Name: Sai Chand Reddy Kamana

SID: 2022785

Date: 06-05-2022

Postgraduate Major Project – MOD002726

Diabetic Retinopathy Detection using inception-v3 model

# Declaration

I hereby declare that this project represents my work which has been done after registration for the Master of Science degree at Anglia Ruskin University, and the thesis entitled ‘Diabetic Retinopathy Detection’ is no more than 15,000 words in length including tables, figures, and references. The thesis contains no work that has been submitted previously, in whole or in part, for the award of any other academic degree or qualification and cited the sources according to the Harvard referencing.

I have read the University’s current research ethics guidelines, and done the research as per the university guidelines. The support provided during the work, including the assistance from my supervisor, has been indicated in the acknowledgment section.

**Date:** 06-May-2022 **Signature:** Sai Chand Reddy Kamana

# Acknowledgment

I would like to thank my dissertation supervisor, Dr. Mahdi Maktabdar Oghaz, a Lecturer in the School of Computing and Information Science, for his continuous support and invaluable advice. This project would not have been completed without his invaluable supervision. The Anglia Ruskin University Faculty of Computer Science Department also deserves my gratitude, with the entire faculty's continuing support, assistance, and academic advice constituting an important part of my dissertation.

I'd want to express my gratitude to everyone involved.

Table of Contents

[Declaration 1](#_Toc101550459)

[Acknowledgment 2](#_Toc101550460)

[List of Figures 5](#_Toc101550461)

[List of Tables 6](#_Toc101550462)

[List of Equations 7](#_Toc101550463)

[List of Abbreviations 8](#_Toc101550464)

[Abstract 10](#_Toc101550465)

[1. Introduction 11](#_Toc101550466)

[1.1. Overview 11](#_Toc101550467)

[1.2. Problem Background 12](#_Toc101550468)

[1.3. Research Aim 13](#_Toc101550469)

[1.4. Research Objectives 14](#_Toc101550470)

[1.5. Research Scope 14](#_Toc101550471)

[1.6. Methodology 14](#_Toc101550472)

[1.7. Contribution 14](#_Toc101550473)

[2. Literature Review 15](#_Toc101550474)

[2.1. Overview 15](#_Toc101550475)

[2.2. Diabetic Retinopathy Detection 15](#_Toc101550476)

[2.2.1. Binary Classification 15](#_Toc101550477)

[2.2.2. Multi-Level Classification 16](#_Toc101550478)

[2.2.3. Lesion-Based Classification 17](#_Toc101550479)

[2.2.4. Vessels-Based Classification 19](#_Toc101550480)

[2.3. Literature Summary 20](#_Toc101550481)

[3. Research Methodology 24](#_Toc101550482)

[3.1. Overview 24](#_Toc101550483)

[3.2. Research Framework 24](#_Toc101550484)

[3.2.1. Data Collection & Preparation 25](#_Toc101550485)

[3.2.2. Image Pre-processing 26](#_Toc101550486)

[3.2.3. Modelling 28](#_Toc101550487)

[3.2.4. Model Evaluation & Testing 31](#_Toc101550488)

[3.2.5. Conclusion 32](#_Toc101550489)

[3.3. Data Collection 32](#_Toc101550490)

[3.4. Evaluation Metrics 34](#_Toc101550491)

[3.4.1. Accuracy 34](#_Toc101550492)

[3.4.2. Recall 34](#_Toc101550493)

[3.4.3. Precision 35](#_Toc101550494)

[3.4.4. F1-Score 35](#_Toc101550495)

[3.4.5. Confusion Matrix 35](#_Toc101550496)

[4. Results 36](#_Toc101550497)

[5. Discussion 37](#_Toc101550498)

[6. Conclusion 38](#_Toc101550499)

[7. References 39](#_Toc101550500)

# List of Figures

[Figure 1 Normal Retina Vs Diabetic Retinopathy (Image Credits - kcretina, 2017) 12](#_Toc102131361)

[Figure 2 Machine Learning Architecture 13](#_Toc102131362)

[Figure 3 Research Flow Diagram 26](#_Toc102131363)

[Figure 4 Train and test Datasets 27](#_Toc102131364)

[Figure 5 Original and processed image of 20\_left.jpeg 28](#_Toc102131365)

[Figure 6 Inception Module - naive version - Szegedy et al (2015) 29](#_Toc102131366)

[Figure 7 Inception module with dimension reductions - Szegedy et al (2015) 29](#_Toc102131367)

[Figure 8 Inception V1 architecture 30](#_Toc102131368)

[Figure 9 Inception V3 architecture 30](#_Toc102131369)

[Figure 10 Diabetic Retinopathy Detection System - User Interface 32](#_Toc102131370)

[Figure 11 Dataset - Images in DR Stages 33](#_Toc102131371)

[Figure 12 Initial Model - Accuracy and Loss graphs 36](#_Toc102131372)

[Figure 13 Initial Model - test dataset - confusion matrix 36](#_Toc102131373)

[Figure 14 Initial model - APTOS dataset - confusion matrix 37](#_Toc102131374)

[Figure 15 Accuracy and Loss graph of Inception-V3 model 37](#_Toc102131375)

[Figure 16 Fine-tuned model - test dataset - confusion matrix 38](#_Toc102131376)

[Figure 17 Fine-tuned Model - APTOS dataset - confusion matrix 38](#_Toc102131377)

# List of Tables

[Table 1 - Abbreviations 9](#_Toc102127033)

[Table 2 - Literature Summary 23](#_Toc102127034)

[Table 3 - Before Oversampling - Total number of images - Each Class 26](#_Toc102127035)

[Table 4 - After Oversampling - Total number of images - Each Class 26](#_Toc102127036)

[Table 5 - List of publicly available datasets 32](#_Toc102127037)

[Table 6 - Inception V3 and fine-tuned Inception-V3 Results 38](#_Toc102127038)

# List of Equations

[Equation 1 - Accuracy 33](#_Toc102125331)

[Equation 2 - Recall 33](#_Toc102125332)

[Equation 3 - Precision 34](#_Toc102125333)

[Equation 4 - F1-Score 34](#_Toc102125334)

# List of Abbreviations

|  |  |
| --- | --- |
| SHORTENED FORM | FULL FORM |
| DR | Diabetic Retinopathy |
| DME | Diabetic Macular Edema |
| MA | Microaneurysms |
| HA | Haemorrhages |
| GDM | Gestational Diabetes Mellitus |
| WHO | World Health Organization |
| NHS | National Health Service |
| VEGF | Vascular Endothelial Growth Factor |
| AI | Artificial Intelligence |
| SVM | Support Vector Machine |
| kNN | k-Nearest Neighbors |
| CNN | Convolutional Neural Networks |
| AM-FM | Amplitude Modulation – Frequency Modulation |
| VGG | Visual Geometry Group |
| DCCMED | Densely Connected and Concatenated Multi Encoder-Decoder |
| IDRiD | Indian Diabetic Retinopathy Image Dataset |
| DIARETDB | Diabetic Retinopathy Database |
| DRIVE | Digital Retinal Images for Vessel Extraction |
| HRF | High-Resolution Fundus |
| ROC | Retinopathy Online Challenge |
| STARE | Structured Analysis of the Retina |
| MESSIDOR | Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology |
| HSV | Hue Saturation Value |
| RGB | Red Green Blue |
| BGR | Blue Green Red |
| CLAHE | Contrast-Limited Adaptive Equalization |
| ReLU | Rectified Linear Activation Unit |
| FOV | Field of View |
| ROI | Region of Interest |
| AUC | Area Under the Curve |
| ROC | Receiver Operating Characteristic |
| APTOS | Asia Pacific Tele-Ophthalmology Society |
| ADAM | Adaptive Moment Estimation |
| ILSVRC | ImageNet Large Scale Visual Recognition Challenge |

Table - Abbreviations

# Abstract

Diabetic Retinopathy is an eye disease that can cause blurry vision and sometimes blindness in patients having diabetes for more than 20 years. Diabetic Retinopathy is classified into 4 stages, i.e., Mild, Moderate, Severe Non-proliferative, and Proliferative. This eye condition doesn’t have any symptoms at an early stage but develops a few symptoms such as floating spots or fluctuating vision in later stages. To avoid the time and effort consuming manual diagnosis process, most of the researchers developed custom deep learning models to detect DR in fundus images. In machine learning, transfer learning is an approach where the knowledge of one model is implemented in another model to solve a specific problem. Using the transfer learning approach for image classification requires less training and effort in building model architecture. So, in this paper, a pre-trained model known as Inception-V3 is used to detect retinal fundus images. To train and test the Inception-V3 model, a Kaggle dataset is used. The images in the dataset are pre-processed using a few image pre-processing techniques that are later explained in the report. The proposed method achieved an accuracy of 90.39% on the test dataset and 54.22% on ‘APTOS 2019 Blindness Detection’ dataset.

# Introduction

## Overview

In the human system, Insulin is a hormone produced by the pancreas and released into the bloodstream. The insulin produced will helps the system to digest the food and help the body to manage glucose (sugar) levels. If a sufficient amount of insulin is not produced in the body, the blood sugar levels in the body become high leading to diabetes. Diabetes is a serious health condition that leads to serious damage to different organs in the human system including the heart, kidney, eyes, nerves, and feet. There are three types of diabetes – type 1, type 2, and gestational. Type 1 diabetes occurs due to little or no production of insulin in children whereas Type 2 diabetes occurs more in adults if the produced insulin is not used by the body. Gestational diabetes is also known as the GDM occurs due to high glucose levels during the pregnancy and disappears after the pregnancy. As per the World Health Organization (WHO), in 2014, around 422 million people in the world are suffering from diabetes and it became the 9th leading cause of death in 2019 with an estimation of 1.5 million deaths (World Health Organization, 2021).

Diabetic Retinopathy, also known as DR, is a complication of diabetes that occurs when diabetes damages the tiny blood vessels in the patient’s eye. This disease may cause blindness in type 1 and type 2 diabetes patients if not treated at an early stage and affects 80% of the patients having diabetes for more than 20 years. The retina at the back of the eye is the light-sensitive layer that is responsible for the vision and needs continuous blood supply through tiny blood vessels. The effect of Diabetic Retinopathy on blood vessels is categorized into 3 stages – background, pre-proliferative and proliferative retinopathy (NHS Choices, 2019).

During the background retinopathy stage, tiny red spots surrounded by the yellow rings known as ‘Microaneurysms’ sometimes known as ‘MAs’ are formed due to the leakage of blood from the vascular. This is common in people suffering from diabetes. At this stage, the patient’s vision is not affected and can prevent blindness in the future by taking precautions suggested by doctors. The second stage is the pre-proliferative stage also referred to as ‘severe non-proliferative or ‘none-proliferative’. In this stage, more severe changes can be seen in the retina including venous bleeding, hard exudates, multiple cotton wool spots – retinal damage areas and blot haemorrhages – blood flecks, and new blood vessels arise or abnormal blood vessels on the optic disc (Leigh, 2022). The final stage of Diabetic Retinopathy is known as the ‘proliferative stage’. In this stage, a new blood vessel known as ‘Neovascularization’ is formed on the surface of the retina. These new blood vessels can lead to retinal detachment causing blindness in the patients. This stage is irreversible, and patients who are experiencing this may lose their vision completely.

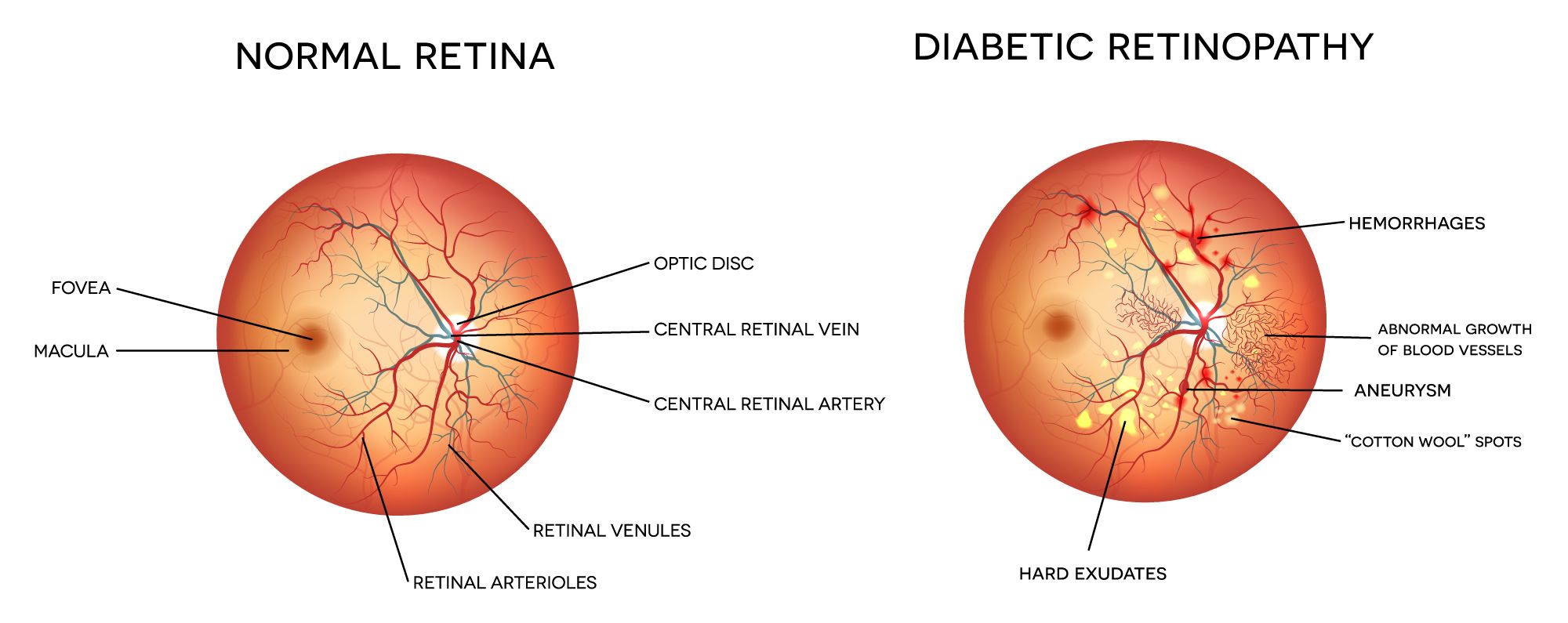


Figure 1 Normal Retina Vs Diabetic Retinopathy (Image Credits - kcretina, 2017)

Currently, there is no cure for diabetic retinopathy but it can be treated to prevent or delay vision loss in the future. Below is the list of treatments available for diabetic retinopathy (Diabetic retinopathy - Diagnosis and treatment - Mayo Clinic, 2022).

1. **Eye Injection –** Toreduce the formation of fluids and new blood vessels, VEGF inhibitors are injected into the vitreous of the eye.
2. **Focal Laser Treatment –** To reduce the leakage from the abnormal blood vessels.
3. **Pan Retinal Photocoagulation –** To shrink the abnormal blood vessels in the eye.
4. **Vitrectomy –** To remove blood and scar tissues from the middle of the eye.

The treatments explained above are costly and available only in a few places in under-developed countries. To reduce costly treatments and avoid vision loss in diabetic retinopathy patients, Diabetic Eye Screening is introduced. During the screening, Ophthalmologists examine the patient's eye and search for diabetic retinopathy signs such as MAs, HAs, soft and hard exudates, and new blood vessels. This type of screening is a time- and effort-consuming process.

Given the current scenario, the need for early detection of diabetic retinopathy has encouraged software engineers to develop image analysis software based on machine learning techniques. These machine learning models detect signs of diabetic retinopathy in eye photographs and help ophthalmologists to avoid the time- and effort-consuming process.

## Problem Background

Artificial Intelligence, or 'AI,' is an area of computer science that allows a machine to mimic human behavior, whereas Machine Learning and Deep Learning is a subset of AI that allows a machine to learn from data without the need to be programmed explicitly. So, with the help of machine learning, many researchers have already developed custom models to detect the signs of diabetic retinopathy in the retina fundus images. These models include SVM (Support Vector Machine), Decision Tree, k-Nearest Neighbours also known as kNN, Naïve Bayes, and more. The below figure shows the typical architecture of most machine learning models.

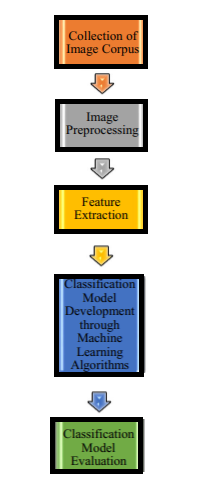


Figure 2 Machine Learning Architecture

Machine learning algorithms will have low accuracy without feature extraction when compared to Convolutional Neural Networks. When implementing machine learning algorithms to classify DR stages, a feature extraction phase is necessary since it extracts important features based on the color, shape, structure, and intensity of MA's, HA's, soft and hard Exudates, which improves the algorithms in achieving good results. Convolutional Neural Networks, on the other hand, is a deep learning algorithm that extracts low-level and high-level features from images automatically.

Using deep learning algorithms will eliminate the need for feature extraction. But it takes a huge amount of time and retinal fundus images to train Convolutional Neural Networks from scratch. To speed up training and improve the performance of deep learning algorithms, the transfer learning approach is implemented in the proposed system. This approach will help the system to train the model in less time and by using a small amount of data that is publicly available on different websites.

## Research Aim

According to WHO statistics, around 422 million people worldwide suffer from diabetes in the year 2014, and approximately 33 million individuals with diabetes are treated with diabetic retinopathy in 2010, with one-third of them having vision-loss DR, i.e., the last stage of Diabetic Retinopathy. As per Congdon, Zheng, and He (2012), by the year 2030, the DR cases are expected to reach 191 million.

The most effective way of reducing the number of DR cases is to get diagnosed early with the help of the screening process. But this method is quite expensive and inaccessible in some under-developed countries. To overcome this issue, the project aims to implement a cost-effective solution for detecting Diabetic Retinopathy in patients at an early stage using electronic devices, such as Mobile phones and laptops.

## Research Objectives

* Implementing a user interface that enables users to upload the retinal scan images.
* Detecting Diabetic Retinopathy in uploaded retinal fundus images using Inception-V3 model.
* Classifying the stage of Diabetic Retinopathy in fundus images, i.e., Normal, Mild, Moderate, Severe, and Proliferative representing the severity of the disease.

## Research Scope

Most of the research papers concentrated on the binary classification of diabetic retinopathy, i.e., classifying the image into normal fundus image or diabetic retinopathy fundus image. The model implemented in the proposed system classifies the image into stages of DR, i.e., Normal, Mild non-proliferative, Moderate non-proliferative, Severe non-proliferative, and proliferative diabetic retinopathy.

The proposed system concentrates only on the image pre-processing and pre-trained model to achieve good results and doesn’t perform any feature extraction related to other parts of the eye like optic-disc removal, blood vessel removal, or other classifications like lesion-based classification, vessel-based classification.

## Methodology

The methodology implemented in this report consists of five different stages, i.e., Data Collection and Preparation, Image pre-processing, training model, model evaluation & testing, and conclusion. In the first step, retinal fundus data is collected from the dataset known as the ‘Diabetic Retinopathy Detection’ dataset that is publicly available on the Kaggle website. This dataset contains a total of 35,126 fundus images. Once the required data is collected, a few image pre-processing techniques are performed to improve the quality of the images.

Once all images are processed, these images are divided into train and test datasets. The training dataset is fed as input to the well-known pre-trained model known as the ‘Inception-V3’ model. Based on the results obtained by the model, it is later fine-tuned to achieve better results than the previous version. Later, the pre-trained model is evaluated using the test dataset and also tested on a different dataset known as ‘APTOS 2019 Blindness Detection’. Once the model is trained and tested, it is used in the Python Django project to classify the retinal images uploaded by the users.

## Contribution

The proposed model obtained an accuracy of 90.39%, Precision of 0.9025, Recall of 0.90386, and F1-Score of 0.9029 on the test dataset. To evaluate the performance of the proposed model on a different dataset, a widely used Kaggle dataset known as the ‘APTOS 2019 Blindness Detection’ dataset is used. The proposed model achieved an accuracy of 54.22%, Precision of 0.4686, Recall of 0.53641, and F1-Score of 0.44126 on the APTOS 2019 Blindness Detection dataset.

# Literature Review

## Overview

Diabetes Mellitus, commonly known as Diabetes is a very common disease all over the world. This disease occurs due to an increase in sugar or glucose levels in the blood and can be seen in all age groups. Over a period of time, this disease can cause many health conditions related to the eyes and other parts of the human system. Diabetes can lead to blindness in patients aged between 20 and 75. This uncontrolled disease that affects the patient’s eye is known as Diabetic Retinopathy or DR in simple terms.

Diabetic Retinopathy is a severe eye condition that can lead to blurry vision or complete vision loss if not treated early. Diabetic Retinopathy is classified into four stages, i.e., Mild Non-proliferative, Moderate Non-proliferative, Severe Non-proliferative and Proliferate Diabetic Retinopathy. In the early stage of DR, i.e., Mild Non-proliferative, tiny bulges, also known as Microaneurysms or MAs, are formed at the end of blood vessels. These tiny bulges can cause blood leakage in the later stages. Patients who are suffering from Mild DR don’t have any vision problems and can control the growth of the DR by consulting a doctor regularly. During the moderate stage, the retina of the eye is affected by a severe condition known as Diabetic Macular Edema. This condition can cause physical changes in the retina due to swelling of blood vessels. In the Severe stage, due to the blockage of existing blood vessels, new blood vessels are formed in the retina. In the final stage of DR, leakage of blood from newly formed blood vessels triggers issues related to vision like blurred vision or even complete loss of vision (Healthline, 2021).

To prevent vision loss, patients with diabetic retinopathy can follow precautions suggested by the doctors or with the help of eye treatments such as Eye injection, Focal Laser Treatment, Pan Retinal Photocoagulation, and Vitrectomy.

With the help of Artificial Intelligence, diabetic retinopathy can be detected at an early stage. This approach eliminates the expensive- and time-consuming process. Many researchers already developed a few machine learning and deep learning techniques to detect diabetic retinopathy at an early stage. This section explains a few techniques suggested by the authors.

## Diabetic Retinopathy Detection

In this section, the detection of diabetic retinopathy is categorized into four different classifications, i.e., Binary Classification, Multi-Level Classification, Lesion-Based Classification, and Vessel-Based classification. Each classification is explained using a few different approaches.

## Binary Classification

The binary Classification approach is the simplest in machine learning. In this method, the dataset or any kind of data is categorized into two values, i.e., 0 or 1 or in simple terms, True or False. The images in the diabetic retinopathy datasets are classified into ‘DR’ and ‘No DR’ images. Before training and testing different types of CNN architectures, these images are preprocessed using a few image pre-processing techniques such as Gaussian filter, and Image Normalization.

The study suggested by Xu, Feng, and Mi (2017) automatically classifies retinal fundus images of the Kaggle dataset into normal or DR images. The dataset used in this research contains only 1000 retina fundus images. So, to increase the performance of the proposed algorithm, the author implemented data augmentation techniques such as translation, rescaling, stretching, flipping, and rotation to the labeled dataset. The CNN architecture implemented by the author consists of 8 Convolutional Layers, 4 Max-Pooling Layers along with 2 Fully Connected Layers. To classify the DR images, the author used the SoftMax function in the last layer. The proposed method obtained an accuracy of 94.5%.

The study suggested by Esfahani, Ghaderi, and Kafiyeh (2018) automatically classifies retinal fundus images of the Kaggle dataset into healthy or DR images using a well-known pre-trained model known as ‘ResNet34’. The Kaggle dataset used consists of 36,000 healthy and DR images collected from healthy and diabetes people. Before passing images to the CNN, to eliminate irrelevant information and defects in the image, Gaussian Blur filter, weighted addition along with normalization technique are applied to the 512 X 512 images. The CNN architecture proposed by the author consists of 22 Convolutional Layers, 1 Max-Pooling Layer, and one Average Pooling Layer. The approach suggested by the author obtained an accuracy of 85% and sensitivity of 86%.

The study suggested by Jiang et al. (2019) automatically classifies the fundus images into normal or DR images using three different pre-trained models, i.e., ‘Inception V3’, ‘ResNet152’, and ‘Inception-ResNet-V2’. The dataset used in this research is private and consists of around 30,000 images. Before feeding images as input to the model, the original images are cropped and resized to 520X520 pixels. Once the images are resized, an effective image enhancement approach known as the ‘Ben Graham processing’ technique is applied to the images to obtain higher results. In this approach, the author implemented an algorithm known as the ‘Adaboost’ algorithm. This algorithm combines several weak classifiers to form a strong classifier. After training the pre-trained models, this proposed approach obtained an accuracy of 88.21% and an AUC of 0.946.

## Multi-Level Classification

The multi-level classification is a different type of approach where the DR images are classified into five different classes. These classes are labeled as 0, 1, 2, 3, 4, or in simple terms, these classes are labeled as Normal, Mild, Moderate, Severe, and Proliferate DR. Each class determines the stage of Diabetic Retinopathy. Many authors already suggested different approaches to detect the stage of Diabetic Retinopathy disease in fundus images.

The approach implemented by Pratt et al. (2016) automatically classifies fundus images into five DR stages, i.e., Normal, Mild, Moderate, Severe and Proliferate using customized CNN. The dataset used for training and testing contains over 80,000 images of all five DR stages. Before passing these images as input to the CNN, to improve the discriminative performance of CNN, the Colour Normalization technique is applied to images. Once all images are processed, to reduce computation time and power while training CNN, the images are resized to 512 X 512 pixels. The proposed CNN architecture consists of 10 Convolutional Layers, 8 Max Polling Layers, and 3 Fully Connected Layers. And to classify the DR images, the SoftMax activation function is used as a classifier in the last layer. During the training process, L2 regularization and Dropout techniques are implemented to reduce the overfitting. The approach suggested by the author obtained an accuracy of 75%, Specificity of 95%, and Sensitivity of 30%.

The study suggested by Wang et al. (2018) automatically classifies fundus images into five DR stages using a group of deep Convolutional Neural Network models, i.e., VGG16, AlexNet, and InceptionNet V3 models. Due to a lot of noise in images obtained from the Kaggle dataset, the author gathered a total of 166 high-quality images out of 35,000 images. In these 166 images, there are a total of 31 healthy images, 30 Mild images, 50 Moderate, 31 Severe, and 24 Proliferative DR images. Using a limited number of DR images, the author trained three pre-trained models to produce models using the cross-validation process. This approach obtained an average accuracy of 50.03% in VGG16, 63.23% in InceptionNet V3, and 37.43% in AlexNet.

The study suggested by Mobeen-ur-Rehman et al. (2019) automatically detects five DR stages in retinal fundus images using three different pre-trained models, i.e., AlexNet, VGG-16, and SqueezeNet, and also a customized 5 layered CNN model. To improve efficiency, and accuracy and to reduce execution time, these 1200 images from the Kaggle dataset are cropped and resized to 244 X 244 pixels. To provide the required contrast between the illuminated retina and other objects like blood vessels and nerve fibers, Histogram equalization is applied on resized images. The customized CNN architecture consists of 2 Convolutional Layers, 2 Max Pooling Layers, and 3 Fully Connected Neural Layers with each layer having its specifications. The 2 Convolutional Layers are responsible to avoid the overfitting problem whereas the 3 fully connected layers have 100, 50, and 10 neurons respectively. The custom CNN model obtained an accuracy of 98.15%, a sensitivity of 98.94%, and a specificity of 97.87%.

The study suggested by Harangi et al. (2019) automatically detects five DR stages in retinal fundus images using modified AlexNet – a pre-trained model. In this suggested approach, the author combined extracted 68 conventional features along with self-extracted CNN-based features into a single framework to achieve good classification performance. The conventional features are extracted using AM-FM Based Image level extraction and Lesion Specific Feature Extraction. The modified AlexNet is trained and tested using Kaggle and IDRiD datasets respectively. The proposed approach obtained an accuracy of 90.07%.

## Lesion-Based Classification

This section explains the approaches suggested by different authors to classify different types of lesions. From the early stage to the proliferative stage, several types of lesions are formed in the retina of the eye. At the early stage of DR, the first type of lesions known as Microaneurysms or MAs are formed. These lesions are tiny circular with distinct margins. At the second stage of DR, i.e., the Moderate stage, different types of lesions including MAs will appear. These new lesions are known as Retinal Haemorrhages, Hard Exudates, and Cotton wool spots also known as Soft Exudates. Each lesion has a different shape, color, and structure when compared with other types of lesions. In the severe stage, the count of these lesions will increase slowly. At the last stage of Diabetic Retinopathy, i.e., Proliferate DR stage, new vessels aka ‘Neovascularization’ are the first type of lesions that will start to grow.

Yan, Gong, and Liu (2019) propose a method for detecting microaneurysms in retinal fundus images by integrating improved pre-trained LeNet architecture and hand-crafted features. To train and evaluate the model, this approach makes use of the public dataset DIARETDB1. Microaneurysms can be seen in the green channel of the image. So, in the first step of the pre-processing stage, the green channel of the images is extracted. Later, the region of interest (ROI) is extracted by applying the threshold segmentation. Once the ROI is extracted, the Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the processed image to improve the brightness of the hidden MAs. To reduce noise and to detect candidates in the processed images, the Gaussian filter and morphological method are used. The proposed version of LeNet architecture consists of 4 Convolutional Layers, 3 Max-Pooling Layers, and one Fully Connected Layer. Before feeding images to the improved LeNet architecture, the blood vessels are separated from the images using U-net CNN. Due to fewer data available, the Random Forest classifier is implemented for classification purposes. The suggested approach obtained a sensitivity of 48.71%.

The approach suggested by Sharath Kumar et al. (2013) implements a new method for pre-processing and eliminating false-positive when detecting the true exudates. In this approach, the author gathered a total of 158 fundus images of which 69 images are annotated by an ophthalmologist, and the remaining 89 retinal fundus images are taken from the DIARETDB1 database. Before converting the color images to Gray-scale images, Hue Saturation Value or HSV is applied to correct the brightness in all images. Later, with the help of Histogram analysis, the author detected the optic disc and exudates in the fundus images. According to the intensity distribution in the fundus images, the author chooses the third peak as the region of the exudate and applied simple binary thresholding to detect possible exudates. These possible exudate regions include true exudates and some regions of the optic disc as false exudates. So, to remove false-positive exudates, the author implemented a multi-channel histogram analysis method. This approach obtained a sensitivity of 88.45% and specificity of 95.5%.

The algorithm proposed by Franklin and Rajan (2014) automatically detects exudates in the retinal fundus images and helps ophthalmologists to diagnose faster. The images used in the pre-processing stage are gathered from the public dataset known as the ‘DIARETDB1’ dataset. In the image processing stage, to eliminate the issues related to the greyscale methods and to replace the luminosity layer, the author converted RGB colored images to Lab colored images. Before applying the mean filtering, the images are reverted to the original color space. To enhance the contrast in the fundus images, the Contrast Limited Adaptive Histogram Equalisation also known as CLAHE is applied. With the help of exudate characteristics, the author has chosen a total of 15 distinctive features from Luv color space to classify the regions as true positive and false positive regions. Later, the neural network trained for classification purposes consists three-layer perceptron having 15 input nodes, one hidden layer with 20 hidden neurons, and a single output node. The proposed method obtained an accuracy of 99.7% using a lesion-based evaluation criterion.

The approach proposed by Chudzik et al. (2018) uses the concept of batch normalization and Dice loss along with the fully convolutional neural network to detect microaneurysms in fundus images. The proposed method was evaluated using three publicly available datasets, i.e., E-Ophtha, ROC, and DIARETDB1. The pre-processing of images involves three stages, i.e., Pre-processing, Patch Extraction, and pixel-wise classification. In the pre-processing stage, the green channel of the images is extracted. Once the green panel is extracted, the mask is generated using Otsu thresholding. Before subtracting the Local Mean Colour, the processed images are resized and cropped based on Field-of-View (FOV). In the pixel-wise classification stage, all possible regions in the processed images are classified as 0 and 1, where 0 is considered as the non-MA and 1 is considered as the MA. In the patch extraction phase, using the sliding window approach, MA patches, as well as non-MA patches, are extracted. Before training CNN architecture, MA patches were extracted and data augmentation is performed to increase the variety in the training dataset. The CNN architecture proposed by the author includes 18 Convolutional Layers, Except the final classification layer with sigmoid function, each layer is followed by the Batch Normalization layer, three Max-pooling layers, and three simple up-sampling layers. The convolutional layers implemented in the architecture use 3x3 filters and ReLU activation functions. The proposed approach achieved a ROC of 0.355.

## Vessels-Based Classification

Vengalil et al. (2016) proposed a modified version of deeplab architecture for blood-vessel segmentation without any image pre-processing techniques. The proposed architecture of a convolutional neural network includes 8 convolutional layers and 3 max-pooling layers. The architecture is categorized into three stages, i.e., stage 1 and stage 2 contain 2 cascaded convolutional layers followed by the ReLU layers, and stage 3 contains 3 convolutional layers followed by a max-pooling layer with stride 1x1. The outputs of layers in stages 1 and 2 are passed to the max-pooling layer with strides 2x2 and 1x1 respectively. Finally, the input layer takes RGB images of size 512x512 whereas the output layer, i.e., the final convolutional layer outputs the segmented Gray-scale images. The proposed method is categorized into two phases, i.e., training and testing phases. In the training phase, the extracted 512x512 image patches are passed as input to the proposed CNN architecture. To evaluate the model, two publicly available datasets, i.e., DRIVE and HRF are used. This approach achieved ROC of 0.894 and 93.94% accuracy.

Budak et al. (2020) study introduced a custom CNN architecture known as the Densely Connected and Concatenated Multi Encoder-Decoder network and a patch-based learning strategy to extract blood vessels from the retinal fundus images. The proposed CNN architecture consists of three encoder-decoder blocks and 2 convolutional layers between them. Each encoder-decoder block consists of 8 convolutional layers with ReLU activation function, 8 Batch Normalization layers, 2 Max Pooling, and 2 Max Unpooling layers. To solve the issue related to the size of the dataset, the author implemented a patch-based data augmentation strategy by using techniques such as flipping and rotation. The proposed architecture obtained an accuracy of 0.9685 and an AUC of 0.9822 while training the CNN using the DRIVE dataset with a batch size of 128, and an accuracy of 0.9735 and an AUC of 0.9868 when training using STARE datasets with a batch size of 64.

## Literature Summary

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **Method Description** | **Pros** | **Cons** |
| Xu, Feng, and Mi (2017) | This study uses a binary classification approach to identify Diabetic Retinopathy in the fundus images. To increase the performance of the model, data augmentation techniques such as stretching, flipping, translation, and rotation are applied. | The proposed CNN achieved an accuracy of 91.5% without data augmentation and 94.5% accuracy with data augmentation. And the trained model can detect diabetic retinopathy in less than 1 second. | A limited number of retinal fundus images are used to train the model, i.e., only 1000 images from the Kaggle dataset are used. |
| Esfahani, Ghaderi and Kafiyeh, (2018) | The study uses a binary classification approach to categorize the images into two groups, i.e., DR images and healthy images. This study involves an image pre-processing step before the actual images are fed to the pre-trained model. | The proposed pre-trained model achieved an accuracy of 86% and sensitivity of 85% when tested on 2000 images. These results are better when compared with the Pratt et al. (2016) model which is tested on 5000 test images. | The suggested approach is not tested on different datasets that are publicly available. And also, the trained model can only categorize the images into two groups, i.e., DR images or normal images. |
| Jiang et al. (2019) | The study implements a binary classification approach using multiple well-trained deep learning models. Before training the models, the images are pre-processed and augmented. | The proposed integrated model achieved an accuracy of 88.21%, specificity of 90.85%, and sensitivity of 85.97%. And this experiment is done with the help of cloud servers. | The suggested approach involves a manual process to eliminate the images with low quality. And can able to detect only a few DR images. The proposed method is not tested on a different dataset. |
| Pratt et al. (2016) | The study implements a multi-level classification approach using custom CNN architecture to identify intricate features such as micro-aneurysms, exudate, and HAs in diabetic retinopathy. The classifications are classified as 0 – No DR, 1 – Mild, 2- Moderate, 3 – Severe, 4 – Proliferative DR. | The proposed method achieved an accuracy of 75%, and 95% of specificity when training the model. | The proposed model achieved only 30% of sensitivity and was not able to detect all stages of Diabetic Retinopathy. |
| Wang et al. (2018) | The study implements a multi-level classification approach using three different pre-trained models, i.e., AlexNet, VGG16, and InceptionNet V3 to detect the stages of diabetic retinopathy in the retinal fundus images. | The average accuracy of three pre-trained models, i.e., VGG16, AlexNet, and InceptionNet V3 is 50.03%, 37.43%, and 63.23%. | The proposed model was only trained and tested using a total of 166 high-quality images and was not able to detect all stages of Diabetic Retinopathy. |
| Mobeen-ur-Rehman et al. (2019) | The proposed study implements four CNN models where three models are well-known pre-trained models and one is a newly trained custom CNN model. Before passing images as input to the model, the images are gone through some image pre-processing techniques. | The proposed approach achieved an accuracy of 93.46%, 91.82%, 94.49%, and 98.15% using AlexNet, VGG16, SqueezeNet, and the 5-layered model respectively. | The proposed models were trained and tested only on MESSIDOR 2 which contains a total of 1200 colored fundus images. And not tested on different datasets like DIARETDB1 etc. |
| Harangi et al. (2019) | The proposed approach is a combination of Convolutional Neural Networks (CNN) and traditional hand-crafted features to classify the retinal fundus images into 5 stages of DR. The classification is done based on the severity level of Diabetic Macular Edema (DME) and Diabetic Retinopathy (DR). | The proposed approach utilized two different datasets, i.e., the Kaggle dataset (22,700 images) and the IDRiD dataset (516 images) for training and testing purposes, and achieved an accuracy of 90.07% for the 5-class DR challenge, and an accuracy of 96.85% for 3-class DME. | The proposed method is not able to detect lesions in the DR images. |
| Yan, Gong, and Liu (2019) | The proposed method implements lesion-based classification using features extracted by LeNet CNN architecture and hand-crafted features and the candidate regions are classified using a random forest classifier. | The approach implemented by the author shows an improvement when compared with other methods. | The proposed approach obtained a low accuracy for detecting lesions in the DR images. |
| Sharath Kumar et al. (2013) | The study detects the true positive exudates by extracting the optic disc of the retinal from the normalization fundus images. | The proposed model achieved a sensitivity of 88.45% and specificity of 95.5%. | This experiment used a total of 158 fundus images where 69 images are obtained from RIO and the remaining 89 images are collected from the DIARETDB1 dataset and no other datasets are used to evaluate the performance of the model. Moreover, the model is not able to detect 10% of exudates. |
| Franklin and Rajan (2014) | The study implements lesion-based classification by detecting the exudates in the fundus images by using color normalization and a three-layer feed-forward neural network. | The proposed algorithm perfectly-identified exudates in retinal fundus images which consists of other lesions such as MAs and Has. | The proposed algorithm is not able to classify the images into stages of DR. |
| Chudzik et al. (2018) | This study proposed lesion-based classification by implementing a different approach that consists of three stages, i.e., pre-processing, patch-extraction, and pixel-wise classification. | This approach can detect MAs in the Diabetic Retinopathy images and requires only three stages when compared with studies that use 5 stages to detect MAs. | The proposed method is not able to detect all MAs in the Diabetic Retinopathy and takes around 220 seconds to process a single retinal fundus image. |
| Vengalil et al. (2016) | The study proposed a simple approach to segment retina blood vessels without any image pre-processing stage. The proposed method is applied to two publicly available datasets, i.e., DRIVE and HRF. | The model achieved a ROC of 0.894 and an accuracy of 93.94% on the test set. | The proposed method is not able to extract finer blood vessels in the retinal fundus images. And not able to detect any DR lesions. |
| Budak et al. (2020) | The study extracts blood vessels from the retinal fundus images using a novel CNN model known as the Densely Connected and Concatenated Multi Encoder-Decoder CNN model (DCCMED). This approach extracts blood vessels without any pre-processing steps. | The proposed approach implemented patch-based data augmentation technique to overcome the issues related to the large datasets. This approach achieved high AUC scores for 2 different publicly available datasets, i.e., DRIVE and STARE when compared with existing methods. | This study successfully extracted all blood vessels in the retinal fundus images. However, this method is not able to detect DR stages of DR lesions such as MAs, HAs, or exudates in the retinal fundus images. |

Table - Literature Summary

# Research Methodology

## Overview

Diabetic Retinopathy is a serious eye disease that affects the eyes of the patients suffering from diabetes for around 20 years. Patients suffering from Diabetic Retinopathy have different forms of lesions such as MAs, HAs, Exudates, Cotton wool spots, and abnormal blood vessels, and the stages of Diabetic Retinopathy are classified as Normal, Mild, Moderate, Severe, and Proliferative DR. The signs such as MAs and new blood vessels started to appear in the early and final stage of DR respectively.

Ophthalmologists examine the retinal fundus images manually to detect the symptoms of DR. This process takes time and effort for Ophthalmologists when detecting the tiny MAs. The computer-aided diagnostic technique has achieved a breakthrough in the medical industry to overcome this process. The sub-domain of computer vision known as image classification deals with the classification of objects in images using pre-defined labels. With the help of software that uses image pre-processing techniques and deep learning models, Ophthalmologists can easily detect the tiny symptoms of DR as well as DR stage with no effort.

Machine Learning is a subset of Artificial Intelligence that can imitate the behavior of humans with the help of machines. There are many machine learning techniques available to develop a model that can detect signs of DR. But this approach takes a huge amount of time to properly train the models. To overcome this issue, the transfer learning approach is discovered. This approach eliminates the time to train the models. The main aim of the transfer learning approach is to make use of pre-trained models that are developed to solve a particular problem and can also be used to solve similar tasks by changing the weights or by freezing the last fully connected layer.

As explained in the literature review section, many researchers already implemented custom deep learning models and pre-trained models to detect diabetic retinopathy in the retinal fundus images. These approaches are categorized into a few classifications based on their functionality.

In the proposed system, a transfer learning approach is implemented by using a well-known pre-trained model known as ‘Inception-V3’. The retinal fundus images used to train the ‘Inception-V3’ model are collected from the large publicly available dataset known as the ‘Diabetic Retinopathy Detection’. Before feeding the fundus images as input to the ‘Inception-V3’ model, few image pre-processing and data augmentation techniques are performed to eliminate the noise and increase the size of the dataset respectively. Once the model is trained, it is later used to detect the DR stage in the image uploaded by the user.

## Research Framework

This section explains the steps carried out while developing the proposed system. There are a total of 5 steps involved in this process, i.e., Data Collection & Preparation, Image Pre-processing, Modelling, Model Evaluation & Testing, and Conclusion. Each step in this process is later explained in the below sub-sections. Below is the diagram describing the sequential flow of these steps.

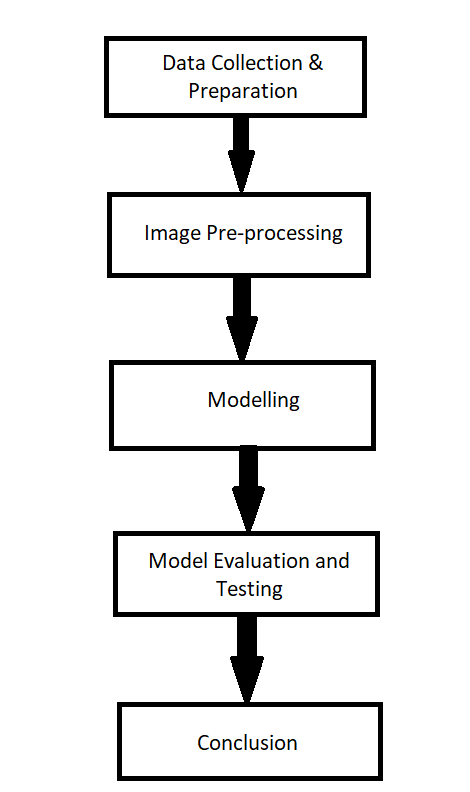


Figure 3 Research Flow Diagram

## Data Collection & Preparation

In the many articles, most of the authors have used the diabetic retinopathy datasets that are publicly available and few other researchers gathered data directly from recognized eye hospitals. Currently, there are many fundus image datasets available online for public use. A few of them are ROC (Retinopathy Online Challenge), E-Ophtha, Kaggle, DRIVE, STARE, DIARETDB0, DIARETDB1, and more. Most of the authors have developed their systems using a well-known dataset, i.e., Kaggle. Kaggle is a publicly available dataset that contains images provided by EyePACS. In the proposed system, the Kaggle dataset known as ‘Diabetic Retinopathy (resized)’ is used to train and test the ‘Inception-V3’ pre-trained model. This dataset is the resized version of the original dataset available in Kaggle known as the ‘Diabetic Retinopathy Detection’ dataset. In the selected dataset, there are a total of 35,126 retinal scan images and can be used with many readily available pre-trained models. The below table shows the number of images available in each class.

|  |  |
| --- | --- |
| Class (DR Stage) | Total Number of Images |
| Normal | 25810 |
| Mild | 2443 |
| Moderate | 5292 |
| Severe | 873 |
| Proliferative | 708 |
| Total | **35126** |

Table - Before Oversampling - Total number of images - each class

From the above-mentioned table, it is clear that the number of samples for the normal class is more when compared to other classes and the number of sample images is quite low for the proliferative class. To increase the samples of minority classes, a random oversampling technique is implemented. The purpose of random oversampling is to create duplicate samples of minority classes. The below table shows the list of samples available in each class after oversampling.

|  |  |
| --- | --- |
| Class (DR Stage) | Total Number of Images |
| Normal | 25810 |
| Mild | 8603 |
| Moderate | 8603 |
| Severe | 8603 |
| Proliferative | 8603 |
| Total | **60222** |

Table - After Oversampling - total number of images - Each Class

After oversampling the original dataset, 80% of the oversampled dataset is used as the training dataset whereas the remaining 20% of the oversampled dataset is used as the testing dataset. Within the training dataset, 10% is considered as the validation dataset which is later used to calculate the performance of the model. Figure 4 is the graphical representation of the number of samples considered as the training and testing datasets.



Figure Train and test Datasets

## Image Pre-processing

The main purpose of image pre-processing is to improve the image quality before feeding it to the pre-trained models, resulting in better DR stage classification. The Kaggle dataset used in this approach contains images of different sizes ranging from 289x433 pixels to 1024x1024 pixels. And also, these images are captured with different cameras under different light conditions. These changes will affect the performance of the CNN model. To overcome this issue, a library of Python known as ‘OpenCV-Python’ is used for image processing. Following image pre-processing steps are implemented before passing the images to the CNN model.

1. First, the original fundus images are converted to Gray-scale images using the built-in function, i.e., cv2.cvtCOLOR.
2. The black background of the image does not contain any useful information, so the background is cropped out from the image.
3. To maintain the constant shape of the retina in all images, the mask is applied to the cropped image using built-in functions such as cv2.circle and cv2.bitwise\_and.
4. Applying Ben Graham method on cropped images.
   1. To blur the images, the Gaussian Blur filter is applied to the images with a Gaussian Kernel Size value of (0,0) and SigmaX value of 10.
   2. To blend original image with the blurred image, the following values are used.
      * src1 = cropped\_image
      * alpha = 4
      * src2 = blurred\_image
      * beta = -4
      * gamma = 128
5. In the final step of image pre-processing, the processed image is resized to 512x512 pixels.

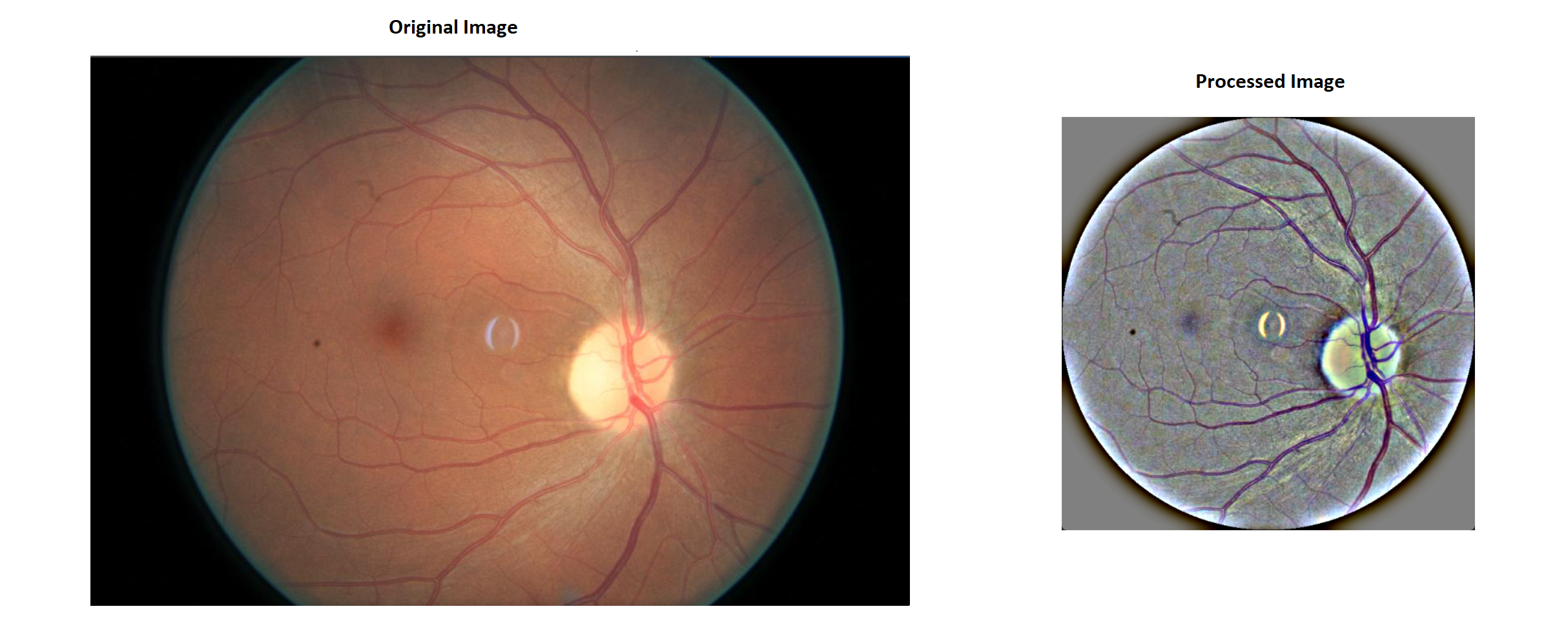


Figure 5 Original and processed image of 20\_left.jpeg

Figure 4 shows the original version and processed version of the image with the name ‘20\_left.jpeg’. The size of the original image is 683x1024 pixels and the size of resized image is 512x512 pixels.

## Modelling

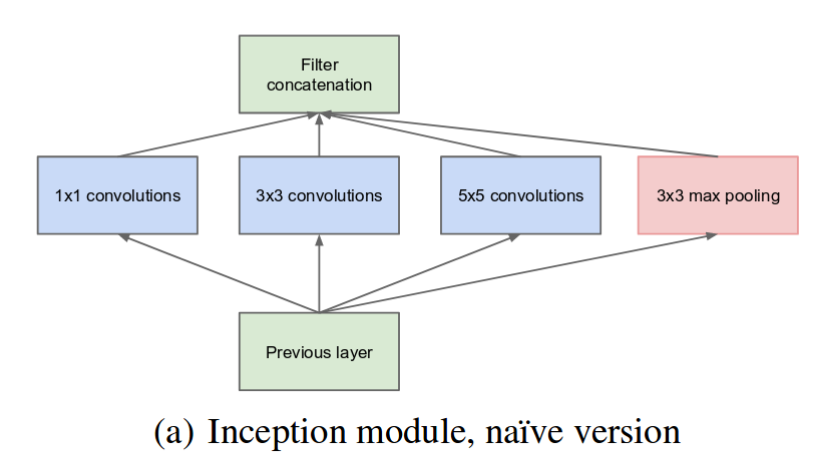


Figure Inception Module - naive version - Szegedy et al (2015)

The first version of Inception was introduced by Szegedy et al (2015). The main aim of the inception model is to use multiple types of filter sizes instead of sticking to the fixed filter size to solve variations in the important information locations in the images. Figure 5 shows the naïve version of Inception. This version consists of 3 different filter sizes, i.e., 1x1, 3x3, 5x5, and one 3x3 max-pooling layer. To eliminate the expensive computation, the author performed dimension reduction by adding 1x1 convolutional layers after the max-pooling layer instead of 5x5 convolutional layers. Figure 6 shows the reduced dimension version of the Inception model.

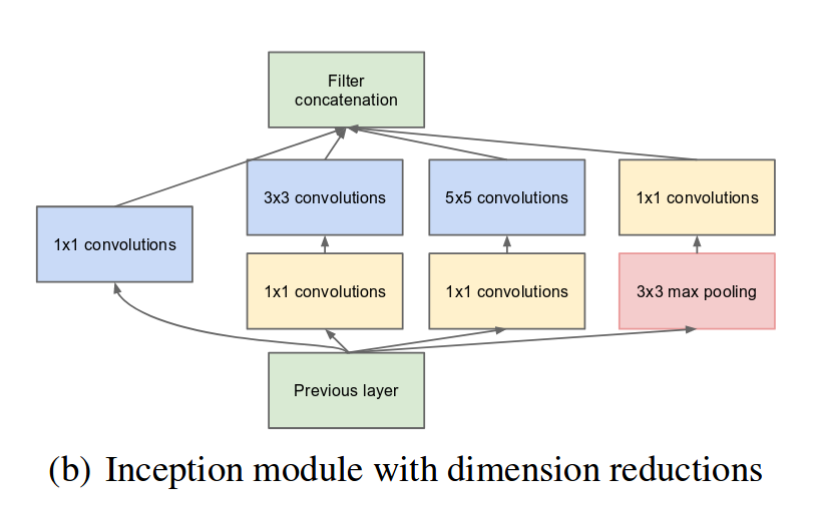


Figure Inception module with dimension reductions - Szegedy et al (2015)

By using the above inception module with dimension reductions, a new deep neural architecture known as ‘Inception V1’ was introduced. The proposed architecture consists of a total of 27 layers. Figure 7 shows the architecture of the Inception V1 model.

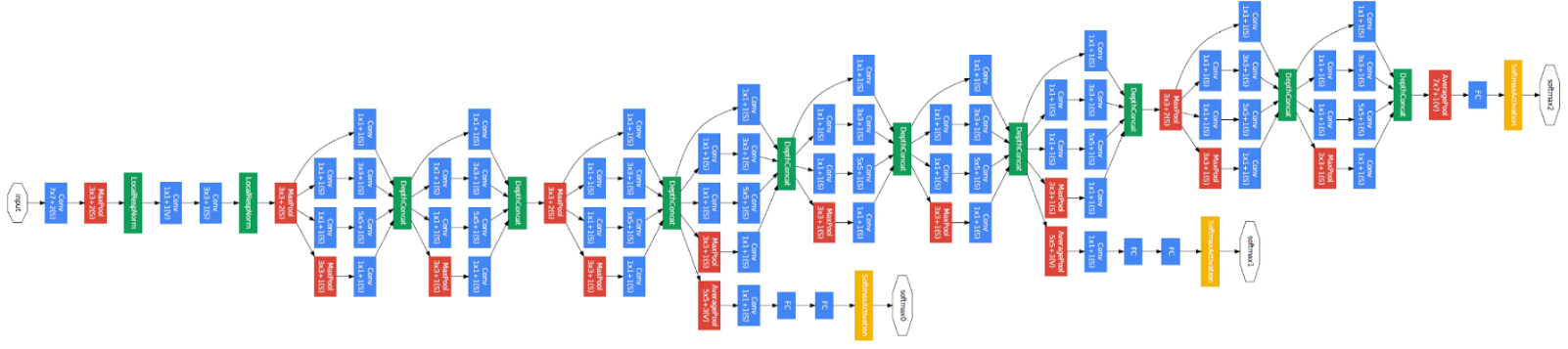


Figure Inception V1 architecture

Later, Szegedy et al. (2015) introduced an updated version of Inception by modifying previous versions. The updated version of inception is known as ‘Inception V3’. The main aim of the Inception V3 is to reduce the usage of computational power when compared with previous versions. To optimize the network, the author performed several techniques including factorized, smaller, and asymmetric convolutions, regularization, and more. The Inception V3 has less error rate when compared with other models and gained popularity in the ILSVRC 2012 classification challenge.

The Inception-V3 model is trained on millions of images from the ImageNet database and can classify thousands of objects from the images. Moreover, the Inception-V3 model is very good at image recognition and classification tasks and also very good at learning features from different sets of images. This model consists of a total of 42 layers, i.e., separate blocks of convolutional layers, average pooling layer, max-pooling layer, and fully connected layer. The default input size of the Inception V3 model is 299x299. Figure 8 shows the architecture of Inception V3.

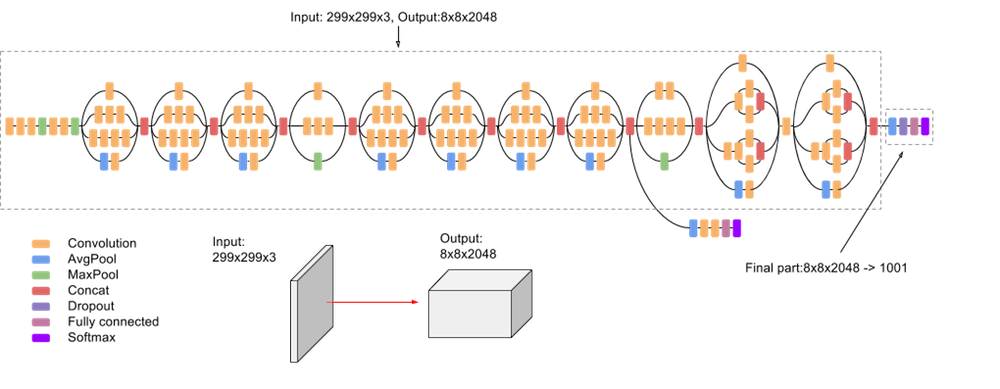


Figure 9 Inception V3 architecture

In the proposed method, the updated version of Inception, i.e., Inception V3 is implemented using a highly powerful and dynamic framework known as ‘Keras’. Keras has been developed by 'Francois Chollet,' a Google artificial intelligence researcher. This framework runs on top of other open-source machine libraries such as TensorFlow.

Initially, the Inception-V3 model is instantiated by loading the ‘imagenet’ weights, selecting input size as ‘(224, 224, 3)’ and ‘global average pooling’ as the optional pooling mode. Later, the Inception-V3 architecture is slightly modified by replacing the last fully connected layer (that is specific to the ImageNet competition) with 2 fully connected layers with 1024 nodes with the ‘ReLU’ activation function, and 1 dropout layer with a value of 0.5 to avoid overfitting and the classification or logistic layer with 5 classes (Normal, Mild, Moderate, Severe, and Proliferative) with SoftMax activation function. During the training of the Inception-V3 model, all other layers are frozen by setting the trainable value as False.

In the initial training, a widely used optimizer known as ‘ADAM (Adaptive Moment Estimation)’ with a default learning rate, i.e., 0.001 is used along with ‘categorical\_crossentropy’ as the loss function. And the initial training is performed with a batch size of 64, epochs of 50. Moreover, to avoid overfitting, early stopping is implemented with a patience value of 10.

Later, to improve the accuracy of the trained model, fine-tuning is performed by freezing only the first two inception blocks of Inception-V3, i.e., first 249 layers of Inception-V3 network, increasing the batch size to 80, and changing the learning rate of the optimizer from 0.001 to 0.0001. Figure 11 and Figure 12 show the accuracy and loss of fine-tuned model respectively.

## Model Evaluation & Testing

Before training the Inception-V3 model, the original dataset is divided into training, validation, and testing datasets by keeping the ratio of 70-10-20 respectively. Once the Inception-V3 model is trained on the training dataset, it is tested on the test dataset and also tested on the different datasets known as the ‘APTOS 2019 Blindness Detection’ Dataset. The proposed Inception-V3 model is evaluated using different evaluation metrics such as Confusion Matrix, Precision, Recall, and F-1 Score. These metrics help to understand the performance of the model.

## Conclusion

The trained Inception V3 model is saved using the ‘load\_model’ function imported from the ‘keras.models’ package. Later, the saved model is used to predict the severity of Diabetic Retinopathy in the retinal fundus image uploaded by the user. To achieve this, the Django web framework is used to design and develop a user interface for the Diabetic Retinopathy Detection – web application.

Django is one of the best Python web frameworks that allow developers to use Model-Template-Views (MTV) architecture for web applications. This framework enables fast and secure development. The proposed system consists of one models.py, views.py, model.h5, and home.html files. The home.html file in the project is responsible for the user interface where the user is allowed to upload the retinal fundus image. The ‘process\_image’ function in the views.py is responsible for retrieving and storing the image uploaded by the user whereas the ‘process\_image’ in the models.py file is responsible for processing the image and loading the trained inception model. Before predicting the severity of diabetic retinopathy in the uploaded fundus image, the image is pre-processed using the steps explained above and fed to the saved model by using the ‘predict’ function.

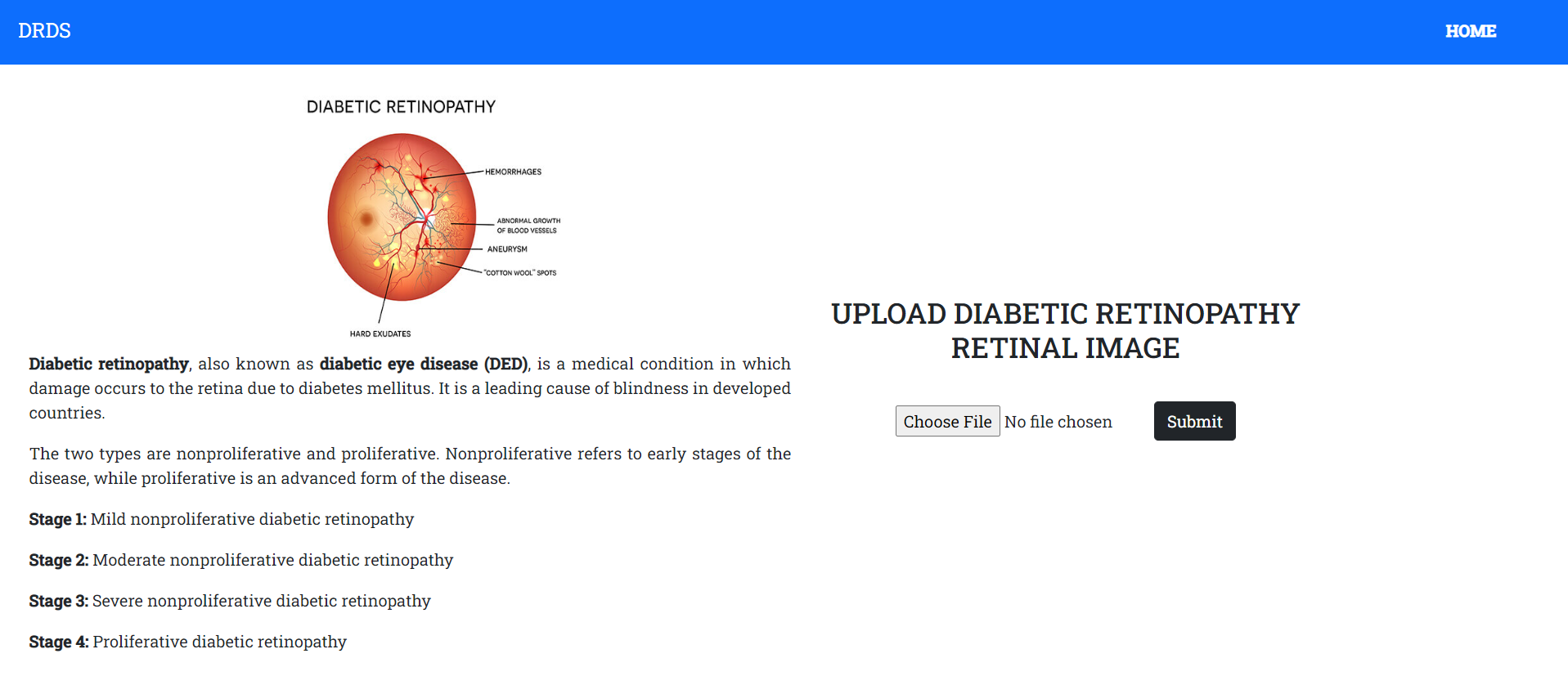


Figure 10 Diabetic Retinopathy Detection System - User Interface

Figure 13 shows the user interface of the application. With the help of the form on the homepage, the user can diagnose the image without the need for Ophthalmologists.

## Data Collection

Currently, many datasets are publicly available to train, validate and test the custom and pre-trained models. Each dataset contains retinal fundus images captured using different cameras at different angles and different lighting conditions. The images in most of the datasets show all signs of diabetic retinopathy and can be used to detect and extract lesions such as MAs, HAs, Exudates, and blood vessels. The below table shows a list of datasets that are currently available for public use.

|  |  |  |
| --- | --- | --- |
| Dataset | Total Number of Images | Image Size |
| DIARETDB1 | 89 images | 1500 x 1152 pixels |
| Kaggle | 88,702 images | 433 x 289 pixels to 5184 x 3456 pixels |
| E-Ophtha | 463 images | 2544 x 1696 pixels and 1440 x 960 pixels |
| DDR | 13,673 images | Different Image Resolutions |
| DRIVE | 40 images | 565 x 584 pixels |
| HRF | 45 images | 3504 x 2336 pixels |
| MESSIDOR | 1200 images | 1440 x 960 pixels |
| MESSIDOR-2 | 1748 images | 1440 x 960 pixels, 2240 x 1488 pixels, 2304 x 1536 pixels |
| STARE | 20 images | 700 x 605 pixels |
| CHASE DB1 | 28 images | 1280 x 960 pixels |
| IDRiD | 516 images | 4288 x 2848 pixels |
| ROC | 100 images | 768 x 576 pixels, 1058 x 1061 pixels, 1389 x 1383 pixels |
| DR2 | 435 images | 857 x 569 pixels |

Table - List of publicly available datasets

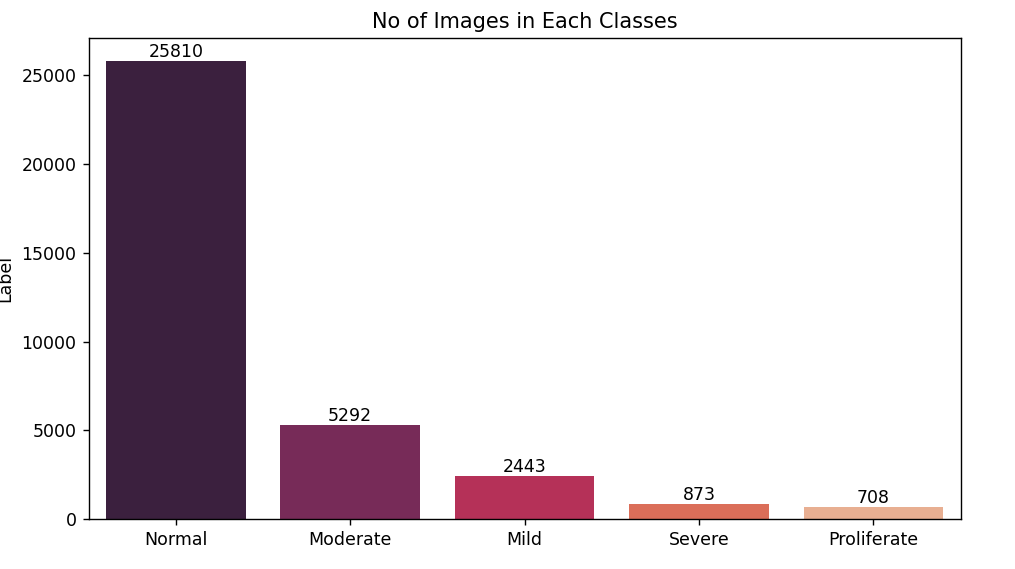
In the proposed approach, one of the largest retina datasets, i.e., the Kaggle dataset is used to train the pre-trained Inception-V3 model. Figure 4 shows the total images available in each stage.

Figure 11 Dataset – Images in DR Stages

The original Kaggle dataset contains a total of 35,126 retinal fundus images in which 25810 images are considered normal images with no DR signs, 2443 images are considered Mild and 5292, 873, and 708 are considered Moderate, Severe, and Proliferative respectively.

## Evaluation Metrics

Evaluating the proposed model is an important part of constructing a machine learning model. The evaluation metrics tell how accurate the proposed model can able to predict the images into 5 classes, i.e., Normal, Mild, Moderate, Severe, and Proliferative. In this project, four evaluation metrics are considered to check the performance of the proposed model. To calculate the performance of the machine learning models, four parameters are considered, i.e., True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN).

**True Positive –** also known as TP value. The multi-class classification model correctly predicted the images as the diabetic retinopathy images, i.e., Mild, Moderate, Severe, or Proliferative DR.

**True Negative –** also known as TN value. The multi-class classification model correctly predicted the images as normal images.

**False Positive –** also known as FP value. The multi-class classification model wrongly predicted the normal images as the diabetic retinopathy images, i.e., Mild, Moderate, Severe, or Proliferative DR.

**False Negative –** also known as FN value. The multi-class classification model wrongly predicted the DR images as normal or healthy images.

Using the above-explained parameters, each metric is explained clearly in the below sub-sections.

## Accuracy

Accuracy is one of the important evaluation metrics in machine learning. This ensures that the proposed model is predicting the DR stages of images correctly without any mistakes. The formula for calculating the accuracy is given below. It reaches 1 as its best value and 0 as its worst value.

Equation - Accuracy

where   predicted value and is the actual value of the i-th sample and are the total number of samples.

## Recall

The recall is a metric that calculates the percentage of a certain class identified correctly out of all samples of that certain class. It reaches 1 as its best value and 0 as its worst value.

Equation - Recall

Where L is the set of labels, y is the set of predicted (sample, label) pairs, is the set of true pairs, is the subset of y with label , is the subset of .

## Precision

Precision is a metric that calculates the ratio between True Positive and all the Positives of that certain class. It reaches 1 as its best value and 0 as its worst value.

Equation - Precision

Where L is the set of labels, y is the set of predicted (sample, label) pairs, is the set of true pairs, is the subset of y with label , is the subset of .

## F1-Score

F1-score is a measurement of both precision and recall. It is the harmonic score of both precision and recall. It reaches 1 as its best value and 0 as its worst value.

Equation - F1-Score

## Confusion Matrix

A confusion Matrix, also known as an error matrix, is a specific table layout of all correct and incorrect classifications predicted by the proposed model. This metric is used to evaluate the performance of the machine learning model from the values of Precision, Recall, F1 score, and accuracy.

# Results

This section is sub-categorized into different sections to illustrate the performance of the proposed model. The fine-tuned model obtained better results when compared with the initial model. Also, while training, the fluctuations in val\_accuracy and val\_loss are smoothened with the help of the learning rate. This helps the fine-tuned model to learn more details from the retinal fundus images.

# Initial Inception-V3

The Initial Inception-V3 trained on an oversampled dataset for 50 epochs. Each epoch took an average of 3546s to complete. After completion of the 50th epoch, the accuracy and val\_accuracy of the model increased to 0.7523 and 0.7100 whereas the loss and val\_loss reduced to 0.5690 and 0.7524 respectively. The below Figure shows the Inception-V3 accuracy and loss graphs.

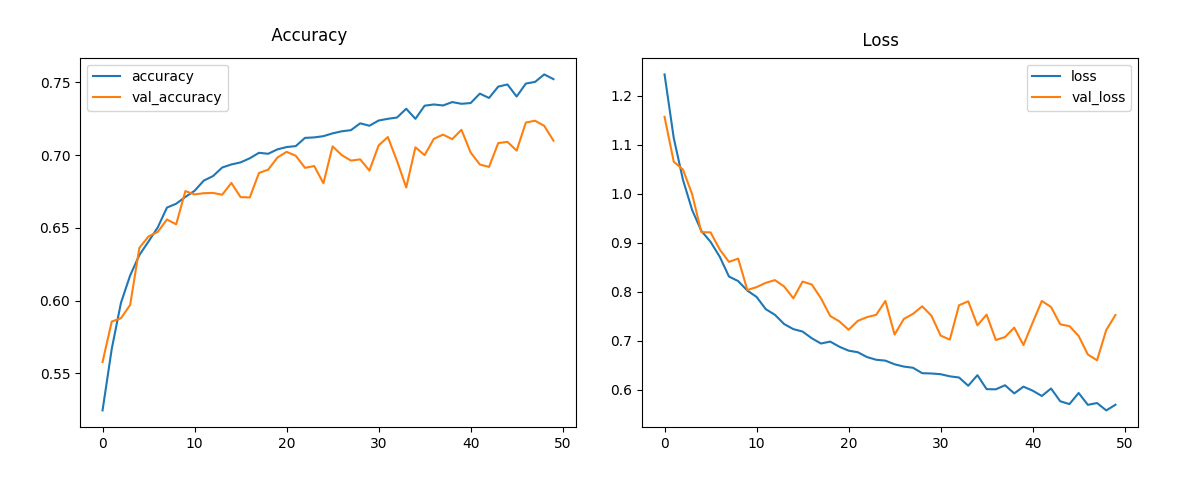


Figure Initial Model - Accuracy and Loss graphs

The Initial Inception-V3 model with all layers frozen achieved an accuracy of 71.23% on the test dataset with a test loss of 0.75042 and with Precision, Recall, and F1 scores of 0.68867, 0.71233, and 0.6805 respectively. Figure 12 shows the all correct and incorrect classifications of Inception-V3 on the test dataset.

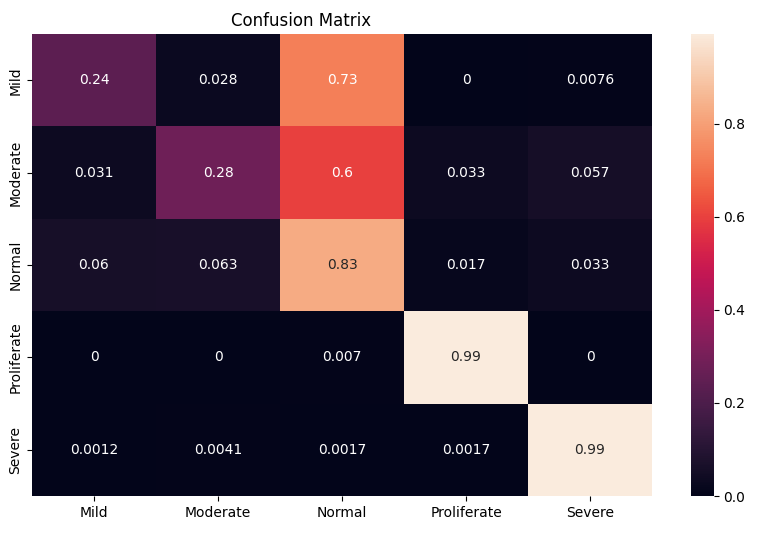


Figure Initial Model - test dataset - confusion matrix

To evaluate the performance of the initial model, a well-known dataset known as ‘APTOS 2019 Blindness Detection’ from the Kaggle dataset is used. On this dataset, the proposed model obtained an accuracy of 39.47% with a test loss of 4.74626 and with Precision, Recall, and F1 score of 0.36501, 0.40413, and 0.36548 respectively. Figure 13 shows all correct and incorrect classifications of Inception-V3 on the APTOS 2019 Blindness Detection dataset.

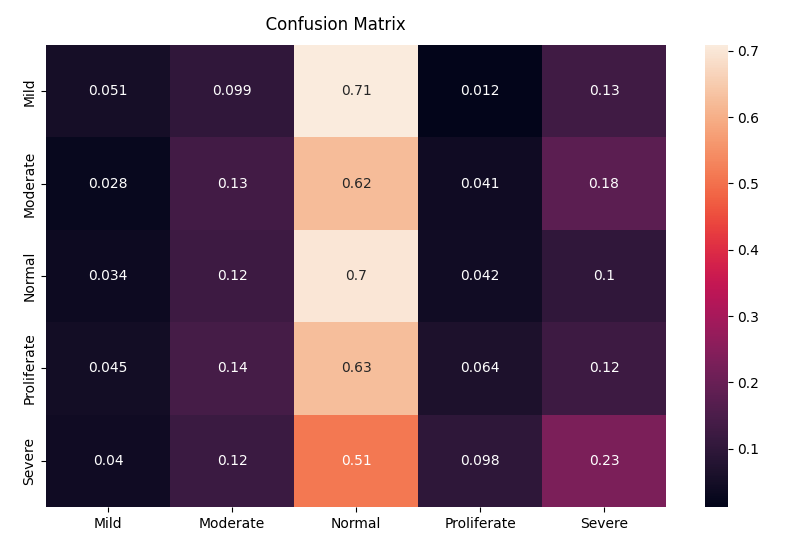


Figure Initial model - APTOS dataset - confusion matrix

# Fine-tuned Inception-V3

To improve the performance of the Inception-V3 model, the first two inception blocks are frozen while the remaining layers are unfrozen by setting the trainable value to False. And also, the number of epochs is increased from 50 to 80 whereas the learning rate of the ‘ADAM’ optimizer is reduced from 0.001 to 0.0001.

The fine-tuned model has better val\_accuracy and val\_loss when compared with the initial Inception-V3 model. Figure 15 shows the accuracy and loss graph of fine-tuned model.

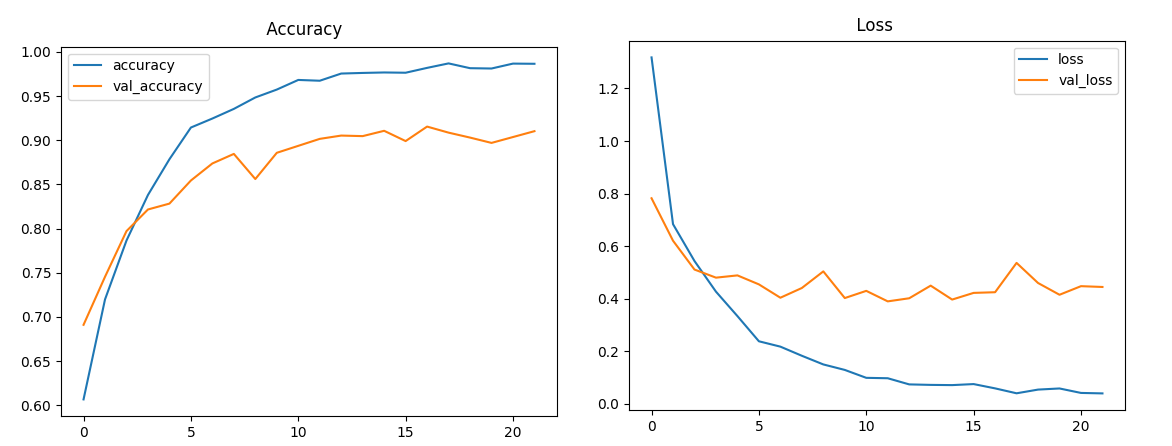


Figure Accuracy and Loss graph of Inception-V3 model

This modification increased the accuracy of the model on the test dataset from 71.23% to 90.39% with the test loss reduced to 0.40295. This fine-tuned model achieved precision, recall, and F1 scores of 0.9025, 0.90386, and 0.9029 respectively.

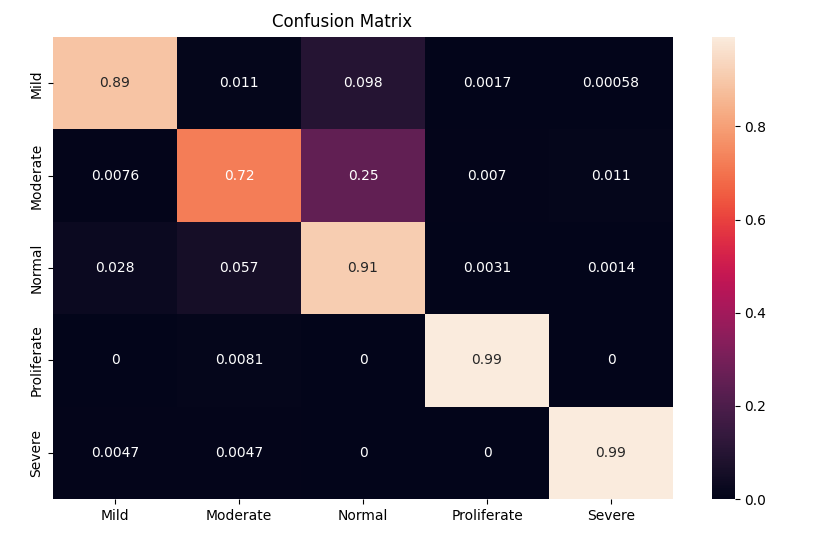


Figure Fine-tuned model - test dataset - confusion matrix

To evaluate the performance of the fine-tuned model on a different dataset, the proposed model is tested on the ‘APTOS 2019 Blindness Detection dataset. The proposed model obtained an accuracy of 54.22% with a test loss of 4.17766 and achieved precision, recall, and F1 score of 0.4686, 0.53641, and 0.44126 respectively. Figure 16 shows the confusion matrix of the APTOS dataset.

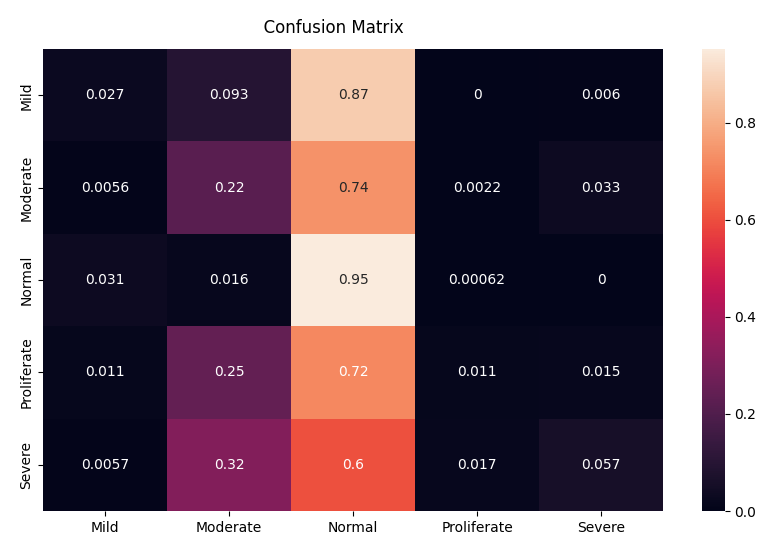


Figure Fine-tuned Model - APTOS dataset - confusion matrix

From the above-mentioned results, it is clear that the fine-tuned model has better when compared with the initial Inception-V3 model. The below table compares the results of both the initial Inception-V3 and fine-tuned Inception-V3 model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Number of Training parameters** | **Training**  **Accuracy** | **Training**  **Loss** | **Validation**  **Accuracy** | **Validation**  **Loss** | **Testing**  **Accuracy** | **Testing**  **Loss** |
| Inception V3 | 3,152,901 | 0.7523 | 0.5690 | 0.7100 | 0.7524 | 71.23% | 0.75042 |
| Fine-tuned Inception-V3 | 14,267,781 | 0.9866 | 0.0403 | 0.9103 | 0.4450 | 90.39% | 0.40295 |

Table - Inception V3 and fine-tuned Inception-V3 Results

# Discussion

The discussion chapter explains more about the results obtained by initial Inception-V3 and fine-tuned Inception-V3. This chapter is sub-categorized into two sub-sections, i.e., Initial Inception-V3 and fine-tuned Inception-V3.

# Inception-V3

The Initial Inception-V3 achieved a low accuracy when compared with the fine-tuned version. Figure 13 is the confusion matrix of the inception-V3 on the test dataset. From the figure, it is clear that the model is not able to detect few stages of DR, i.e., Mild and Moderate more accurately. However, this model can able to detect the last stages of DR, i.e., Severe and Proliferative. So, this model is only useful when the patient is in the last stage of diabetic retinopathy. The main aim of the project is to detect diabetic retinopathy in the retinal fundus image at an early stage. So, this helps the patient to reduce the chances of losing vision.

Figure 12 shows the accuracy and loss graphs of the initial Inception-V3 model. From the graphs, it is clear that, due to the high learning rate, the model has more fluctuations in val\_accuracy and val\_loss. To overcome fluctuations and less accuracy, the initial model is fine-tuned by freezing only the top 2 inception blocks of the model. This helps the model to learn more features from the early stages of retinal fundus images, i.e., Mild and Moderate.

# Fine-tuned Inception-V3

To fine-tune the inception-V3 model, a few layers are unfrozen, and also the learning rate is reduced to 0.0001. This helps the model to learn more features from all stages of the DR images. Once the model is fine-tuned and tested on the test dataset, the model obtained high accuracy when compared to the previous model. When the confusion matrix of both models is compared, it is clear that the fine-tuned model can detect all stages of DR more accurately.

Finally, both models, i.e., initial and fine-tuned models are tested on a different dataset that is available on the Kaggle website. The ‘APTOS 2019 Blindness Detection’ dataset is used to test both models. This dataset contains a total of 3,662 images of all DR stages. This dataset is hugely imbalanced, i.e., around 80% of images are considered as normal images. Due to an imbalanced dataset, both models are not able to detect the images accurately. However, the fine-tuned model has better accuracy when compared with the initial Inception-V3 model.

# Conclusion

From the literature review, it can be seen that many existing systems can able to detect diabetic retinopathy from retinal fundus images through different classification techniques, i.e., binary classification, multi-level classification, lesion-based classification, and vessel-based classification.

The existing binary classification models can able to classify the images into either healthy images or diabetic retinopathy images. These models did not classify the images into 5-classes, i.e., Normal, Mild, Moderate, Severe, and Proliferative. The existing multi-level classification models can able to detect stages in the retinal fundus images. However, some of these models involve manual diagnosis whereas the remaining models are trained using fewer data and not tested on different datasets that are currently available to the public. Moreover, some of these models are not able to detect all stages of the DR in the retinal fundus images. The lesion-based systems and vessel-based systems can able to extract lesions such as Microaneurysms, Haemorrhages, Exudates, and blood vessels in the retinal fundus images respectively. These systems did not classify the retinal fundus images into 5-class DR images.

Moreover, all the above-mentioned systems take huge time, data, and computational power to train the convolutional neural networks from scratch. To overcome all these problems, the proposed system is developed. The main aim of the proposed system is to classify the retinal fundus images into 5 stages, i.e., Normal, Mild, Moderate, Severe, and Proliferative. The proposed system uses the concept of transfer learning. Transfer learning is a research problem that helps to solve a particular problem using the models that are trained to solve a different problem. This technique is widely used to overcome time- and cost-consuming issues and to train the models with the limited dataset.

The proposed system utilized a well-known pre-trained model known as the ‘Inception-V3’ model. Inception-V3 is better at extracting features from the different types of images when compared to other well-known pre-trained models such as VGG16, Xception, InceptionResNetV2, and MobileNet. To train and test the Inception-V3 model, a popular Kaggle dataset known as ‘Diabetic Retinopathy (resized)’ is used. This dataset is the resized version of the original dataset available in Kaggle known as ‘Diabetic Retinopathy Detection’.

The final version of the trained Inception-V3 model achieved an accuracy of 90.39% with a test loss of 0.40295. To examine the performance of the trained Inception-v3 model, the fine-tuned model is again tested on a different dataset, i.e., the ‘APTOS 2019 Blindness Detection’ dataset. When tested on the different datasets, the trained model obtained an accuracy of 54.22% with a test loss of 4.17766.

However, the proposed model is not able to classify images into 5 stages more accurately on different datasets. So, to improve the accuracy of the model on different datasets, the present work can be improved by implementing different lesion-based techniques such as detecting Microaneurysms, Exudates, and Haemorrhages. And also, a more balanced dataset can able to improve the performance of the model. Currently, the dataset used is hugely imbalanced. Most of the images in the dataset are considered healthy images. So, to overcome this issue, data augmentation is implemented on the original dataset. Training the model on a dataset that has more images of other stages, i.e., Mild, Moderate, Severe, and Proliferative will help to classify the DR images more accurately.

# References

World Health Organization (2021). Diabetes. [online] World Health Organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/diabetes> [Accessed 29 March 2022].

NHS Choices (2019). Overview - Diabetic retinopathy. [online] NHS. Available at: <https://www.nhs.uk/conditions/diabetic-retinopathy/> [Accessed 29 March 2022].

Leigh, R., 2022. Pre-proliferative and proliferative retinopathy. [online] Oxford Medicine Online. Available at: <https://oxfordmedicine.com/view/10.1093/med/9780199544967.001.0001/med-9780199544967-chapter-7> [Accessed 29 March 2022].

kcretina, 2017. Diabetic-Retinopathy. [image] Available at: <https://d3b3by4navws1f.cloudfront.net/328135916-Diabetic-Retinopathy.jpg> [Accessed 29 March 2022].

Mayoclinic.org. 2022. Diabetic retinopathy - Diagnosis and treatment - Mayo Clinic. [online] Available at: <https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/diagnosis-treatment/drc-20371617> [Accessed 29 March 2022].

Congdon, N., Zheng, Y. and He, M. (2012). The worldwide epidemic of diabetic retinopathy. Indian Journal of Ophthalmology, 60(5), p.428. [Accessed 29 March 2022].

Healthline. (2021). Diabetic Retinopathy Stages: The 4 Stages and What to Do. [online] Available at: <https://www.healthline.com/health/diabetes/diabetic-retinopathy-stages> [Accessed 29 March 2022].

Xu, K., Feng, D. and Mi, H. (2017). Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image. Molecules, 22(12), p.2054. [online] Available at: <https://doi.org/10.3390/molecules22122054> [Accessed 29 March 2022].

Esfahani, M., Ghaderi, M. and Kafiyeh, R. (2018). Classification of diabetic and normal fundus images using new deep learning method. [online] Available at: <http://lejpt.academicdirect.org/A32/233_248.pdf> [Accessed 29 March 2022].

Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., and Qian, W. (2019). An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/abstract/document/8857160?casa_token=SxXs99KdJu0AAAAA:PtsUX0tWteE9STUeQG7Fbtd_mljTVYJ14UOqY5fZkZk5MceMm5MaZeBP4Jqgkd3V6wSaIKzwCM9b9g> [Accessed 29 March 2022].

Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P. and Zheng, Y. (2016). Convolutional Neural Networks for Diabetic Retinopathy. Procedia Computer Science, 90, pp.200–205 [Accessed 29 March 2022].

Wang, X., Lu, Y., Wang, Y. and Chen, W.-B. (2018). Diabetic Retinopathy Stage Classification Using Convolutional Neural Networks. 2018 IEEE International Conference on Information Reuse and Integration (IRI) [Accessed 29 March 2022].

Mobeen-ur-Rehman, Khan, S.H., Abbas, Z. and Rizvi, S.M.D. (2019). Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/abstract/document/8701231/> [Accessed 29 March 2022].

Harangi, B., Toth, J., Baran, A. and Hajdu, A. (2019). Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/document/8857073> [Accessed 02 April 2022].

Yan, Y., Gong, J. and Liu, Y. (2019). A Novel Deep Learning Method for Red Lesions Detection Using Hybrid Feature. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/document/8833190> [Accessed 02 April 2022].

Sharath Kumar, P.N., Kumar, R.R., Sathar, A. and Sahasranamam, V. (2013). Automatic detection of exudates in retinal images using histogram analysis. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/abstract/document/6745487> [Accessed 02 April 2022].

Franklin, S.W. and Rajan, S.E. (2014). Diagnosis of diabetic retinopathy by employing image processing technique to detect exudates in retinal images. IET Image Processing, 8(10), pp.601–609 [Accessed 02 April 2022].

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 [Accessed 03 April 2022].

Vengalil, S.K., Sinha, N., Kruthiventi, S.S.S., and Babu, R.V. (2016). Customizing CNNs for blood vessel segmentation from fundus images. [online] IEEE Xplore. Available at: <https://ieeexplore.ieee.org/abstract/document/7746702> [Accessed 03 April 2022].

Budak, Ü., Cömert, Z., Çıbuk, M. and Şengür, A. (2020). DCCMED-Net: Densely connected and concatenated multi-Encoder-Decoder CNNs for retinal vessel extraction from fundus images. Medical Hypotheses, 134, p.109426 [Accessed 03 April 2022].

www.kaggle.com. (n.d.). Diabetic Retinopathy (resized). [online] Available at: <https://www.kaggle.com/datasets/tanlikesmath/diabetic-retinopathy-resized> [Accessed 01 April 2022].

Alyoubi, W.L., Shalash, W.M. and Abulkhair, M.F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. Informatics in Medicine Unlocked, 20, p.100377 [Accessed 03 April 2022].

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going Deeper with Convolutions. Cv-foundation.org, [online] pp.1–9. Available at: <https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html> [Accessed 03 April 2022].

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. [online] arXiv.org. Available at: <https://arxiv.org/abs/1512.00567> [Accessed 03 April 2022].