result in different classification results. Therefore, any proposed algorithm must take combination of these features into account and hence image-preprocessing techniques must be applied so that complex features are easier to detect and improves classification accuracy.

3. **Choosing appropriate method for high level performance**

   There is a wide range of DL based methods used for data and image analysis classification task. Hence, the proposed model should be robust and

computationally efficient to be used for further research and clinical processes. Ideally, the outcome should match or beat the agreement between graders.

2. Related Work

Retinal features are very helpful to detect diabetic retinopathy and different approaches have been applied to identify the those features on the surface of the eye.

(Williamson et al., 1996) used feature engineering approach to build an artificial neural network (ANN). The network was trained on 83 labeled images to detect features like vessels, exudates, and hemorrhages. The model was then tested on 100 images to determine if a fundus image contains those features of disease. The reported accuracy for the detection of vessels, exudate and hemorrhages were 91%, 93%, and 74% respectively. The network achieved a sensitivity and specify of 88.5% and 83.5% respectively for the detection of DR. The test set comprised small number of images, and thus it can be argued is not representative of an ideal screening dataset. However, the work showed that ANN's were able to detect features of DR although more robust models are required.

(Acharya et al., 2009) created a five-class classification model by detecting retinal features as hemorrhages, micro-aneurysms, exudate and blood vessels from eye images using image segmentation technique. These retinal features were then fed into a support vector machine (SVM) classifier for multiclass classification. This model produced a sensitivity and specificity of 82% and 86% respectively for PDR and NDPR stages and an average accuracy of 85.9% on the five-class classification problem. This method was tested on a small dataset containing approximately 20 images per class.

(Adarsh and Jeyakumari, 2013) used feature engineering to identify retinal feature and fed them to (SVM) classifier to diagnose DR in to five classes. Different image pre-processing methods were applied to detect retinal blood vessels, exudate, micro-aneurysms, and texture features. The lesion area and texture features were used to build a feature vector for a multi-class SVM task. This model reported accuracies of 96% and 94% for DR classification on the image dataset DIARETDB0 and DIARETDB1 containing 89 and 130 images respectively.

All the above-mentioned work relies on feature engineering methods. Additionally, all the methods described above were trained and tested on smaller dataset of approximately 100 images or less and the reported results do not consider any aspects of time spent in diagnosis. Since feature engineering process is time consuming process and for automated

methods to be efficient and accurate, the image analysis techniques and models must generalize over a large dataset to prove its robustness and stability. CNN's can be seen as an alternative to the above-mentioned methods for image classification tasks as they have the capacity to automatically detect features of images and are new state-of-art.

(Xu et al., 2017) classifies images into DR and non-DR grades and compares traditional feature extraction-based approach with CNN model approach on 800 labelled images of Kaggle dataset. Four different feature engineering approaches were used to identify hard exudates, retinal lesions, MA's, and blood vessels respectively. The images were resized to 224*224 pixel and were passed as an input to a 13-layer CNN model. The result shows that CNN based methods are superior to feature extraction approach and reported accuracy of CNN model is 94.5 % while those of feature extraction approaches are below 90%. However, the paper does not present any evaluation of sensitivity and specificity scores. Therefore, it is not possible to determine how well the model has learned complex features of DR.

A similar work to this on developing CNN model with dropout techniques but with more layers was presented by (Chandore, 2017) in which they implemented a deep CNN network which contained 24 layers. The model used images of 448 × 448 pixels and was trained on a KD of 61,000 eye images for DR classification. The model was validated on 14,000 eye images and achieved and accuracy of 85% with reasonable sensitivity and specificity values of 81% and 88% respectively.

(Gulshan et al., 2016) used transfer learning methodology using Inceptionv3 model pre-trained on ImageNet dataset to classify images into normal and DR images. An ensemble of 10 networks was used for classification the final prediction was calculated by taking a linear average over all predictions given by the ensemble. The dataset that was used in training contained 128,175 eye images. The observed results were validated on two test datasets, Eye-PACS-1 with sizes 9963 images and Messidor-2 with size of 1748 images, had sensitivities of 90.3 % and 87%, respectively while specificity was 98% and 98.5% respectively. The area under receiver operating curve is0.991 for Eye-PACS-1 and 0.990 for Messidor-2. There were 22 million parameters in the model and hence the model was heavy weight requiring high training time and resources.

A comparison of transfer learning models was performed by (Lam et al., 2018) in which they compared five pre-trained model that were a part of ILSVRC challenge – VGG-16,

AlexNet, Resnet, GoogleNet and InceptionV3 for binary and multi-class classification. Images were resized to 2048*2048 pixels and the approach was to use the Kaggle dataset to train the model and tested the proposed approach on 195 images of E-Optha dataset. The authors considered a multi-class classification task with five DR grades. The model which outperformed other models was InceptionV3 with a multi-class accuracy of 96% and a binary-class accuracy of 98%.

A transfer learning method was used by (Ashikur et al., 2020) in which involves using pre-trained models on natural image datasets and retraining the weights towards a different medical image dataset. Pre-trained models that were user are VGG-16, AlexNet and GoogleNet . The results showed that GoogleNet outperformed other models and reported sensitivity and specificity of 95% and 96% respectively. The paper also showed that the results for classifying the earlier stages of DR were still less with 29% sensitivity reported for the mild stage of DR.

(Pérez et al., 2020) argued that transfer learning methods based on pre-trained models such as DenseNet, AlexNet and Inception that were a part of ILSVRC Image Net challenge contains huge number of parameters and thus are not suitable for quality assessment on light weight devices. They presented a lightweight CNN model named MFQ-NET trained on Kaggle dataset suitable to run on mobile devices. The main idea is to train an initial smaller model on image patches and later extend it to full images. The network consists of a fifteen layers CNN, which is structured on two main blocks: patch feature extraction (PFE) block and the image classification (IC) block. The first block is pretrained with 224 × 224 × 3 patches extracted from the original images. The second block takes the output of the first block, which is extended from patches to full images, and makes the prediction for a full 896×896×3 image. The results in the paper show that the MFQ-Net was as effective as other pre-trained deep models but with a size which is one to two orders of magnitude less than these models and achieved an accuracy of 92% and 85% for binary and three class classifier. The precision for binary and three class classifiers was 0.87 and 0.85 respectively and recall was 0.94 and 0.85 respectively.

Table 2 shows summarizes different methodologies and obtained results for the past work done in DR classification.

Year	Author	Methodology	Outcome
1996	(Williamson et al., 1996)	Feature Extraction	Sensitivity – 88.40% Specificity – 83.50%
2009	(Acharya et al., 2009)	Feature Extraction	Sensitivity – 82% Specificity – 86%
2013	(Adarsh and Jeyakumari, 2013)	Feature Extraction	Accuracy -96%. No sensitivity and specificity reported
2017	(Xu et al., 2017)	CNN	Accuracy -94.5%. No sensitivity and specificity reported
2017	(Chandore, 2017)	CNN	Sensitivity – 81% Specificity – 88%
2016	(Gulshan et al., 2016)	Inceptionv3	Sensitivity – 90.3% Specificity – 87%
2018	(Lam et al., 2018)	Comparison of following models – AlexNet VGG16 GoogleNet Resnet Inceptionv3	Best model was Inceptionv3 with- Accuracy – 98%
2020	(Ashikur et al., 2020)	Alexnet VGGNet GoogleNet	GoogleNet performed best with- Sensitivity – 95% Specificity – 96%
2020	(Pérez et al., 2020)	CNN (MFQ-Net)	Sensitivity –87% Specificity – 94%

Table 2.1: Literature Review

3. Research Question

The following questions needs to be explored from the research-

1. **Can we suggest that transferring features from ImageNet model trained on natural images are suitable for eye images?**

   ImageNet pre-trained model is mainly trained using natural images which contains more information than medical images. Because of the big difference between natural images and medical images, we need to fine-tune our networks. Additionally, the models based on ImageNet have millions of parameters involved. Transfer learning may have a very limited effect when we switch the domain from one type to another. Hence, this case may be no better than training from scratch, as the networks learn very different high-level features in the two tasks. Certainly, we know if we have enough data, training from scratch is a feasible approach. Hence, we need to explore if competitive results could be achieved by training CNN's from scratch rather than using ImageNet based pre-trained model. Additionally, we would also want to compare the parameters involved in both type of learning.

2. **Is CapsNet more suitable than CNN's for classifying retinal fundus images?**

   The aim of the research is to perform a 5-class classification for DR and classification rules mentioned in Table 1 are complex and dependent on the presence of features in accordance with the quadrants of eye image. The closer a feature is to the center of the eye i.e., macula, the area in the center of the retina, the more effect it has towards the severity of DR classification. This makes our research more challenging as CNN's are not spatially invariant. Instead, they rely on pooling layers to achieve translation invariance, and on data-augmentation to handle rotation invariance. Additionally, CNN requires lot of data for training process other it suffers with model overfitting issues. Recently, CapsNet (Sabour et al., 2017) was introduced as an alternative DL based architecture and training approach to overcome the disadvantages on CNN's and model the spatial variance of a feature or an object in the image. This property is known as equivariance and

this type of learning is also called as one-shot learning type of vision. The research conducted will answer if CapsNet are more suited to perform a 5-class classification on retinal fundus images than CNN's.

4. Aims and Objectives

1. **To identify the most suitable image pre-processing technique that can be effectively utilized making retinal features more visible**

   The dataset that will used in the research comes from different models and variants of cameras so the images would be affected by color and . These variations might not let the features be visible and hinder with classifier performance. Therefore. it is imperative to remove any variation and noise from images so that the model could identify DR features more accurately. Several experiments need to be performed on eye images to identify which pre-processing technique gives better results.

2. **To determine the most effective method that addresses the class imbalance issue appropriately**

   Any medical image analysis task generally needs to take account of the fact that it will see more negative cases than the positive ones and hence the tasks should be able to learn from minimal amounts of data with respect to the class which has less images. From a medical point of view, clinicians would expect model that have a high sensitivity meaning a smaller number of false negatives. A classification model for DR that has a high sensitivity indicates that the technique is robust and will not miss any positive case during screening. Hence, the deep learning model to be implemented for the classification of diabetic retinopathy screening must provide high sensitivity.

3. **To determine the most robust and efficient DL based method that can be utilized for multi class classification for DR**

   The research will aim to explore three DL based techniques. The most appropriate deep learning method to extract the features of DR from image data and to accurately classify into 5 stages needs to be identified.

4. **To suggest a method to interpret and explain the prediction**

The application of DL based models is widely perceived to be black box, i.e., non-interpretable, in the clinical community. It is very difficult to identify that if the model built for DR severity classification is adhering to grading framework as shown in Table 1 or if the model has learned its own framework for classification task. Hence, we need to answer if DL based models can provide regional and pixel-based prediction interpretations. We need to explore and explain that what are the predictors how the model has reached to is prediction.

5. Scope and Significance of Study

The increasing rate of diabetes is well known to everyone around the world. 422 million people are now living with some type of diabetes, and this number is projected to increase rapidly in future (WHO, 2018).

The research will be significant in below mentioned cases-

1. **Screening programs covered by public and private authorities**

   With the advent of high-resolution retinal imaging system many countries in the world have implemented screening programs to cover their citizens for DR detection at an early stage to minimize the risk of vision loss. The impact of this screening program is clear that from the fact that DR is no longer the leading cause of vision impairment in the UK(Liew et al., 2014). However, conducting screening programs are complex and costly and required highly trained graders. These programs require at least annual screening of people with diabetes and as a result produce large quantities of digital images that need to be studied. The screening process is also costly and repetitious which puts pressure on graders to produce high quality results over a longer period. Additionally, given the increasing rise of diabetes, country-wide screening programs might struggle to meet the ever-rising demand using manual grading approach. Hence DR diagnosis through automated screening will be helpful to cover large mass in screening program with cost-effective and standardized way.

2. **Availability in areas which lack clinical resources**

   Advances methods and technology could not be available in rural areas and hence delayed process in detection of DR can lead to propagate the disease to higher stages and introducing more complications and cost in treatment. Rural areas can benefit from automated detection is several ways. With application of an AI system in those areas, the photos could be uploaded to an online server and the server will process the images and provide results in less time. Moreover, DR could be detected at early stages and necessary steps could be taken to prevent its

advancement to higher stages. Hence, automated detection can be identified as a cost-effective, less time-consuming use of health service resource in remote areas which lack resources and facilities.

Due to time limitations, the following points were not considered in scope of the research-

1. The data in the research is taken from Kaggle dataset provide by Eye-PACS, an organization which provides end-to-end services to implement a successful blindness prevention program. Hence validation on images and their grades will not be a part of this research.

2. As AI based deep learning methods are a part of ongoing research, there are wide number of techniques and models available in deep learning. We will limit our techniques to only those that will answer our questions given in section 3 and will help us to achieve objectives given in section 4. The research will be limited to the application of CNN's models, pre-trained CNN models that were a part of ILSVRC challenge and Capsule networks.

6. Research Methodology

The research methodology is a process that will help to achieve our objective discussed in section 4 and answer questions discussed in section 3 of this proposal. The initial work was to collect data which is complete. Figure 2 depicts flow of how the research will be conducted from data loading to evaluation and interpretation.

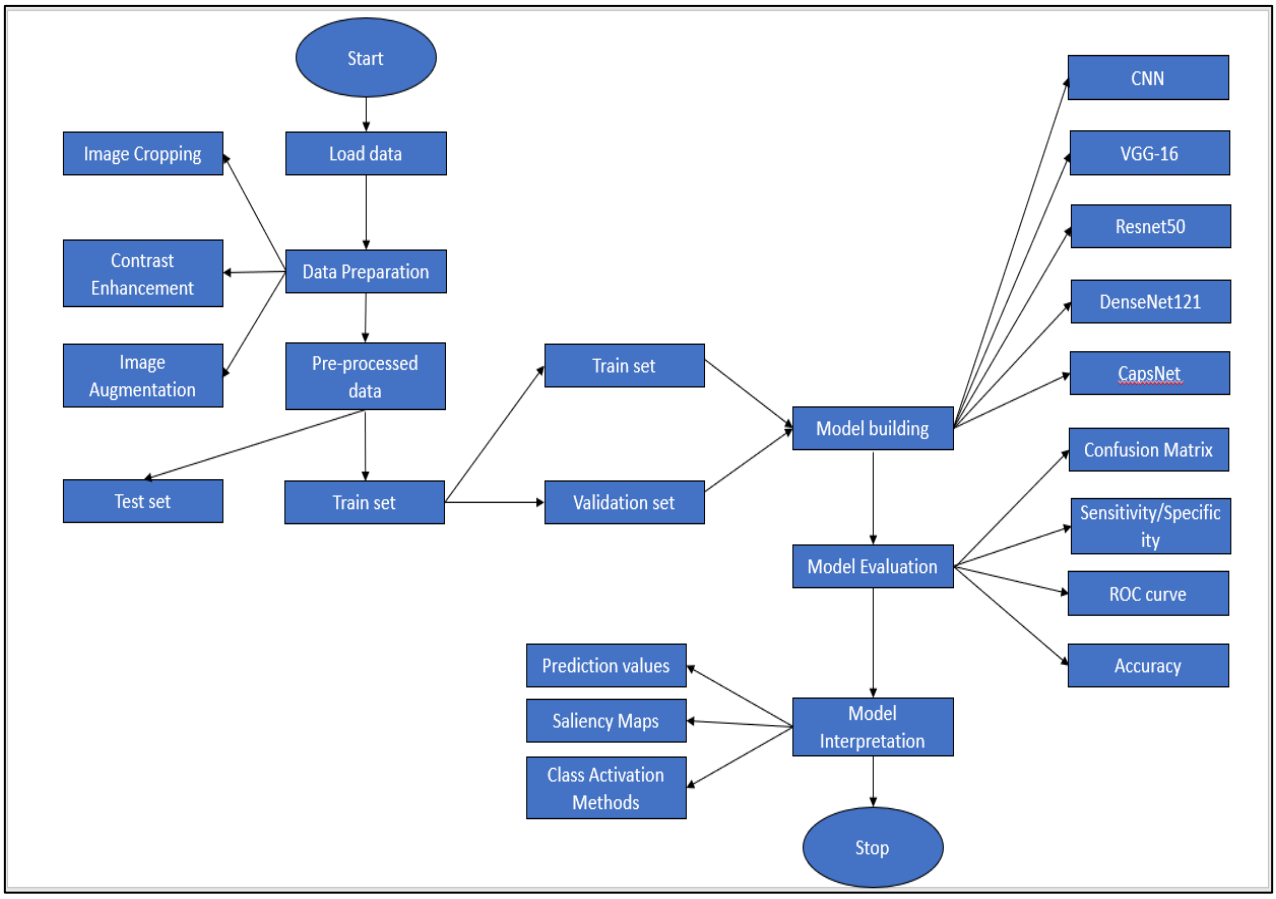


Figure 6.1: Research Flow

The following sections discuss in detail about data understanding and preparation, model building and evaluation statistics.

6.1 Dataset Description:

Kaggle dataset (KD) is a well-known and widely used for the detection of diabetic retinopathy. This dataset contains the total number of 88,702 retinal fundus photographs. This dataset was produced by Eye-PACS to facilitate the researchers without any cost. In this dataset, 35,126 fundus photographs were assigned for training purposes and 53,576 for

the testing. A grader has graded the level of DR in each patient's eye using the fundus images and according to the five point scale presented earlier in Table 1. The total number of images in each class is given in Table 3.

Grades	Train Images	Test Images
0	25,810	39,534
1	2443	3,762
2	5292	7,861
3	873	1,214
4	708	1206

Table 6.1: Dataset overview

6.2 Data Pre-processing:

The color retinal images are many times degraded by several problems like noise, uneven illumination, poor contrast, and variation in capturing. Hence, image preprocessing needs to be done to remove the noise and enhance the image so that DR features are more visible. The following steps would be taken to pre-process image.

- **Cropping**

  Since the original images are large (say, 3000x2000 pixels on average) and most of them contained a large significant black border. We will start by removing most of these black borders. We require square matrix images as the input of our network, the images will be first resized to say 3000 x 3000 (in the case of 3000 x 2000) by adding extra black borders and then resizing these images to 192*192 pixels for Caps-Net and 256*256 pixels for CNN's.

- **Contrast Improvement**

  Retinal features are very important in classification of DR. Different combination and complex features leads to various grade identification in DR. Hence identifying those grades accurately becomes important. Images in the dataset come from different models and types of cameras, which can affect the visual appearance. Thus, we need to apply techniques that would improve the contrast on eye images so that retinal features are clearly identified. (Ramasubramanian and Selvaperumal,

2016) presented a review on several pre-processing techniques on eye fundus images. We will experiment with some of the techniques using OpenCV library to see which method could be a best representation of retinal features in the eye image.

The criteria for identifying the best technique will be the one which will give the smallest loss after 25 epochs will be used for modeling.

- **Augmentation**

As seen in Table 2, the data is highly imbalanced and most of the cases belong to Grade 0. Additionally, deep CNN's are designed to be spatially invariant, that is - they are not sensitive to the position and need huge amount of data to prevent overfitting. To address all these challenges, we need to apply data augmentation techniques like flipping, shifts and rotation etc.

6.3 Model Building

The three modelling techniques to be considered are:

1. **Deep CNN's to be trained from scratch**

A convolutional neural network consists of an input layer, hidden layers and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a convolutional neural network, the hidden layers include layers that perform convolutions. Typically this includes a layer that does multiplication or other dot product, and its activation function is commonly ReLU. This is followed by other convolution layers such as pooling layers, fully connected layers and normalization layers. The architecture could be understood from Figure 3.

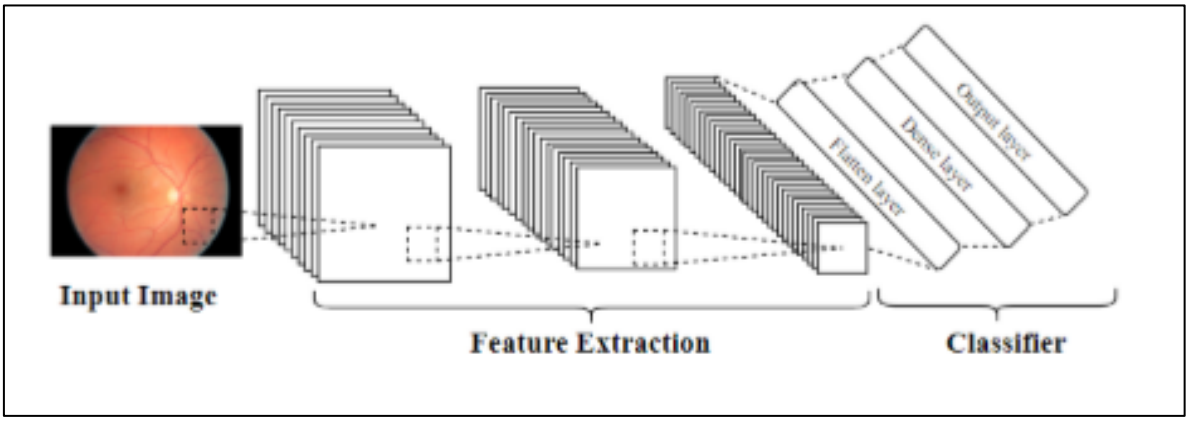


Figure 6.2: CNN Architecture

2. Transfer learning using pre-trained model.

The transfer learning (TL) approach deals with training a base network on one task and then using that network for some other task. The first n-layers of the model can then be freezed and the user can choose to backpropagate the errors back into the networks up to that n-layer. The choice of choosing n depends upon the size of dataset and number of parameters in that layers. The major advantage of transfer learning is that it requires less training data as it is difficult and time consuming to train a model from scratch. Additionally, using TL training could start from pre-trained weights instead of randomly initialized weights which gives many advantages in training phase

This TL process will tend to work if the features are general, meaning suitable to both base and target tasks, instead of specific to the base task. Hence, we will need to analyze if features are transferrable from pre-trained network based on natural images to the task of classifying medical images. This will provide answers to our first question in section 3. The pre-trained model that we would use are as follows:

- **VGG-16**

  VGG16 provided an accuracy rate of 91.90% in the ILSVRC competition in 2014. It is a 16-layer deep network which network has an image input size of 224-by-224 .The network has 138,355,752 parameters and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images.

- **Resnet50**

  ResNet, which stands for residual network, was introduced by (He et al., 2016) in 2015 and achieved an accuracy of 94.29% in ILSVRC challenge. It has a total of 25,000,000 parameters. In comparison with other models, the ResNet has residual connection that makes sure that during backpropagation, the weights learned from the previous layers do not vanish and are easier to optimize and can gain accuracy from considerably increased depth. The main benefit of this model is the use of residual connections so that it is possible to make network deeper.

- **DenseNet121**

  DenseNet (densely connected convolutional networks) architecture (Huang et al., 2017) has 8,062,504 parameters and was inspired by ResNet, but instead

of the residual connections, the model uses dense blocks. The dense block consists of sequential convolution layers, like VGG, but each layer has connection to all other layers. The main idea behind in this approach is to get information from all previous layers so as to minimize information loss. The model achieved a 93.34% accuracy rate on the ILSVRC challenge.

3. Capsule Networks(CapsNet)

Capsule Networks (CapsNet) introduced by (Sabour et al., 2017) are the networks that are able to fetch spatial information and more important features so as to overcome the loss of information that is seen in pooling operations in CNN's. Capsule networks achieved a state-of-art performance in identifying digits on MNIST dataset. Figure 4 provides architecture of CapsNet used in MNIST dataset (Sabour et al., 2017).

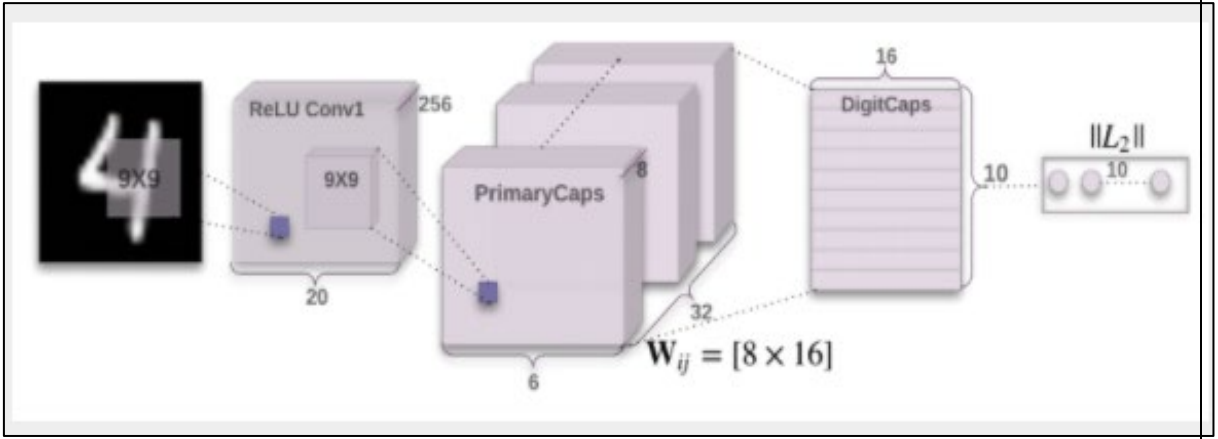


Figure 6.3: CapsNet Architecture

The three modelling techniques will be investigated in the following four-step manner:

1. Development of each model on training images.

2. Experiments to increase the performance of the classification model based on the output of validation images.

3. Analysis of the results obtained from step 2 followed by model tuning and hyperparameter optimization of the technique to further improve the results.

4. Comparison of different models to identify which model is best suitable for DR classification into multiple stages.

6.4 Model Evaluation

A one versus all mechanism will be adopted to find the confusion matrix i.e. we need to compute a confusion matrix for every class and consider that class as the positive class and all other negative class. The confusion matrix is shown in Figure 5. We will use sensitivity, specificity, accuracy, and AUC metrics as evaluation benchmark. If P' represents the positive predicted binary class and N' represents the negative predicted binary class. Similarly, for the ground truth, P represents the positive case and N represents the negative case.

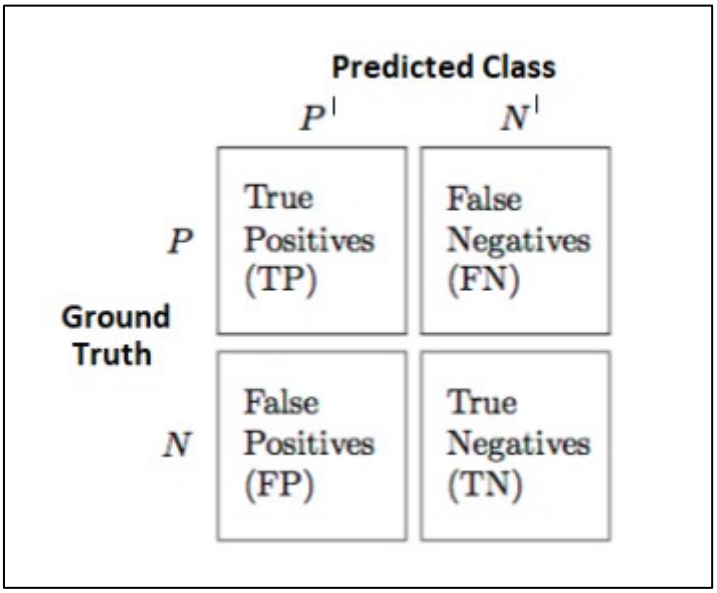


Figure 6.4: Confusion Matrix

The formulas for the same are given below:

Accuracy:

$$Correctly\ Classified\ Images/All\ Images$$

$$(TP\ +\ TN)/(TP\ +\ TN\ +\ FN\ +\ FP)$$

Sensitivity:

$$Correctly\ classified\ positive\ images/All\ positive\ images$$

$$TP/TP\ +\ FN$$

Specificity:

$$Correctly\ classified\ negative\ images/All\ negative\ images$$

$$TN/TN + FP.$$

6.5 Model Interpretation

The outcome of the DL based approach is uninterpretable. It is not clear that which factors are responsible which lead to the classification result. Hence, to explain the prediction of models we will use different technique that will help to understand on how the model reached to its prediction. Different approaches like saliency maps, prediction values and class activation methods (CAM's) needs to be explored to explain the model prediction.

7. Expected Outcome

The expected outcomes from the research will be:

1. Efficient and robust model to classify DR diseases into different grades using DL based approach.

2. Answers to all research questions and objectives discussed in section 3 and 4.

3. Interpretability of model to determine which pixels are more relevant for output prediction.

4. Thesis report document containing all experiments conducted and observations in detail.

5. Video/ PowerPoint presentation summarizing whole research.

6. Opportunity for other researchers for future work in medical image analysis techniques.

8. Required Resources

The software and hardware requirements to be used in the research in given in Table 4

Requirement Type	Details
Software Requirements	Python v3.x
	Libraries - pandas, numpy, sklearn, matplotlib/seaborn
	OpenCV v4.5.1
	Keras v2.3.0
	Tensorflow v2.0
Hardware Requirements	Intel i7 6-core CPU with 32 GB RAM with 1TB Hard disk
	Nvidia Quadro P1000 GPU with 20GB VRAM

Table 8.1: Required Resources

9. Research Plan

Table 5 shows detailed research plan to be carried out.

	Task name	Start date	End date	Duration
		26/01/2021	10/05/2021	15w
1	**Data Preparation**	**26/01/2021**	08/02/2021	2w
1.1	Cropping	26/01/2021	27/01/2021	2d
1.2	Contrast Enhacement techniques	28/01/2021	04/02/2021	1w 1d
1.3	Data augumentation	03/02/2021	08/02/2021	4d
	Add a task | Add a milestone			
2	**Model Development and training**	**09/02/2021**	29/03/2021	7w
2.1	CNN	09/02/2021	22/02/2021	2w
2.2	VGG-16	23/02/2021	01/03/2021	1w
2.3	ResNet50	02/03/2021	08/03/2021	1w
2.4	DenseNet121	09/03/2021	15/03/2021	1w
2.5	CapsNet	16/03/2021	29/03/2021	2w
2.6	Hyperparameter Tuning	09/02/2021	29/03/2021	7w
	Add a task | Add a milestone			
3	**Model Evaluation and Comparison**	**30/03/2021**	12/04/2021	2w
3.1	Evaluate Modelling Results	30/03/2021	12/04/2021	2w
3.2	Compare Results	30/03/2021	12/04/2021	2w
	Add a task | Add a milestone			
4	**Model Interpretation**	**13/04/2021**	26/04/2021	2w
4.1	Prediction Values	13/04/2021	16/04/2021	4d
4.2	Saliency Maps	19/04/2021	22/04/2021	4d
4.3	CAM's	23/04/2021	26/04/2021	2d
	Add a task | Add a milestone			
5	**Documentation**	**27/01/2021**	10/05/2021	14w 4d
5.1	Pre-processing results	27/01/2021	08/02/2021	1w 4d
5.2	Modelling Results	09/02/2021	29/03/2021	7w
5.3	Prepare Research Artifacts	27/04/2021	30/04/2021	4d
5.4	Documentation Review	27/04/2021	10/05/2021	2w
5.5	Documentation Improvements	27/04/2021	10/05/2021	2w

Table 9.1: Research Plan

10. Risk and Contingency Plan

The risks and their mitigation plan are discussed in Table 6

Sno	Risk	Mitigation
1	Loss of document or jupyter notebook changes due to system crashes	Sync machine with One Drive and make sure sync is always on
2	Unavoidable health circumstances due to pandemic or other serious disease	Buffer time maintained in research plan
3	Report not in direction with research plan due to time limitations	Periodic discussion and feedback from thesis supervisor.

Table 10.1: Risk and Mitigation

References

1. Acharya, U.R., Lim, C.M., Ng, E.Y.K., Chee, C. and Tamura, T., (2009) Computer-based detection of diabetes retinopathy stages using digital fundus images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2235, pp.545–553.
2. Adarsh, P. and Jeyakumari, D., (2013) Multiclass SVM-based automated diagnosis of diabetic retinopathy. *International Conference on Communication and Signal Processing, ICCSP 2013 - Proceedings*, pp.206–210.
3. Ashikur, M., Arifur, M. and Ahmed, J., (2020) Automated Detection of Diabetic Retinopathy using Deep Residual Learning. *International Journal of Computer Applications*, 17742, pp.25–32.
4. Chandore, V., (2017) Automatic Detection of Diabetic Retinopathy using deep Convolutional Neural Network. 3, pp.633–641.
5. Gritsevskiy, A. and Korablyov, M., (2018) Capsule networks for low-data transfer learning. *arXiv*, pp.1–11.
6. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L. and Webster, D.R., (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA - Journal of the American Medical Association*, 31622, pp.2402–2410.
7. Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li and Li Fei-Fei, (2009) ImageNet: A large-scale hierarchical image database. pp.248–255.
8. Krizhevsky, B.A., Sutskever, I. and Hinton, G.E., (2012) Cnn实际训练的. *Communications of the ACM*, 606, pp.84–90.
9. Lam, C., Yu, C., Huang, L. and Rubin, D., (2018) Retinal Lesion Detection With Deep Learning Using Image Patches. *Investigative opthamology and visual science*, 59, pp.590–596.
10. Liew, G., Michaelides, M. and Bunce, C., (2014) A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open*, 42, pp.1–6.
11. Pérez, A.D., Perdomo, O. and González, F.A., (2020) A lightweight deep learning model for mobile eye fundus image quality assessment. January 2020, p.67.

12. Pratt, H., (2019) Deep Learning for Diabetic Retinopathy Diagnostics. *Liverpool University*. [online] Available at: https://livrepository.liverpool.ac.uk/3046567/.
13. Ramasubramanian, B. and Selvaperumal, S., (2016) A comprehensive review on various preprocessing methods in detecting diabetic retinopathy. *International Conference on Communication and Signal Processing, ICCSP 2016*, pp.642–646.
14. Sabour, S., Frosst, N. and Hinton, G.E., (2017) Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 2017-DecemNips, pp.3857–3867.
15. World Health Organization, W.H.O., (2018) *GLOBAL REPORT ON DIABETES*.
16. Williamson, T.H., Gardner, G.G., Keating, D., Kirkness, C.M. and Elliott, A.T., (1996) Automatic detection of diabetic retinopathy using neural networks. *Investigative Ophthalmology and Visual Science*, 373, pp.940–944.
17. Xu, K., Feng, D. and Mi, H., (2017) Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules*, 2212.