Diabetic Retinopathy using Morphological Operations and Machine Learning

Mr. JaykumarLachure
M.Tech.of CSE
GCOEAmravati,
jaykumarlachure@gmail.com

Mr.A.V.Deorankar
HOD of IT
GCOEAmravati,
avdeorankar@gmail.com

Mr. Sagar Lachure
Asst.Profof CSE
YCCENagpur,
Sagarlachure1@gmail.com

Miss. Swati Gupta
M.Tech.of CSE
RKNEC, Nagpur,
sswatigupta19@gmail.com

Mr. Romit Jadhav
ME of CSE
PR Pote COEM
Amravati,
romitjadhav@gmail.com

Abstract - Diabetic Retinopathy that is DR which is a eye disease that affect retina and further later at severe stage it lead to vision loss. Early detection of DR is helpful to improve the screening of patient to prevent further damage. Retinal micro-aneurysms, haemorrhages, exudates and cotton wool spots are kind of major abnormality to find the Non- Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR). The main objective of our proposed work is to detect retinal micro-aneurysms and exudates for automatic screening of DR using Support Vector Machine (SVM) and KNN classifier. To develop this proposed system, a detection of red and bright lesions in digital fundus photographs is needed. Micro-aneurysms are the first clinical sign of DR and it appear small red dots on retinal fundus images. To detect retinal micro-aneurysms, retinal fundus images are taken from Messidor, DB-rect dataset. After pre-processing, morphological operations are performed to find micro-aneurysms and then features are get extracted such as GLCM and Structural features for classification. In order to classify the normal and DR images, different classes must be represented using relevant and significant features. SVM gives better performance over KNN classifier.

Keywords –diabetic retinopathy, micro-aneurysms, exudates, SVM, KNN, NPDR, PDR.

I. INTRODUCTION

Now a days, there are a lot of Diabetic patients that are suffered from diabetic retinopathy (DR). Most of patients got blindness due to DR. DR is the most common disease of eye which could affects 80-85 percents of the patients who have suffered from diabetics for more than ten years. According to clinical test more than 10% patients with diabetic have vision problem. The clinical test report said that early detection of DR could reduce the chance of vision loss of the patients at least 90 percents, if they are get treat their eyes properly. If the patients get suffer from diabetes long time, then there is higher chances to suffer from DR.[1]

In biomedical application, automatic diagnosis of DR using fundus image could help ophthalmologists easier for detection of diabetic retinopathy severity level. In traditional method it is very hard to detect DR and patient also suffered as the conventional method takes long time. Diabetic Retinopathy is diseases that further lead to blindness. DR is occurred due to the leakage of protein and blood in the blood vessels. It also caused due to high glucose level occurred in blood which can damage the small blood vessels that provide nutrients and oxygen to the retina. Diabetic Retinopathy can be classified into two main categories which are Non-Proliferative Diabetic Retinopathy (NPDR), which is an early stage of DR. In this NPDR, blood vessels inside the retina leak blood or fluid which can cause swell in the retina and the blood vessels which get weak and next stage is Proliferative Diabetic Retinopathy (PDR). This stage can lead to blindness in diabetic patients. DR get affects the blood circulation system of body so, large numbers of new blood vessels are grow around the retina. The new blood vessels are weak and fragile that affects the retina. Further, the new blood vessels have scar tissue that cause detachment or wrinkle to retina. This stage can lead to damages in the optic nerve. Optic nerve is a nerve that carries image from eyes to brain.

This paper is also focus on detection of exudate and micro-aneurysms. That can help for detection of early diabetic retinopathy. Exudates are normally appeared nearer to the damaging capillaries within the retina. The exudates occurred in retina due to the leakage of proteins and lipids from the bloodstream into the retina [3]. In mild NPDR Micro-aneurysms are occurred, it could occurred due to swelling of tiny blood vessels called small area of balloon. The conventional screening process involves excessive dilation of pupil [3]. So, automatic screening process is required for easily detection of DR. In this project we use messidor dataset and diaretdb1 dataset for early detection of DR. In this we are using morphological operations. The figure 1 shows that normal image and image that cause with DR.

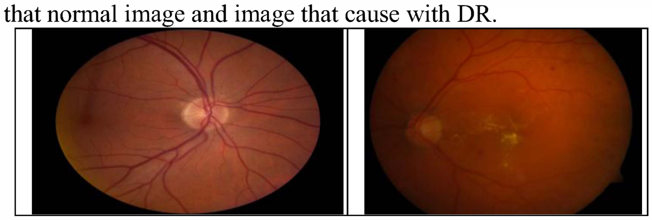


Fig. 1. Normal Image Fundus diabetic retinopathy image

Sohini Roychowdhury *et.al.*[1] proposed three stage algorithm for automatic detection and classification. For automated detection, novel two-step hierarchical binary classification is used. For classification purposed GMM, SVM, KNN and ADABOOST methods are used. They take 30 top features like area, variance of Ired channel, Igreen chaneel, I sat of object, major and minor axis length, Mean pixels for Igreen, Ired and Intensity, solidity etc.

Mahendran Gandhi *et.al.* [2] has proposed diagnosis of DR using morphological operations like erosion followed by dilation for detection of exudates. Then GLCM features are calculated. Before that they have used preprocessing operation like color space conversion, image restoration and enhancement operation .After feature extraction image get segmented and which was applied to svm and knn classifier to classify the image according to its severity grade. In this SVM classifier is used to evaluate training data to find a best way to classify images into different cases like mild, moderate or severe retinal images. This diagnosis has been performed on images, which is in .jpg format captured by retinal fundus cameras. And result was matched with that manually outlined by the ophthalmologist.

Li Tang *et.al.* [3] introduced A novel splat feature classification method which was used for the hemorrhage detection in retinal fundus images. The whole retinal color images are covered and partitioned into non-overlapping segments. Each partition is called as splat which contains pixels with spatial location and same color. Each splat consists of retinal features which is extracted and are compared with variety of filter bank interactions with neighboring splats and texture information and shape. Splat features is selected by a filter approach and then apply wrapper approach. Given splats along with their associated feature vectors and reference standard labels, these are passed to classifier for training purpose to detect required objects. Messidor publically available dataset used in this project. Through this 0.96 was achieved for area under the receiver operating characteristic curve at the splat level and also 0.87 was achieved at the image level.

Dr. R. Geetha Ramani *et.al.* [4] proposed a comparative study between two algorithms and compares the result of both algorithms. This comparison was happened between two data mining algorithm *i.e.*. C4.5 decision tree algorithm verses random tree algorithms. In this comparison decision tree gives better result than random tree algorithm. C4.5 algorithm gives 72% accuracy and random tree gives 65% accuracy.

Jagadish Nayak *et.al.* [5] Proposed comparative classifier using two classifier that are Bayesian statistical classifier and artificial neural network for classification of funds images in dataset. It used neural network for classification of severity of deceases and it gives better result than Bayesian statistical classifier.

L. Giancardo *et.al.* [6] introduced a new algorithm which gives good result and its accuracy. This algorithm was the reconstruction algorithm which contain three steps that was Preprocessing containing common operation like grey scale conversion, noise filter, edge detection, histogram acquisition. Second step that it was used that restoration and the last step contain Naïve- Height map reconstruction step. This paper was work on Reconstruction concept using registration method.

Luca Giancardo *et.al.* [7] introduced a new methodology for diagnosis of Diabetic macular edema (DME) with help of a new set of features which are based on color wavelet decomposition and automatic lesion segmentation. The single feature vector generated for each image for the DEM diagnosis purpose and thus the feature vector created is based on three types of analysis: Exudates probability map, Color Analysis and Wavelet Analysis. These features are play a key role to train a classifier which is able to automatically diagnose DME through the presence of exudation. The accuracy obtained using proposed algorithm is in between 88% to 94%.

Corla Agurto *et.al.* [8] have Proposed AM- FM texture feature extraction method which in new than usual method. In usual method segmentation process has performed for feature extraction. In this structure get classified according to type of lesions. And In this accuracy has been calculated according to distance method. This method achieves accuracy up 92% and gives good sensitivity and spasticity.

Marwan D. Saleh, C. Eswaran *et.al.* [9] provides an automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection. In this paper we extracts some important features, such as optic disc, fovea, and blood vessels for accurate segmentation of dark spot lesions in the fundus images. Dark object segmentation approach is used to locate abnormal regions such as MAs and HAs. Based on the number and location of MAs and HAs get used to evaluated the extremity level of Diabetic Retinopathy *i.e.* type of DR. The dataset used consist of 98 color images to evaluate the performance of the given work. The proposed system achieves 84.31% and 87.53% values in terms of sensitivity for the detection of MAs and HAs and specificity the system achieves are 93.63% and 95.08% values in terms specificity respectively. In the development of automated screening systems the given system leads to Reliable detection of retinal hemorrhages.

LI Yafen *et.al.* [10] have proposed a new method using different image processing techniques such as image enhancement, morphological image processing and texture analysis. For the classification purposed SVM classifier used. It gives accuracy of 89% then the sensitivity is 90% and specificity is 95%.Proposed paper worked on accuracy of classifier and for that they using Directdb dataset for fundus image .

Keerthi Ram *et.al.* [11] In this paper they mainly focused on clustering based methods to segment the exudates like feature in fundus image. And extract the features were multi space and color space feature. In this paper mainly focus on processing time factor so due to this its speed is faster than any other techniques. In this it archives accuracy up to 89.6% and gives positive prediction values gives 87%.

Alireza Osareh *et.al.* [12] Approach of this paper is same but in this paper they are performing all operations on a color image. In this paper they has been used the c-mean clustering and color normalization technique for preprocessing operations. In this color image get segmented using fuzzy and feature of retinal image get extracted using genetic base algorithm. This approach gives 93.6% accuracy and 92.2% sensitivity which is good.

A new contextual clustering algorithm have been proposed by C. JayaKumari, and R. Maruthi *et.al.*[13] to detect the presence of hard exudates in the fundus images. First the pre-processing stage, then segment the exudates has been done through proposed algorithm. Features extraction is done from the segmented regions which results into the standard deviation, mean, intensity, edge strength and compactness. These extracted features are given as inputs to Echo State Neural Network (ESNN) to differentiate between the normal and pathological image. The dataset consists of a total 50 images have been used to find the exudates. Out of 50, 35 images used to train the ESSN which consist of both normal and abnormal and the remaining 15 images are used to test the neural network. The proposed algorithm achieved 93.0% sensitivity and 100% specificity in terms of exudates based classification.

The remaining section describes: Section II overall methodology for detection of exudates and micro-aneurysms through preprocessing and morphological operations on image. Severity of diseases gets classified into different grade using SVM and KNN. Sections III represent the conclusion of this paper. In Proposed work we are performing following steps which are as follow:

II. METHODOLOGY

In given proposed methodology as shown in Fig. 2 shows that image is get input from given data set pre-processing methods are applied then morphological operations are performed to identify exudates and micro-aneurysms. Finally, by applying multiclass SVM and KNN classifier giving severity or grade of abnormality. For given methodology the input images are taken from MESSIDOR, Diabeticret DB1.

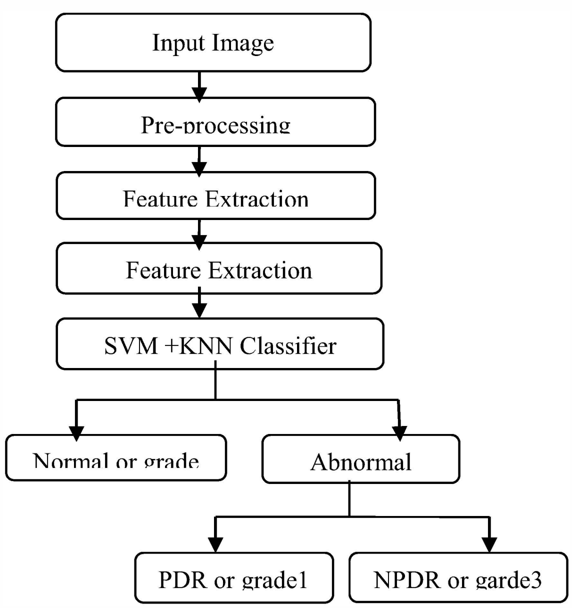


Fig 2. Proposed method development diagram.

A. Preprocessing

The input image of size 2240 X 1488 pixels in .tiff format is used for pre-processing stage and then applied to segmentation method. In preprocessing stage, the image is get rectified from the problems such as blurring, non-clarity and size. In this stage resizing of the image is done and then color space conversion problem, image restoration and finally enhance the image. In color space conversion the input color fundus image is converted into hsi model . HSI means hue saturation and intensity. In this color model space, the intensity component get decoupled from color which carry information (hue and saturation) in color image hence it is an ideal tool for image processing. Gray scale model for processing purposed and all the futures are get extracted from gray color fundus images. gray image is more suitable than color image as intensity adjustment problem.The transformation equations used in the conversion of RGB to HSI are

$$h' = \begin{cases} undefine\ if\ c = 0 \\ \frac{Green-Blue}{C} mod\ 6\ if\ M = R \\ \frac{Blue-Red}{C} + 2 \quad if\ M = R \\ \frac{Red-Green}{C} + 4 \quad if\ M = B \end{cases} \qquad \{1\}$$

$$h = 60^{\circ} * h' \qquad \{2\}$$

and the saturation component is given by,

S = 1-3/[(Red +Green +Blue)*min(Red, Green, Blue)] {3}

Finally intensity component is given by:

I = 1/ [3*(Red + Green+ Blue)] {4}

The converted images are then filtered to remove noise like pepper and salt occur during image acquisition using hybrid median filter. Hybrid median filter smooth the quality of image and reduced the noise appears due to thickness and thinness of boundaries of features and also provide better edge corner prevention. The CLACHE means the contrast limited adaptive histogram equalization is performed after filtering for contrast enhancement, which improve the quality of images. The figure below shows the histogram equalization and image after preprocessing.

Fig. 3. Original Fundus Image

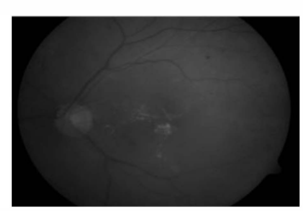


Fig.4. HSI model image

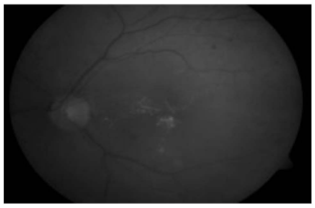


Fig.5. Pre-processed Fundus Image

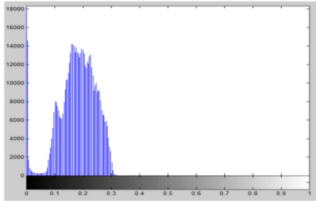


Fig.6. Histogram of pre-processed Image

B. Detection of Micro-aneurysms and Exudates

1) Candidate extraction: In candidate extraction Morphological operations such as erosion, dilation, closing, and opening are performed for finding micro-aneurysms. Then invert image method is applied to for inverting the image. Then holes are filled in the image.

2) Optical Disc Elimination: The optical disc is brightest part of the eye in fundus images and approximately its shape is just like oval or elliptical. In color images optical disc appears like yellowish or white part region. Exudates somewhat have similar intensity value like optical disc. So, it is very essential to remove the optical disc from fundus image. The optical disc then masks and removed using region properties and area finding. After preprocessing edge detection algorithm is applied for detection of optical disc and blood vessels. Canny edge detection is used for counter detection [3]. The canny edge detects the edges where intensity of image is getting changes. Canny edge detection algorithm preserve all local maxima the gradient for enhances blurred edges, through this it detects optimal boundaries of features. After that the mask image is created and then mask image is get subtracted from the edge detected image. The logical black and white function is used to create and then invert the image to create mask image as shown in Fig.7

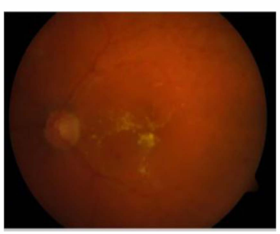


Fig. 7. Dilated Image

3) *Blood Vessels Extraction and Removal*: For detection of exudate remove blood vessel and optical disc is must from retinal image, because it have similar concentration level like exudates. By mistake ophthalmologists can take blood vessel as exudates. Dilation operation on the intensity image helps to remove high level of contrasts vessels. By using Structure element dilation operator is used to fill the small holes. Structure elements (SE) are in different shapes like diamond, disc, round, rectangle *etc.*, but here SE described flat disc shaped structure which is used to remove the optical disc and blood vessels. If SE start from bright pixel, there will be no change and it move to next pixel to next pixel. But if SE start from black pixels then image is covered by SE Dilation which adds pixels to pixels and create boundaries of objects in an image. To remove pixels on object boundaries erosion method is used. Erosion followed by dilation using the same structuring element for both operations is called opening operation.

$$X \oplus Y = \{Z|(Y)_Z \cap X \neq \phi\} \qquad \{5\}$$

Disk Shaped SE (Q) on P, Erosion operator is used to remove completely blood vessels from images

$$X \odot Y = \{Z|(Y)_z \leq X\} \qquad \{6\}$$

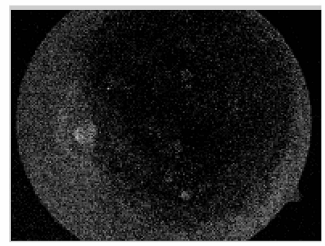


Fig. 8. Blood Vessels Removed Image

4) Identify Exudates: After removing blood vessels and optical disc from image, exudate is detected by using morphological operation closing. This closing operation applied on eroded image. In closing dilation is followed by erosion operator. So it detects the exudates.

$$X \cdot Y = (X \oplus Y) \odot Y \qquad \{7\}$$

The result shows that exudates are extracted in given fig. by using morphological operations.

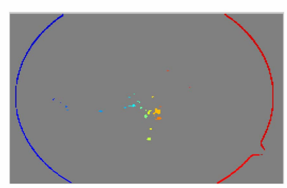


Fig. 9.Exudate area detected image

5) Detection of micro-aneurysms: For micro-aneurysms detection opening operation is performed in which erosion is followed by dilation where as the micro-aneurysms are appearing as red spot which get swell in the retina. Invert image method is used and then morphological opening operation is performed for detection of micro-aneurysms. So, we can easily count the micro-aneurysms values.

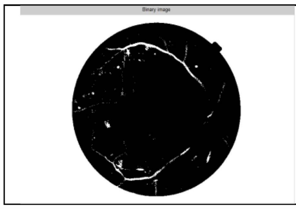
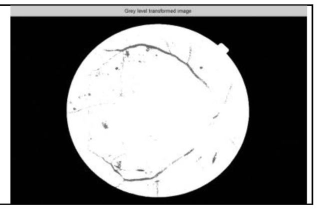
 

Fig.10. Binary imageFig.11. Gray level transformed image

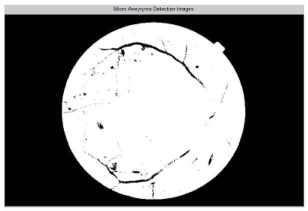


Fig.12.Micro-aneurysms Detection Image

C. Classification

After detecting exudates and micro-aneurysms in color fundus image, the features get extracted. Splat, GLCM (Gray Level Co-occurrence Matrix) and calculated are applied to SVM and KNN classifier. SVM classifier gives better results than KNN classifier. In this we are using splat features and GLCM features likes Energy, contrast, entropy, homogeneity, and area part of exudates and . The formulas for extracted features are given below:

1) *Entropy:* A scalar value representing the entropy of gray-scale retinal image. Entropy feature is a statistical measure of randomness that can be used to characterize the texture of the input retinal image. It is defined as

= sum (p.*log2(q))

Where q is the histogram counts.

$$\text{Entropy}=\sum_i \sum_j G(i,j)\log(G(i,j)) \qquad \{8\}$$

2) *Contrast:* The contrast function enhances the contrast of an `image`. It creates a new gray`cmap`,color map which has an approximately equal to intensity distribution. All of the three elements in each row are identical.`cmap =` contrast`(X)` returns a gray colormap that is the same length as the current colormap.

$$\text{Contrast}=\sum_i \sum_j (i,j)^2 (G(i,j)) \{9\}$$

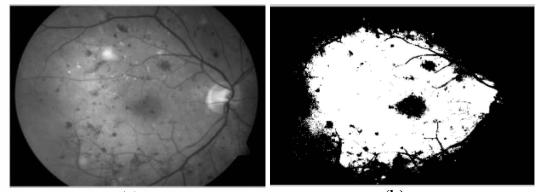
3) *Energy:* Energy is used to describe a measure of "information" when formulating an operation under a probability framework such as MAP (maximum a priori) estimation in conjunction with Markov Random Fields.

$$\text{Energy}=\sum_i \sum_j (G(i,j))^2 \qquad \{10\}$$

4) *Homogeneity:*

$$\text{Homogeneity}=\sum_i \sum_j (G(i,j))^2/[1+|i-j|] \quad \{11\}$$

These features are selected for reducing noise and enhancing the result of classifier accuracy. The result is shown in fig 12. If the MA >15 then it treated as severe and if total 5<MA<=15 so it treated as moderate. If the value is less than 5 then it is normal image [1].



(a) (b)

4Fig.13. (a) original image (b) Severe NPDR

Table I show the diabetic retinopathy severity or grade as normal, mild and serve.

1	MA<5	Normal
2	5<MA<15	Mild
3	MA>15	serve

D. SVM Classifier

The SVM approach basically used for binary classification problems. SVM used for many practical applications though we can solve a multi-class pattern recognition problem. There are two types to solve multi-class problems by using binary SVM classifiers: a) The first one is "one-against-one" approach in which for each possible pair of classification a binary classifier is calculated, each classifier is trained on a subset of the training set containing only training examples of the two involved classes. If n is the number of total classes, all (n-1)*n/2 classifiers are combined through a majority elective system to estimate the final classification. b) Other approach is the "one-against-rest" in this method n different classifiers are constructed, one for each class. Here, the i^{th} classifier is trained on the whole training data set in order to classify the members of class 1 against the rest of the classifier. In the classification stage, the classifier with the maximal output states the estimated class label of the current input vector. While former has too many computations, we adopt the latter one for DR. Though, these are not very sophisticated approaches to solve multiclass problems. The constructing multiclass SVMs is a better alternative provided. In this study, we were construct a two-class classifier over a feature vector $\varphi(\bar{p}, q')$derived from the pair consist of the input features and the class of the data. At testing time the classifier is get choose from the class.

$$y = argmax_{q'} \bar{w}^{T}\varphi(\bar{p}, q')$$

The margin through training is the gap between this value for the correct class and for the nearest other class. These general approaches which can be used to extend and give a multiclass formulation of various kinds of linear classifiers. It is a simple instance of a generalization of classification where the classes are not just a set of independent categorical labels, but might be arbitrary structured objects with relationships defined between them.

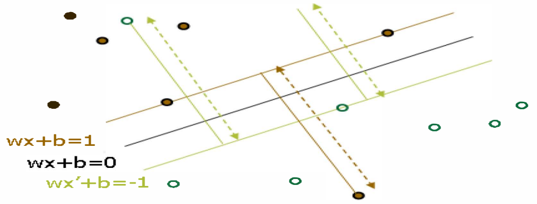


Fig.14. Hyper plane representation for multi-classification

KNN classifier-The k-nearest-neighbor classifier is based on the Euclidean distance between a sample of testing and the training samples. Let x_i $(x_{i1},x_{i2},\dots,x_{ip})$ be an input sample with p features , n(i=1,2,…,n) be the total number of input samples and p (j=1,2,…,p) the total number of features. So, the Euclidean distance between sample xi and xl (l=1, 2,…, n) are defined as

$$d(x_i,x_l) = \sqrt{(x_{i1}-x_{l1})^2 + (x_{i2}-x_{l2})^2 + \cdots + (x_{ip}-x_{lp})^2} \quad \{12\}$$

A Voronoi cell is used to encapsulates all neighboring points that are nearest to each sample and it is defined as

$R_i=\{\mathbf{x}\in R^p: d(\mathbf{x},\mathbf{x}_i)\le d(\mathbf{x},\mathbf{x}_m), \forall i\neq m\}$

Where, R_i is the Voronoi cell for sample $\mathbf{x}_i$, and $\mathbf{x}$ signifies all possible points within Voronoi cell R_i. Characteristic used the KNN classification rule is to assign a test sample from the majority category label of its k nearest training samples.

III. CONCLUSION

In this proposed method both exudates and micro-aneurysms are detected. For exudate detection optical disc and blood vessels are extracted for avoiding false problem to ophthalmologists. For detection of exudates morphological operations are performed like closing. Dilation and erosion operators are used. For micro-aneurysms detection count the number for MA occurred in the image so we can decide the grade of the system. Then features are calculated and feed to both SVM and KNN classifier. SVM classifier is better classifier than KNN. So from the extracted feature it directly concludes the disease grad as normal, moderate and severe. So earlier detection and diagnosis of Diabetic retinopathy help the patients from vision loss and also the severity of disease can be decreases. As combined dataset our specificity is 100% and sensitivity is more than 90% for SVM.

IV. REFERENCES

[1] Sohini Roychowdhury, Dara D. Koozekanani, and Keshab K. Parhi, “DREAM: Diabetic Retinopathy Analysis Using Machine Learning ” IEEE Journal of Biomedical and Health Information, VOL. 18 pp. 1717 - 1728 ,No. 5, September-2014.

[2] Mahendran Gandhi, and Dr. R. Dhanasekaran, “Diagnosis of Diabetic Retinopathy Using Morphological Process and SVM Classifier” Communications and Signal Processing (ICCSP), 2013 International Conference on pp. 873 – 877, 3-5 April 2013.

[3] Li Tang, Meindert Niemeijer, Joseph M. Reinhardt, Mona K. Garvin, and Michael D. Abràmoff, “Splat Feature Classification withApplication to Retinal Hemorrhage Detection in Fundus Images”, IEEE Trans Med Imaging. 2013 Feb;32(2):364-75. doi: 10.1109/TMI.2012.2227119. Epub 2012 Nov 15.

[4] Dr.R.Geetha Ramani1 et al. “Data Mining Method of Evaluating Classifier Prediction Accuracy in Retinal Data” Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on 18-20 Dec. 2012 pp.1 – 4.

[5] Jagadish Nayak *et.al.* “Automated Identification of Diabetic Retinopathy Stages Using Digital Fundus Images”, Journal of Medical Systems, April 2008, Volume 32, Issue 2, pp 107-115.

[6] L. Giancardo, *et.al.*, “Textureless macular swelling detection with multiple retinal fundus images,” IEEE Trans. On Biomedical Engineering., Vol. 58, No. 3, pp. 795–799, Mar. 2011.

[7] Luca Giancardo, Fabrice Meriaudeau, Thomas P. Karnowski , *et.al.*, “Exudate-based diabetic macular edema detection in fundus images using publicly available datasets Medical Image Analysis: Volume 16, Issue 1, January 2012, Pages 216–226.

[8] C. Agurto, *et.al.* “Multiscale am-fm methods for diabetic retinopathy lesion detection,”IEEE Trans Med Imaging. 2010 Feb;29(2):502-12. doi: 10.1109/TMI.2009.2037146.

[9] Marwan D. Saleh, and C. Eswaran, “An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and Has Detection,” Computer Methods and Programs in Biomedicine, 108 (1) 186-96 . DOI : 10.1016/j.cmpb .2012.03.004

[10] LI Yafen*et.al.* “Automated Identification of Diabetic Retinopathy Stages Using SupportVector Machine” proceeding of the 32nd Chinese control conference july 26-28,2013, Xi’an, china

[11] Keerthi Ram and JayanthiSivaswamy, “Multi-space clustering for segmentation of Exudates in Retinal Colour Photographs,” Conf Proc IEEE Eng Med Biol Soc. 2009;2009:1437-40. doi: 10.1109/IEMBS.2009.5332911.

[12] Alireza Osareh, Bita Shadgar, Richard Markham ,”A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images,” IEEE Transactions on Information Technology in Biomedicine - Special section on biomedical informatics ,Volume 13 Issue 4,Pages 535-545, , July 2009.

[13] C.JayaKumari, and R.Maruthi , “Detection of Hard Exudates in Color Fundus Images of the Human Retina,” International Conference on Communication Technology and System Design 2011, Vol.30, pp. 297 – 302,2012.