Grading the severity and Segmenting Lesions in Diabetic Retinopathy

Submitted By

Aritra Sen (C22004)

Ashish Jadhav (C22005)

Joy Bhowmick (C22012)

Kavya Pari (C22013)

Kushagra Mohan Prasad (C22014)

Shubham Koshti (C22023)

Sohan Pant (C22025)

Under the guidance of

Dr. SUBHASIS DASGUPTA

(Associate Professor, Department of Data Science)

Submitted to Praxis Business School as

Capstone Project in partial fulfilment of the requirements

of Post Graduate Program in Data Science, Spring 22

Bakrahat Road, Rasapunja, 24 Parganas South, Kolkata

OCT 2022

ABSTRACT

Diabetic Retinopathy is a complication of diabetes that affects the eyes due to diabetes mellitus. DR is considered as the most prevalent cause of avoidable vision impairment, mainly affecting the working-age population in the world. Fundus photography technique was used to take these photographs of the retina. Pre-trained Convolutional Neural Network (CNN) VGG-16 has been used to detect the severity of Diabetic Retinopathy from the image. A computerized method was used for the detection of diabetic retinopathy through the analysis of the retina image which would help people with diabetes to detect the symptoms at its earlier stage. The basis of the classification of different stages of diabetic retinopathy is the detection and quantification of the severity of Diabetic Retinopathy. Segmenting lesions from retina images is an essential prerequisite for accurate severity assessment of diabetic retinopathy. Because of variation in morphologies, number and size of lesions, the manual grading process becomes really challenging and time-consuming. This necessitates the need of an automatic segmentation system that can accurately explain the region of interest boundaries and assist ophthalmologists in speedy diagnosis along with diabetic retinopathy severity grading. Single Shot Detection technique has been implemented to detect lesions. This project outlines the challenge, its organization, the dataset used, evaluation methods and results of top-performing participating solutions. The major aim of the project is to contribute towards enhancing the diagnosis of Diabetic Retinopathy by providing deep learning methods to facilitate lesion segmentation and early detection.

ACKNOWLEDGEMENT

Nothing worth having comes easy since success demands hard work and determination. In the present world of competition there is a race of existence in which those have the will to come forward, succeed. A capstone project is a bridge between theoretical and practical working. Though this capstone is a group work, we could have never been able to explore the depths without the help, support, guidance and efforts of a lot of people. We owe gratitude to all those who have made this project possible and because of whom our experience has been such that we will cherish forever.

During the course of our project, we were recipients of a good amount of help from numerous people whose contributions were invaluable and helped in the success of our project. Hence, we are highly indebted to them and are expressing our sincere thanks to them.

We are very grateful and thankful to **Dr. Sourav Saha**, Dean Academics, Data Science, Praxis Business School, who has been instrumental in constantly motivating us for successful completion of our task.

We wish to express our deep gratitude and sincere thanks to our guide **Dr. Subhasis Dasgupta**, Associate Professor, Department of Data Science, Praxis Business School, who from the conceptual stage of the project to the final phase provided us with valuable help and active encouragement in all spheres of our project.

Words can hardly convey our profound indebtedness to our families whose love, encouragement and all types of support that let us go on in any circumstances throughout our lives and that is what made this work possible. Hence, we wish to thank our beloved parents and friends for their enduring patience, support, love and affection for constantly standing by us during the entire period of our project.

Thank you all

Place: Kolkata

Date: 21st October 2022

CONTENTS

Topics	Page Number
1. Introduction	**4**
1.1) Background	**4**
1.2) Classification of Diabetic Retinopathy	**6**
1.3) Problems and Preventions	**6**
2. Literature Review	**7**
2.1) Disease Grading	**7**
2.2) Lesion Segmentation	**7**
3. Data Description and Rationale	**9**
3.1) Problem Statement	**9**
3.2) Research Aim	**9**
3.3) Objective of Study	**9**
3.4) Dataset	**9**
3.4.1) Disease Grading	**10**
3.4.2) Segmentation	**10**
4. Methodology and Results	**11**
4.1) Disease Grading	**11**
4.1.1) Method 1 (Using Keras)	**12**
4.1.2) Results (Using Keras)	**13**
4.1.3) Method 2 (Using Image Data Generator)	**15**
4.1.4) Results (Using Image Data Generator)	**16**
4.2) Lesion Segmentation	**18**
4.2.1) Annotations (Bounding Box)	**18**
4.2.2) Single Shot Detection (SSD)	**21**
4.2.3) Training an Object Detection SSD model	**21**
4.2.4) Results	**22**
5. Conclusion, Limitations and Scope	**25**
5.1) Conclusion and Discussions	**25**
5.2) Limitations	**26**
5.3) Scope of Future Study	**26**
References	**27**

CHAPTER 1: INTRODUCTION

1.1) BACKGROUND

Diabetic Retinopathy (DR)and Diabetic Macular Edema (DME) are the most common sight-threatening medical conditions caused due to retinal microvascular changes triggered by diabetes (Reichel and Salz, 2015), predominantly affecting the working-age population in the world (Atlas, 2017). It is one of the main causes of blindness in human eyes, and regular fundus screening is an effective way to discover the location of disease (S. Wang et al., 2020). As diabetes progresses, the disease slowly affects the circulatory system including the retina and occurs as a result of long term accumulated damage to the blood vessels, declining the vision of the patient leading to diabetic retinopathy. After 15 years of diabetes about 10% of people become blind and approximately 2% develop severe visual impairment. According to an estimate by WHO, more than 220 million people worldwide have diabetes (WHO, 2011).

Diabetes or more precisely Diabetes Mellitus (DM) is a metabolic disorder happens because of high blood sugar level in the body. Over the time, diabetes creates eye deficiency also called as Diabetic Retinopathy (DR) causes major loss of vision (S. Dutta et al., 2017). There are three stages to Diabetic Retinopathy:

a) **NPDR (Non-Proliferative Diabetic Retinopathy) -** This is the early stage of diabetic eye disease. Many people with diabetes have it. With NPDR, tiny blood vessels leak, making the retina swell. When the macula swells, it is called Macular Edema. This is the most common reason why people with diabetes lose their vision.

b) **PDR (Proliferative Diabetic Retinopathy) -** PDR is the more advanced stage of diabetic eye disease. It happens when the retina starts growing new blood vessels. This is called neovascularization. These fragile new vessels often bleed into the vitreous. You might see a few dark floaters if they only bleed a little. If they bleed a lot, it might block all vision.

c) **DME (Diabetic Macular Edema) -** It is an additional category in the field of Diabetic Retinopathy is DME, which is an important representation of this disease that can be detected at any severity level of DR both NPDR and PDR and it manifest the most common cause of vision loss in patients with DR, it is usually induced from diabetes-induced breakdown of the blood-retinal barrier (BRB), with continues fluid and circulating proteins leakage. The abnormal thickening in the retina comes as result of excessive leakage of fluid into the neural retina and often cystoid Edema of the macula.

Lesion segmentation of Diabetic Retinopathy is the prerequisite work for screening and diagnosis, and it also lays a foundation for the subsequent grading of the severity of Diabetic Retinopathy. Segmentation in general means dividing into separate parts. Image Segmentation is to partition an image into multiple parts or regions based on the characteristics of the pixels in the image. Lesion segmentation is employed on a retinal image that has abnormalities that occur in the different stages of DR, viz., NPDR and PDR. Generally, common diabetic retinopathy consists of Microaneurysms (MA), Hard Exudates (EX), Soft Exudates (SE), and Haemorrhage (HE).

a) **Haemorrhages** generally appear as 'dot' or 'blot' or 'flame-shaped' depending upon their depth within the retina.

b) **Microaneurysms** are round localized capillary dilations and are the earliest clinically visible signs of DR. The dilations generally appear as small red dots in clusters or in isolation but do not affect eye vision. Haemorrhages and microaneurysms are collectively known as red lesions, while, soft and hard exudates are known as bright lesions.

c) **Soft exudates** appear as discoloured greyish-white patches with indistinct edges in the nerve fibre layer.

d) **Hard exudates** are clusters of waxy yellow-white intra-retinal deposits that are made up of extracellular lipid leaked from abnormal retinal capillaries. In extreme cases, these may evolve into rings known as circinate, which significantly deteriorate the eye vision (Joshi S. et al., 2017).

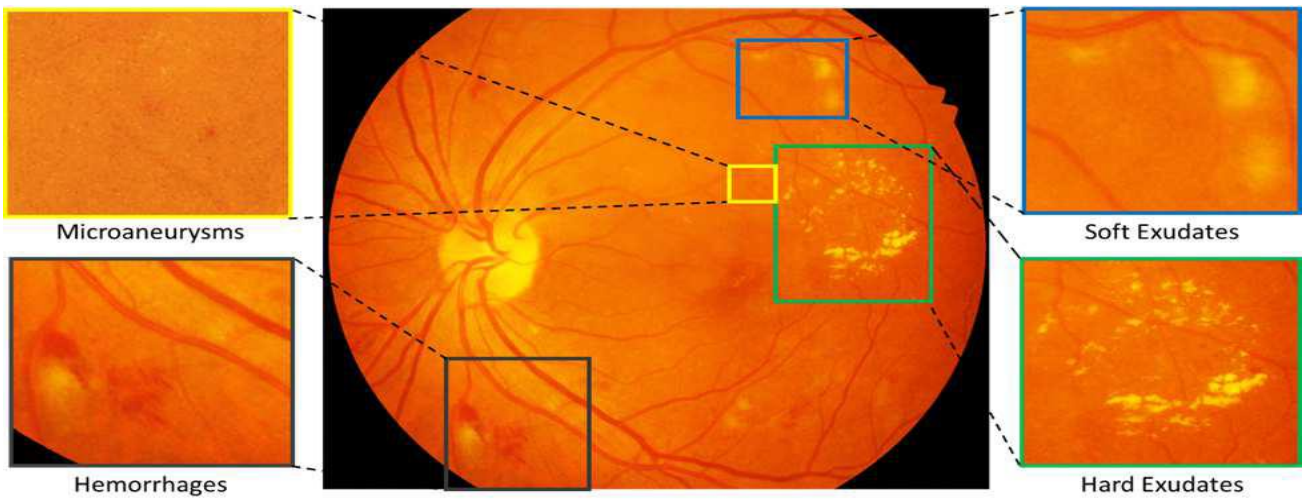

Fig 1: Colour fundus photograph containing different retinal lesions associated with Diabetic Retinopathy

1.2) CLASSIFICATION OF DIABETIC RETINOPATHY

I. Non-proliferative DR (NPDR):
   a. No DR
   b. Mild
   c. Moderate
   d. Severe
   e. Very Severe

II. Proliferative DR (PDR):
   a. Mild to Moderate PDR
   b. High Risk PDR

III. Clinically significant macular oedema (CSME):
   a. May exist by itself or along with NPDR and PDR

1.3) PROBLEMS AND PREVENTIONS

I. Symptoms:
   a. Blurred vision
   b. Distorted vision
   c. Poor night vision
   d. Partial or total loss of vision

II. Complications:
   a. Vitreous Haemorrhage
   b. Tractional retinal detachment
   c. Glaucoma
   d. Blindness

III. Risk Factors:
   a. Duration of Diabetes
   b. Hypertension
   c. Obesity
   d. Smoking
   e. Pregnancy

IV. Prevention:
   a. Early intuition
   b. Appropriate referral to ophthalmologist

CHAPTER 2: LITERATURE REVIEW

2.1) DISEASE GRADING:

Diabetic retinopathy (DR) is considered one of the serious complications of diabetes, which is responsible for 2.6% of overall blindness (El. Houby, E., 2021). Therefore, an intelligent automated method for early and accurate detection of DR is required to manage the progress of the disease and thus guarantee appropriate treatment. Classification of DR includes the weighting of many features and finding the position of these features. Great work is achieved on detecting DR automatically using tradition methods such as K-Nearest Neighbour (K-NN) and support vector machine (SVM) that depend on hand-crafted features extraction and then classifying different cases depending on the selected features (Philip, S. et al., 2007, Mookiah, M. R. K. et al., 2013). In contrast, features can be learned automatically from the original images through the training phase using deep learning (Litjens, G. et al., 2017).

There are many high-performing pre-trained models such as VGG-16, that can be imported and used for image recognition. Examples of these models from the published literature are visual geometry group (VGG) (Simonyan K. et al., 2014), inception modules (GoogleNet) (Szegedy C, et al. 2015, 2016), residual neural network (ResNet) (He K, et al., 2016) and neural architecture search network (NasNetLarge) (Zoph B, et al., 2018) etc. These models were trained using ImageNet data which consists of 1,000,000 images with 1,000 classes, so they have learned to detect generic features and their learned weights are provided and used in similar problems. They achieved state-of-the-art performance and when used to develop other image recognition tasks, they remain effective. CNN was trained on the Kaggle fundus images dataset to classify DR and achieved an accuracy of 75%, sensitivity of 30% and specificity of 95% (Pratt H., et al., 2016). TL based on pre-trained models was used; GoogLeNet and AlexNet were applied on Kaggle and Messidor-1 datasets. The achieved accuracies were 74.5%, 68.8% and 57.2% for 2-ary, 3-ary and 4-ary classification models, respectively (Lam C., et al., 2018). AlexNet, VGG-16 and SqueezeNet were applied on 1200 images of MESSIDOR dataset to classify the severity level of DR. The achieved accuracies were 93.46%, 91.82% and 94.49% respectively; specificity of 94.53, 88.54 and 94.54, respectively; and sensitivity of 92.38, 93.47 and 94.47 (Rehman M.U., et al., 2019).

2.2) LESION SEGMENTATION:

In the early years, the medical researchers focused on the segmentation of diabetic retinopathy based on traditional digital image processing methods. The various segmentation approaches can be grouped into four types mainly: Mathematical morphology, Region growing, Thresholding and Machine learning-based techniques. Kaur and Mittal, 2015, employed a dynamic region growing

method achieving sensitivity and specificity of 97.9% and 99.4% respectively. Despite the overall process being efficient and inexpensive, the performance of the proposed technique is heavily affected by irregular illumination. Fleming et al. used morphological operations and Gaussian matched filters to extract candidate regions of microaneurysms. Antal and Hajdu adopted an ensemble learning strategy to integrate a series of image pre-processing approaches to improve final segmentation of microaneurysms. Kavitha and Duraiswamy extracted exudate features using a multilayer threshold method, but this model has requirements for the input image quality. Most existing models with excellent performance in medical image segmentation tasks are reconstructed based on FCN or UNet. In FCN (E. Shelhamer, et al., 2015, 2017), the last full connection layer was replaced with a convolution layer.

Lian et al. 2015, proposed a global and local residual U-Net to overcome the drawback of global image pre-processing in retinal vessel segmentation. A fully convolutional neural network to simultaneously segment arterioles and venules directly from the retinal images has been proposed by Xu et al., 2018. The model achieved sensitivity and specificity of 87% and 98% on the DRIVE dataset. Wang et al., 2019, proposed a dense U-Net and patch learning-based strategy for the segmentation of retinal vessels using DRIVE and STARE dataset. The model obtained an accuracy of 95.11% on the DRIVE dataset and 95.38% on the STARE dataset. V. Badrinarayanan et al., 2017 and G. Lin, et al., 2017 advanced in U-Net by using max-pooling indices and multipath input, respectively. Sambyal et al., 2020, presented a modified U-Net architecture based on the residual network and employ periodic shuffling with subpixel convolution initialized to convolution nearest neighbour resize. Yifei Xu, et al. 2015, used a baseline U-Net, to improve FFU-Net segmentation performance by 11.97%, 10.68%, and 5.79% on metrics SEN, IOU, and DICE, respectively.

CHAPTER 3: DATA DESCRIPTION AND RATIONALE

3.1) PROBLEM STATEMENT:

I. Lesion Segmentation from fundus retinal images using Deep Learning Techniques to identify and diagnose the disease

II. Computerized method for detection of DR through the analysis of the retinal image.

III. Applying CNN techniques which takes an image as the input and classifies it into the appropriate category. The output is obtained from the classification layer.

3.2) RESEARCH AIM:

The primary aim of the proposed work is to contribute towards enhancing the diagnosis of Diabetic Retinopathy (DR) by providing a Deep Learning method for early detection and classification of the severity of DR.

3.3) OBJECTIVE OF STUDY

I. Disease Grading: -
   a. To build a classification model which looks into the images and classify into one of the 5 classes i.e., 0, 1, 2, 3, 4.
   b. Detects the presence of Diabetic Retinopathy.

II. Segmentation: -
   a. Annotating images with the help of Bounding box.
   b. Identifying retinal lesions: Haemorrhages, Hard Exudates and Soft Exudates.

3.4) DATASET:

The fundus images were acquired from an Eye Clinic located in Nanded, Maharashtra, India. Retinal images of humans affected by diabetes were captured with 39 mm distance between lenses and examined eye using non-invasive fundus camera having xenon flash lamp. This dataset provides two types of annotations, namely pixel level annotations of lesions, image level DR and DME grading.

The Indian Diabetic Retinopathy Image Dataset (IDRiD) is a publicly available retinal fundus image dataset consisting of 516 images categorized in two parts:

- Retinal images with the signs of DR and/or DME.
- Normal retinal images (without signs of DR and/or DME).

The dataset provides ground truths associated with the signs of Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) and normal retinal structures.

3.4.1) Disease Grading:

Fig 2: Disease Grading Dataset Information

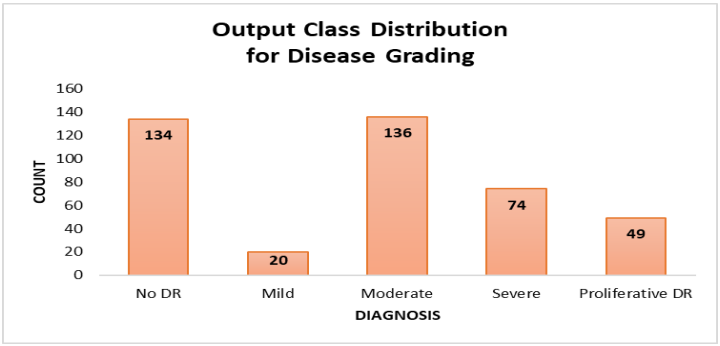

Fig 3: Output Class Distribution for Disease Grading

3.4.2) Segmentation:

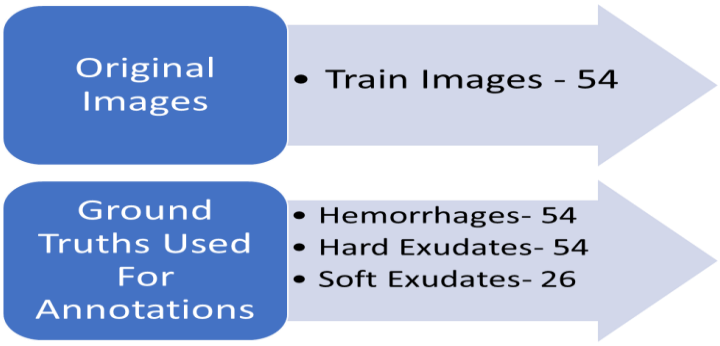

Fig 4: Dataset Information for Training

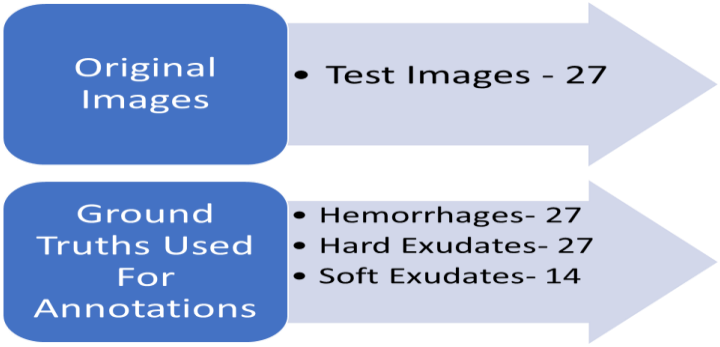

Fig 4: Dataset Information for Testing and validation

CHAPTER 4: METHODOLOGY AND RESULTS

4.1) DISEASE GRADING:

To understand the methodology and working, the VGG-16 model becomes important.

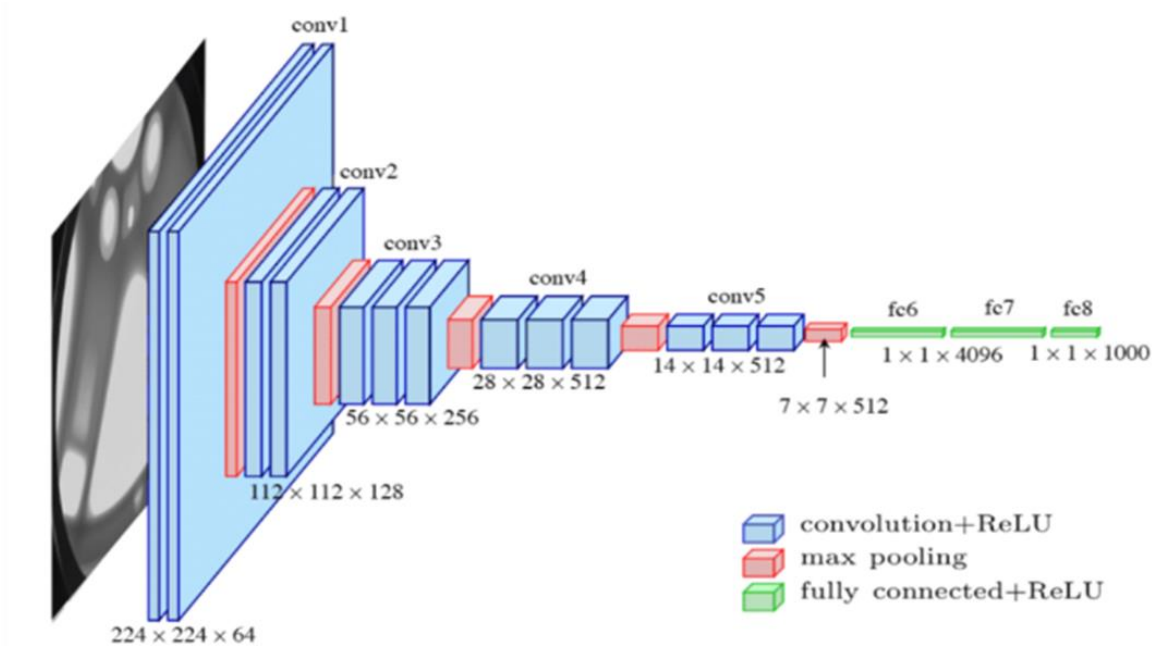

Fig 5: Architecture of VGG-16

VGG refers to Visual Geometry Group Oxford, and 16 refers to 16 layers it has in its architecture.

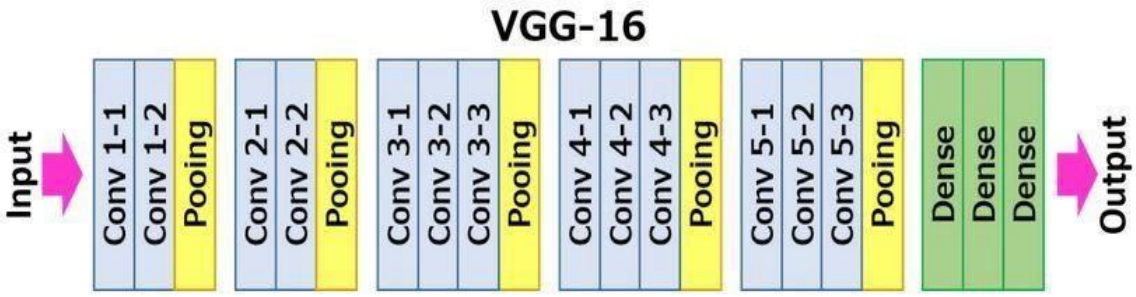

Fig 6: The 16 layers of VGG-16

VGG16 takes input of tensor size as 224, 244 with 3 RGB channel.

It has five parts: The 1st two parts are the 2 convolution layers and latter 3 parts have 3 convolution layers followed by a max pool layer after which there are fully connected layers. Conv-1 Layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters.

Most unique thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with stride 1 and always used the same padding and max pool layer of 2x2 filter of stride 2.

4.1.1) Method 1 (Using Keras):

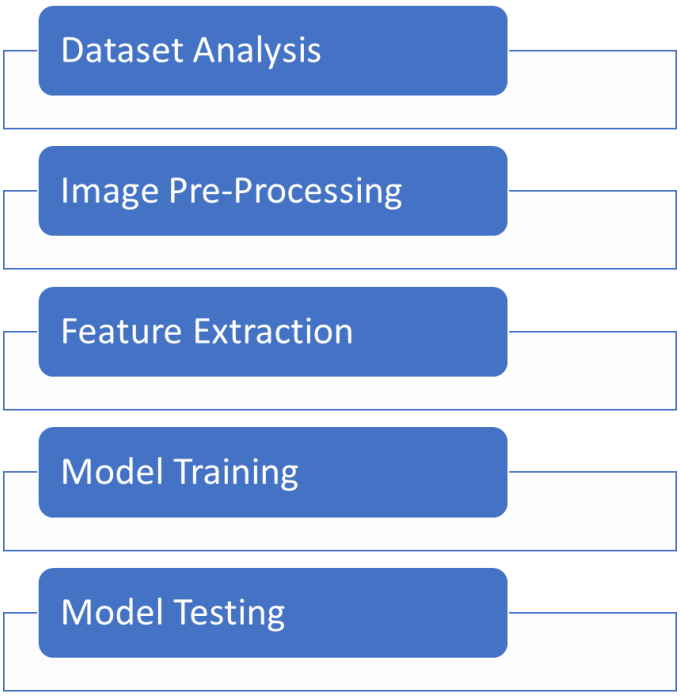

Figure 7: Methodology flow chart

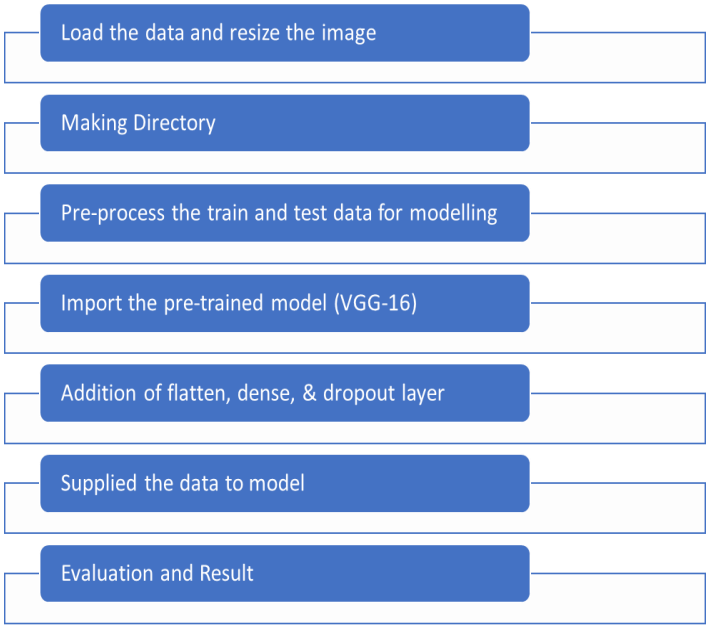

Figure 8: Methodology flow chart using Keras

The most important task is to load the dataset and connect this with their respective Ground truth. The original image size was very high which couldn't be processed to our machine so we reduce the image size to 356*536. We make 3 array to connect images with their grade and risk.

We are using pretrained VGG-16 model which was trained on image net dataset because we have very less data and with this, it is not possible to update the weight of neural net so we have used weight of vgg-16 model to train this.

Now the top layer is added to this model because learning of neural nets occurs in dense layer following that we have added 1 flatten layer, 4 dense layers, 1 dropout layer with dropout rate 20%, and prediction layer. In all the layers Activation function used is ReLu except prediction layer where softmax has been used to get the probability of prediction.

Now model is ready to implement. Thus, compilation is done using Adam optimizer and loss is 'Categorical Cross Entropy' because this is multiclass prediction. Now the data is supplied to model and executed for 15 epochs.

4.1.2) Results (Using Keras):

Dataset

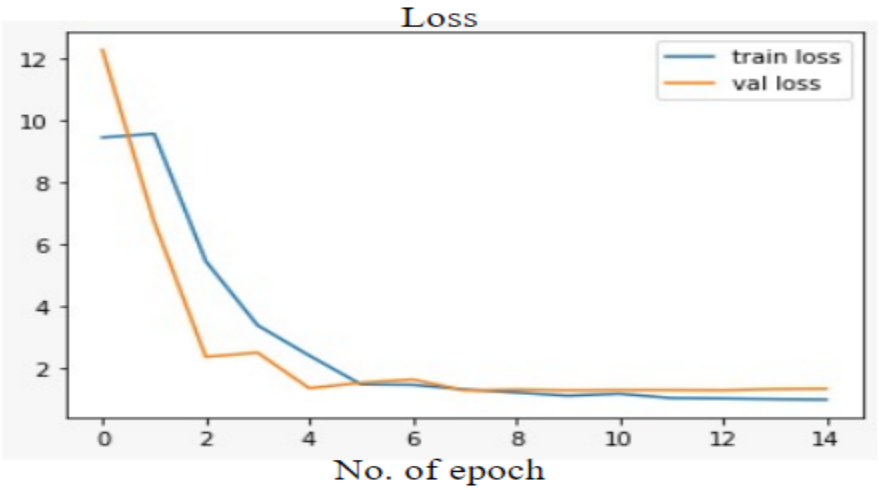

Fig 9: Decrease in Loss with increasing Epochs

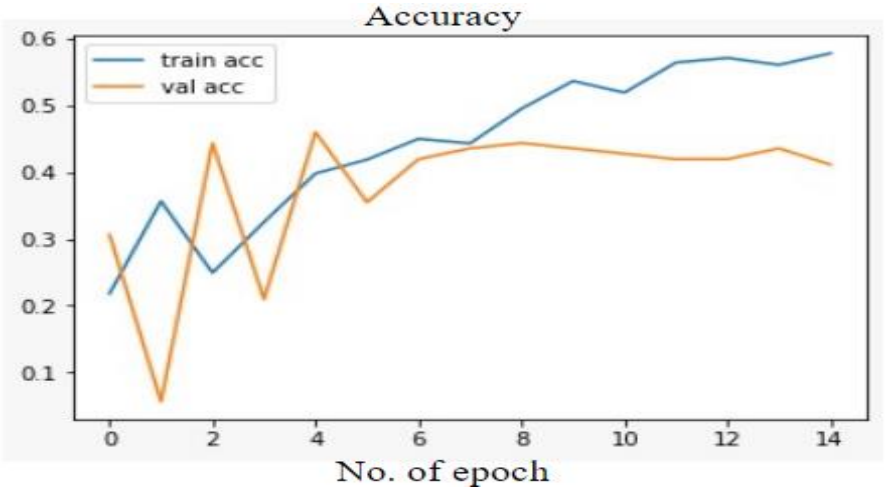

Fig 10: Increase in Loss with increasing Epochs

The results come to be at par with:

- Training accuracy – 56%
- Validation accuracy – 44%
- Test Accuracy – 39%

The following plots show that the Validation loss drops drastically after 2 epochs and the training loss shows similar trend after 5 epochs. Subsequently, both the training accuracy and validation accuracy show steady increase after 6 epochs.

Classification Report

	precision	recall	f1-score	support
0	0.55	0.35	0.43	34
1	0.00	0.00	0.00	5
2	0.33	0.84	0.48	32
3	0.00	0.00	0.00	19
4	0.00	0.00	0.00	13
accuracy			0.38	103
macro avg	0.18	0.24	0.18	103
weighted avg	0.28	0.38	0.29	103

Fig 11: Accuracy Scores for classification

From the Report it is evident that model is able to classify the images belonging to class 0 and class 2 but unable to classify the remaining classes. A plausible reason pertaining to this outcome is the availability of only 100 images belonging to the test set, out of which 34 and 32 belongs to class 0 and class 2 respectively making the dataset highly imbalanced.

Confusion Matrix

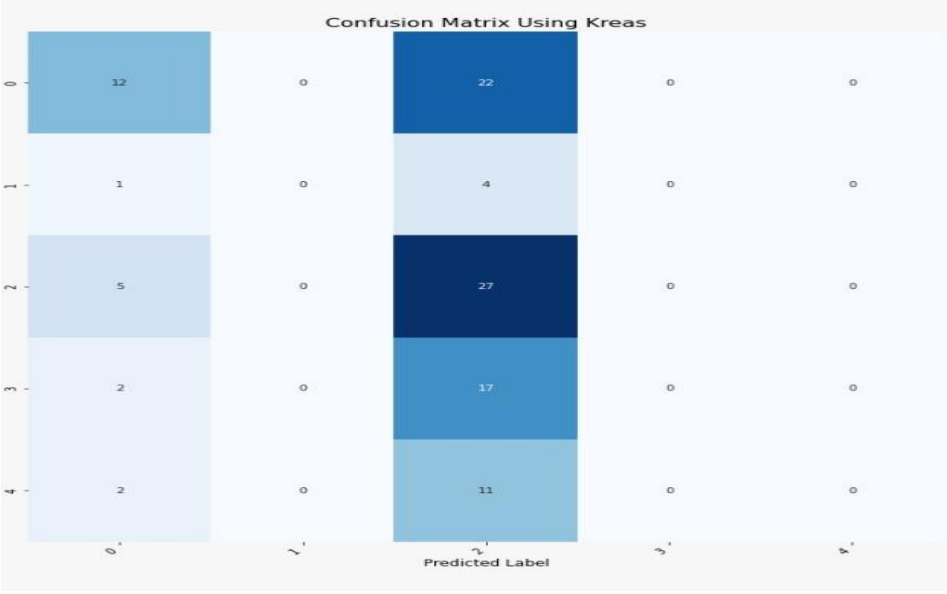

Fig 12: Confusion Matrix for classification

The confusion matrix shows numerous mis-classification. The images being nearly indistinguishable, it is unusually difficult for humans to classify them, making it pretty evident for the machine to find it hard to classify them correctly

Thus, an alternate approach deems necessary.

4.1.3) Method 2 (Using Image Data Generator):

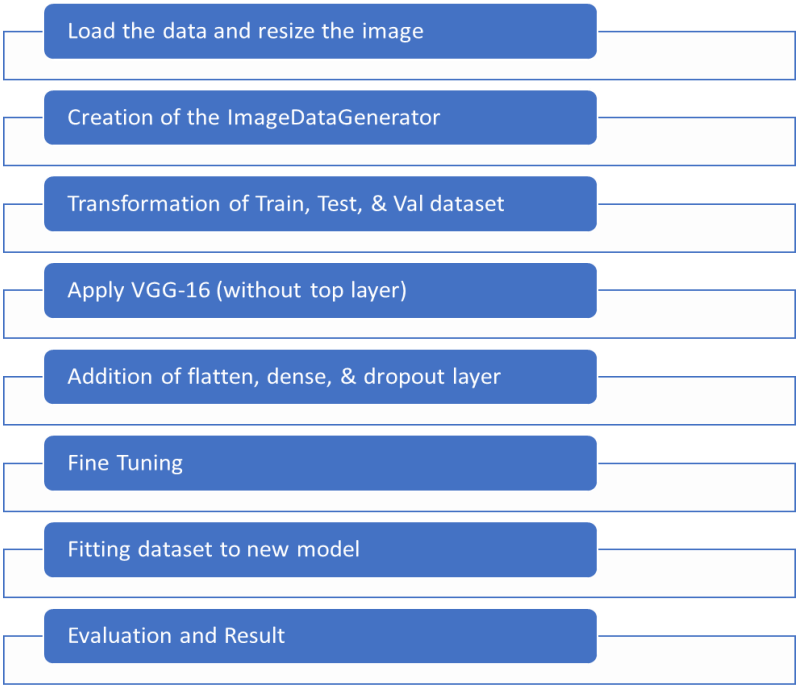

Fig 13: Methodology flow chart using Image Data Generator

To improve upon the previous model to classify diabetic retinopathy, two new models were used to see the improvement they might bring about in the training and testing score of the models. To do the same, before the data was set as an input to the model it was transformed using a data generator. The images were manipulated by rotating, flipping, increasing or decreasing their brightness, and so on. This allowed the model to be more robust as it now has more data to work with and thus improves its learning for the segregation of the class the image belongs to.

After the data was transformed it was split into training and validation from the training images in a 75-25% split. The testing images were then used for testing the model. Following that VGG16 was used without the top layer. The top layer was added which contained two dense layers, one drop-out layer followed by a dense output layer.

The data was supplied in batch mode over 100 epochs to the model. This was not adequate enough and thus to improve upon it, fine-tuning was used to improve the score. A fine-tuning of two

was selected, which meant that two layers unfrozen of the VGG-16 model so that weights can be updated of the VGG-16.

Thus, the model with data image generators for transformation and fine-tuning of two gave us a much better model which can give a decent accuracy even on this small dataset and an unbalanced dataset. The steps involved in building the architecture were:

I. Segregation of Images according to class.
II. Train and Test Image Data Generator were created.
III. Splitting the Train data into Train and Val.
IV. Creation of top CNN Model.
V. Transfer learning (VGG-16)
VI. Combat lack of data, over-fitting
    a. Dropout, Early stopping
    b. Data augmentation (flips, rotation)
VII. Fit the model
VIII. Evaluation using Confusion Matrix

4.1.4) Results (Using Image Data Generator):

Dataset

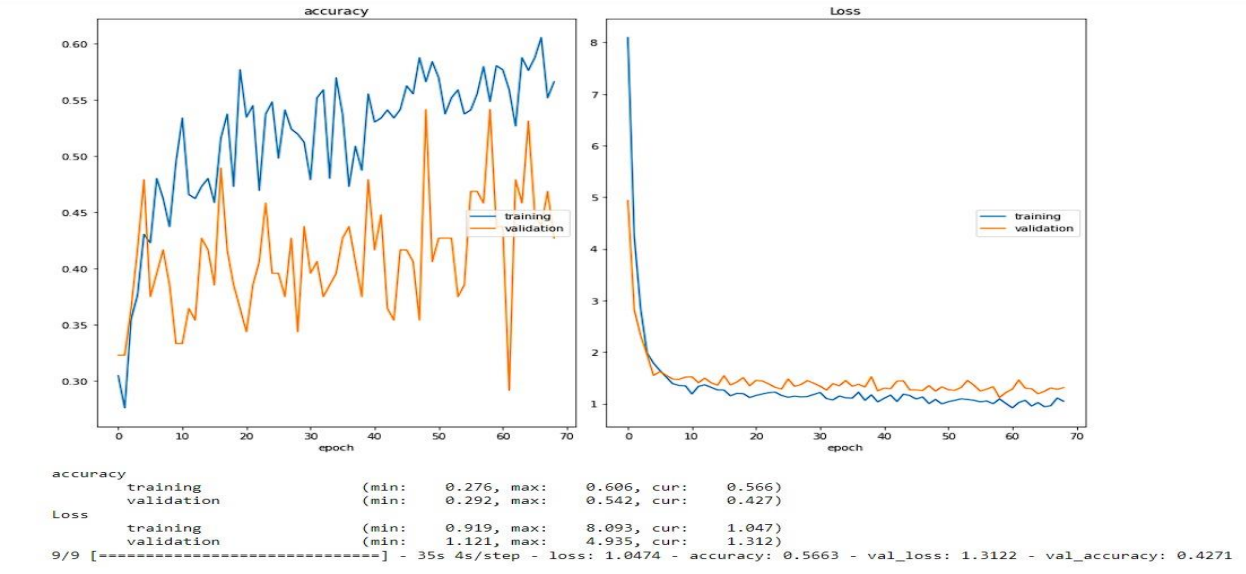

Fig 14: Increase in Accuracy and Decrease in Loss with increasing epochs

This first model gave us an accuracy of 54.5% in training and 48.54% in testing. With the adjustments, the improved model gave an accuracy of 60.6% in training and 51.5% in testing. These results were achieved on a mere 413 training images and 103 testing images.

Classification Report

	precision	recall	f1-score	support
0	0.71	0.65	0.68	34
1	0.00	0.00	0.00	5
2	0.43	0.75	0.55	32
3	0.50	0.26	0.34	19
4	0.40	0.15	0.22	13
accuracy			0.51	103
macro avg	0.41	0.36	0.36	103
weighted avg	0.51	0.51	0.48	103

Fig 15: Accuracy Scores for classification

From the classification report, we can clearly see that class 1 is having 0 F1 scores. This was so because there were only twenty images of that class. Class 0 and class 2 has the highest F1 score as these had the highest number of images; thus, the algorithm could learn the maximum from these classes.

Confusion Matrix

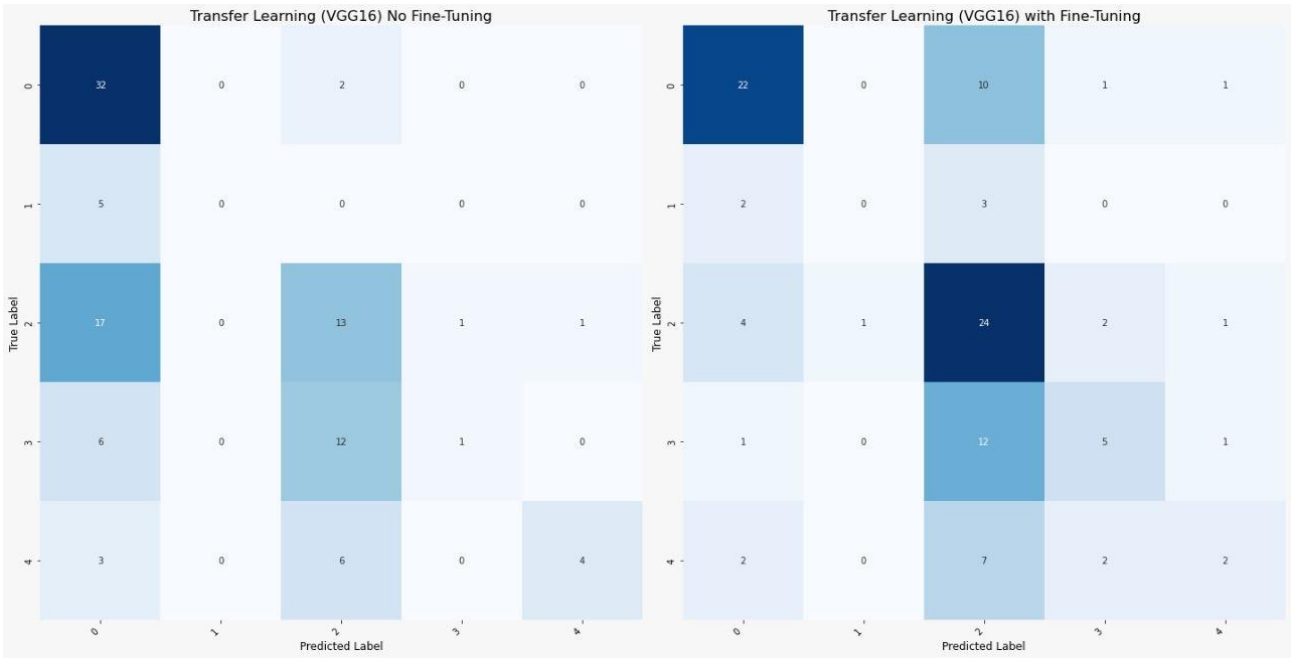

Fig 16: Confusion Matrix for classification (With and Without Fine-Tuning)

From the Confusion Matrix, we can see that the two classes 0 and 2 were better classified than other classes. This can be improved with increase the number of images and thus reducing misclassification between classes.

4.2) LESION SEGMENTATION:

Segmentation is a technique used in digital image processing and analysis to partition an image into multiple parts or regions based on the characteristics of the pixels in the image. The segmentation dataset consists of Eighty-one colour fundus photographs along with ground truths that were separated into a training set and a test set. For images with pixel-level annotations, data was separated as 2/3 for training and 1/3 for testing.

The ground truths were provided for Microaneurysms, Soft Exudates, Hard Exudates and Haemorrhage. The study focuses primarily on the 3 lesions: Soft Exudates, Hard Exudates, and Haemorrhages. Microaneurysms are difficult to locate and have a similar appearance to Haemorrhages.

4.2.1) Annotations (Bounding Box):

Bounding box annotation is one of the most basic techniques that help in calculating of attributes easier for computer vision-based models. The bounding box annotated images are used to train autonomous vehicles for detecting different objects on the streets such as traffic, signals, lanes, potholes, and so on.

In this Case we will use bounding box to locate abnormalities like:

1. Haemorrhages
2. Hard Exudates
3. Soft Exudates

Computer Vision Annotation Tool (CVAT) is a free, open source, web-based image and video annotation tool which is used for labelling data for computer vision algorithms. Originally developed by Intel, CVAT is designed for use by a professional data annotation team, with a user interface optimized for computer vision annotation tasks.

CVAT supports the primary tasks of supervised machine learning: object detection, image classification, and image segmentation. CVAT allows users to annotate data for each of these cases.

While Working with CVAT it is essential to provide 2 inputs:

a) Labels:

Three Labels with colours to be given are constructed and supplied into the system. For this purpose, following code is used in the raw input:

```
[
  {
    "name": "Hemorrhages",
    "id": 63394,
    "color": "#5f09f3",
    "type": "any",
```

```
    "attributes": []
   },
   {
    "name": "Hard_exudates",
    "id": 63395,
    "color": "#ecec0b",
    "type": "any",
    "attributes": []
   },
   {
    "name": "Soft_exudates",
    "id": 63396,
    "color": "#04f304",
    "type": "any",
    "attributes": []
   }
]
```


b) Images to be labelled:

Total Images to be labelled is supplied (81 Images). Below figure refers the supplied image:

Fig 17: Original Images to be labelled

For making the bounding boxes. We took the reference of Ground Truths and then with the help of their locations we have drawn the bounding boxes.

Below figure shows the Ground Truth Images:

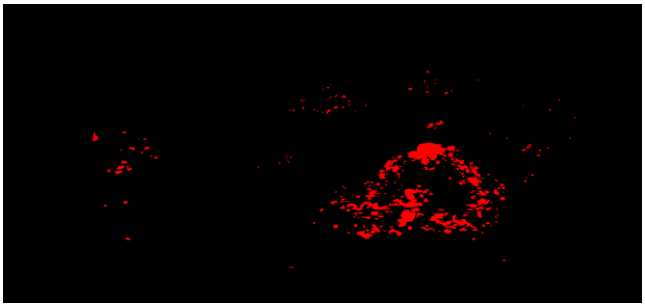

Fig 18: Hard Exudates

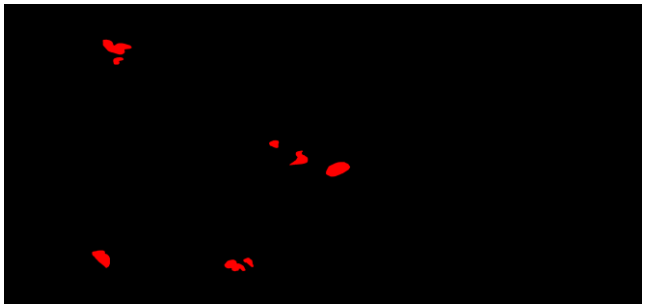

Fig 19: Soft Exudates

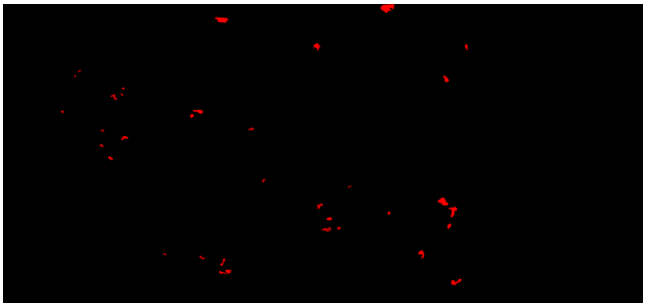

Fig 20: Haemorrhages

The annotations give an output in the form of .xml files which stores the information of the bounding boxes in the image that was annotated.

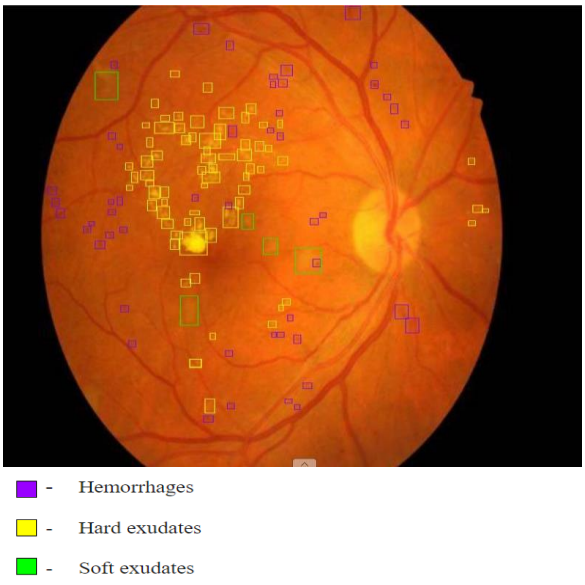

Fig 21: Annotated images specifying lesions

```
<annotation>
  <folder>IDRiD</folder>
  <filename>IDRiD_01.jpg</filename>
  <source>
    <database>Unknown</database>
    <annotation>Unknown</annotation>
    <image>Unknown</image>
  </source>
  <size>
    <width>800</width>
    <height>600</height>
    <depth />
  </size>
  <segmented>0</segmented>
  <object>
    <name>Hemorrhages</name>
    <truncated>0</truncated>
    <occluded>0</occluded>
    <difficult>0</difficult>
    <bndbox>
      <xmin>484</xmin>
      <ymin>171</ymin>
      <xmax>495</xmax>
      <ymax>179</ymax>
    </bndbox>
    <attributes>
      <attribute>
        <name>rotation</name>
        <value>0.0</value>
      </attribute>
    </attributes>
```

Fig 22: Snippet of the .xml file with the dimensions

4.2.2) Single Shot Detection (SSD):

Single Shot Detector takes only one shot to detect multiple objects present in an image using multibox**.** High detection accuracy in SSD is achieved by using multiple boxes or filters with different sizes, and aspect ratio for object detection. It also applies these filters to multiple feature maps from the later stages of a network. This helps perform detection at multiple scales (C. Szegedy et al., 2016).

The underlying architecture of an SSD network can be visualized as:

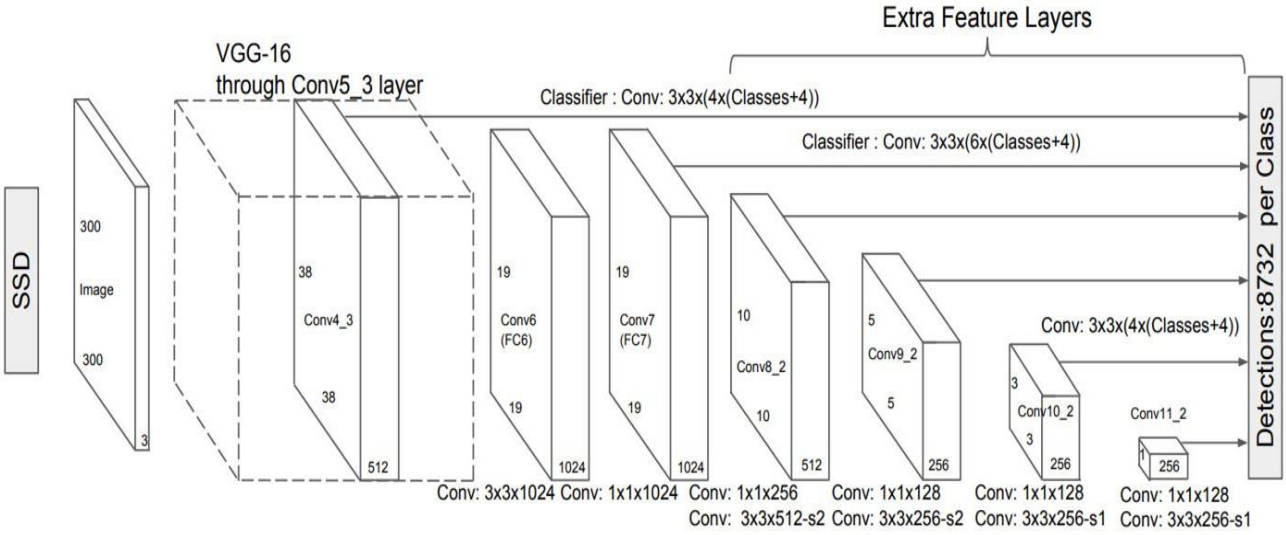

Fig 23: Architecture of Single Shot Detector

1. SSD has a base VGG-16 network followed by multibox convolutional layers.
2. VGG-16 is used for feature extraction.
3. On top of this additional convolutional layers are added
4. Convolutional layers at the end of the base network decrease in size progressively helping detection of objects at multiple scales.
5. Convolutional layer has multiple features with different scale and hence it is able to detect objects in multiple scales better

4.2.3) Training an Object Detection SSD model:

The model takes as input the original images that has already been resized into 800X600 dimension along with their respective .xml files which will have the bounding box annotations for the lesions to be segmented. The steps in training an SSD model include:

a. Preparing the .csv file:

The training and test .xml annotation files need to be converted into two single .csv files that will be passed to the model.

```
<annotation>
  <folder>IDRiD</folder>
  <filename>IDRiD_01.jpg</filename>
  <source>
    <database>Unknown</database>
    <annotation>Unknown</annotation>
    <image>Unknown</image>
  </source>
  <size>
    <width>800</width>
    <height>600</height>
    <depth />
  </size>
  <segmented>0</segmented>
  <object>
    <name>Hemorrhages</name>
    <truncated>0</truncated>
    <occluded>0</occluded>
    <difficult>0</difficult>
    <bndbox>
      <xmin>484</xmin>
      <ymin>171</ymin>
      <xmax>495</xmax>
      <ymax>179</ymax>
    </bndbox>
    <attributes>
      <attribute>
        <name>rotation</name>
        <value>0.0</value>
      </attribute>
    </attributes>
```

	Image_Name	Abnormalities	xmin	xmax	ymin	ymax
0	IDRiD_01.jpg	Hemorrhages	309.67	316.67	145.76	152.76
1	IDRiD_01.jpg	Hemorrhages	234.62	243.62	62.70	75.71
2	IDRiD_01.jpg	Hemorrhages	264.84	272.84	80.72	90.72
3	IDRiD_01.jpg	Hemorrhages	62.50	72.51	297.86	308.87
4	IDRiD_01.jpg	Hemorrhages	120.54	129.55	338.88	348.89
...	...	...	...	...	...	...
2963	IDRiD_54.jpg	Hard_exudates	325.15	331.91	117.40	124.77
2964	IDRiD_54.jpg	Hard_exudates	288.88	299.95	226.80	233.56
2965	IDRiD_54.jpg	Hard_exudates	314.08	326.99	205.91	216.35
2966	IDRiD_54.jpg	Hard_exudates	330.06	338.05	215.13	223.73
2967	IDRiD_54.jpg	Hard_exudates	334.37	342.36	86.66	98.96

2968 rows × 6 columns

Fig 24: .xml files converted and stored in .csv file

b. Downloading the Base Model and Configuring the training pipeline:

The Base model is downloaded from "http://download.tensorflow.org/models/object_detection/". The Pipeline is then configured by tuning the batch size, number of steps and number of classes on the SSD Mobilenet V2 Configuration for MSCQO dataset.

c. Training the model:

The model is trained first on 10 epochs, then on 100 epochs and lastly on 500 epochs to see the increasing accuracy and decreasing loss. The .pb file having the model is then saved and executed on the test data by "Running Inference Test".

4.2.4) Results:

The results show evident difference as the number of epochs are increased from 10 to 500.

A. <u>With 10 epochs</u>:

The loss after 10 steps decreased to 19.0012 which is much higher than the optimal level as we is evidently visible from the accuracy of the bounding boxes.

```
I1021 02:38:48.586054 139903987341184 learning.py:512] global step 1: loss = 27.2370 (20.225 sec/step)
INFO:tensorflow:global step 2: loss = 27.2081 (5.534 sec/step)
I1021 02:38:54.484476 139903987341184 learning.py:512] global step 2: loss = 27.2081 (5.534 sec/step)
INFO:tensorflow:global step 3: loss = 27.2158 (5.426 sec/step)
I1021 02:38:59.911451 139903987341184 learning.py:512] global step 3: loss = 27.2158 (5.426 sec/step)
INFO:tensorflow:global step 4: loss = 24.0889 (5.551 sec/step)
I1021 02:39:05.463985 139903987341184 learning.py:512] global step 4: loss = 24.0889 (5.551 sec/step)
INFO:tensorflow:global step 5: loss = 23.0186 (5.397 sec/step)
I1021 02:39:10.862523 139903987341184 learning.py:512] global step 5: loss = 23.0186 (5.397 sec/step)
INFO:tensorflow:global step 6: loss = 23.2656 (5.422 sec/step)
I1021 02:39:16.285372 139903987341184 learning.py:512] global step 6: loss = 23.2656 (5.422 sec/step)
INFO:tensorflow:global step 7: loss = 19.7853 (7.591 sec/step)
I1021 02:39:23.877090 139903987341184 learning.py:512] global step 7: loss = 19.7853 (7.591 sec/step)
INFO:tensorflow:global step 8: loss = 21.4634 (8.128 sec/step)
I1021 02:39:32.005843 139903987341184 learning.py:512] global step 8: loss = 21.4634 (8.128 sec/step)
INFO:tensorflow:global step 9: loss = 20.7697 (5.443 sec/step)
I1021 02:39:37.449683 139903987341184 learning.py:512] global step 9: loss = 20.7697 (5.443 sec/step)
INFO:tensorflow:global step 10: loss = 19.0012 (5.442 sec/step)
I1021 02:39:42.892317 139903987341184 learning.py:512] global step 10: loss = 19.0012 (5.442 sec/step)
```

Fig 25: Decreasing loss with increasing steps/epochs

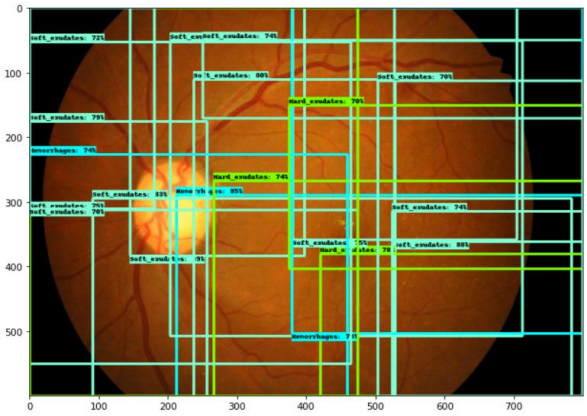

Fig 26: Annotation on the test based on the trained model

The training accuracy is not able to achieve the optimal level, thus failing to annotate the bounding boxes on the test images. The preliminary bounding box that is assigned is recorded as best, thus making overlapping boxes with indistinguishable boundaries.

B. With 100 epochs:

The loss after 100 steps came to be 7.9205 which is less than half of what the loss was achieved with 10 epochs. This gives the motivation to execute the model on test images to check the accuracy.

```
I1021 06:40:49.453769 139725438064512 learning.py:512] global step 95: loss = 7.7443 (7.014 sec/step)
INFO:tensorflow:global step 96: loss = 7.5808 (6.932 sec/step)
I1021 06:40:56.387348 139725438064512 learning.py:512] global step 96: loss = 7.5808 (6.932 sec/step)
INFO:tensorflow:global step 97: loss = 7.3386 (6.940 sec/step)
I1021 06:41:03.328602 139725438064512 learning.py:512] global step 97: loss = 7.3386 (6.940 sec/step)
INFO:tensorflow:global step 98: loss = 7.6263 (6.985 sec/step)
I1021 06:41:10.315210 139725438064512 learning.py:512] global step 98: loss = 7.6263 (6.985 sec/step)
INFO:tensorflow:global step 99: loss = 7.4097 (6.984 sec/step)
I1021 06:41:17.299939 139725438064512 learning.py:512] global step 99: loss = 7.4097 (6.984 sec/step)
INFO:tensorflow:global step 100: loss = 7.9205 (7.132 sec/step)
I1021 06:41:24.432908 139725438064512 learning.py:512] global step 100: loss = 7.9205 (7.132 sec/step)
INFO:tensorflow:Stopping Training.
```

Fig 27: Decreasing loss in the last 5 steps.

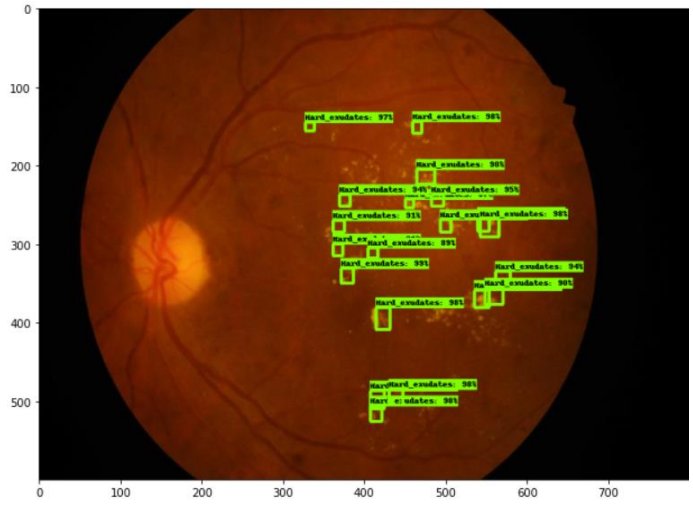

Fig 28: Annotation on test image based on trained model on 100 epochs

The accuracy of the bounding boxed has largely increased. The Hard Exudates being present in the maximum quantity is predicted most accurately. But there is evidence of prominent false negatives present in the annotations and a number of lesions have been missed.

C. With 500 epochs:

The loss after 500 steps is 4.2681, assuring the fact that with increasing epochs, it is possible to achieve better results. This is even less than half of what was achieved with 100 epochs.

```
I1022 22:45:07.027691 139976660690816 learning.py:512] global step 493: loss = 4.2015 (5.046 sec/step)
INFO:tensorflow:global step 494: loss = 5.7560 (5.046 sec/step)
I1022 22:45:12.074838 139976660690816 learning.py:512] global step 494: loss = 5.7560 (5.046 sec/step)
INFO:tensorflow:global step 495: loss = 4.8916 (5.076 sec/step)
I1022 22:45:17.151779 139976660690816 learning.py:512] global step 495: loss = 4.8916 (5.076 sec/step)
INFO:tensorflow:global step 496: loss = 4.0674 (5.020 sec/step)
I1022 22:45:22.173178 139976660690816 learning.py:512] global step 496: loss = 4.0674 (5.020 sec/step)
INFO:tensorflow:global step 497: loss = 3.7415 (4.995 sec/step)
I1022 22:45:27.169133 139976660690816 learning.py:512] global step 497: loss = 3.7415 (4.995 sec/step)
INFO:tensorflow:global step 498: loss = 3.8968 (4.998 sec/step)
I1022 22:45:32.167974 139976660690816 learning.py:512] global step 498: loss = 3.8968 (4.998 sec/step)
INFO:tensorflow:global step 499: loss = 4.3051 (4.984 sec/step)
I1022 22:45:37.153080 139976660690816 learning.py:512] global step 499: loss = 4.3051 (4.984 sec/step)
INFO:tensorflow:global step 500: loss = 4.2681 (5.077 sec/step)
I1022 22:45:42.231202 139976660690816 learning.py:512] global step 500: loss = 4.2681 (5.077 sec/step)
INFO:tensorflow:Stopping Training.
```

Fig 29: Decreasing loss from 493rd step to 500th step

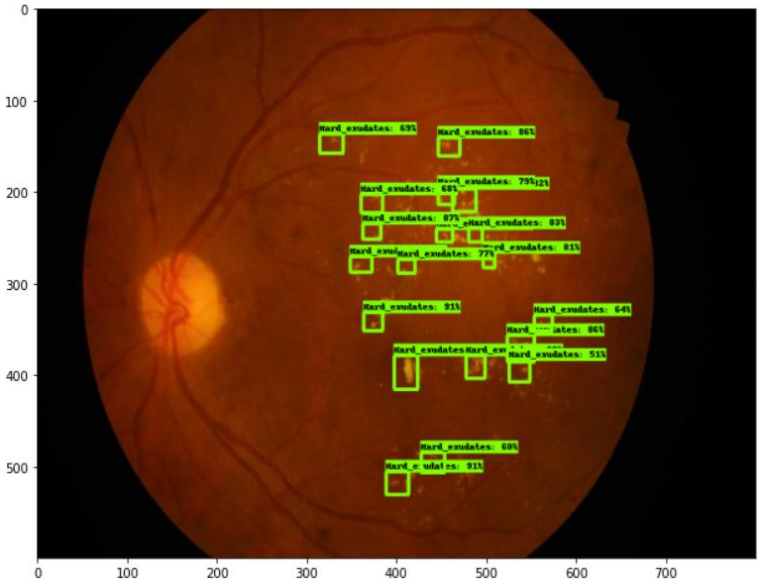

Fig 30: Annotation on test image based on trained model on 500 epochs

The bounding boxes have almost accurately covered all the objects in the image. Hard Exudates being dominantly present makes the data imbalanced resulting in less accurate annotation of Haemorrhages and Soft Exudates. The amount of false negatives have also largely decreased, increasing the accuracy of the annotations on the test images.

CHAPTER 5: CONCLUSION, LIMITATIONS AND SCOPE

5.1) CONCLUSION AND DISCUSSIONS:

Diabetic Macular Edema (DME) is an advanced symptom of diabetic retinopathy and can lead to irreversible vision loss. In this study, a computerized assist system for automated macular edema grading is presented. The CNN model with pre-trained VGG-16 framework has been able to achieve an accuracy of 51.5% on the training dataset and 44% accuracy on the testing data. The model detects Diabetic Retinopathy and also gives information regarding the severity of the disease. Sinthanayothin et al., (2003) reported sensitivity of 80.21% and specificity of 70.66% while differentiating diabetic retinopathy from normal images. Similar results have been reported by Larsen et al., (2003), where haemorrhages and microaneurysms were detected to diagnose diabetes. Their method had 71.4% specificity and 96.7% sensitivity in detecting diabetic retinopathy. In both cases the dataset available was more than 3000 images. The present work has been able to achieve above 50% accuracy with less than 500 images at hand.

This model can be helpful in assisting doctors for a faster diagnosis of diabetic retinopathy disease. Classification assists doctors to find the grade of disease, on the other hand segmentation can further give clarity of the diagnosis.

The loss was brought down to less than 5 as the number of epochs in the Single Shot Detector model was increased from 10 to 500. The loss initially was 19.0012. Further processing and tuning of the SSD configuration and execution on more epochs will further increase the accuracy. The SSD model detected Hard Exudates with more than 85% accuracy for most of the spots in the original image. The presence of false negatives decreased hugely and more lesion spots were annotated accurately. The weights generated through training on the present data can be deployed as a pre-trained model on larger datasets, saving time for manual annotations and helping doctors and medical practitioners to detect present of lesions more accurately and in shorter period of time facilitating better diagnosis at the early stage.

Improvement on the base model and deployment of similar models will be efficient in detecting and diagnosis of other retinal lesions through fundus images at hand. Diabetic Retinopathy is one of the most common and avoidable retinal abnormality, which can be cured if diagnosed and treated at an early stage.

Segmentation of lesions and grading them according to the severity of DR posed the primary motivation of the present work which has been achieved overcoming number of limitations, which if solved can have significant contribution in the health and medical industry.

5.2) LIMITATIONS:

The present work faced a number of limitations and hurdles in trying to achieve its goal. The dataset at hand had less 500 images as data points for disease grading and only 81 images to build a segmentation model. The data was highly imbalanced with disease class 0 and 2 predominantly present while other classes were less than half in number rendering the disease grade classification model not able to predict the classes 1, 3 and 4 accurately.

Hard Exudate data points were almost 3 times the number Soft Exudates, posing the similar problem for the SSD model in segmentation. Computational constraints posed an unavoidable hurdle. The computational strength required to train the images for 5000 epochs, which according to literature is an adequate number to get the best accuracy, was not available deeming it impossible to move above 500 epochs to train the model.

5.3) SCOPE OF FUTURE STUDY:

With availability of upgraded computational facility, U-Net and Faster RCNN can be employed to study the improved accuracy of the segmentation annotations. Localization of the lesions can further facilitate accurate and faster diagnosis of the severity of Diabetic Retinopathy. Removing the imbalance in the data will make classification as well as segmentation of all the classes accurate and help in getting the best architecture.

Medical industry requires models with nearly 100% accuracy to be deployed. With the limitations resolved, it will be possible to work towards achieving that goal. Improvisation and selection of the best model can identify the patient with diabetic retinopathy at their earlier stage using the classification model.

REFERENCES

- Atlas, I. D. F. D., 2017, Brussels, Belgium: International diabetes federation Int Diabet Federat (IDF) http://diabetesatlas.org/resources/2017-atlas.html
- B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," IEEE transactions on biomedical engineering, vol. 59, no. 6, pp. 1720–1726, 2012.
- Chudzik P, Al-diri B, Caliv F, Ometto G, Hunter A. Exudates segmentation using fully convolutional neural network and auxiliary codebook. 2018 40th Annu Int Conf IEEE Eng Biol Soc; 2018.
- E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 4, pp. 640–651, 2017.
- Fleming, A., S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," IEEE Transactions on Medical Imaging, vol. 25, no. 9, pp. 1223–1232, 2006.
- G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: multi-path refinement networks for high-resolution semantic segmentation," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1925–1934, Honolulu, HI, USA, 2017.
- He K, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
- https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid
- https://www.geeksforgeeks.org/vgg-16-cnn-model/
- https://github.com/debayanmitra1993-data/Blindness-Detection-Diabetic-Retinopathy-/blob/master/1_eda.ipynb
- https://www.academia.edu/42908775/Literature_Review_on_the_Diabetic_Retinopathy_in_Retinal_Images
- https://towardsdatascience.com/faster-r-cnn-for-object-detection-a-technical-summary-474c5b857b46
- https://www.dlology.com/blog/how-to-train-an-object-detection-model-easy-for-free/#disqus_thread
- https://blog.roboflow.com/training-a-tensorflow-faster-r-cnn-object-detection-model-on-

your-own-dataset/

- https://towardsdatascience.com/ssd-single-shot-detector-for-object-detection-using-multibox-1818603644ca
- https://developers.arcgis.com/python/guide/how-ssd-works/#:~:text=SSD%20has%20two%20components%3A%20a,classification%20layer%20has%20been%20removed.
- J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440, Boston, MA, USA, 2015.
- Joshi S, Karule PT. A review on exudates detection methods for diabetic retinopathy. Biomed Pharmacother 2017; 97:1454–60.
- Kaur J, Mittal D. Segmentation and measurement of exudates in fundus images of the retina for detection of retinal disease. J Biomed Eng Med Imaging 2015;2(1):27–38.
- Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012; 25: 1097-105.
- Lam C., et al. Automated detection of diabetic retinopathy using deep learning. AMIA summits on translational science proceedings. 2018; 2018: 147.
- Lian S, Li L, Lian G, Xiao X, Luo Z, Li S. A global and local enhanced residual U-Net for accurate retinal vessel segmentation. IEEE/ACM Trans Comput Biol Bioinforma 2015;14:1–10.
- Litjens G., et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017; 42: 60-88.
- Liu, W., Anguelov, D., Erhan, D., SSD: Sigle Shot MultiBox Detector, arXiv:1512.02325v5 [cs.CV] 29 Dec 2016
- Mookiah M.R.K., et al. Computer-aided diagnosis of diabetic retinopathy: a review. Comput Biol Med. 2013; 43(12): 2136-55.
- N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified u-net architecture for semantic segmentation of diabetic retinopathy images," Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1094–1109, 2020.
- Philip S., et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. Br J Ophthalmol. 2007; 91(11): 1512-7.
- Pratt H., et al. Convolutional neural networks for diabetic retinopathy. Procedia Comp Sci. 2016; 90: 200-5.
- Rehman M.U., et al. Classification of diabetic retinopathy images based on customised CNN

architecture. 2019 Amity International Conference on Artificial Intelligence (AICAI). IEEE; 2019.

- Reichel, E., Salz, D., 2015, Diabetic retinopathy screening In Managing Diabetic Eye Disease in Clinical Practice, Springer, pp.25–38.
- Russakovsky O., et al. Imagenet large scale visual recognition challenge. Int J Comput Vis. 2015; 115(3): 211-52.
- S. Kavitha and K. Duraiswamy, "Automatic detection of hard soft exudates in fundus images using color histogram thresholding," European Journal of Scientific Research, vol. 48, no. 3, pp. 493–504, 2011.
- S. Wang, X. Wang, Y. Hu et al., "Diabetic Retinopathy Diagnosis Using Multichannel Generative Adversarial Network With Semi-supervision", IEEE Transactions on Automation Science and Engineering, pp. 1–12, 2020.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. University of Oxford and Google DeepMind. arXiv preprint arXiv:1409.1556; 2014.
- Szegedy C, et al. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
- Szegedy C, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
- V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.
- Wang C, Zhao Z, Ren Q, Xu Y, Yu Y. Dense U-net based on patch-based learning for retinal vessel segmentation. Entropy 2019;21:1–15.
- WHO, 2011. http://www.who.int/mediacentre/factsheets/fs312/en/
- Xu X, Wang R, Tan T, Xu F. An improved U-net architecture for simultaneous arteriole and venule segmentation in fundus image. Med Image Underst Anal 2018;1–10.
- Zoph B, et al. Learning transferable architectures for scalable image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.