Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 3, 2026

IMPROVED AUTOMATED DETECTION OF DIABETIC RETINOPATHY ON A PUBLICLY AVAILABLE DATASET THROUGH DIFFERENT MACHINE LEARNING ALGORITHMS

Bijjar Aziz[1,2], Noor Uddin[*2,3], Sohrab Khan[2], Noor Ahmed[2], Ghanwa Haider[4] and Inayat Ullah[5]

[1,] *Computer Science Department Lasbela University of Agricultural Water and Marine Science Uthal 90150, Pakistan*
[*2,] *Department of Computer Systems Engineering and Sciences, Balochistan University of Engineering and Technology, Khuzdar 89100, Pakistan*
[3,] *Department of Software Engineering, Balochistan University of Engineering and Technology, Khuzdar 89100, Pakistan*
[4,] *Department of Computing Institute of Space Technology Islamabad, Islamabad 44000, Pakistan*
[5,] *Department of computer science, University of Management and Technology, Lahore 45000, Pakistan*

[1,2] bijjarazizbaloch111@gmail.com

DOI: https://doi.org/10.5281/zenodo.22057011

Keywords
Intrusion Detection System, Software Defined Networking, Deep Learning, ANN, Ensemble Machine Learning, Anomaly Detection

Article History
Received: 25 January 2026
Accepted: 10 March 2026
Published: 24 March 2026

Corresponding Author: *
Engr. Dr Nooruddin

Abstract
Diabetic retinopathy is a common complication of diabetes that can lead to vision loss if left untreated. Early detection and treatment are therefore essential to prevent irreversible visual impairment. In this study, we explore the use of machine learning algorithms for detecting diabetic retinopathy. Specifically, we compare the performance of five popular algorithms Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Trees (DT), Random Forest (RF), and Gradient Boosting Classifier (GB) using hyperparameter tuning via Grid Search Cross-Validation.
Our experiments were conducted on a dataset of retinal images obtained from diabetic patients. The images were preprocessed to extract relevant features, including blood vessel density and lesion characteristics. These features were then combined with clinical patient information to train and test the models.
The results demonstrate that all five algorithms were capable of accurately detecting diabetic retinopathy from retinal images. However, the SVM algorithm outperformed the other four in terms of both accuracy and computational efficiency. Overall, our findings suggest that machine learning algorithms can serve as valuable tools for the early detection of diabetic retinopathy. Notably, the Random Forest algorithm also showed strong performance and may be particularly well suited for this task.

INTRODUCTION

Diabetic Retinopathy (DR) is a medical condition caused by diabetes mellitus that can cause retinal damage and blood leaks. If not treated promptly, this condition can cause symptoms ranging from mild vision problems to total blindness. Unfortunately, the process of DR is not reversible, and treatment only sustains vision [4,6]. Early detection and treatment of DR can avoid the risk of loss of vision significantly. Unlike diagnosis system with the help of computer, the manual diagnosis of retina funds image by doctor's time, effort and cost consuming, And there's also the risk of misdiagnosis [1,2]. Machine learning has become one of the most popular techniques to improve performance in in various fields, especially medical images analysis and classification [10]. Machine learning method in medical image analysis and they are highly effective and machine learning methods are being used in medical image analysis [15]. For this article, we are going to propose a machine leering model of DR images detection and classification using Machine learning techniques have been reviewed and analyzed. Difference challenging issues that require more investigation are also discussed.

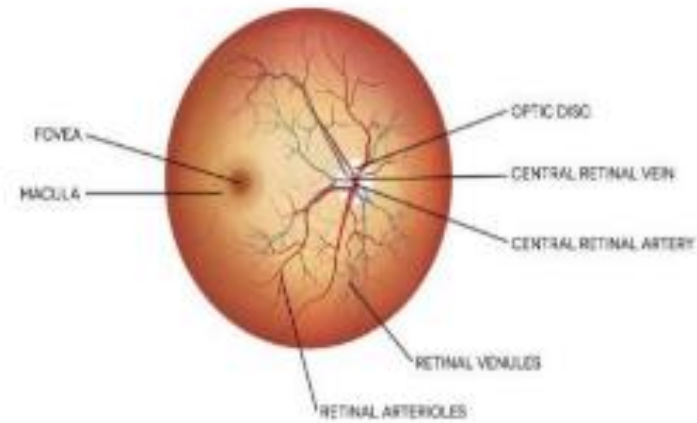


Figure 1

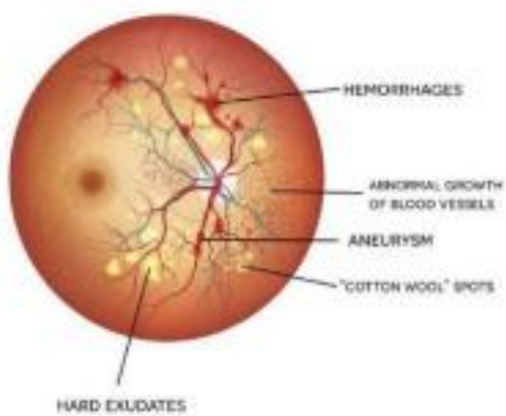


Figure 2

PROBLEM BACKGROUND:

Diabetic Retinopathy is a lethal disease due to which the patient can become blind It is very difficult to detect DR in the beginning. It can be detected by screening method by doctors. Doctors use special cameras to take images of eye retina [3]. that cameras doctor's us are very high coasting and not available ever hospitals in Pakistan specially in Balochistan. The method doctors use is very time consuming and high costing. Due to lack of doctors in most areas This disease is difficult to detect So there must be an automatic way for detection this disease. which save the patients from blindness.

Diabetic Retinopathy is one of the most serious problems affecting the whole world [1]. The attention of several researchers is focused on finding better solutions for early detection of DR, many studies have been conducted and continues in this area with an aim to ease and dearly detection of this disease.

PROBLEM STATEMENT:

One of the most dangerous side effects of diabetes is diabetic retinopathy, which, if unchecked, can result in lifelong blindness. Early detection is one of the biggest problems because it's crucial for effective treatment. Unfortunately, correct staging of diabetic retinopathy is notoriously challenging and necessitates skilled human fundus imaging interpretation. Millions of people could benefit from the detecting process being made simpler. Machine Learning (ML) have already been used in many fields that are related. However, the high cost of large labeled datasets and inconsistency in different clinicians hamper the performance of these methods. Author proposes an automatic method based on machine learning to detect DR from individual human fundus images.

OBJECTIVES:

The goal is to develop a machine learning model that can detect Diabetic Retinopathic images its early stage which can reduce the risk of vision loss.

- Process grayscale retinal fundus images for DR detection.
- For pre-processing Extract key features from images.
- Classify whether the Diabetic Retinopathy is presence or not.
- Find which classifier have higer accuracy.

STUDY SCOPE:

Scope of the task is to plan an easy and time saving mothed for Diabetic Retinopathic detection to avoid risk of vision loss.

- Early detection of diabetic retinopathy
- Diabetic Retinopathy funds image classification
- Help doctor for DR detection
- Can be used in home for selfie testing of Diabetic Retinopathy detection

BENEFITS OF PROJECT:

There are several benefits of this project which includes are.

- Early detection of Diabetic Retinopathy.
- Risk reduction.
- time, effort, and cost saving.
- More efficient and accurate.

RELATED WORK

Diabetic retinopathy is a common complication of diabetes and is a leading cause of blindness worldwide. The use of machine learning algorithms for the detection of diabetic retinopathy has been the focus of many research studies in recent years. Several studies have compared the performance of different machine learning algorithms for the detection of diabetic retinopathy.

In a study by Abràmoff et al. (2016), [17] the authors evaluated the performance of several machine learning algorithms, including SVM, Decision Tree, Random Forest, and others, using retinal images from diabetic patients. The authors found that ensemble-based methods, such as Random Forest, performed better than single classifiers, such as SVM and Decision Trees.

Another study by Gulshan et al. (2016) compared the performance of a deep learning algorithm, based on convolutional neural networks, with that of ophthalmologists in the detection of diabetic retinopathy. The authors found that the deep learning algorithm performed as well as, or better than, the ophthalmologists in detecting diabetic retinopathy.

In a recent study by Al-Kharashi et al. (2021), the authors compared the performance of SVM, Decision Tree, and Random Forest algorithms in detecting diabetic retinopathy using OCT (Optical Coherence Tomography) images. The authors found that Random Forest outperformed the other two algorithms in terms of accuracy and AUC (Area Under the Curve).

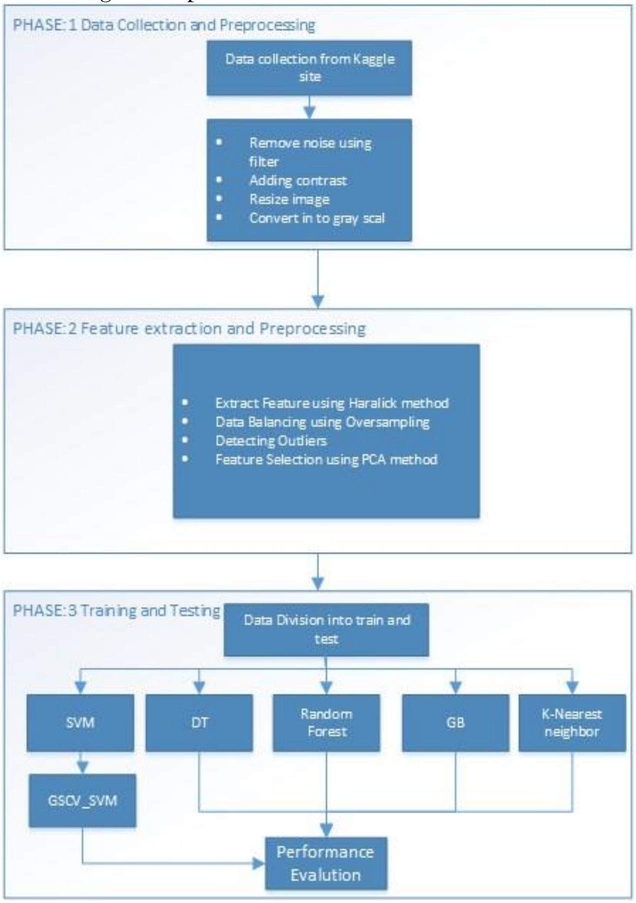
J. Calleja [6] in their work used a two staged method for DR detection, Local Binary Patterns for feature extraction and Machine Learning specifically SVM and RF for classification purpose. They got results by the RF outperformed the SVM with an accuracy of 97.46%.

ANAS BILAL [7] used features like blood vessels, microaneurysms, exudates, and hemorrhages from 500 DR retina images using SVM with an accuracy of 85%.

K. Anant [8] in their t work used texture and wavelet features for DR detection and they achieved 97.95% accuracy.
Gandhi [9] in their work proposed a method for automatic DR detection with SVM classifier by detecting exudates from DR fundus images. Some works try to integrate the manual extraction of features with deep learning feature extraction for DR. one of such work include J. Orlando.et al [10] where CNN with hand crafted feature is used for feature extraction for detecting red lesion in the retina of an eye.

M.Voets [20] in their study work used a Kaggle dataset EyePACS for classification of diabetic retinopathy from retinal fundus images.
Overall, these studies demonstrate the effectiveness of machine learning algorithms in the detection of diabetic retinopathy using retinal and OCT images. While several studies have evaluated the performance of different algorithms, there is still a need for further research to compare the performance of these algorithms on larger datasets and to fine-tune their hyperparameters for optimal performance.

METHODLOGY:
Project Framework follows following three phases.



PHASE: 1
DATA SET:
The used dataset has been downloaded Kaggle website containing total 35000 retinal images from. Kaggle is a public website that allows users to discover and share datasets, explore and develop models in a web-based data science environment, collaborate with other data scientists and machine learning experts, and participate in competitions to solve data science challenges [5]. The size of each image was 1024 X1024 with 3-color channel.

PHASE: 2

PREPROCESSING:

First, researcher have removed the noises using filters then cropped the images and added boundaries at the corners of each image. Then increased the contrast. As the images sizes are images to 512 X 512 then converted them gray scale. Finally used python programing language for coding.

```
In [6]: for x in data:
            img=cv2.cvtColor(x,cv2.COLOR_BGR2RGB)
            plt.imshow(img)
```

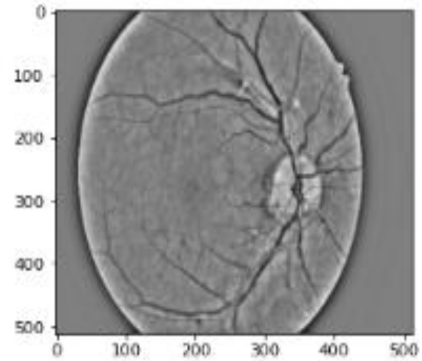


Figure 3

FEATURE EXTRACTION:

There are so many ways to extract features from an image, so far selected Haralick method. In Haralick method the features extracted by four ways to read pixel by pixel,

0 degree

45 degrees

90 degrees

135 degrees

The Haralick feature extraction method returned a total of 14 features of a single image. So, following the Haralick feature extraction from 35000 images I took only 16000 images and extracted features then stored in a single csv file, in csv file every single row consists of a single image feature.The csv file consists 16000 rows and 13 columns, so in this way I completed feature extraction now the data was ready for further cleaning.

For further process, imported the file, read every record line by line, and stored them in data frame.

Out[4]:

	Angular Second Moment	Contrast	Correlation	Variance	Inverse Difference Moment	Sum Average	Sum Variance	Sum Entropy	Entropy	Difference Variance	Difference Entropy	Information Measures of Correlation	Maximal Correlation Coefficient	Cl
0	0.013975	432.245583	0.862999	1578.371849	0.218905	277.857667	5881.241814	7.552682	11.778527	0.000207	4.873644	-0.245760	0.981010	
1	0.013935	423.895577	0.869466	1624.548588	0.221719	278.894712	6074.298774	7.563747	11.751740	0.000210	4.846667	-0.251610	0.982520	
2	0.014364	317.475603	0.908719	1739.971756	0.262472	277.114837	6642.411422	7.303299	11.128170	0.000285	4.486354	-0.282601	0.986739	
3	0.014104	393.084111	0.893128	1839.974581	0.239392	276.602160	6966.814215	7.544245	11.617579	0.000239	4.760562	-0.265472	0.985214	
4	0.002437	355.337411	0.857762	1250.024501	0.159035	277.463441	4644.760594	7.781470	12.444961	0.000159	5.031259	-0.193485	0.963037	
5	0.002123	336.668513	0.871191	1307.641625	0.162263	278.585825	4893.897986	7.831273	12.456961	0.000165	4.993501	-0.200940	0.967143	
6	0.003560	153.117981	0.939120	1258.440172	0.233929	277.295507	4880.642709	7.175625	10.968631	0.000279	4.222967	-0.261204	0.980583	
7	0.004229	185.141837	0.949381	1830.211503	0.251690	285.041829	7135.704174	7.219350	10.860964	0.000297	4.191981	-0.288252	0.986432	
8	0.003323	245.807291	0.923751	1613.237084	0.205297	281.436607	6207.141044	7.462159	11.554202	0.000225	4.575556	-0.246048	0.979485	
9	0.001741	228.926927	0.933745	1728.279835	0.202129	280.568546	6684.192412	7.671974	11.767872	0.000234	4.498089	-0.257674	0.983575	

Figure 4

DATA BALANCING:

First, it has been checked rather data is balanced or not. Further used visualization method and observed that the data consisted with 11649 records against NODR and 4351 of DR so in this case it has been noticed data set is imbalanced. If it considers to take this data set and train, it in resultant model will be a dumb and biased since the majority data is consisting of NODR.

```
<AxesSubplot:>
```

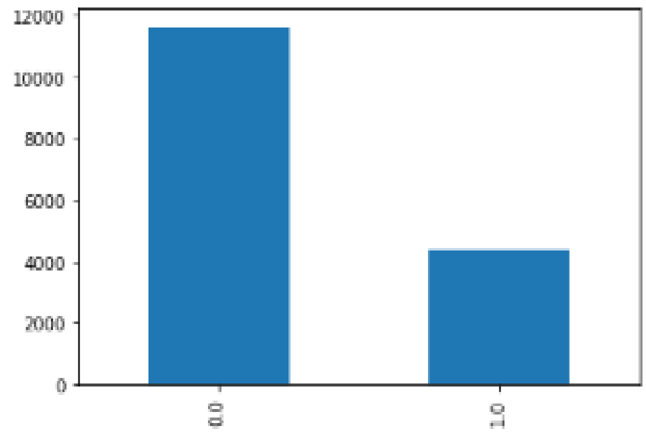


Figure 5

the result show that data sets are not blanced so for this problem we used oversampling method.
In over sampling method it has been checked the highest samples of class, the class NODR had 11649 samples. Now then balance them with ratio of 1:1, so added more samples in class DR. and balanced them. The purpose of balancing of data set is avoiding dumb model and biased model


```
In [28]: pd.value_counts(y_over).plot(kind='bar')
Out[28]: <AxesSubplot:>
```

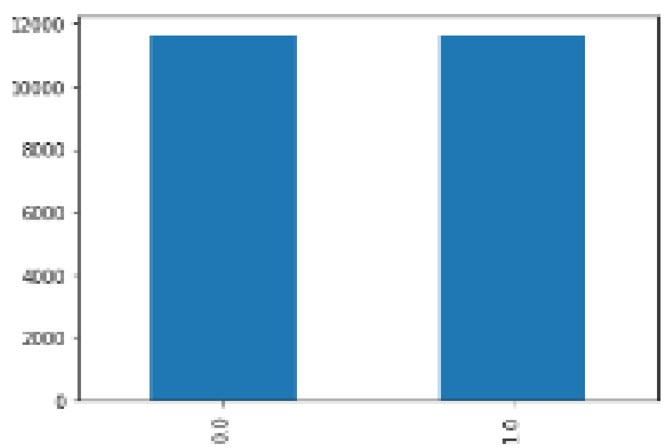


Figure 6

DETECTING OUTLIERS:

What are outliers?

In actually 97-99%, dataset of every feature consists within range. The remaining 1-2%, placed apart from distance such these points called outliers, they are also data points. If these outliers are not treated, they affect the model and main problem model faces overfitting problem. The outliers also cause for overfitting and make the model as dumb model and model becomes complex.

How to detect outlier? The best way to see outliers for human being is visualization; with the help of visualization tool, I can detect the outlier.

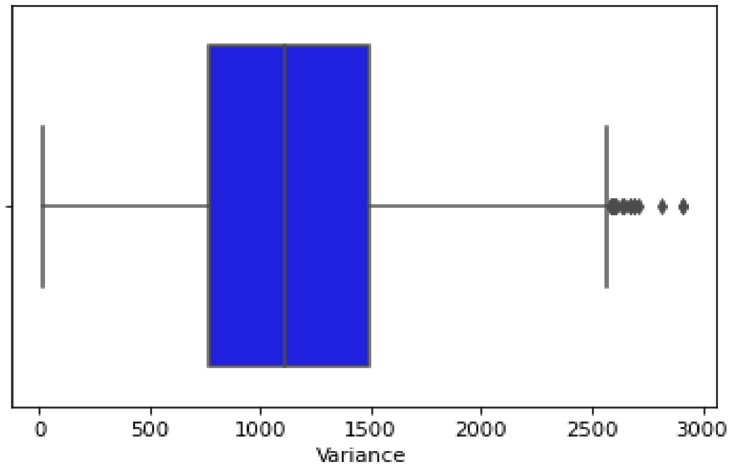


Figure 7

Therefore, researcher used histogram, probability plot and box plot to visualize the features and noticed the outliers. Therefore, in reality every feature will be some outliers so in this case we have to find a technique to resolve the problem of outliers. we used Quantiles. However, why Quantiles? Before explaining Quantiles, let us discuss other for rejection reasons. The IQR and Gaussian used to remove the outliers from data sets, in case of removal the data points the data size will reduce and may be the outliers have good information therefore in order to stop removing any single points since used Quantiles method.

In Quantiles method, it defines the quantiles/limits. Upper limit and lower limit. The lower limit is at fifth Quantile and upper limit is 95th quantile if any point laying below the lower limit, then removed that data points and placed at 5th Quantile same if any point laying at greater the 95th quantile then removed and fit ted at 95th quantile. Therefore, in this way someone will not lose any data points.

So, apply quantile method all outliers are fitted within range and solved the outliers' problems.

```
In [23]: ax = sns.boxplot(x=y["Variance"])
```

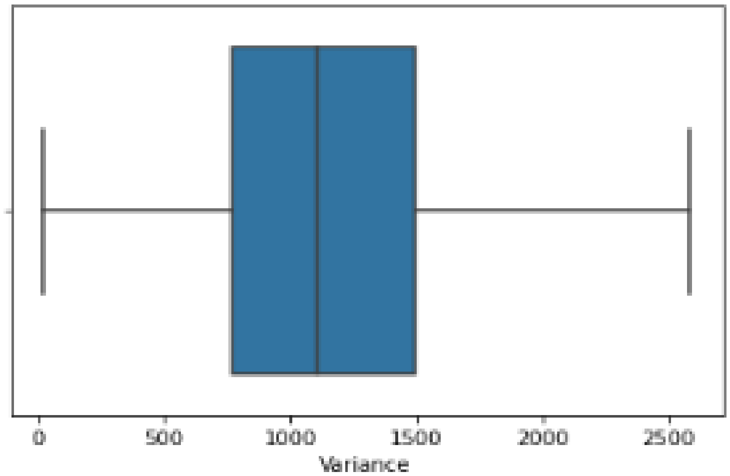


Figure 8

Before detecting outliers

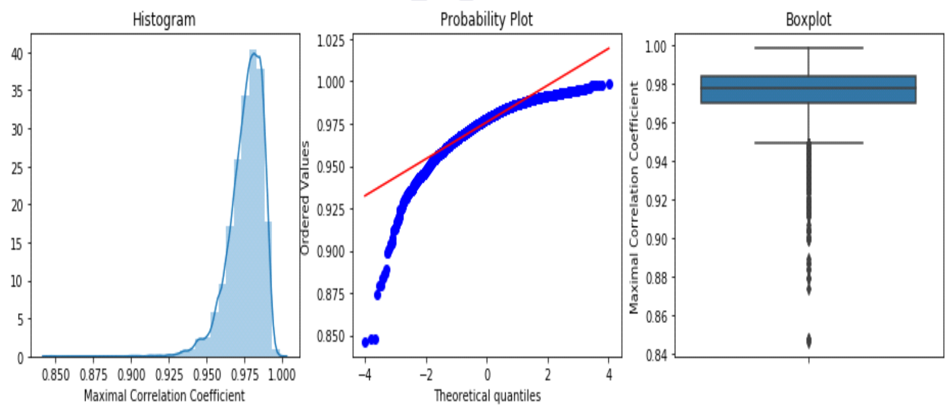


Figure 9

After detecting outliers

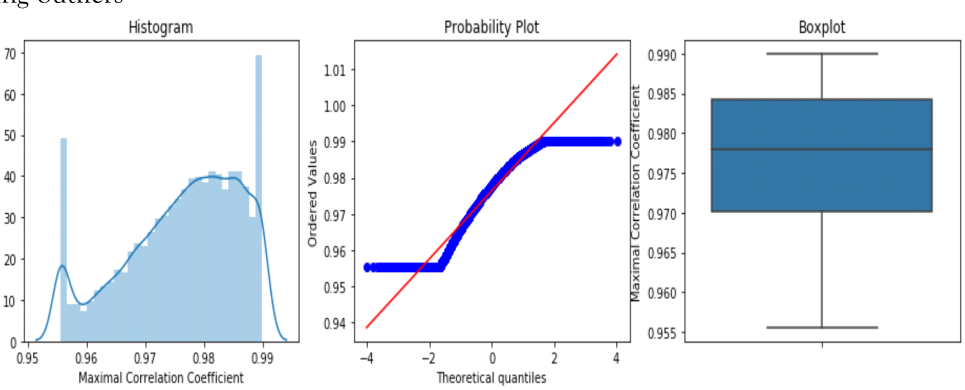


Figure 10

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

FEATURE SELECTION:

One of the problems for overfitting/complex model is using extra features, some features cause multicollinearity due to multicollinearity then used some useless features. In reality systems would not need more computational power to train model.

For this problem, Principal Component Analysis (PCA) feature selection method has been used Applying this, technique it has been noticed the number of features is reduced from 13 to 7.

Before using PCA

```
In [39]: X_over.shape
Out[39]: (23298, 13)
```

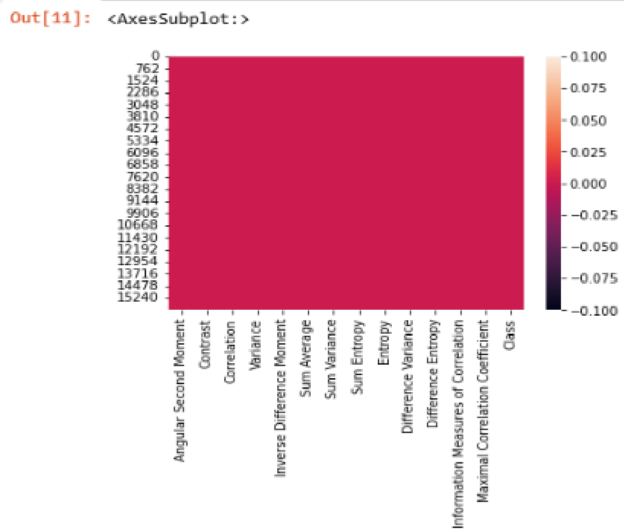
After using PCA

```
In [41]: features.shape
Out[41]: (23298, 7)
```

FINDING NULL VALUES:

```
Out[9]: Angular Second Moment                  0
        Contrast                               0
        Correlation                            0
        Variance                               0
        Inverse Difference Moment              0
        Sum Average                            0
        Sum Variance                           0
        Sum Entropy                            0
        Entropy                                0
        Difference Variance                    0
        Difference Entropy                     0
        Information Measures of Correlation    0
        Maximal Correlation Coefficient        0
        Class                                  0
        dtype: int64
```

Checking Null Values Using Heatmap



The results show that there are no any null values in may data but if there exist then we have removed them as they can comprimse the accuracy of model.

CORRELATION BETWEEN COLUMNS:

Correlation in may data.

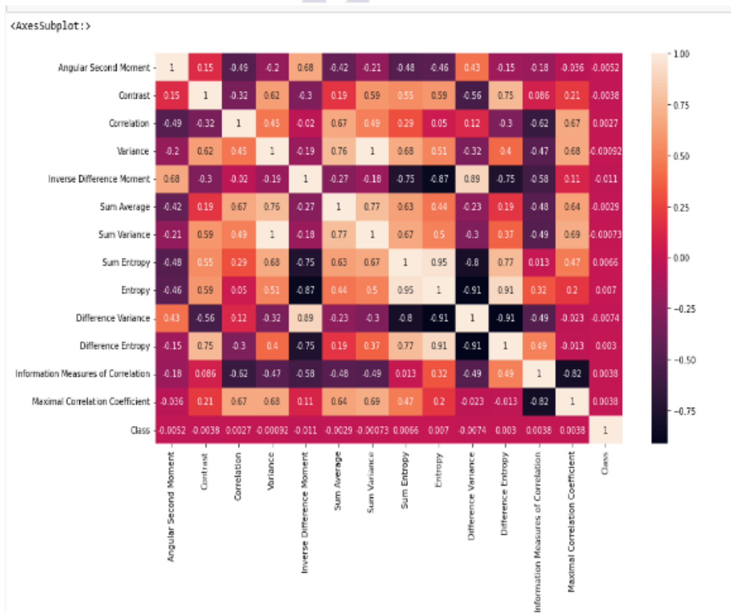


Figure 11

Before training, normalized the data using Min Max method to reduce the dimensionalities and fit the point with range of 0-1.

SCALING USING MINMAX SCALER:

```
In [36]: X[0]

Out[36]: array([0.45443591, 0.7934955 , 0.41629629, 0.60766478, 0.40766645,
                0.59474627, 0.59159011, 0.66138782, 0.68244341, 0.28659708,
                0.76193268, 0.52274557, 0.64944514])
```

DATA DISTRIBUTION

Now the data is completely ready for training. Then divided the data in 75% for training and 25% for testing.

```
In [42]: x_train,x_test,y_train,y_test=train_test_split(features,y_over)
```

UNDER FITTING:

Due to the less data set and insufficient information, the model faces under fitting moreover, it becomes useless and dumb model. The under fitting measured by Bias and variance. Both are errors at some conditions both names are given. At the time of training, the error called Bias and the time of testing the error called variance.

If the model faces high bias and high variance this model called under fitted model in case as someone can see the accuracies are grater then bias and low variance. Therefore, models are not under fitted model.

OVERFITTING:

The main problem in machine learning is overfitting. How to handle this problem to reduce the complexity of model.

What is overfitting? There are several reasons that causes of overfitting. If the model too much learned and never missed single points, it becomes complex model and miss classifies the unseen data. In overfitting conditions are Bias comes less and variance becomes high. How to handle this problem. I used grid search cross validation to made the model more generalize in this way I removed the overfitting problem.

GRID SEARCH CROSS VALIDATION:

It is technique to make the model more generalize form. Grid search actually uses kernel; the kernel increases the low dimension to high dimension. Moreover, separate each class using two points and place the hyperplane between each other.

To finding the two best points, I gave different values to C and gamma, and found the best points. At the end, I got two best points C: 1 and gamma: 10 after that I used five cross validations for improving the models now models are either facing overfitting nor under fitting.

After the training, I tested with SVM model they are giving good result on unseen data.

ROC CURVE (RECEIVER OPERATING CHARACTERISTIC CURVE):

In ML, measuring performance is an essential task. In a classification work, we can rely on an AUC-ROC curve. When we need to test or visualize the performance of a multi-class classification problem, we use the AUC (Area Under the Curve), ROC (Receiver Operating Characteristics) curve. It is one of the most important evaluation measures for testing the learning rate of a classification model. It is also known as AUROC (Area Under the Receiver Operating Characteristics). ROC is a probability curve and AUC represent the degree or measure of separability. It shows the extent to which the model is able to differentiate between classes. The higher the AUC, the better the model is able to predict class 0 as 0 and class 1 as 1. Similarly, the higher the AUC, the better the model can recognize patients with and without disease. An ROC curve (receiver operating characteristic curve) is a diagram showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

- True Positive Rate
- False Positive Rate

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

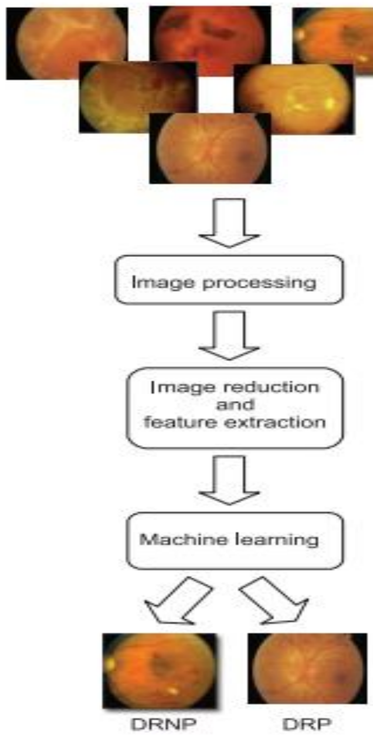


Figure 12

PHASE: 3
TRAINING:
In this study researcher used five machine-learning algorithms to trained different models.

ALGORITHMS:
- Support Vector Machine
- Decision Tree
- Random forest
- Gradient Boosting Classifier
- K-nearest neighbor

SUPPORT VECTOR MACHINE
Result:

	Precision	recall	f1-score	support
0.0	0.56	0.56	0.56	2888
1.0	0.57	0.57	0.57	2937
Accuracy			0.56	5825
Macro avg	0.56	0.56	0.56	5825
Weighted avg	0.56	0.56	0.56	5825

Confusion Metrix:

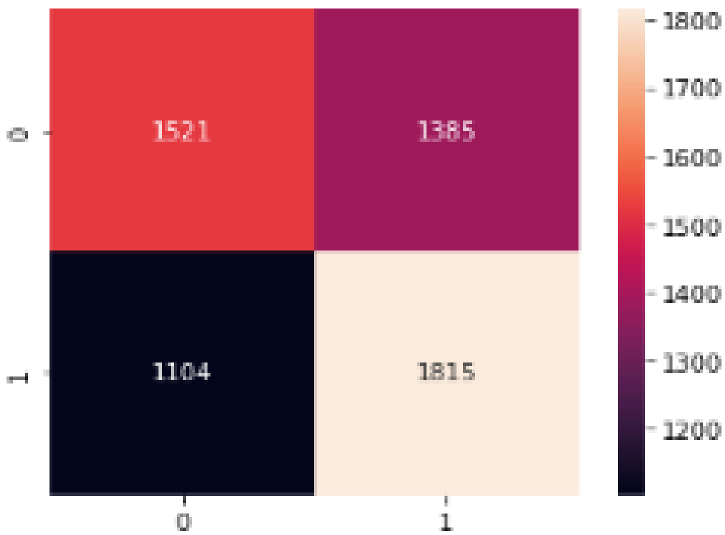


Figure 13

Testing Accuracy: 0.57330
Compression between test and predicted data.

Result:

	Y_test	Y_pred
0	0.0	0.0
1	0.0	0.0
2	0.0	1.0
3	1.0	1.0
4	0.0	1.0
5	0.0	1.0
6	1.0	1.0
7	1.0	0.0
8	1.0	0.0
9	0.0	0.0

Here we check accuracy using legend graph.

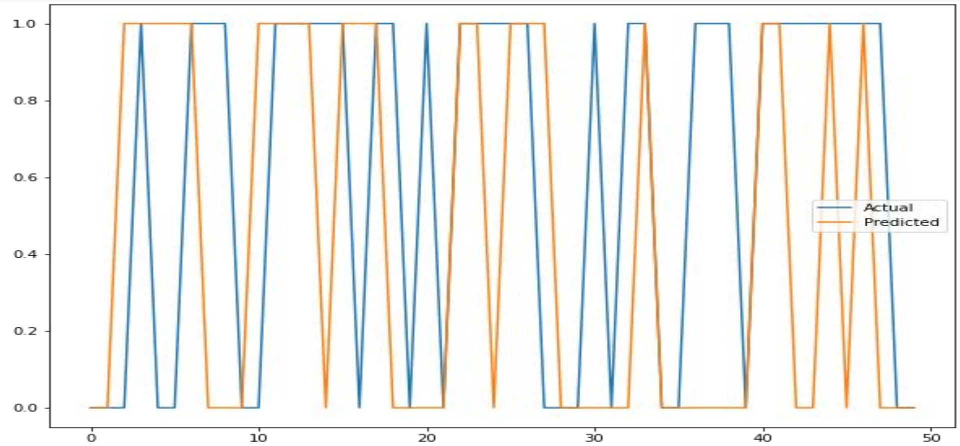


Figure 14

In this graph we can easily see line of actual and predicted are not matching it mean it have low accuracy.

DECISION TREE
Result:

	Precision	recall	f1-score	support
0.0	0.95	0.87	0.91	2867
1.0	0.89	0.96	0.92	2958
Accuracy			0.92	5825
Macro avg	0.92	0.92	0.92	5825
Weighted avg	0.92	0.92	0.92	5825

Confusion Metrix:

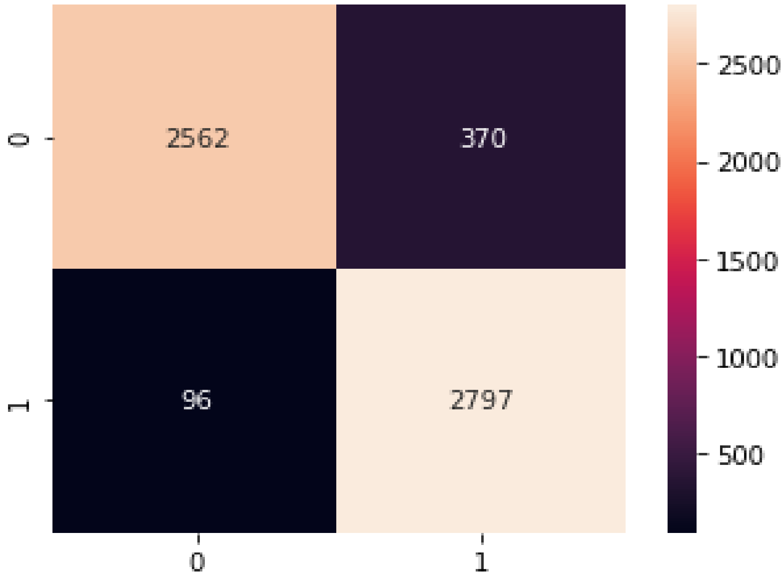


Figure 15

Testing Accuracy: 0.92

Compression between test and predicted data.

	Y_test	Y_pred
0	0.0	0.0
1	0.0	1.0
2	0.0	0.0
3	1.0	1.0
4	0.0	1.0
5	0.0	0.0
6	1.0	1.0
7	1.0	1.0
8	1.0	1.0
9	0.0	0.0

Now we check accuracy using legend graph.

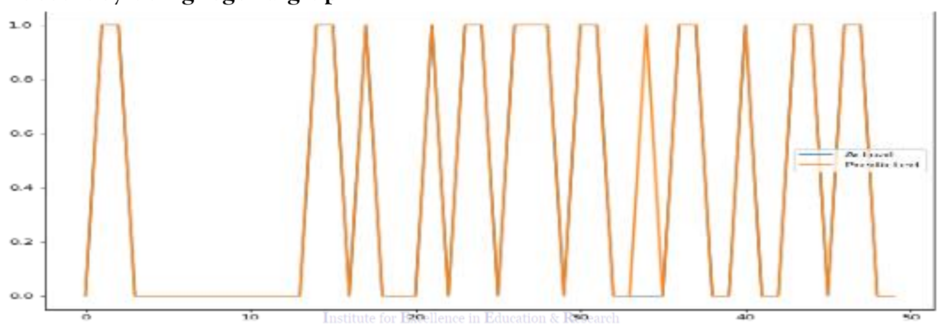


Figure 16

RANDOM FOREST

Result:

	Precision	recall	f1-score	support
0.0	0.95	0.96	0.95	2867
1.0	0.96	0.95	0.96	2958
Accuracy			0.96	5825
Macro avg	0.96	0.96	0.96	5825
Weighted avg	0.96	0.96	0.96	5825

Confusion Metrix

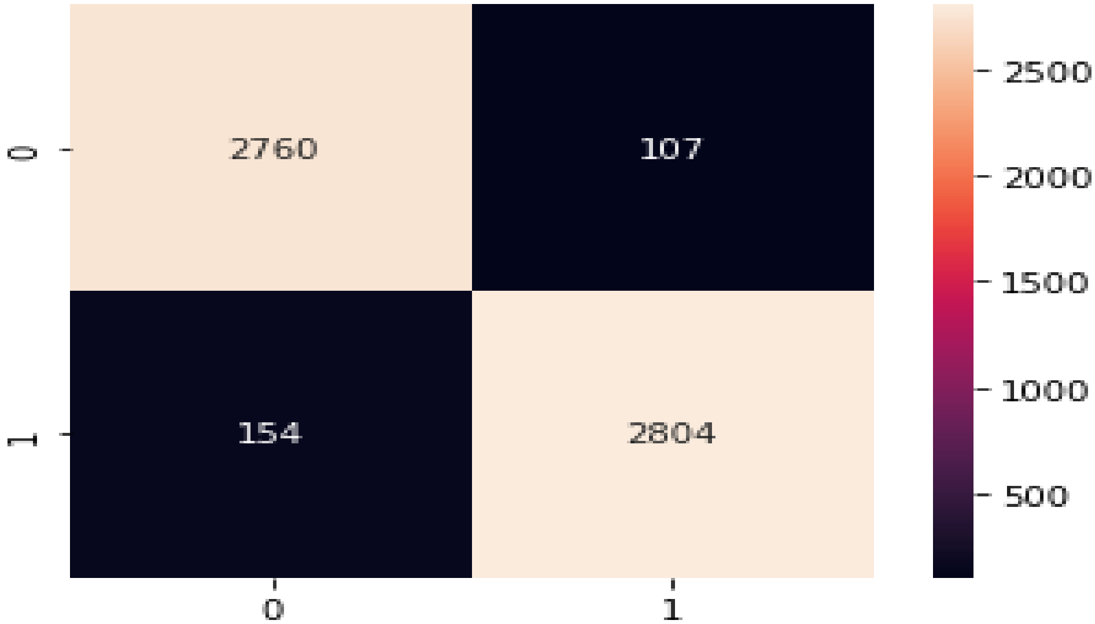


Figure 17

Testing Accuracy: 0.937167381974249
Compression between test and predicted data.

Out[56]:

	Y_test	Y_pred
0	0.0	0.0
1	1.0	1.0
2	1.0	1.0
3	0.0	0.0
4	0.0	0.0
5	0.0	0.0
6	0.0	0.0
7	0.0	0.0
8	0.0	0.0
9	0.0	0.0

Now we visualize accuracy using legend graph.

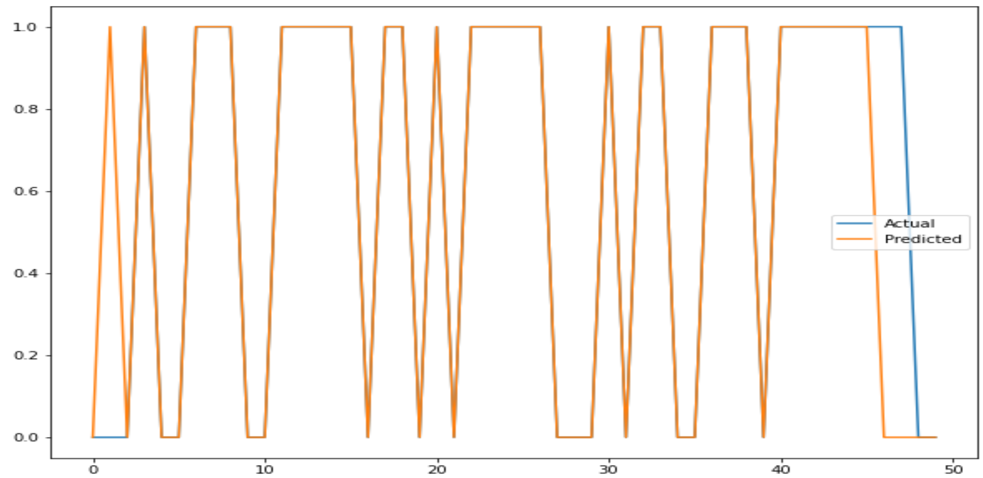


Figure 18

So, in this grape we can see accuracy is much better then previse classifiers

GRADIENT BOOSTING CLASSIFIER

Result:

	Precision	Recall	F1-score	Support
0.0	0.59	0.62	0.60	2867
1.0	0.61	0.59	0.60	2958
Accuracy			0.60	5825
Macro avg	0.60	0.60	0.60	5825
Weighted avg	0.60	0.60	0.60	5825

Confusion Metrix:

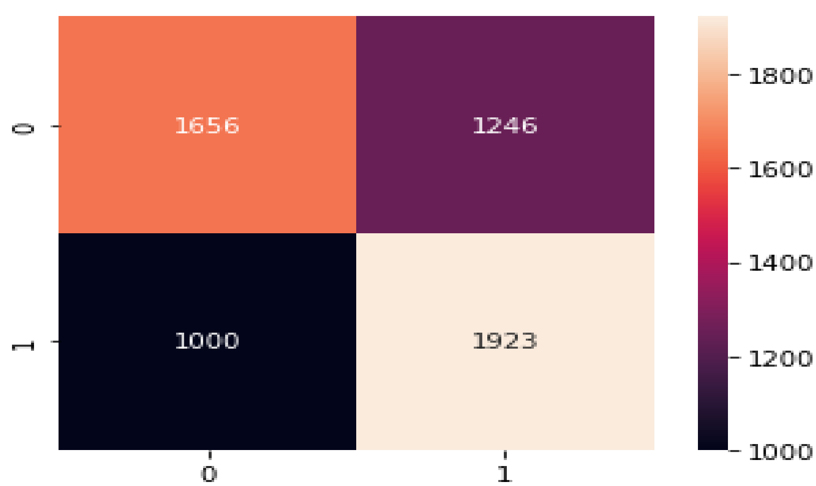


Figure 19

Testing Accuracy: 0.615825
Compression between test and predicted data.

	Y_test	Y_pred
0	0.0	0.0
1	0.0	0.0
2	0.0	1.0
3	1.0	1.0
4	0.0	1.0
5	0.0	0.0
6	1.0	1.0
7	1.0	0.0
8	1.0	1.0
9	0.0	0.0

Now we visualize accuracy using legend graph.

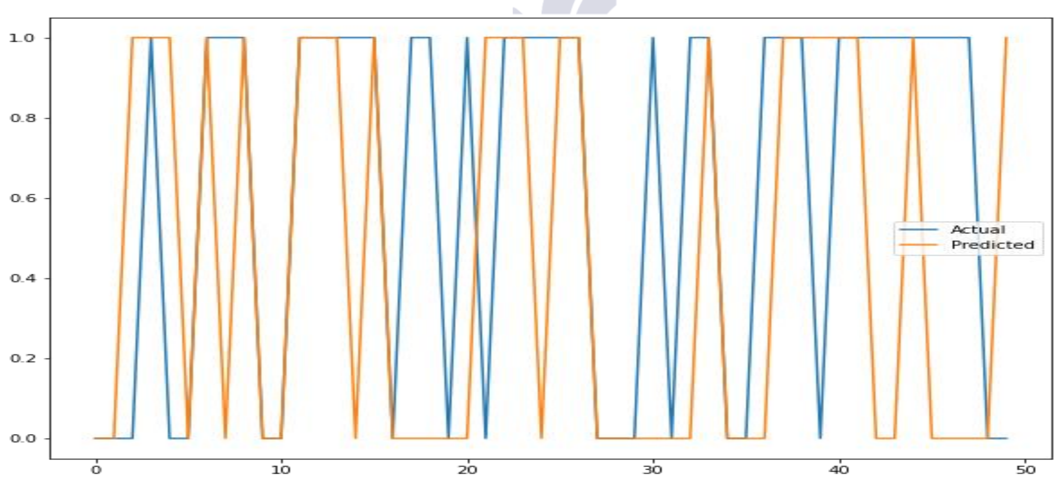


Figure 20

So, in this grape we can see accuracy is much less then decision tree classifier.

K-NEAREST NEIGHBOR

Result:

	Precision	Recall	F1-Score	Support
0.0	0.95	0.87	0.91	2867
1.0	0.89	0.96	0.92	2958
Accuracy			0.88	5825
Macro avg	0.88	0.88	0.88	5825

Confusion Metrix:

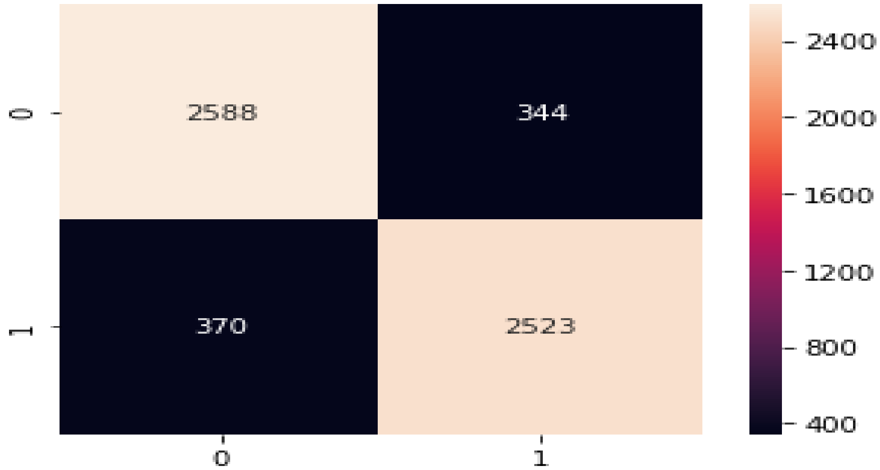


Figure 21

Testing Accuracy: 0.8774248927038627
Compression between test and predicted data.

	Y_test	Y_pred
0	0.0	0.0
1	0.0	1.0
2	0.0	0.0
3	1.0	1.0
4	0.0	0.0
5	0.0	0.0
6	1.0	1.0
7	1.0	1.0
8	1.0	1.0
9	0.0	0.0

we visualize accuracy using legend graph.

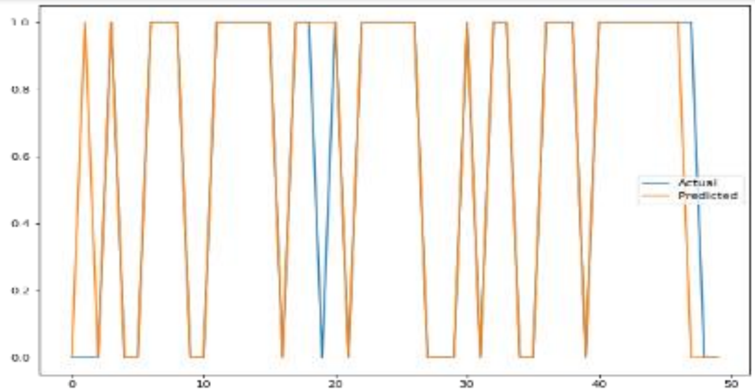


Figure 22

So, in this grape we can see accuracy is much improved.

GRID SEARCH CROSS VALIDATION

It is technique to make the model more generalize form. Grid search actually uses kernel; the kernel increases the low dimension to high dimension. Moreover, separate each class using two points and place the hyperplane between each other.

To finding the two best points, I gave different values to C and gamma, and found the best points. At the end, I got two best points C: 1 and gamma: 10 after that I used five cross validations for improving the models now my models are either facing overfitting nor under fitting.

After the training, I tested with SVM model they are giving good result on unseen data.

GRID SEARCH CROSS VALIDATION USING SVM

Result:

	Precision	Recall	F1-Score	Support
0.0	0.95	1.00	0.98	2902
1.0	1.00	0.95	0.98	2923
Accuracy			0.98	5825
Macro avg	0.98	0.98	0.98	5825
Weighted avg	0.98	0.98	0.98	5825

Confusion Metrix:

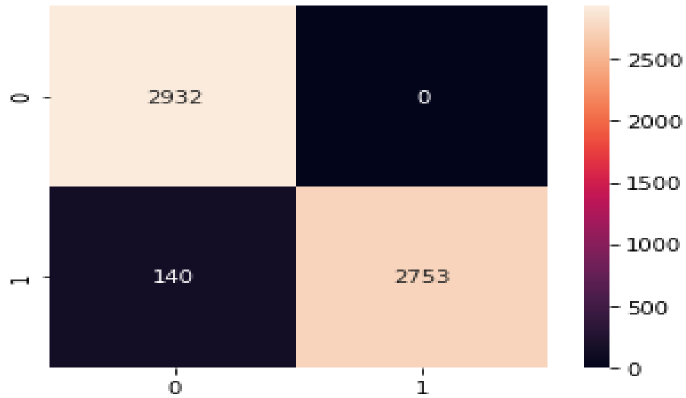


Figure 23

Testing Accuracy: 0.985825

Compression between test and predicted data.

	Y_test	Y_pred
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	1.0	1.0
4	0.0	0.0
5	0.0	0.0
6	1.0	1.0
7	1.0	1.0
8	1.0	1.0
9	0.0	0.0

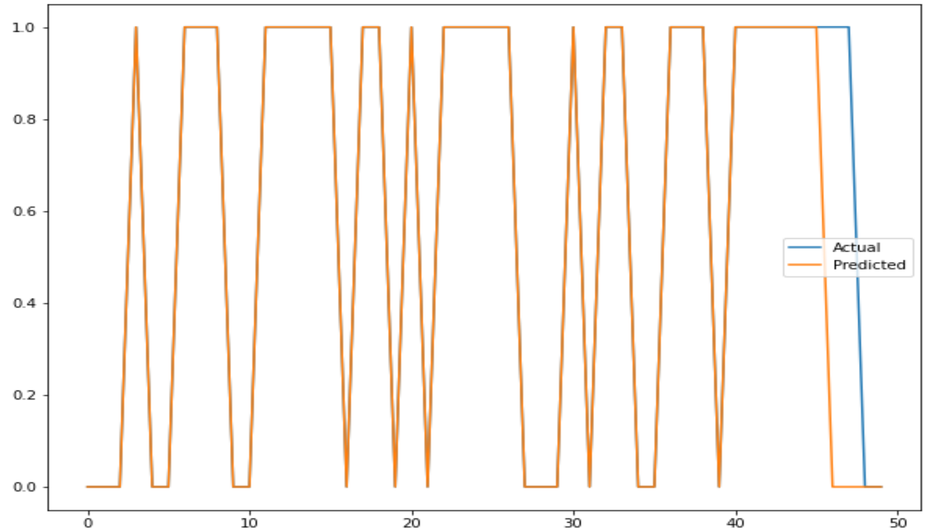


Figure 24

Now we visualize accuracy using legend graph.
Now we can see in this grape after using Grid Search Cross validation filter on SVM geving best accuracy.

ROC CURVE (RECEIVER OPERATING CHARACTERISTIC CURVE)
Result:

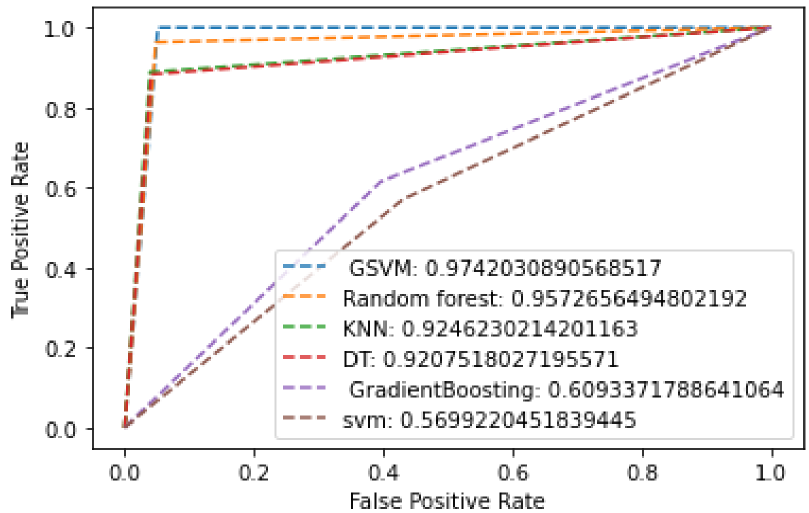


Figure 25

This curve show that the highest accuracy is SVM when we use Grid search cross validation otherwise it is lowest accuracy.

GSVM= 0.97%
RF=0.95%
KN=0.92%
DT=0.92%
GB=0.60%
SVM=0.56

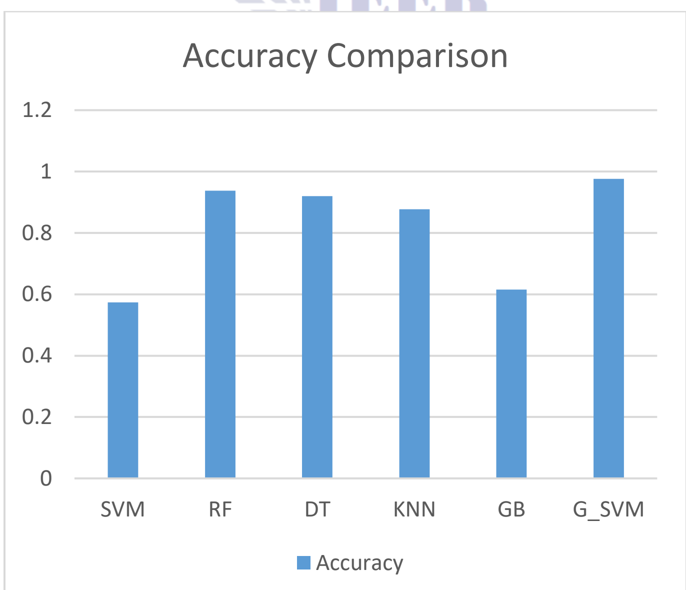


Figure 26

Classifier	Accuracy
SVM	0.5733080357142857
DT	0.92
RF	0.937167381974249

GB	0.615825
K_NN	0.8774248927038627
GSCV_SVM	0.975825

Now, we have classified and compare the result, so from here we know that If filter is used with svm then the result will come high from the rest of classifiers the results. This means that using filter with classifier give us better result.

CONCLUSION:

Due to increase in diabetes current Limits the validity of existing manual testing. New algorithms are taking too long to diagnose any disease. Diabetic retinopathy is a disease Due to which the eyes of a person can be weak or can be without complete Early detection disasters can be of great help to diabetic retinopathy patient and prevent the bad condition like blindness. doctors test this disease through special cameras. With the help of these cameras, doctors take photos of this retina. Retinal fun images can help for automatic diagnosis. There are some symptoms of diabetic retinopathy such as Micro hemorrhages aneurysm and HEM. It is difficult to request them because It's like a normal Latina. We can see diabetic retinopathy with special cameras. With the help of these cameras, we take photos of of the eye retina. Using these retina pictures use can extract feature of DR which use for DR detection.

In this propose work we use five different classifiers comparing which classifiers have better accuracy using these diabetic retinopathy funds images. then we use filter on these classifiers and aging check accuracy rate. finally, we can see using Grid Search Cross validation using SVM has highest result then others. we visualization the accuracy using ROC Curve. The proposed approach is also suitable for small datasets. The new machine learning-based methods are data-intensive but show impressive performance in a variety of classification tasks.

COMPLIANCE WITH ETHICAL STANDARDS

It is declared that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

REFRENCES

Han S., Wang Y., Liao W., Duan X., Guo J., Chen X., H. (2019). The distinguishing intrinsic brain circuitry in treatment-naïve first-episode schizophrenia: Ensemble learning classification. Neurocomputing

G Thippa Reddy., Sweta Bhattacharya., Siva Ramakrishnan S., & Chiranji Lal Chowdhary H. (2020). (ic-ETITE): An Ensemble based Machine Learning model for Diabetic Retinopathy Classification

ANAS BILAL., GUANGMIN SUN., YU LI., SARAH MAZHAR., & ABDUL QADIR KHAN H. (February 10, 2021). (National Natural Science Foundation of China):Diabetic Retinopathy Detection and Classification Using Mixed Models for a Disease Grading Database

Gazala Mushtaq ., & Farheen Siddiqui. (School of Engineering Sciences and Technology, India):Detection of diabetic retinopathy using deep learning methodology H.(November ,15, 2020)

Pilar Perez Conde, Jorge de la Calleja, Antonio Benitez, Ma. Auxilio Medina (Universidad Politecnica de Puebla ´ Puebla, Mexico): Image-Based Classification of Diabetic Retinopathy using Machine Learning H. (April 10, 2018).

G. T. Zago, R. V. Andreão, B. Dorizzi, & E. O. Teatini Salles, (Diabetic retinopathy detection using red lesion localization and convolutional neural networks) Comput. Biol. Med., vol. 116, p. 103537, 2020, doi: 10.1016/j.compbiomed.2019.103537.

Roychowdhury, S., Koozekanani, D. D., & Parhi,(IEEE journal of biomedical and health informatics) : An ensemble-based system for automatic screening of diabetic retinopathy. Knowledge-based systems H.(May 8, 2019)

Arenas-Cavalli, J. T., Rios, S. A., Pola, M., & Donoso .(Procedia Computer Science): (March 12,2015)

Yao, L., Zhong, Y., Wu, J., Zhang, G., Chen, L., Guan, P., ... & Liu, L.(Targets and Therapy): Multivariable Logistic Regression And Back Propagation Artificial Neural Network To Predict Diabetic Retinopathy. (September 10, 2019)

Vaitheeswaran, K., Subbiah, M., Ramakrishnan, R., & Kannan, (Journal of Applied Statistics): Acomparison of ordinal logistic regression models using Classical and Bayesian approaches in an analysis of factors associated with diabetic retinopathy (January 1, 2019)

Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, & F. Meriaudeau, (Indian diabetic retinopathy image dataset (IDRiD)): A database for diabetic retinopathy screening research H. (July25, 2018)

M. A. U. Khan, T. A. Soomro, T. M. Khan, D. G. Bailey, J. Gao, and N. Mir, (Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)): Automatic retinal vessel extraction algorithm based on contrast-sensitive schemes H. (November 5, 2016)

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; (Amrican medical assocation genral)“ Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs(December 13, 2016)

T. A. Soomro, M. A. U. Khan, J. Gao, T. M. Khan, and M. Paul (Signal, Image Video Process., vol. 11, no. 8): Contrast normalization steps for increased sensitivity of a retinal image segmentation method H. (November 20,2017)

S. Gayathri, A. K. Krishna, V. P. Gopi, and P. Palanisamy, (IEEE Access, vol. 8): Automated binary and multiclass classification of diabetic retinopathy using Haralick and multiresolution features H. (December 12, 2020)

Michael David Abramoff,Yiyue Lou,Ali Erginay,Warren Clarida(Invest Ophthalmol Vis Sci): Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning(Octumber 01, 2016)