CHAPTER NO. 1

INTRODUCTION

1.1 Background:

Retinopathy is the formation of various types of lesions in the eye which is usually found in middle aged and old age group of people. While diabetes may be a very common cause of developing retinopathy in patients, medical practitioners have also often found retinopathy condition in people with absence of hyperglycemia. Diabetic retinopathy eventually leads to loss of complete vision but other causes of retinopathy are less severe, causing mild vision impairments.

1.2 Importance of Project:

Previous or existing systems of retinopathy (most commonly diabetic retinopathy or DR) are based on one or two features of the affected eye [1] [9]. Point of concern is using some features to detect retinopathy won't showcase accurate results regarding the disease. Using different features will help in getting more accurate results. In this project four features are utilized for detection. All the features are extracted from images using different image processing methods. Different methodologies gives accurate results in extraction [3]. All these factors help this project to build a system to improve the accuracy of the results and make this automated systems more reliable.

1.3 Motivation:

It has been observed that it is somewhat difficult to identify retinopathy in patients without diabetes. Diabetic patients who have been suggested to test for retinopathy found it tedious and costly to undergo diagnosis. Also, not many doctors are fully informed about the occurrence of retinopathy in non-diabetic patients. It may happen due to ocular and other systematic reasons. In order to control the effect of retinopathy in the vision, it should be treated at early stage [2]. To detect retinopathy at early stage we propose detection of retinopathy using machine learning.

1.4 Scope:

Retinopathy is developed due to carotid disease, severe anaemia, etc and diabetic retinopathy is due to diabetes mellitus, there are chances of more than 35 percent of diabetic patients get affected by DR [4]. To eradicate that at early stage, early detection of the condition is required. For early stage detection main features like blood vessels, micro aneurysms, hemorrhages, exudates are checked for. So, these features are extracted for early stage detection. Algorithms are trained in such a way that it can predict the chances of getting affected by retinopathy in eye using the above symptoms as exudates.

1.5 Objectives:

The main objective is to detect retinopathy universally in diabetic as well as non-diabetic patients using different training algorithms. After training them results are to be compared. Results will tell the effectiveness of algorithms. Algorithms are to be trained with datasets. Datasets and training algorithms are only to get you better results.

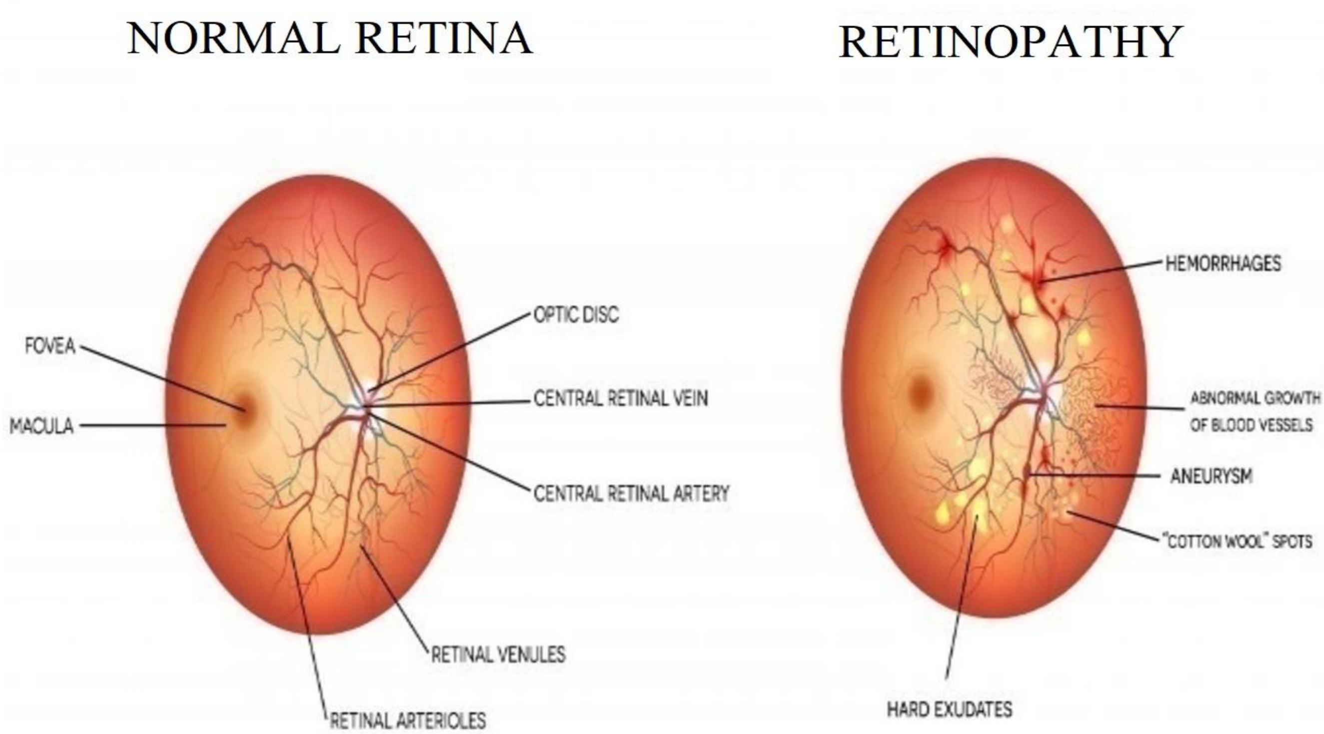

Fig 1: Comparison between healthy and affected eye [8].

CHAPTER NO. 2

LITERATURE SURVEY

2) Literature Survey:

2.1)"Automatic Detection of Diabetic Retinopathy in Large Scale Retinal images", M. Kamaladevi, S. SnehaRupa, T. Sowmya School of computing, SASTRA Deemed to be University. [1]

The given method has following phases: Pre-processing of retinal images, Feature extraction and classification for detecting DR. The pre-processing step is used to avoid or reduce the unnecessary features present in images and to enhance them for further processing. As a part of first phase image undergoes gray-scale conversion, CLAHE, Reduction of noise, Edge detection and binary conversion. Values of extracted features are used for training. A framework is developed for detecting DR automatically. Accuracy of provided model is 97 percent when Random Forest is used. The identification label 1 and label 0 represents whether the patient has DR or note. The model is useful for doctors to help affected patient and avoid vision loss.

2.2)"Diagnosis of Diabetic Retinopathy using Machine Learning", Swati Gupta and Karandikar AM, Department of Computer Science and Engineering, Shri Ramdeobaba College Of Engineering and Management, Nagpur [2].

In this paper, data set of 94 images are used to evaluate the DR detecting system which are fundus images. All the images are stored in same format which is JPG and were resized into 1500 X 1152 intensities. Pre-processing technique enhances the image and hue saturation intensity is used for restoration. Histogram equalization is used for contrast enhancement which follows pixel intensity range 0 and 1.Features like optic disc, blood vessels, micro aneurysms and exudates are identified for feature extraction. As part of feature extraction structural features are also extracted. SVM classifiers are used to classify. Classifiers are "one-against-one" and "one-against-rest". 70-30 training and testing ratio for classification and proposed system detected DR with high sensitivity of 87% and specificity of 100% respectively.

2.3)"Detection of Hemorrhages in Retinal Images using Hybrid Approach for Diagnosis of Diabetic Retinopathy", S.Karkuzhali Assistant Professor Kalasalingam University and Dr. D. Manimegalai National Engineering College, Kovilpatti, Tamilnadu [3].

In this paper, the following method for detection of hemorrhages is used:

The input retinal image is preprocessed with median filter. Kmeans is an iterative procedure and this algorithm clusters the data iteratively by computing a mean intensity for each group. Input data and cluster number are provided. Kmeans algorithm is worked on centroid and distance from centroid. Each pixel distance is measured and cluster is assigned to image.

The k-means algorithm resultant five clusters among them one returns image with high intensity HAs and it is taken for further processing. The morphological operations, the segmented pixels are dilated using structuring elements to acquire the other smaller left out details in the cluster. Dilation makes object to expand and then filling in small holes and connecting disjoint object.

2.4)"Feature Extraction of Diabetic Retinopathy Images", ShradhaMirajkar, M. M. Patil SVERI'S College of Engineering, Pandharpur [4].

Blood Vessel feature extraction according to the paper:

Blood vessels of different thicknesses have been extracted using the well-known kirsch template. After image selection from database, the images are converted into grayscale from RGB. Then, adaptive histogram equalization is applied as it increases the Global Contrast of the images and it acts as a good tool for the enhancement of the edges. A threshold value of the pixels is fixed and then given to kirsch templates for feature extraction. This method has spatial filtering using different orientations. Image segmentation is done in three steps, in first one image is converted into gray scale, in second step edge detection mechanism (krisch) is used and finally filtered gray scale image's value is verified whether it is greater than the given threshold.

CHAPTER NO. 3

RELATED THEORY

AND

PROBLEM DEFINITION

3.1) Related Theory:

The ocular and systematic causes of retinopathy in patients without diabetes are highly diverse. Main symptoms include micro aneurysms, hemorrhages in retina, hard and soft exudates, cotton wool spots, abnormalities of the retinal veins, intraretinal micro vascular abnormalities and formation of new vessels [3]. Diabetic retinopathy is a long-standing issue of diabetes, due to minimum symptoms it is hard to detect. Now a days OCT (Optical Coherence Tomography) is used for detection but the equipment used is expensive [5]. It will be helpful for both patients and ophthalmologists if a system is designed which can deliver accurate results on the data provided. To design a system as proposed, some of the machine learning algorithms are used to choose which delivers the accurate results as per the requirement. We chose SVM (Support Vector Machine), Decision Tree, Random Forest, CNN (Convolutional Neural Networks). Datasets are used to train and test the algorithms. Data is fed in different ratio of training data and testing data (For example: 80% train and 20 % test; similarly 60%, 40%). Kaggle dataset is used, for now algorithms are trained and tested on 2000 images of fundus. More the data fed to algorithm, more the accuracy. Quantity of classified is directly proportional to the accuracy of the system.

3.2) Problem Definition:

India is said to have the maximum number of diabetic patients in the world. In the year 2000, the total number of diabetes patients were around 31.7mn. It was estimated that 3.4mn deaths were caused due to high blood sugar levels. The number of diabetes patients is expected to rise upto 79.4mn by the year 2030, i.e. by more than 100%.

Along with the above issue, many doctors are rarely aware of the retinopathy developed due to ocular causes like retinal vein occlusion, retinal macro aneurysm and systematic causes like hypertension, atherosclerosis, vasculitis, maxillary or cranial radiations, etc.

Detection of retinopathy is a monotonous and a painful job as it is time consuming, expensive process and requires experienced medical professionals. Studying Indian scenario for the effectiveness of existing machine learning methods for image based retinopathy and diabetic retinopathy detection and identify gaps for improving such techniques. Thus is becomes crucial to channelize efforts in leveraging machine learning techniques for automated detection and screening of such patients. The need for an exhaustive and automated method of retinopathy identification has been realized using image classification, pattern recognition, and machine learning with increased amount of efforts to register progress. With color fundus photograph as input, the goal of this project is to extend the abilities of an automated detection system.

Problem Statement:

Detecting retinopathy using machine learning with the help of features like blood vessels, exudates and hemorrhages of eye. This system will help doctors to evaluate and treat patients from early stage of retinopathy.

CHAPTER NO. 4

DESIGN METHODOLOGY

4.1) Preprocessing of Retinal Image

a) RGB to Gray image conversion: Color image is fed to the system as input but for better understanding of the machine it is converted into gray scale image which enhances the brightness. Later the image is resized.

b) Contrast adjustment: The equalization Contrast-limited Adaptive Histogram Equalization (CLAHE) is used to adjust the contrast level of the gray scale image. This functions on small parts of the image called tiles.

c) Binary threshold: This procedure is used to convert the gray scale into binary image. This converted binary image is black and white whose pixels are 0 and 1 respectively. Threshold value is chosen for the training.

d): Morphological operation: Mathematical Morphology for Noise Reduction in images is used to calculate and compute the values of thresholding. Morphological open is used to eliminate the blood vessels and smoothens the inner vessels in the optic disc. Opening operation is used to remove noise.

4.2) Working of the model:

First of all, the respective doctor will upload a color image of fundus on the web page. Secondly, the uploaded image will be initially processed to obtained valuable information from it. The input image will be processed many times to extract the desired features- blood vessels, exudates, hemorrhages and some structural features like homogeneity, contrast, energy and entropy.

Each and every image in the dataset will undergo image processing similar to the steps mentioned above and entry of data of each image is stored in one CSV file. This CSV is used to trained the model using different algorithm and the progess is saved.

On this saved model, the image fed by doctor will be distinguished as retinopathy eye or healthy eye.

4.3) Proposed System Architecture:

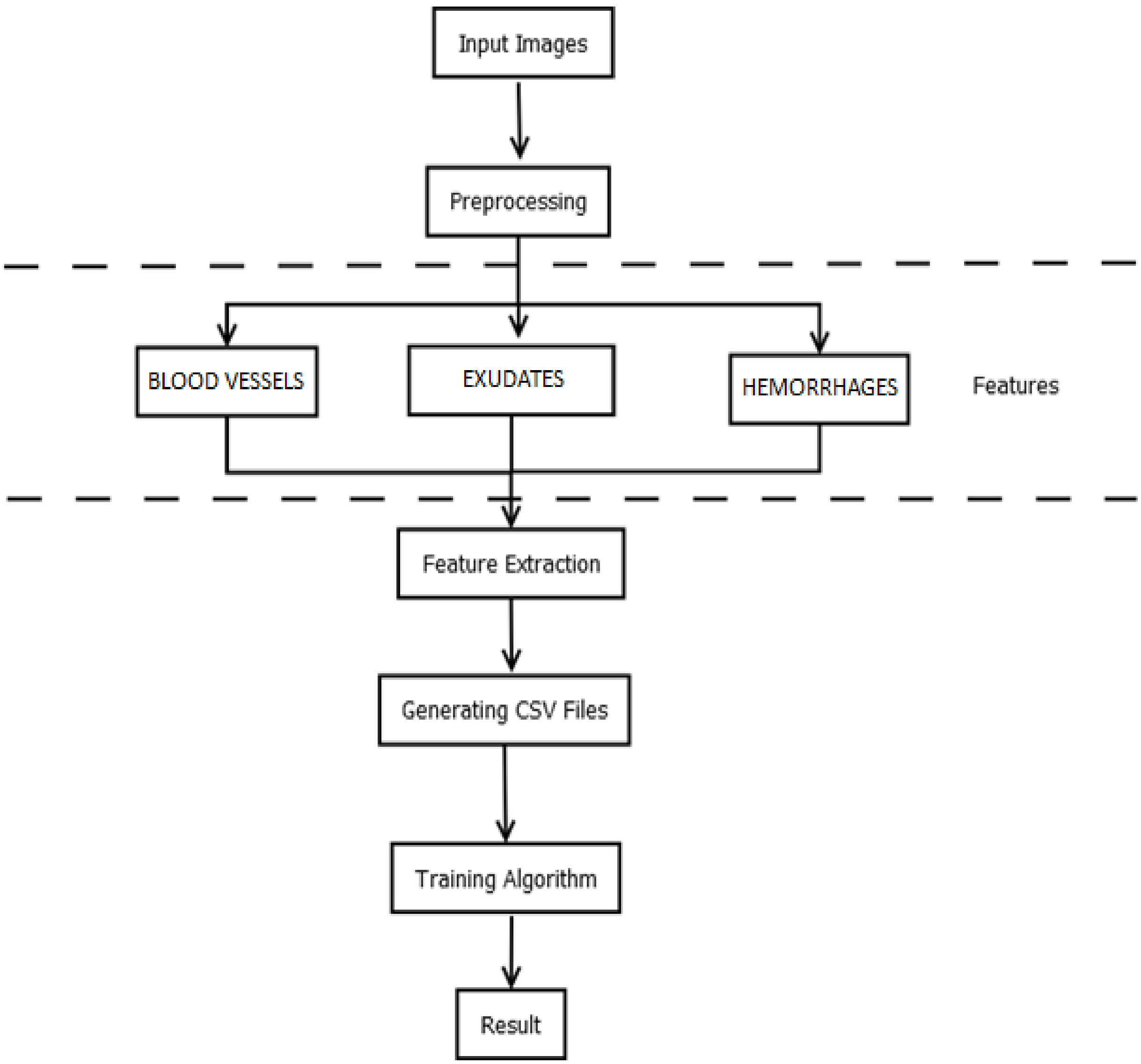

Fig 2: Architecture for detecting retinopathy.

4.3) UML Diagrams

4.3.1) Use case Diagram:

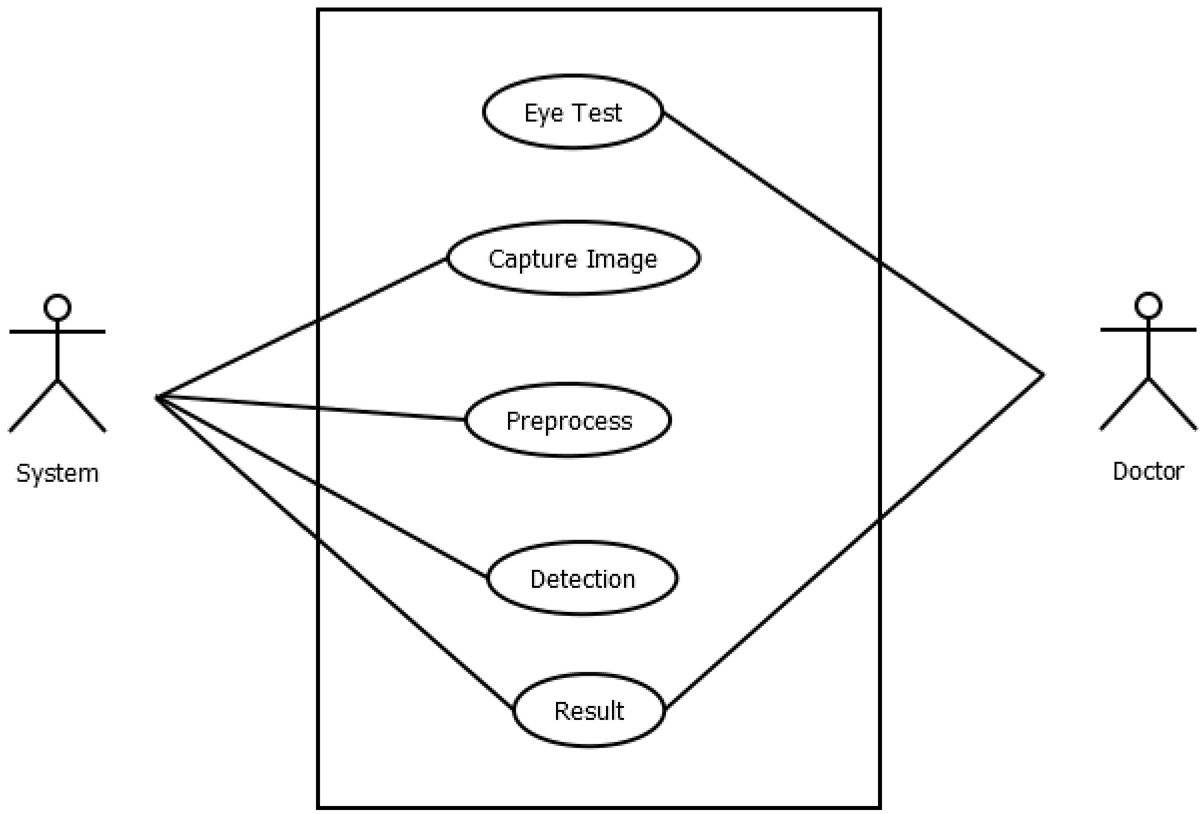

Fig 3: Use case diagram for retinopathy detection.

4.3.2) Activity Diagram:

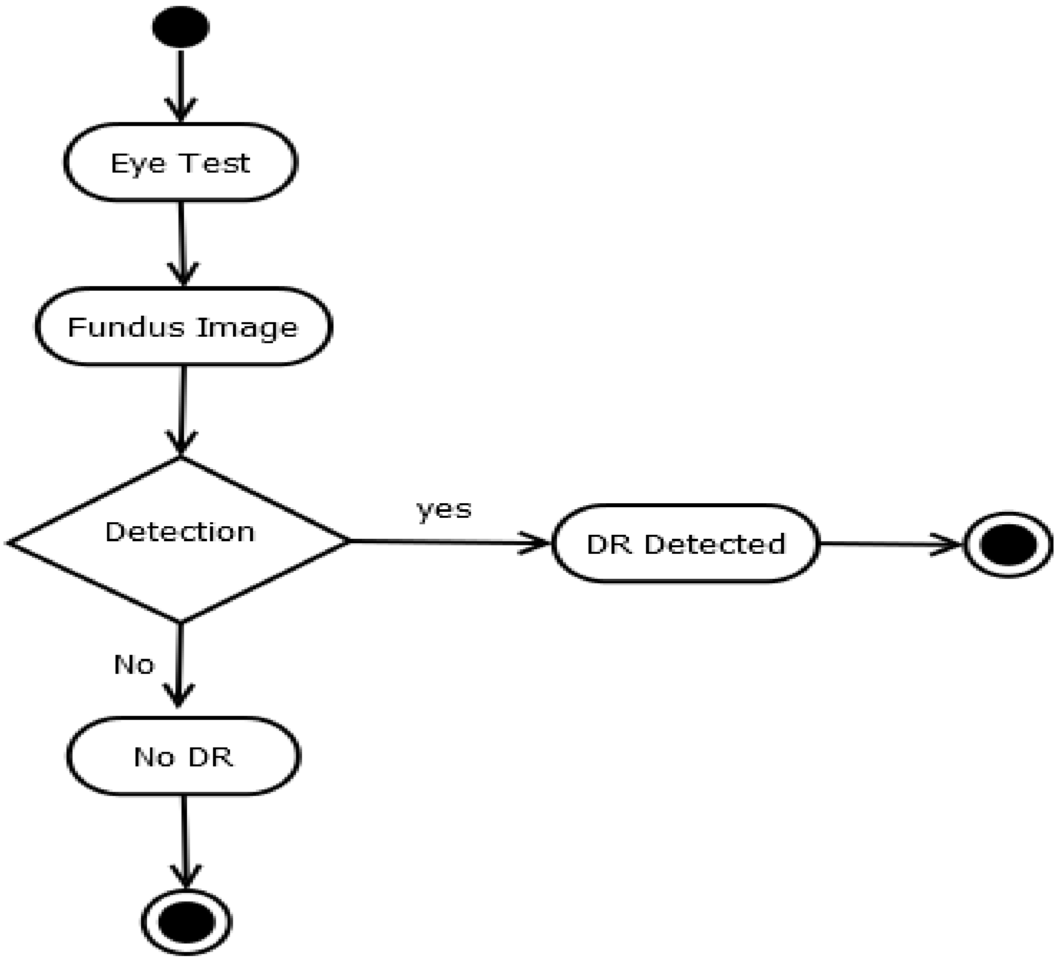

Fig 4: Activity diagram for retinopathy detection.

4.3.3) Sequence Diagram:

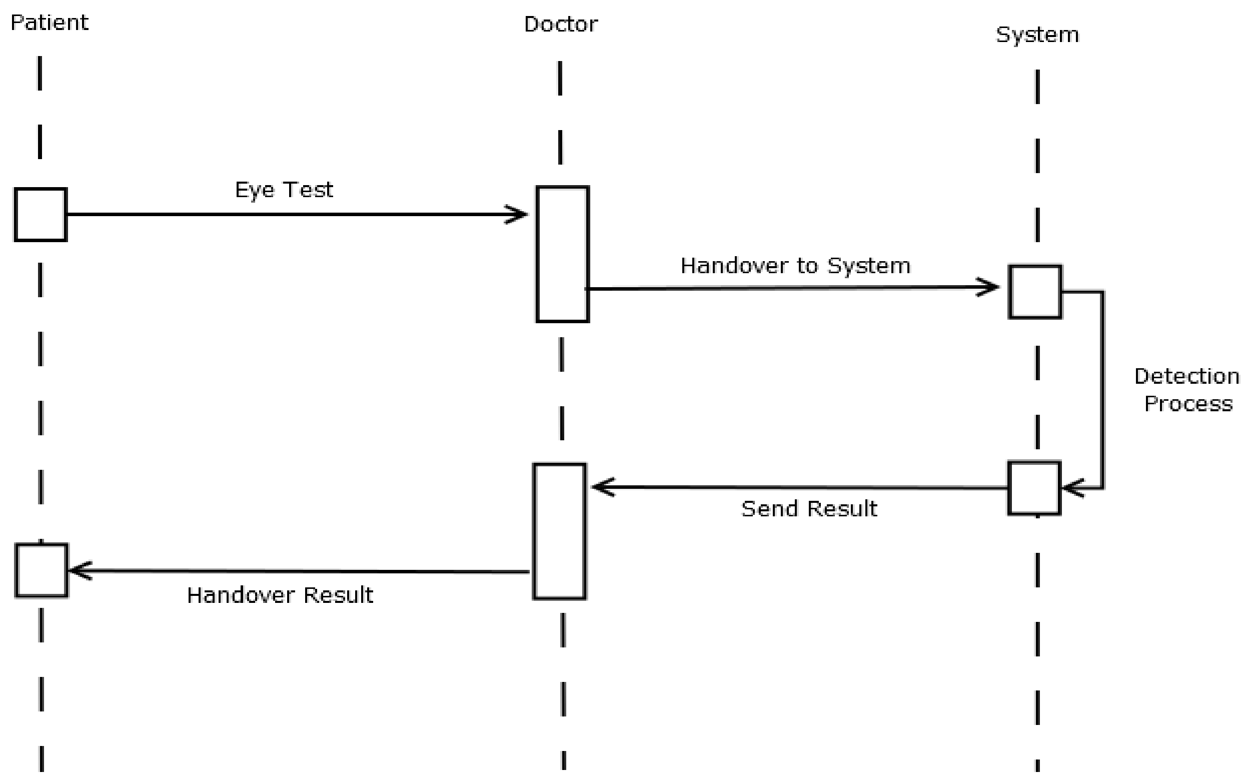

Fig 5: Sequence diagram for retinopathy detection.

4.3.4) Collaboration Diagram:

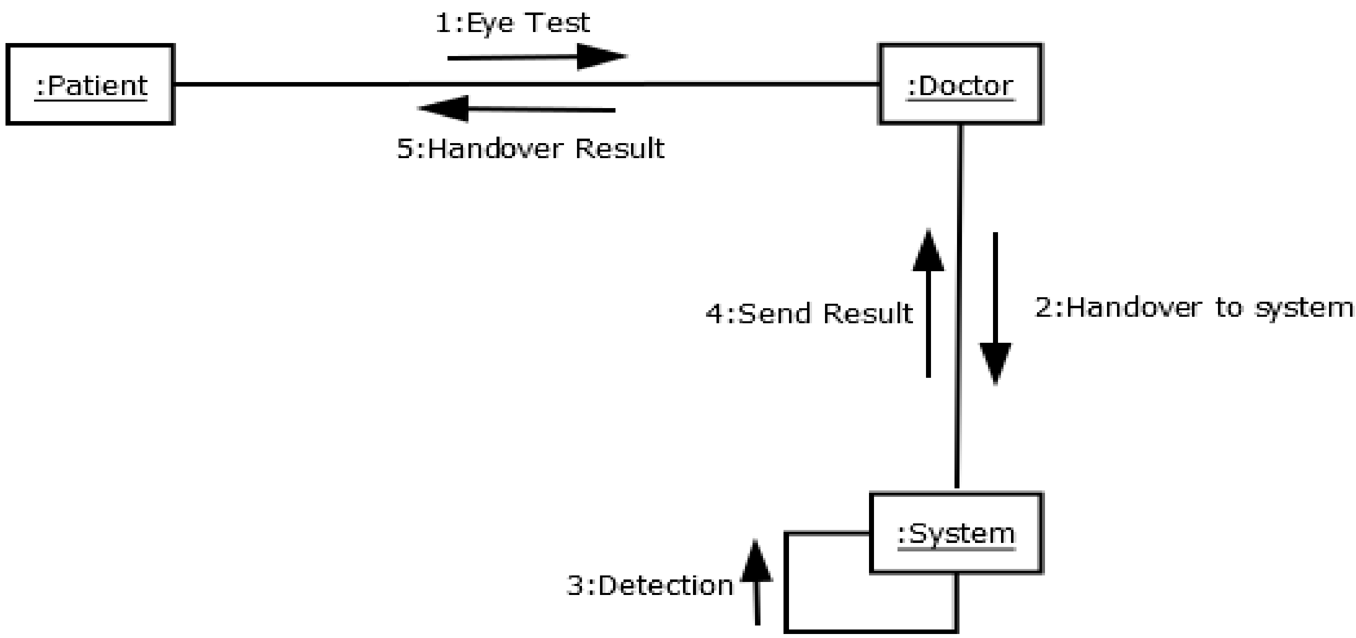

Fig 6: Collaboration diagram for retinopathy detection.

4.4) Technical Specifications:

4.4.1) Software:

Python

Python is programming language which is used for implementation.

Pycharm

Pycharm [5] is ide which is used for development of system. Pycharm is freely available tool which helps for package managing easily because without pycharm. Otherwise command line interface and commands should use for installing package. Pycharm make it easier.

Version-2018.2.5(community Edition)

Build #PC-182.5107.22 build on November 13, 2018

Open CV library

Open computer vision library [7] is python open source library which is available for real-time computer vision. This library helped us in a lot of operations. most of the operation is done by using open CV. Form reading image to every feature extraction is done by using different functions of open CV library. It is Wrapper package for OpenCV python bindings.

Version-4.0.0.21

NumPy Library

NumPy [7] is basic package for array computing with Python. Logical and numerical operations on array can be performed using NumPy.

Version-1.16.2

Pandas Library

It is an open source python library [7]. It provides data structure and data analysis tools for python programming. It provided powerful data structure for data analysis, time series and statistics.

Version-0.24.2

Scikit-learn library

Python library [7] for machine learning which provide different classification, clustering and regression algorithms such as SVM, K-means, Random forest, DBSCAN and gradient boosting.

Version-1.0

4.4.2) Hardware:

Processor Brand: Intel

Name: Intel(R) I7 (seventh generation)

Variant: 7700 HQ

Chipset: Intel HM175

Number of Cores: 4

RAM: 8 GB (DDR4)

Cache: 6 MB

Clock Speed: 2.8 GHz-3.8 GHz

Refresh Rate: 144 MHz

GPU-0:

Processor Brand: Intel

Name: Intel

Variant: Intel(R) HD 630

GPU-1:

Processor Brand: NVIDIA

Name: GeFORCE

Variant: GTX 1050Ti

Memory: 4 GB

Memory type: DDR5

CHAPTER 5

IMPLEMENTATION

5) Implementation:

Whole implementation of this project is carried out in 4 different phases- 1) Pre-Processing 2) Feature Extraction [7] [9] 3) Storing the values 4) Training Algorithm. In the first phase RGB image is taken as input, as the taken image is a normal image, it has to be resized. Image is resized into 512* 512 size. Processing depends on what kind of feature is needed to be extracted. There are 3 type of features to be extracted. First one is blood vessels, second one is exudates and third one is hemorrhages. After pre-processing on RGB image, each processed image gives one feature separately which means 3 type of image pre-processing on every image gives a value. This procedure is called feature extraction. Pre-processing is done on around 405 images [7] in which 200 images are not effected and 205 are affected and data is saved in one comma separate value type file. Machine learning algorithm is used to train and test on the collected data. Algorithm will be ready to predict result once the precision and accuracy are drawn by itself, then it can classify whether the given image is affected with retinopathy or not.

Dataset:

Dataset which was used for training and testing algorithm is available at one repository of GitHub made available by group called Nomik [11]. This dataset was divided into 2 classes- symptoms and no symptoms. The symptoms folder consisted of 596 images and no symptoms consisted of 1468 images. We have to perform supervised classification for which this dataset is appropriate to use. This dataset is actually belongs to Kaggle [12] organization.

On Kaggle, which is an online community of users to find and publish data sets, explore and build models in a web-based data-science environment. We found 2 types zip files, one for training and other one for testing, which further contained 7 and 5 folders respectively. Each of the folder has images accounting to 7 GB of memory. These images are not classified so we have to use curated data provided by Nomik group.

After training and testing of algorithms and development of final system some data is used for real time testing which was acquired from Krishna Medical College, Karad. They provided us with 84 images of affected eye and 30 images of not affected eye.

For implementation we have used 80-20 percent image split in which both affected and not affected images are taken in equal counts. For training purpose we had 953 images of which half were of affected eye and other half of not affected eye.

Remaining 293 images were used for testing purpose by similarly splitting half images of affected eye and other half of not affected eye.

5.1) Image Pre-Processing:

Open CV library

This library helped us for lot of operations. Most of operation are done by using open CV[8]. Form reading image to every feature extraction is done by using different functions of open CV library

cv2.imread()

This is the most basic function which is used for reading actually RGB image.

img = cv2.imread(file)

cv2.resize()

This is second function of open CV which is used for resize the image size. In available dataset images don't have same size some of 1000*2000 or 1345*1200. There is irregularities in size of image that's why resize function is used for making every image of same size and so that work load of machine also reduce while processing.

img = cv2.resize(img, (512, 512), interpolation=cv2.INTER_CUBIC)

cv2.split()

Split function is used to split image into separate Red, Green, Blue channels. For different feature extraction we need different channels. For blood vessels, micro aneurysms we need green channel and for hemorrhage only red channel, that can be achieved by using split function.

b, g, r = cv2.split(img).

Contrast limited adaptive histogram equalization (CLAHE)

We just can't use global histogram technique because it will improve contrast of background of an image but that technique make image very bright so we could lost important area of an image (rog). CLAHE histogram works simply by just divide and concern method in which 8*8 grid is moved over image and if pixel value is above default value that is 40 then pixel is uniformly distributed in grid.

clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8, 8))

claheImg = clahe.apply(g).

Canny edge detection

Canny is the most popular function which is used for detection of edges in an image. It is a multistage process which include following parts

- Noise reduction

Canny edge detection is vulnerable to noise, so noise removing is done in initially by using 5*5 Gaussian filter.

- Finding Intensity Gradient of the Image

Repeat image filter with sobelkernal in both horizontal and vertical directions to get horizontal derivative first and next vertical derivative. From these two images, we can find edge gradient and direction for each pixel as follows:

Edge_Gradient(G)=G2x+G2y−−−−−−−√Angle(θ)=tan−1(GyGx)

Gradient direction is always perpendicular to edges. It is rounded to one of four angles representing vertical, horizontal and two diagonal directions.

Non maximum suppression:

In about step function get gradient magnitude and direction so that unwanted pixel are removed from an image using full scan of an image.

Hysteresis Threshold:

This is last step of canny edge detection in which one min value and max value is fixed that are threshold values. If pixel value is above max value this is sure to be an edge and those pixel intensity is below min value it is discarded from image.

Morphological Transformations [9]

i. **Erosion**

Basic idea of Erosion is decreasing white spaces. First kernel of pixel value 1 is moved along image. Kernel is with Full match, Half match or No match. If it is Full or Half match pixel value of kernel is made 1 if it is No match then pixel value is made to 0 (i.e eroded) After evaluation all the pixels which are near the boundary are discarded depends on size of kernel.

ii. **Dilation**

Dilation is opposite of Erosion. It increases white spaces. Pixel value is 1 if at least one pixel under kernel is 1. In noise reduction Erosion followed by Dilation is used. Erosion eradicates white noises but it also shrink the object. Hence we dilate it .as noise is gone it won't be back and object area is also increases.

iii. **Opening**

Opening is Erosion followed by Dilation. Used in noise reduction cases.

iv. **Closing**

It is reverse of opening. i.e Dilation followed by Erosion. It is useful to remove small black points on object.

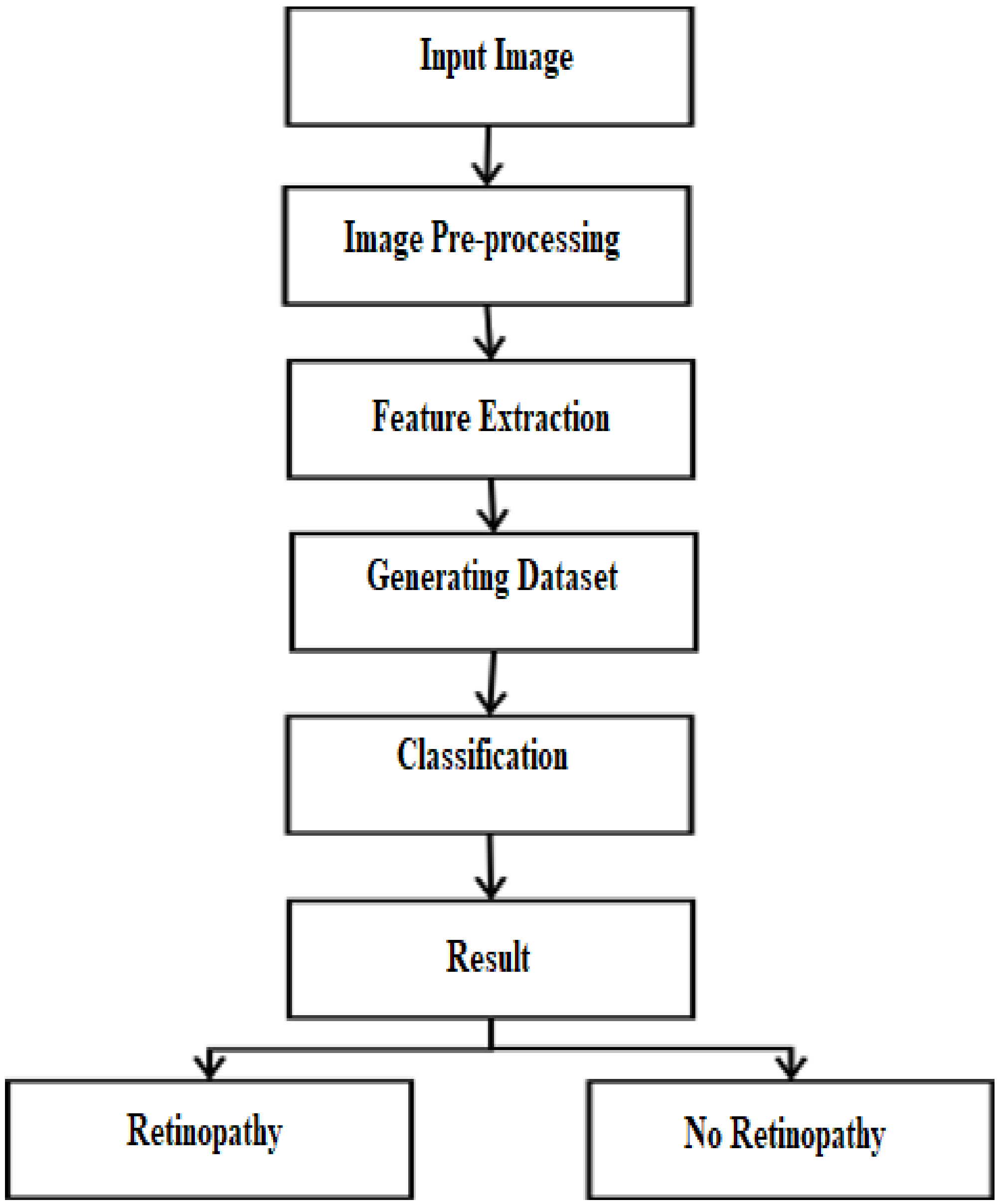

Fig 7: Flow of the process while detecting Retinopathy.

5.2) Feature Extraction:

Blood Vessel Extraction:

Blood vessel extraction [7] is basic and important feature need to extract from image. In affected eye or human who facing problem of diabetic retinopathy has abnormal growth of blood vessels or cut blood vessels in eye structure.

For extraction first from RGB image green channel is selected. Green channel is the best channel in which we can see blood vessels in more detail than red, blue or gray scale mode. Then image is resized into 512*512 pixel. Contrast of a green channel is adjusted using CLAHE histogram adjusting method. Now finally contrast adjusted image is passed through canny edge detection algorithm that detect blood vessels in image. And finally for getting one numerical value for image we have used fast Fourier transformation. That convert image into frequency domain. Fast Fourier transformation gives 512 complex number each number gives information pixel.

Fast Fourier transformation complex number are such that real number is combination of frequency and imaginary part is combination of phase of pixel. So we take a mean of that numbers and divided into 2 part one is real part and other one is imaginary part that gives information about average frequency and average phase of an image [8].

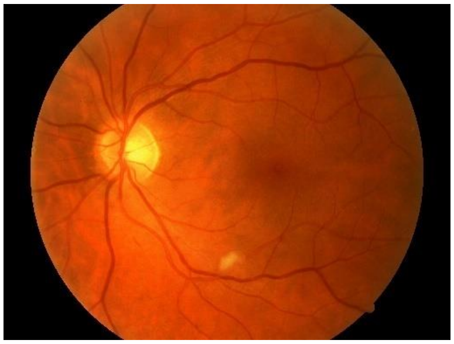

Fig 8: RGB Image for BV Extraction.

Fig 9: Processed Image.

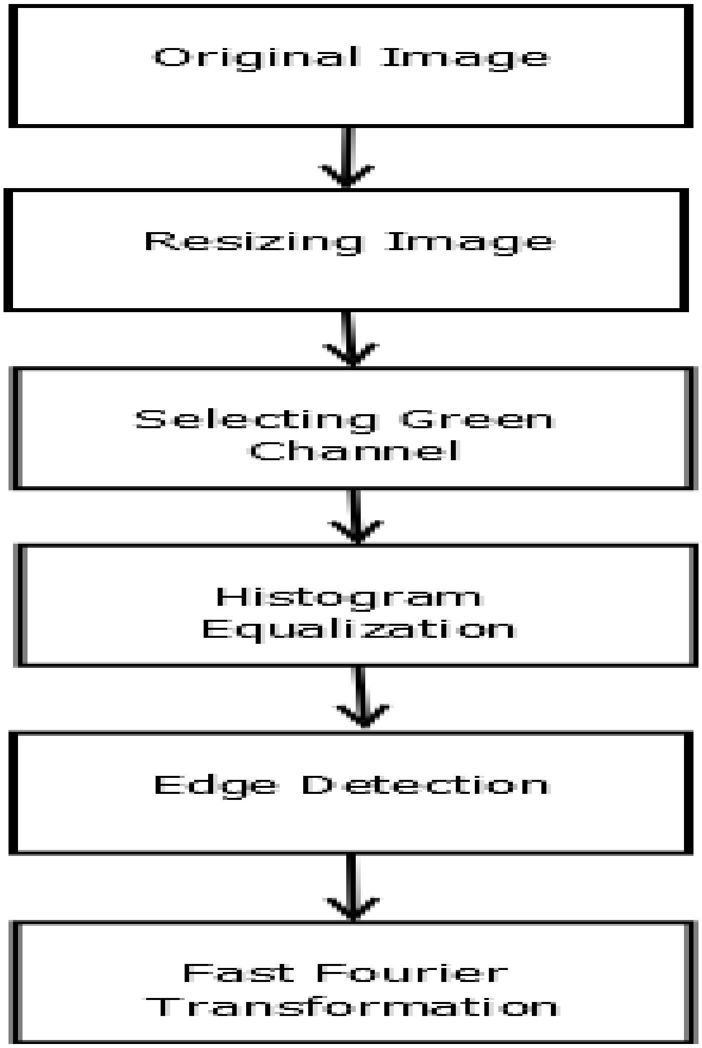

Fig 10: Step by step procedure of Blood Vessel Extraction.

Hemorrhages Extraction:

Hemorrhages is first visible symptoms that can see by naked eye if and only if that Bright enough. As stated early this are bright dot in RGB image. Actually hemorrhages are biological liquid that is came out after breaking of blood vessels while abnormal growth of blood vessels.

For extraction of different preprocessing [5] is used but some basic steps are again followed from blood vessels extraction. Again image is resized into 512*512 pixel then green channel is selected using split function. CLAHE histogram adjusting technique is used to adjust contrast of image and getting better visible of blood vessels and hemorrhages now similarly again edge detection is used image convert into binary image. Binary image we have to implement some morphological operation. That are erosion, dilations and opening operation. Erosion reduce the noise from image and dilations increase white pixel in image.

The most important part of preprocessing is using opening operation it is dilation followed by erosion. Number of iteration and kernel size are major part of this morphological operation. After proper combination we get hemorrhages dots in image and finally we count the dots in image and store the number of dots in particular image [10] [11].

Fig 11: RGB Image for Hemorrhages Extraction.

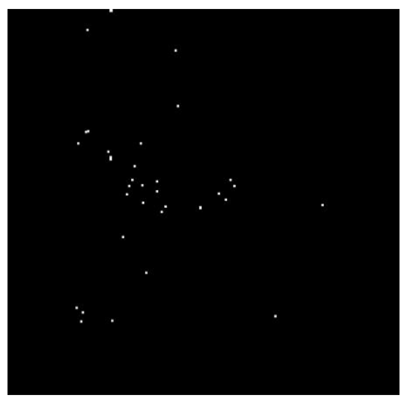

Fig 12: Hemorrhages dots spotted.

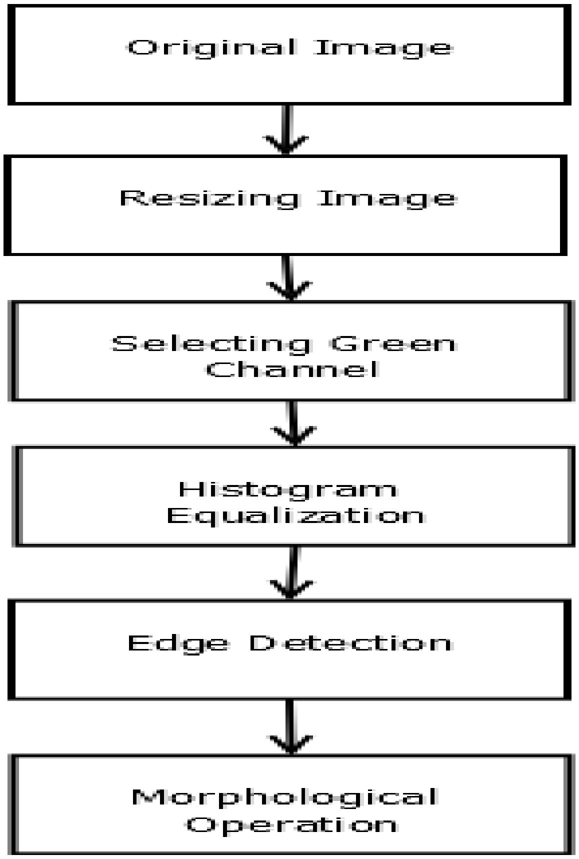

Fig 13: Step by step procedure of Hemorrhages Extraction.

Exudates Extraction:

Exudates [7] are the deposits of lipid fluids that leak from the blood vessels accumulate in the retinal region. They are usually found in oval or round shape and white or creamy white in color. The optic disc is similar to exudates in terms of luminance and color property and hence during clustering, both exudates and optic disk gets clustered. Thus preprocessing is done so that the optic disk is not misclassified as one of the exudates which may lead to wrong diagnosis. Also, major problem in processing of such spherical images is difficult as they are not uniformly illuminated.

The preprocessing for exudates begins with splitting the RGB image into its colour components and using the green channel for further process. This is done because exudates have found to have higher contrast in green component. Next, the contrast is enhanced using CLAHE, in which noise is not amplified. Amongst various methods to obtain uniform illumination, one of the most popular methods is median filtering. So we further apply median filter to enhance its edges.

After preprocessing, edge based segmentation has been used in which morphological operations have been performed in order to get the total number of pixels covered by exudates in the image. First the morphological erosion operation is performed on the preprocessed image and then it is morphologically dilated which makes the edges of the exudates prominent. Next, the morphological closing operation is performed using ellipse (3*3) as the structural element to locate the exudates. Once the exudates are located, we calculate the number of pixels under the exudates found and then thresholding is done to get the number of pixels under the exudates region and then compare it with total number of pixels in the image under examination.

GLCM Structural Features Calculation:

Feature extraction is done to reduce the dimensions to an image as well as to obtain more useful information present in the signals by excluding useless information. In terms of signal processing, the signal is processed to deprive it of noise and interference prior to actual feature extraction.

In image processing, a gray level co-occurrence matrix (GLCM) [2] is used which is a histogram of co-occurring grayscale values at a given offset over an image to describe the texture of an image, it is usual to extract features such as entropy, energy, contrast, correlation, etc. from several co-occurrence matrices computed for different offsets.

Syntax: stats = graycoprops(glcm,properties)

Contrast- Returns a value of the intensity of contrast between a pixels. Contrast is 0 for a constant image.

Energy- Returns the sum of squared elements in the GLCM. Energy is 1 for a constant image.

Homogeneity- Returns a value that measures the proximity of the elements in the GLCM. Homogeneity is 1 for a diagonal GLCM.

Entropy- Refers to disorder or uncertainty and is the average rate at which information is produced.

CHAPTER 6

RESULT AND DISCUSSION

6.1) Result:

Fundus images are used to train the algorithms. All the images are preprocessed before the feature extraction to increase the detailing of features in the images. Different procedures are followed for different features. Every extracted image gives a value for different features. All the values with labels are stored in a CSV file. Now this file is used to train maintaining some ratio as earlier mentioned. All the classification algorithms get trained on data and start self- evaluating the results obtained. Firstly SVM classifier is used which states 62.96% accuracy, second one decision tree is used which states 61.72% , then KNN is used it states 57.12% and finally random forest states62.96%. Data was increased to 1192 images in which equal number of lesion images were present. Feature normalization was done on the values extracted from features. Strong features were identified using ranker and feature selection was done. All the above procedure reported 74% and 71% accuracy for random forest and SVM respectively. Parameter tuning was used to tune down the parameters of classifiers then SVM reported 84.93% accuracy. It is the best accuracy of that model then web interface is created where user is asked to upload the image saved model evaluates the image and reports whether Retinopathy or No Retinopathy.

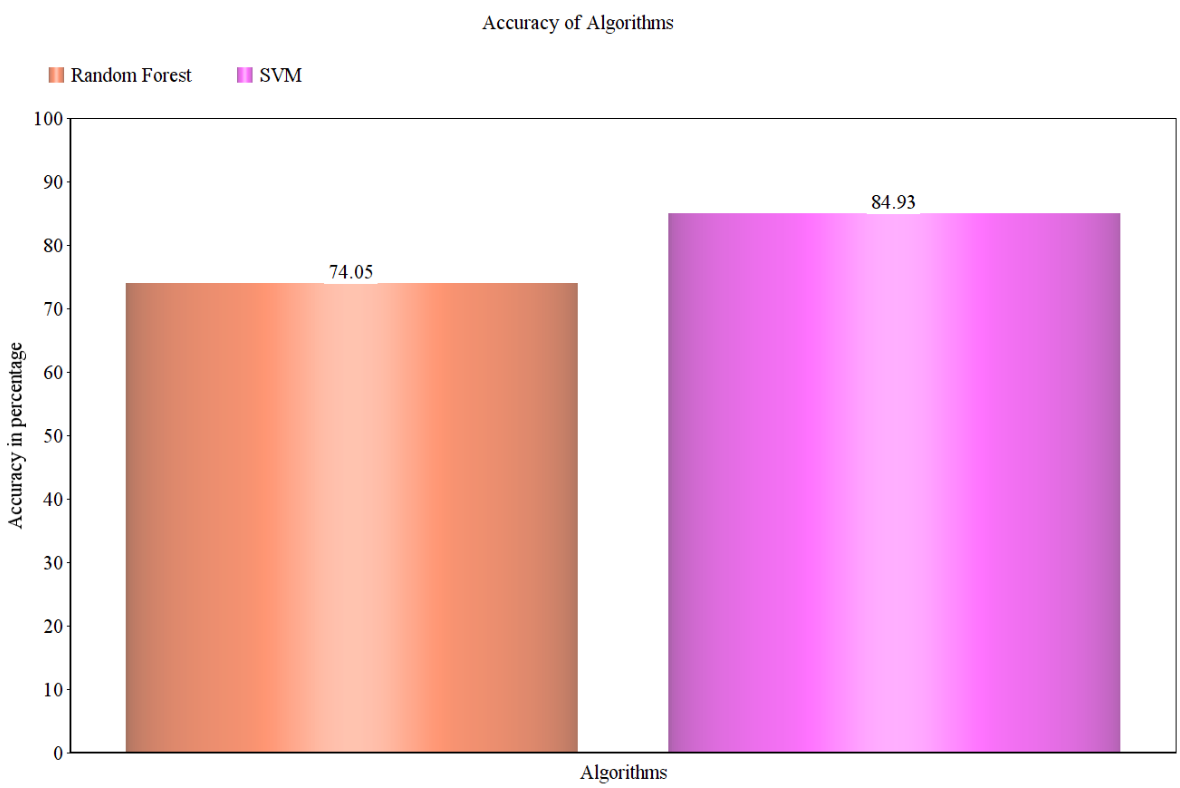

Fig 14: Accuracy of algorithms

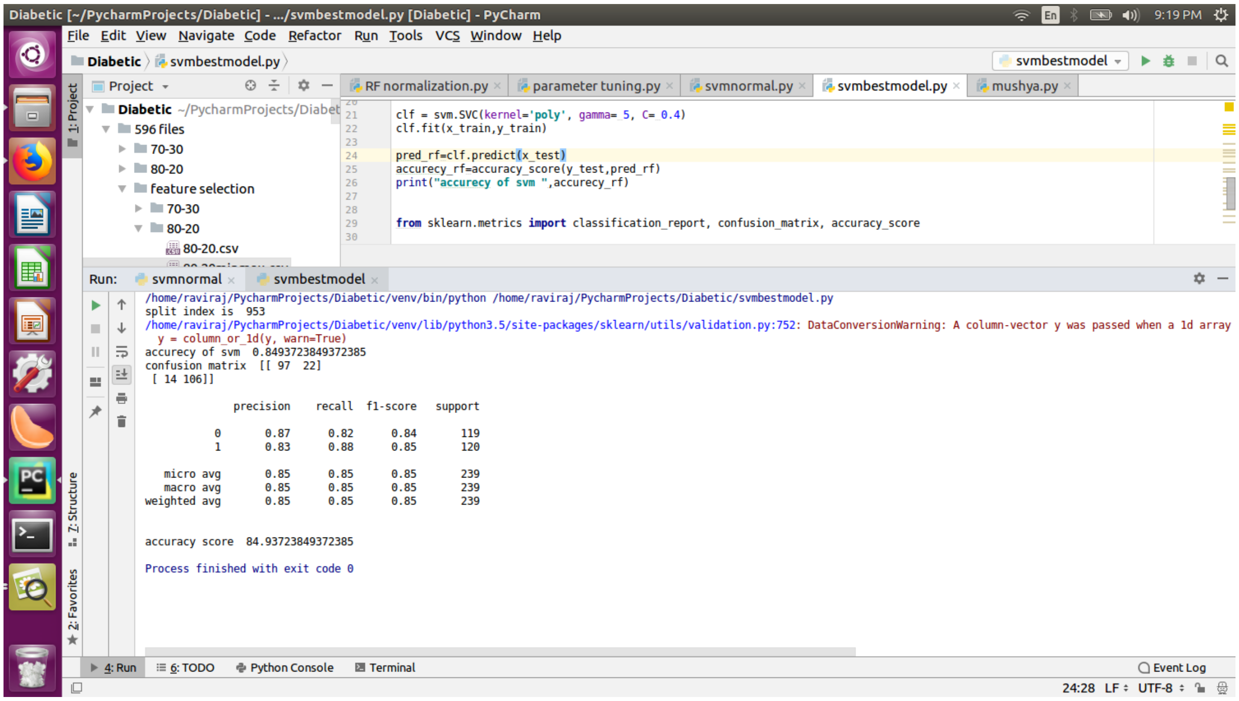

Fig 15: Screenshot of final accuracy.

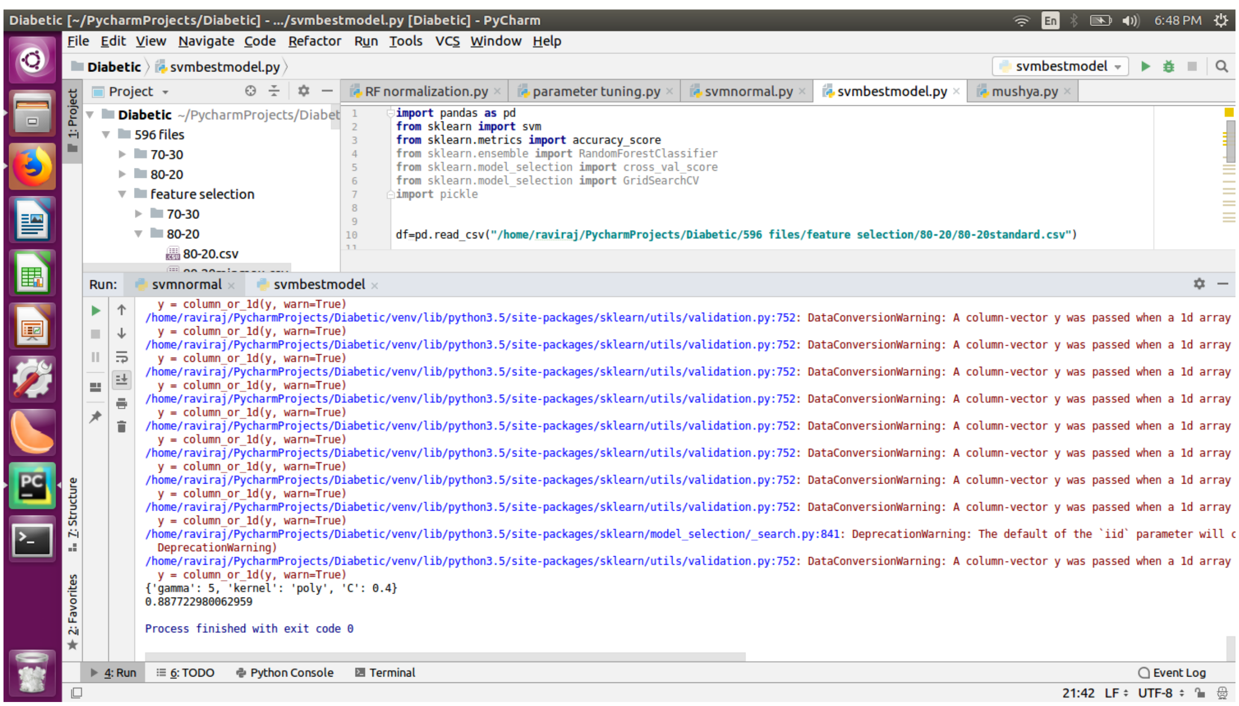

Fig 16: Best accuracy of SVM.

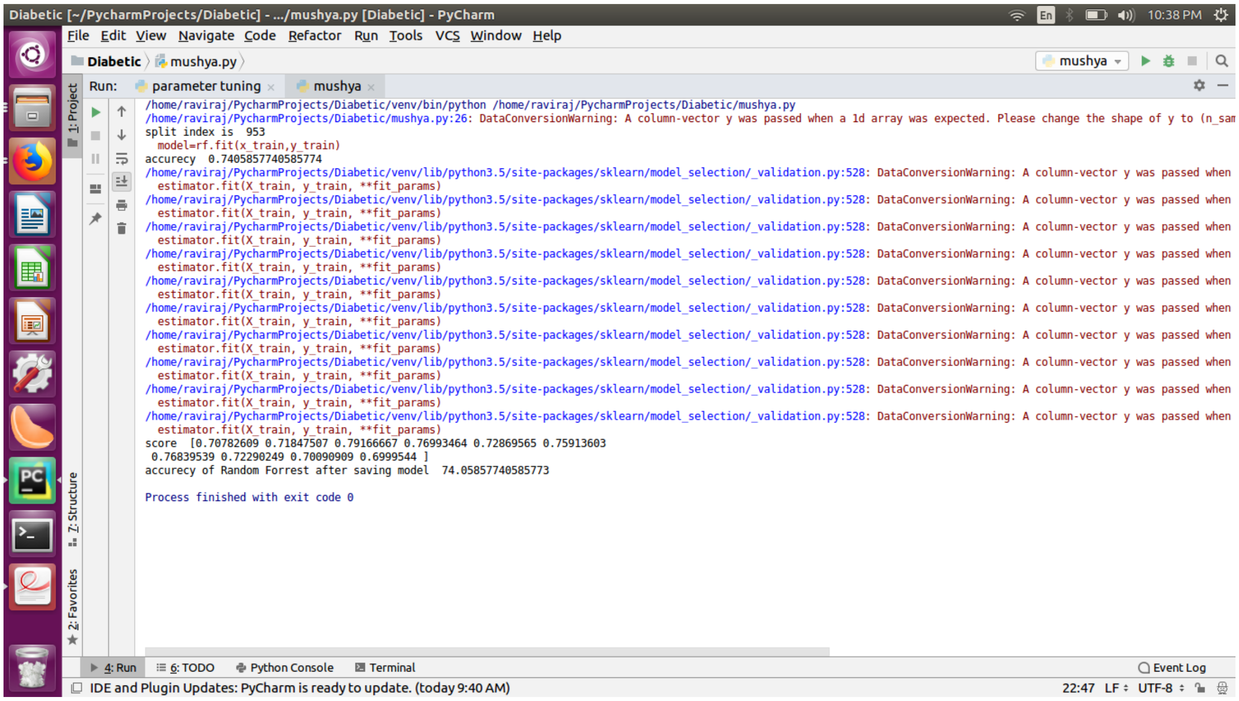

Fig 17: Final accuracy of random forest.

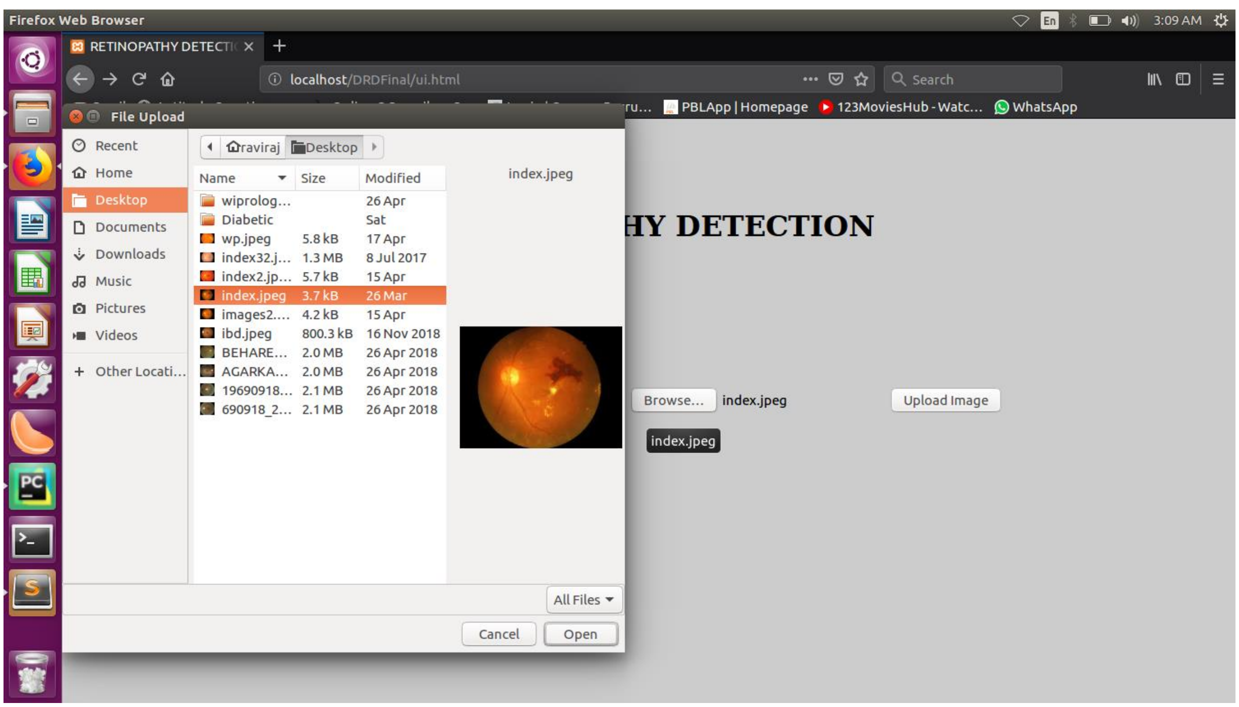

Fig 18: Affected image is chosen for evaluation.

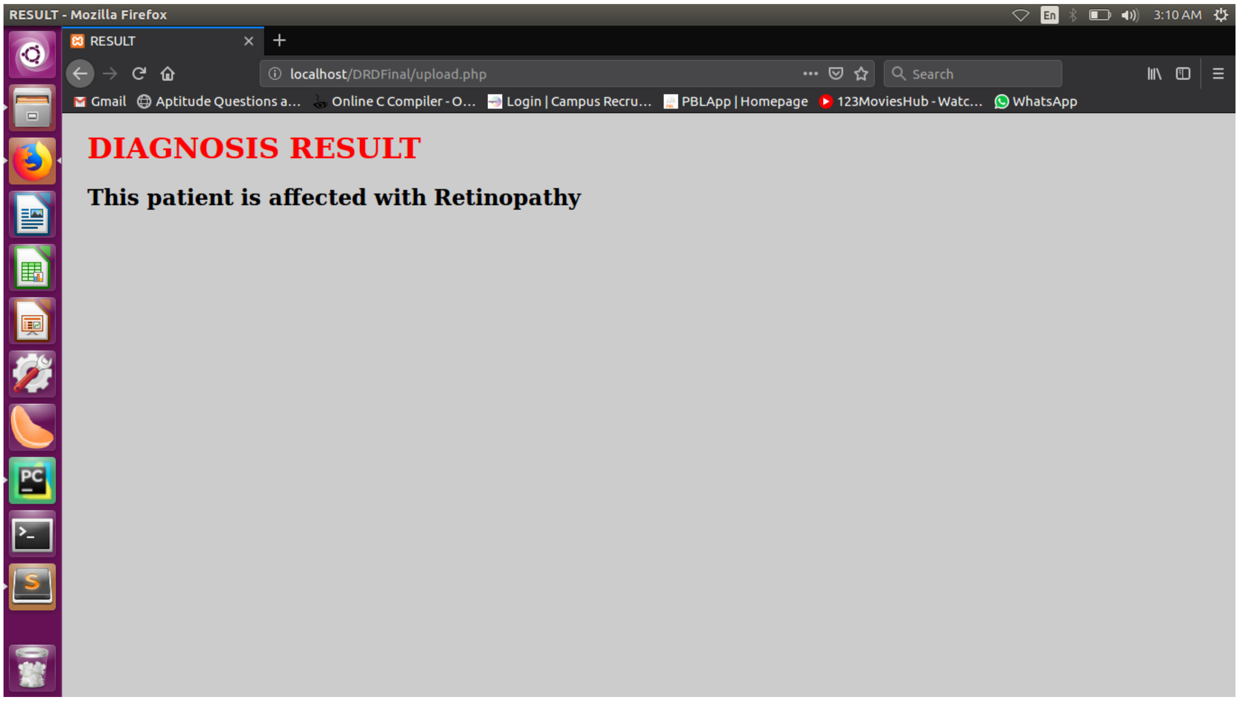

Fig 19: Result of affected image.

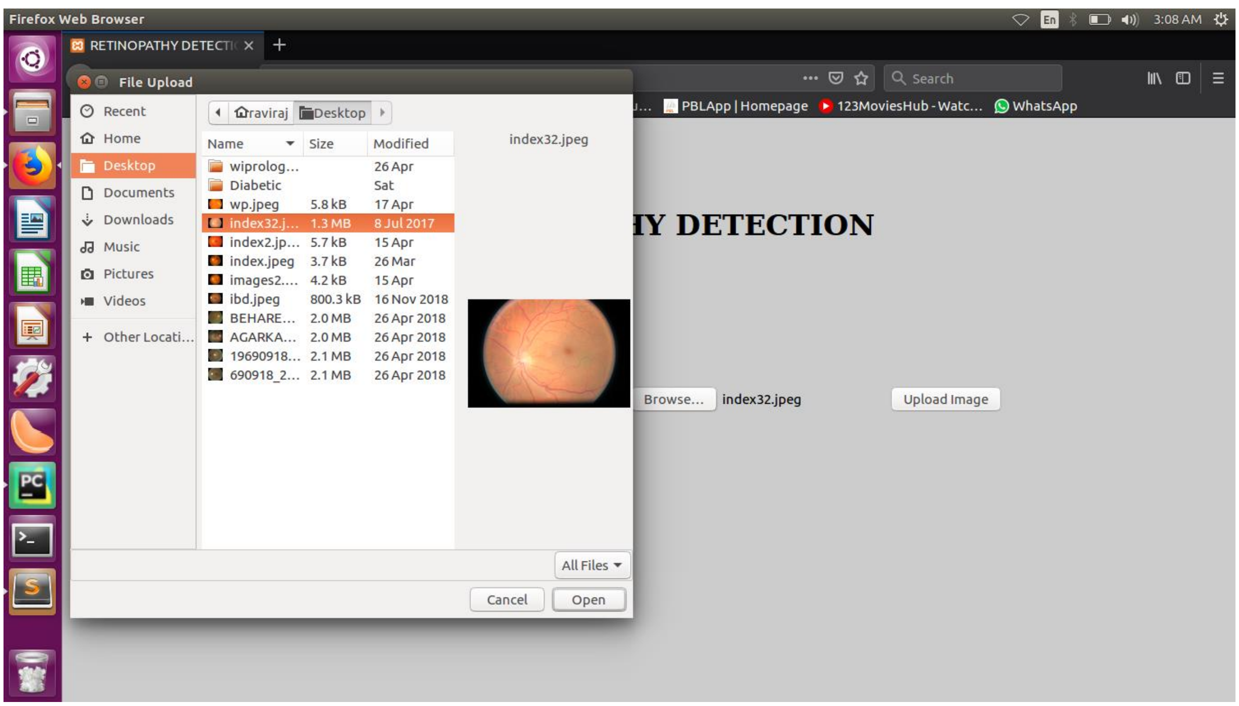

Fig 20: Non-affected image is selected for evaluation.

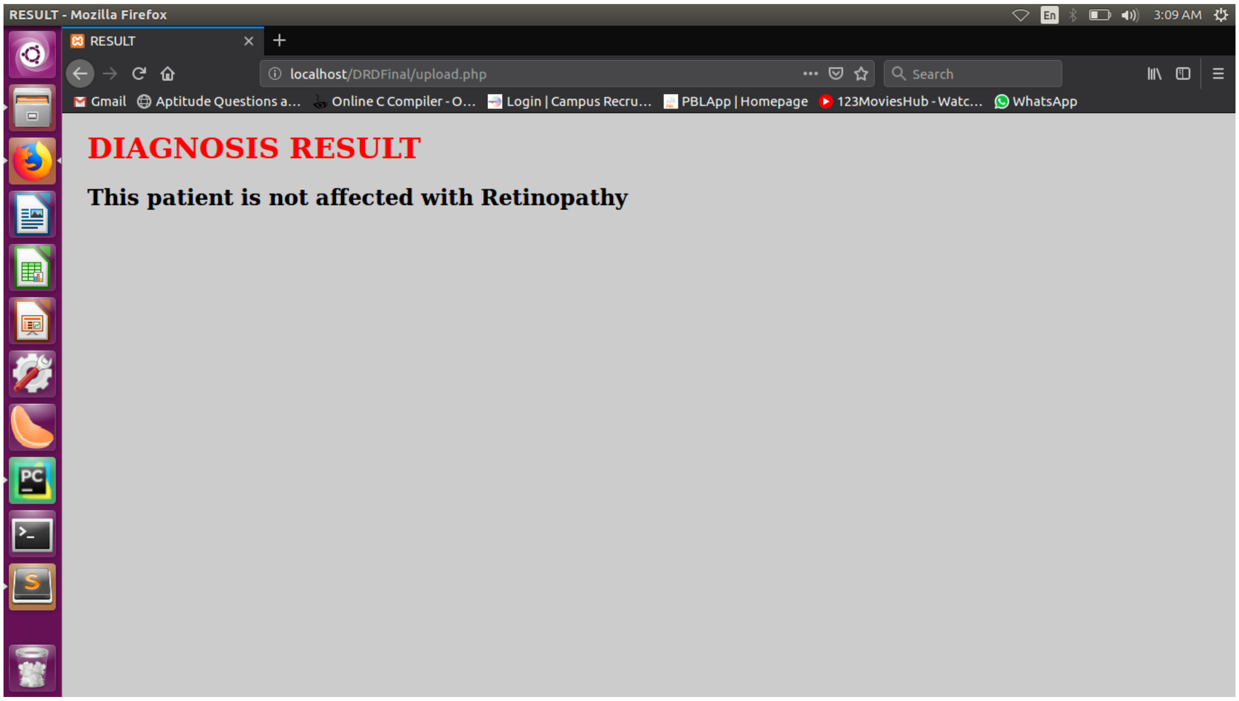

Fig 21: Result of non-affected image.

SVM Kernels	RBF	Linear	Poly
Accuracy	84.93	66.50	66.50

Table1: Accuracy measured on different SVM kernels.

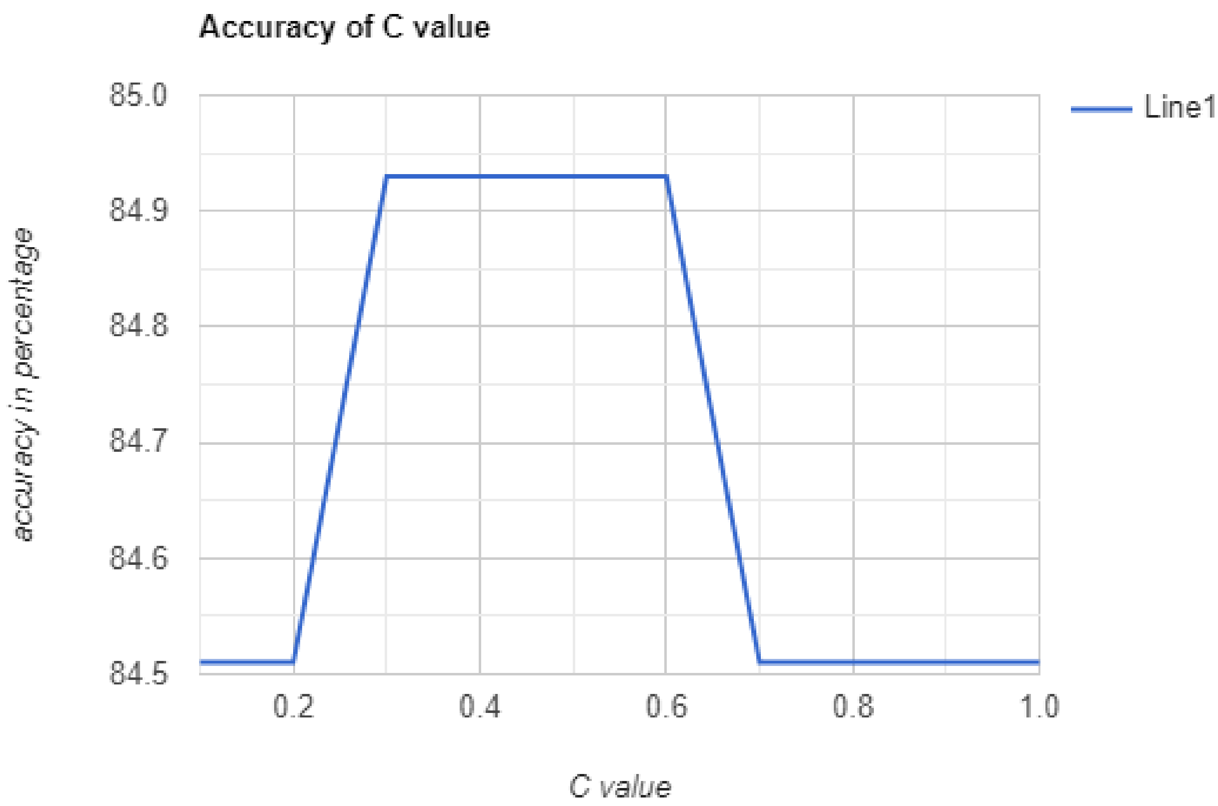

Fig 22: Accuracy of different C values of RBF.

C Values	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Accuracy	84.51	84.51	84.93	84.93	84.93	84.93	84.51	84.51	84.51	84.51

Table2: Accuracy of different C values of RBF.

6.2) Discussion and Analysis:

Training on 405 images of gave the below results. So, to improve the results volume of images was raised to 2000 and 80-20 module was used which delivered as per the graph below. Results may vary when the feature extraction vary, improving the extraction methodologies will give accurate feature value. Optimizing the parameters of the classifier will improve the scope accuracy. Trying to get more features for extraction with enhanced processing of image will directly affect the outcome. Combination plays a vital role because one combination may show that one of the algorithm has high accuracy while a different combination may give a totally different output. To improve the accuracy structural features were extracted, which led to more number of features. Now from those some strong features are to be identified because weak features may bring down the accuracy. So, feature ranker is used to find the best features, 53 combinations are made for feature selection with respect to ranks of features. Even the accuracy gained after training 2000 images is not giving efficient results. It is a biased system because of the dataset contains Retinopathy and Non-Retinopathy images at 1:3 ratio. In order to overcome that problem dataset was made at 1:1 ratio and normalized so that the system would give exact results. After following all the above procedures there were only slight changes in the accuracies, then parameter tuning of algorithm is used. Finally it reported 75% accuracy for random forest and 84.93% for SVM. This proposed system has a limitation, it can classify with either lesion 1 or lesion 2. If more data is used to train the network then they can also be classified in to lesion 3 and 4 as well.

Algorithms	Training data	Testing data	Total volume	Accuracy
RandomForestClassifier	70	30	405	62.96%
Neural_Network	70	30	405	52.90%
SVM	70	30	405	62.96%
DecisionTreeClassifier	70	30	405	61.72%
QuadraticDiscriminantAnalysis	70	30	405	60.23%
KneighborsClassifier	70	30	405	57.12%

Table 3: Accuracies attained on 70% training and 30% testing.

Algorithms	Training data	Testing data	Total volume	Accuracy
RandomForestClassifier	80	20	405	64.19%
Neural_Network	80	20	405	58.02%
SVM	80	20	405	66.66%
DecisionTreeClassifier	80	20	405	64.19%
QuadraticDiscriminantAnalysis	80	20	405	62.12%
KneighborsClassifier	80	20	405	61.43%

Table 4: Accuracies attained on 80% training and 20% testing.

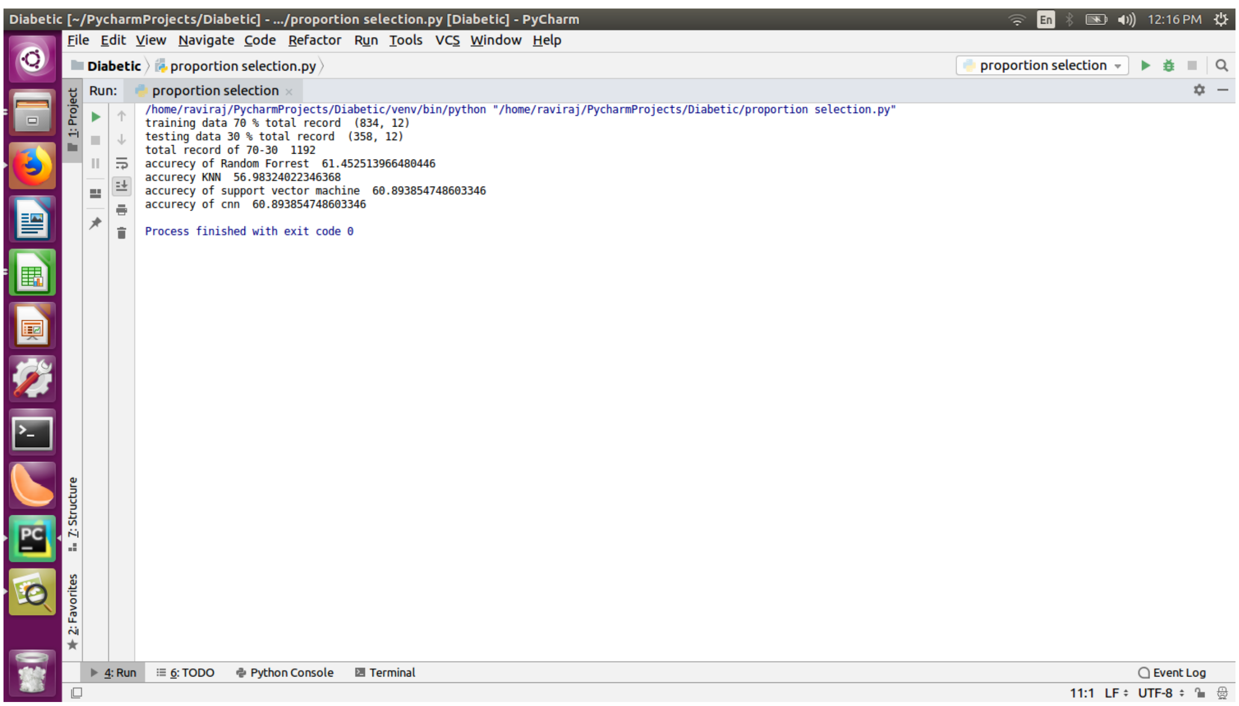

Fig 23: Accuracies of different algorithms trained with 70-30 split ratio for feature normalization.

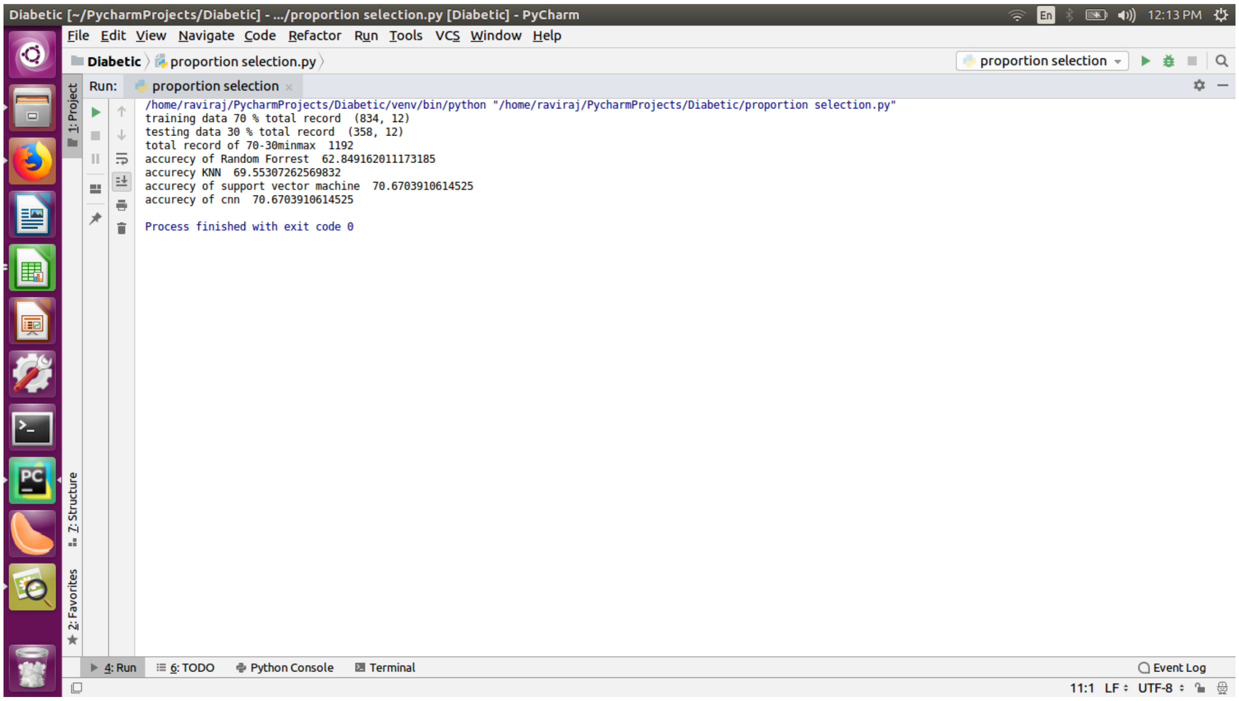

Fig 24: Accuracies of algorithms 70-30 training testing with min-max feature normalization.

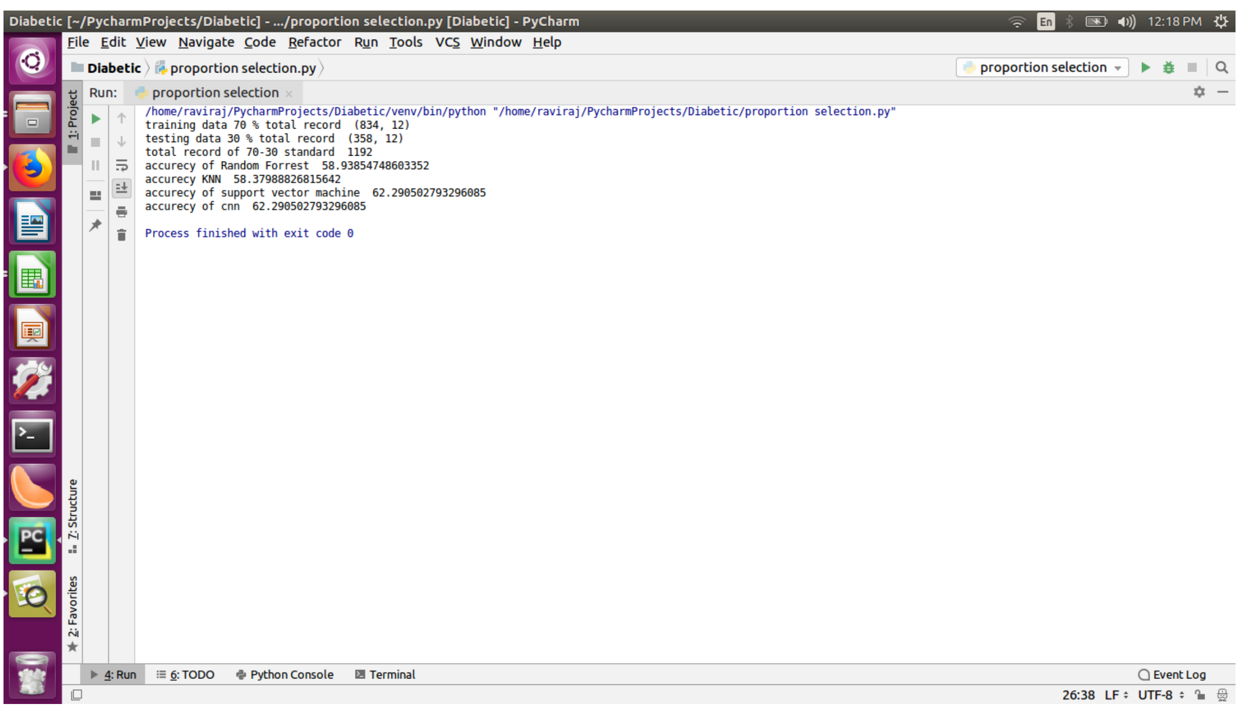

Fig 25: Accuracies of algorithms 70-30 training testing with standard normalization.

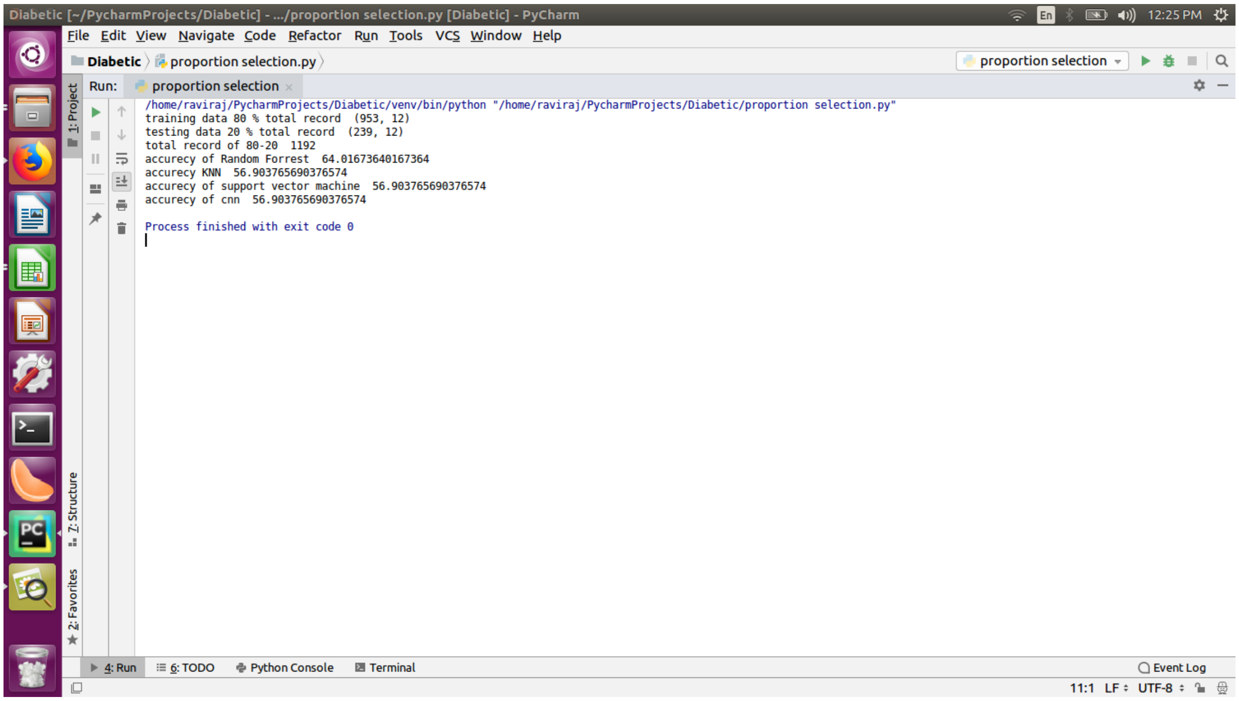

Fig 26: Accuracies of algorithms 80-20 training testing with normalization.

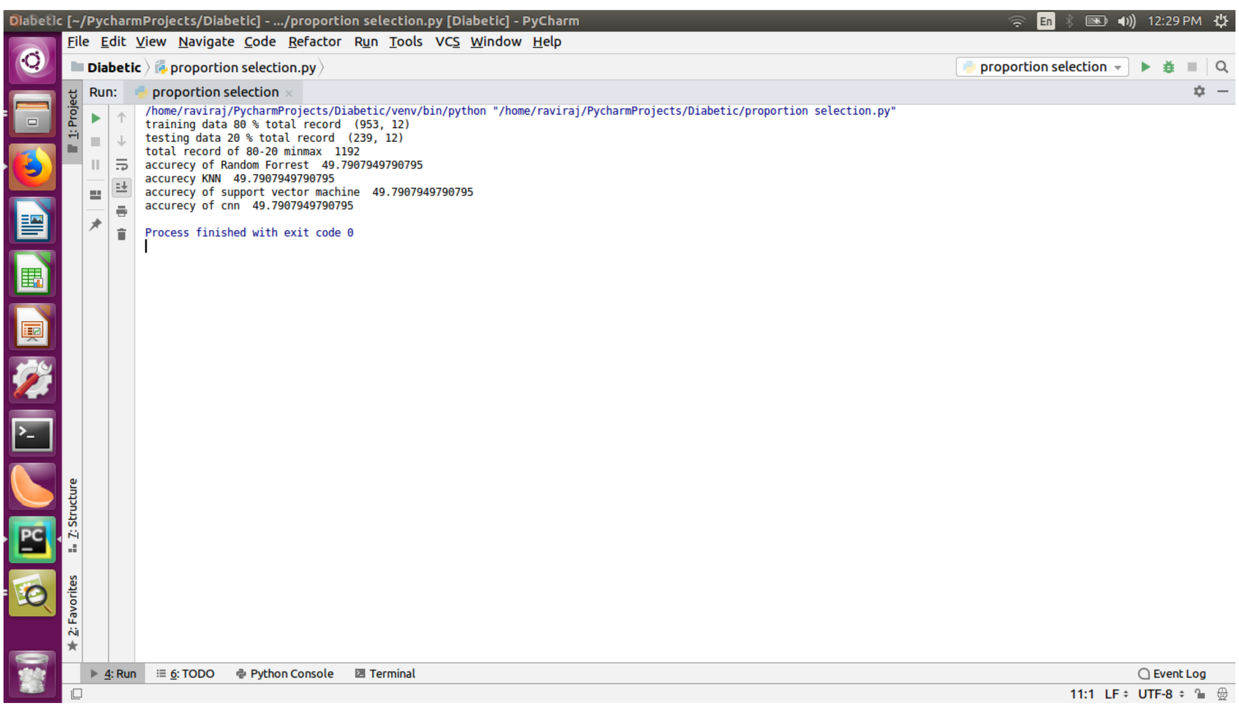

Fig 27: Accuracies of algorithms 80-20 training testing with min-max normalization.

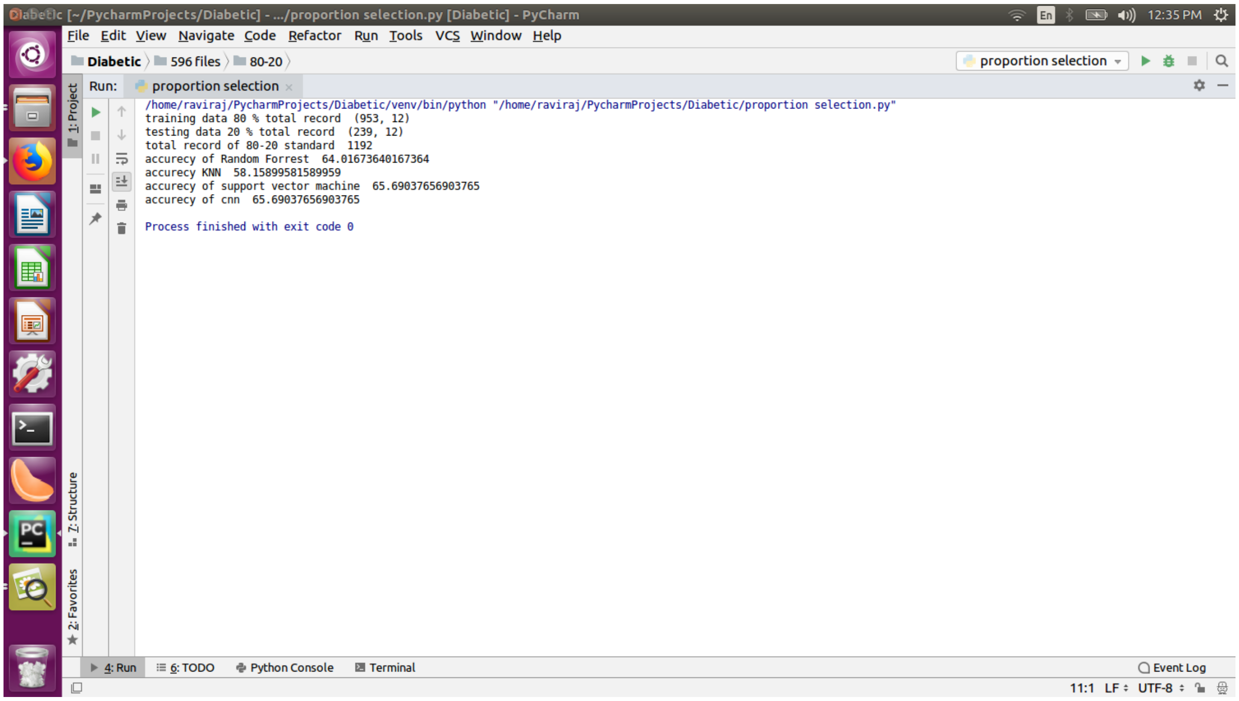

Fig 28: Accuracies of algorithms 80-20 training testing with standard normalization.

All above images are of different algorithms trained on 2000 images. These 6 images contains 2 combinations of training and testing models, first 3 images are of 80-20 and the other 3 are of 70-30 models. Each combination consists algorithms trained on regular data, minmax data and standard normalized data.

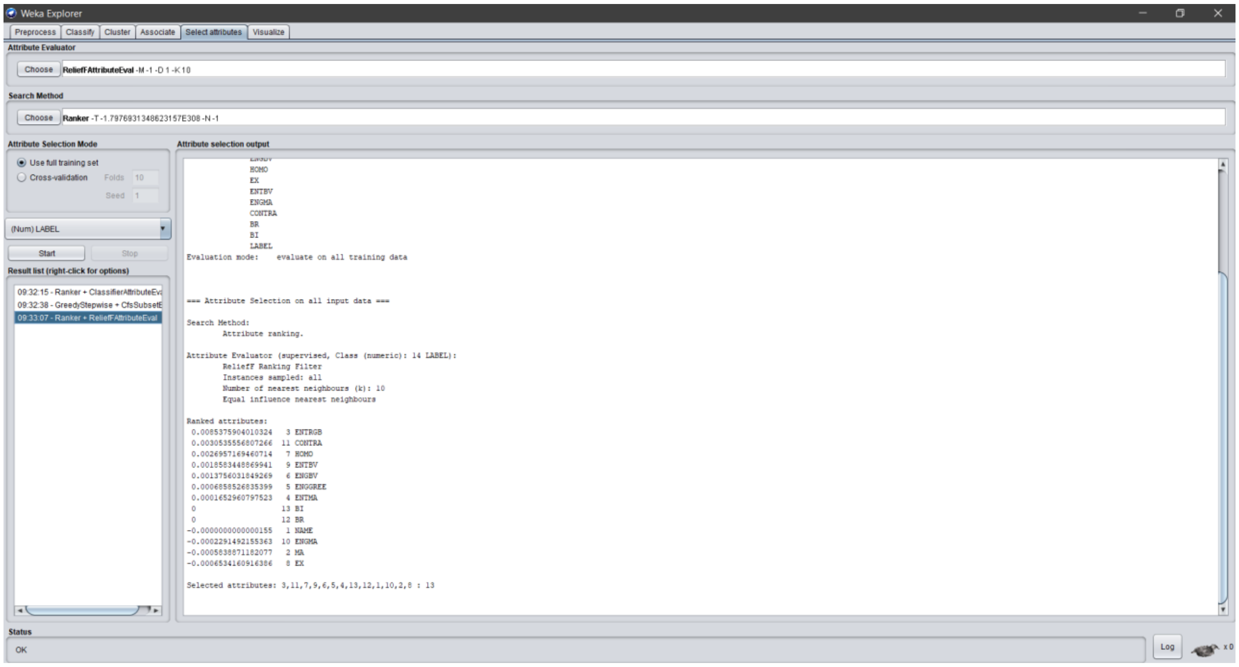

Fig 29: Features are Ranked using Weka.

Features ranked according to their strengths. Relief attribute evaluator and ranker search methods are used to rank the features. 10 strong features according to their ranks are used.

Fig 30: Retinopathy image used for evaluation after training.

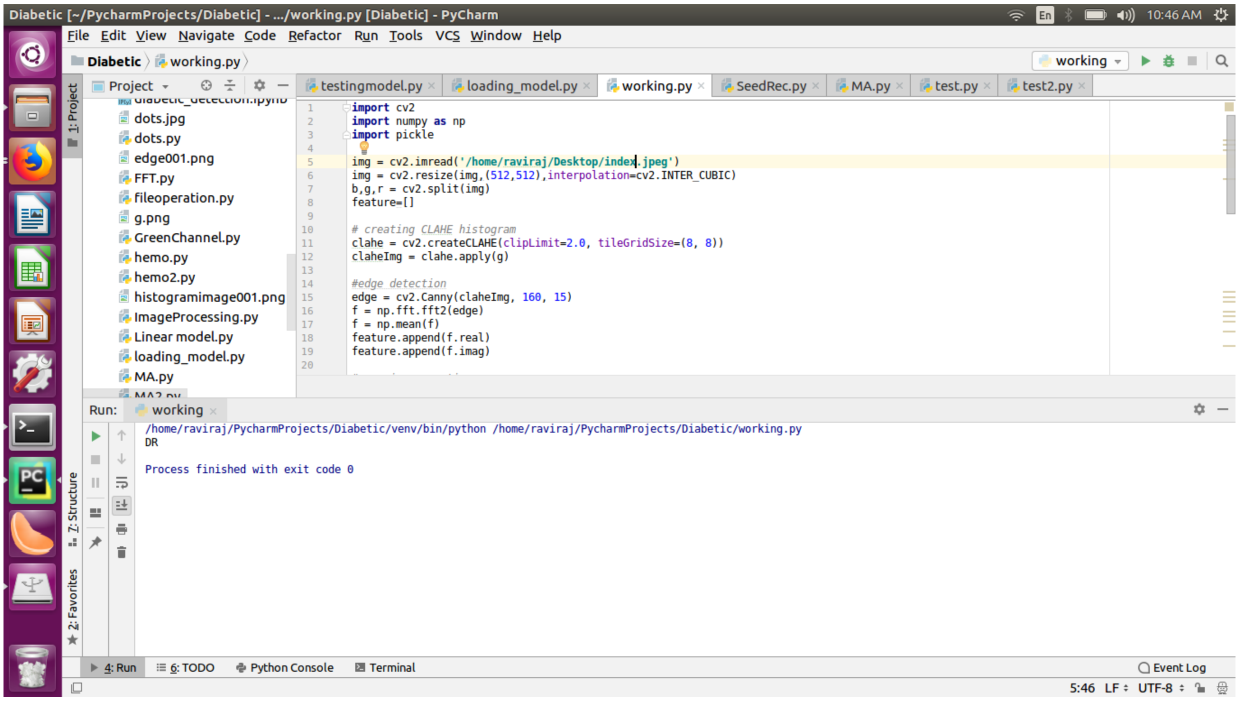

Fig 31: Result after evaluating the Retinopathy image.

Fig 32: Non retinopathy image used for evaluation after training.

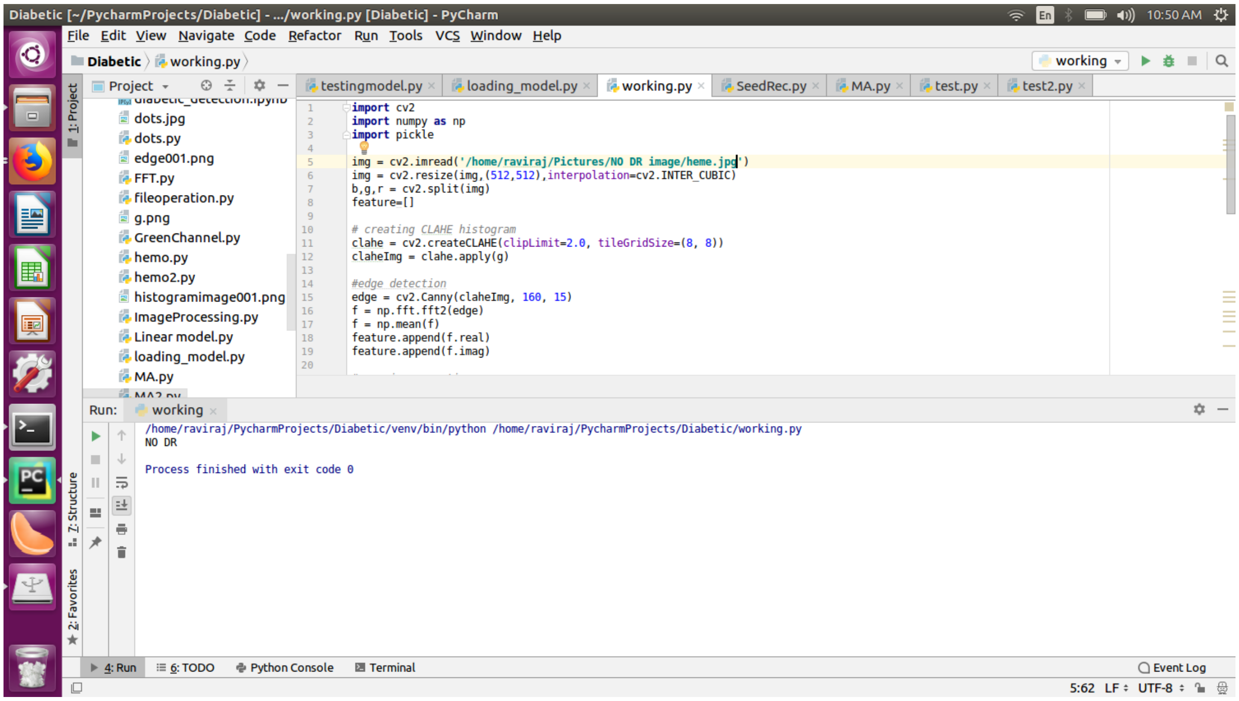

Fig 33: Result After evaluating the non-retinopaty image.

Tables below explain the combinations of features and accuracies of algorithms during feature selection. All algorithms were trained on both types of data which is sequential and shuffled. Training testing model used for the data is 80-20.

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	72.28	72.28	72.28	72.04	72.28	72.77	73.73
CNN	71.08	28.91	71.08	71.08	71.08	71.08	71.08
SVM	71.08	71.08	71.08	71.08	71.08	71.08	71.08

Table 5: Accuracy over shuffled data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	71.8	72.28	72.28	72.28	72.04	72.77	71.08
CNN	71.08	28.91	71.08	71.08	71.08	71.08	71.08
SVM	70.66	70.84	71.08	70.84	71.8	71.56	71.08

Table 6: Accuracy over shuffled standard data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	73.01	72.53	73.25	73.01	73.73	73.73	73.97
CNN	71.08	71.08	71.08	71.08	71.08	71.08	71.08
SVM	72.53	72.53	73.01	73.01	74.21	72.28	73.73

Table 7: Accuracy over shuffled minmax data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	74.21	74.93	74.69	75.18	74.45	75.42	74.93
CNN	71.08	28.29	28.91	71.08	71.08	71.08	71.08
SVM	71.08	71.08	71.08	71.08	71.08	71.08	71.08

Table 8: Accuracy over sequential data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	72.77	72.53	71.8o	71.32	71.8	71.8	70.6
CNN	71.08	71.08	71.08	71.08	71.08	71.08	71.08
SVM	72.77	73.34	74.21	72.53	71.8	71..56	71.08

Table 9: Accuracy over sequential standard data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	74.55	74.45	73.73	74.45	73.73	74.45	74.69
CNN	71.08	71.08	71.08	71.08	71.08	71.08	71.08
SVM	75.42	75.18	75.18	75.9	75.9	74.69	76.38

Table 10: Accuracy over sequential minmax data

No. Of features selected → / Algorithm used ↓	11	10	9	8	7	6	5
RF	74.5	74.45	73.73	74.45	73.73	74.45	74.69
CNN	71.08	71.08	71.08	71.08	71.08	71.08	71.08
SVM	75.42	84.93	75.18	75.9	75.9	74.69	76.38

Table 11: Final Accuracies after parameter tuning on 1192 images(1:1).

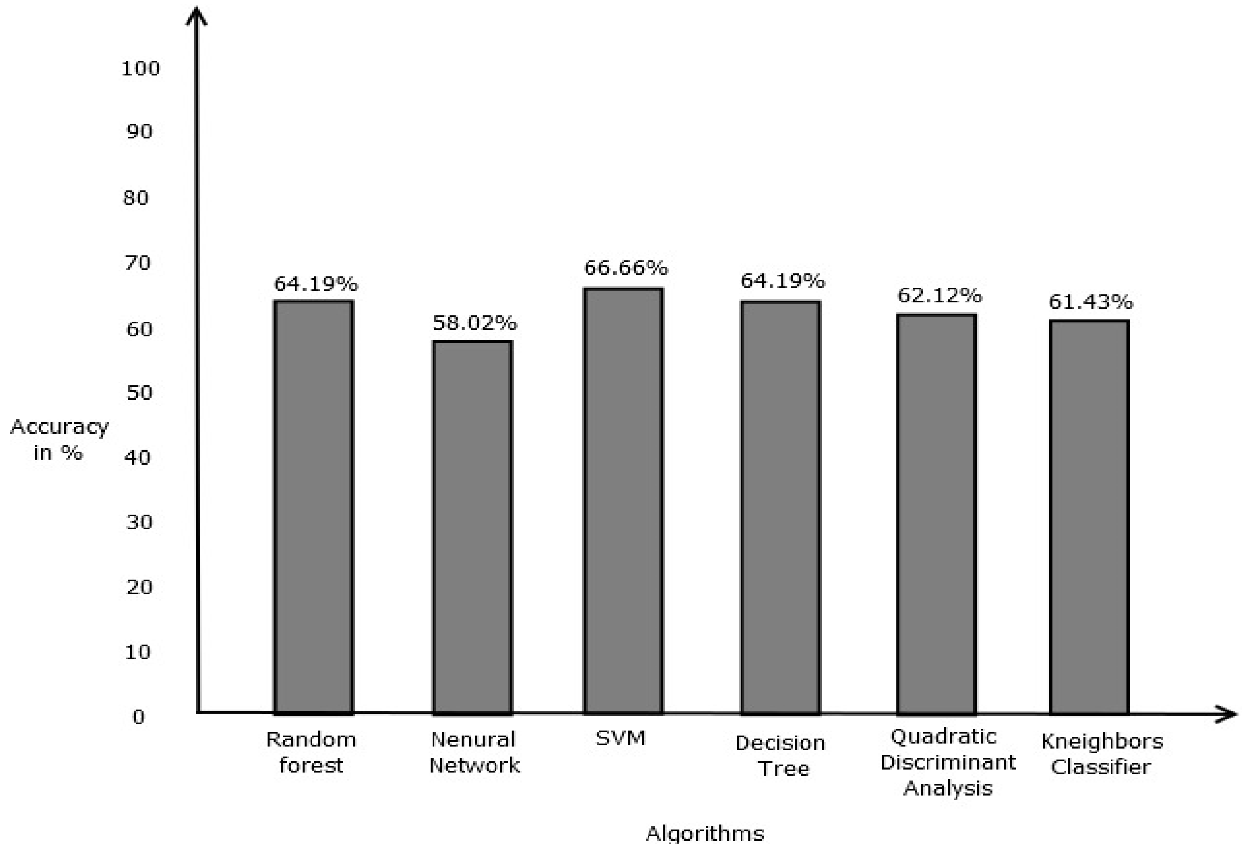

Fig 34: Graph plotted of Algorithms VS Accuracy on 405 images.

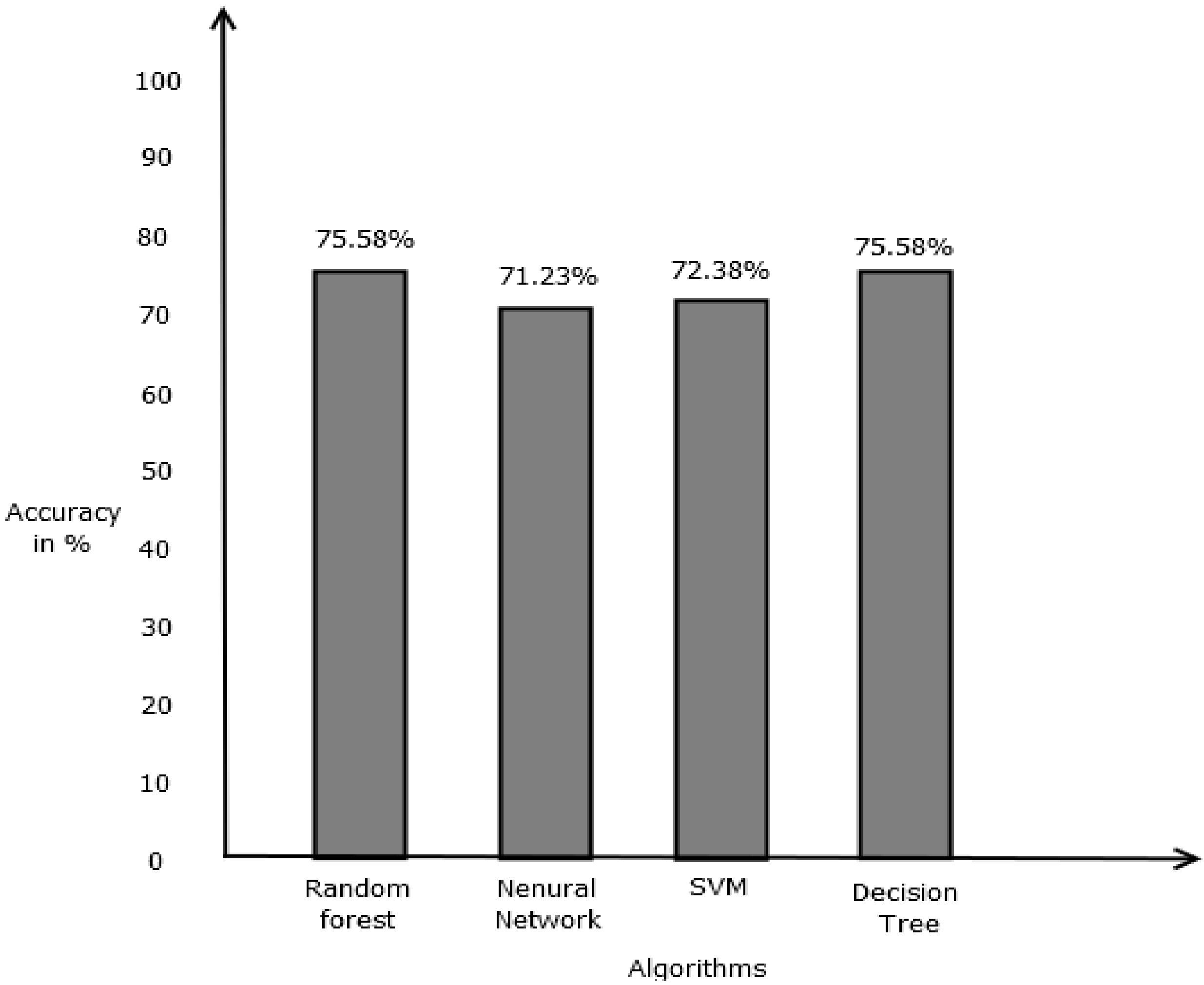

Fig 35: Graph plotted for Algorithms VS Accuracy on 2000 images.

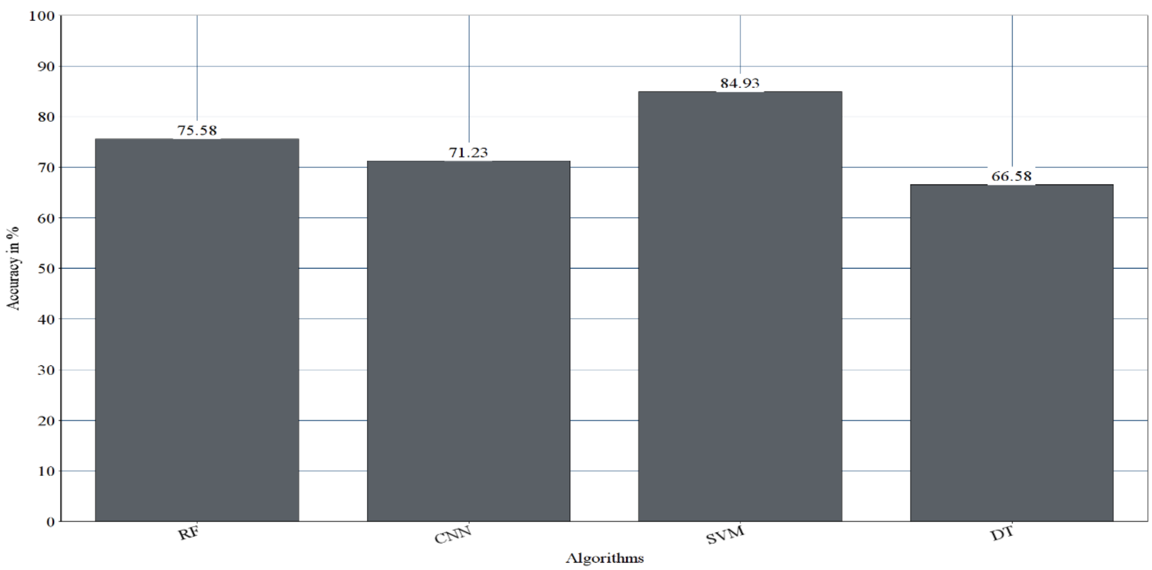

Fig 36: Graph Plotted for Accuracy VS Algorithms for 1192 images

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1) CONCLUSION:

Retinopathy is a disease caused to diabetic and non-diabetic patients which results loss of vision apparently. We have used datasets referring to diabetic retinopathy, of retinal images having diabetic retinopathy or no diabetic retinopathy as diabetes is the major cause of retinopathy and more data is available to train the model. The model has learned from datasets of DR and no DR but as the features we chose are universal for retinopathy, the model stands true in recognizing results for ocular and systematic retinopathy as well. Some of the pre-processing techniques are used to enhance the image for feature extraction. Image is converted into gray scale or one of the channels of RGB is selected according to the feature. CLAHE is used for the enhancement of the contrast levels of image. Image is resized where dimensions are adjusted. Canny edge detection used for detecting edges, finding intensity of image, avoid maximum suppression and binary threshold to convert image into binary image. To extract features morphological transformation is used. Extracted features are used to train the algorithms and classify image. Feature normalization and feature selection are used for improving the accuracy which eventually led to a biased system, then equal number of affected and not affected images were used for training. Finally parameter tuning was used and maximum accuracies were reported by algorithms. These algorithms are capable of evaluating the images.

7.2) FUTURE SCOPE:

In the present proposed system features like blood vessels and hemorrhages were selected for training the algorithms. In future remaining features such as micro aneurysms should also be used to get better results. Even the images used for training are already evaluated. Real time images can also be used for evaluation and training. Segmentation is done using some of the consistent algorithms. So, algorithm which were not used in this system can be used. Generally when evaluating different modules are used, even an automatic screening system with all the modules combined can be designed on the basis of existing system.

REFERENCES

[1] Automatic Detection of Diabetic Retinopathy in Large Scale Retinal images, M. Kamaladevi, S. SnehaRupa, T. Sowmya School of computing, SASTRA Deemed to be University.

[2] Diagnosis of Diabetic Retinopathy using Machine Learning, Swati Gupta and Karandikar AM Department of Computer Science and Engineering, Shri Ramdeobaba College Of Engineering and Management, Nagpur.

[3] Ocular and other systematic causes of retinopathy in non diabetic patients, Jayant Venkatramani and Paul Mitchell.

[4] https://acadpubl.eu/hub/2018-119-12/articles/6/1316.pdf

[5] Detection of Hemorrhages in Retinal Images using Hybrid Approach for Diagnosis of Diabetic Retinopathy, S.Karkuzhali Assistant Professor Kalasalingam University and Dr. D. Manimegalai National Engineering College, Kovilpatti, Tamilnadu.

[6] https://en.wikipedia.org/wiki/PyCharm.

[7] https://github.com/getsanjeev/retina-features/find/master?q=

[8] https://docs.opencv.org/3.0

[9] https://www.surveyophthalmol.com/article/S0039-6257(02)00387-9/abstract

[10]http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.3068&rep=rep1&type=p

[11] https://github.com/Nomikxyz/retinopathy-dataset

[12] https://www.kaggle.com/c/diabetic-retinopathy-detection/data