

Multiple Feature Sets and SVM Classifier for the Detection of Diabetic Retinopathy Using Retinal Images

Ninu preetha N.S

*Department of Medical Electronics,
St.Xavier's Catholic College of Engineering
Kanyakumari, Tamil Nadu, India
ninu.n.s@gmail.com*

Praveena S

*Department of Applied Electronics,
Ponjesly College of Engineering
Kanyakumari, Tamil Nadu, India
Praveena.s.raveendran@gmail.com*

Abstract: Diabetes Mellitus is one of the growing vitally fatal diseases that can affect the patient's sight and its most severe effect is on blood vessels inside the eye called diabetic retinopathy. Due to its significance, a design of an efficient classifier for the detection of Diabetes disease is one of the challenging tasks. In this paper, we have proposed SVM classifier for diagnosing the diabetics from retinal images using two features like optic disc and blood vessel. Initially, the Gaussian filter is used for performing the pre-processing phase. Once the noise free image is generated, the segmentation processed is applied for detecting the both optic disc and blood vessel areas. Then, the relevant features are extracted from the optic disc and blood vessel such as mean, variance, perimeter, diameter, maximum intensity and minimum intensity. Then, the diabetic images are classified from the input images using the proposed SVM classifier. Finally, the experimentation results of the proposed classification technique is carried out using the Stare database, which shows that the SVM classifier can be successfully classifies the diabetic images with better classification accuracy of 96%.

Keywords: *Diabetic retinopathy, fundus image, Optic disc detection, Support Vector Machine (SVM), Gaussian filter.*

1. Introduction

Diabetic retinopathy (DR) is a relentless eye disease which occurs because of diabetes that caused by complete lack of insulin and hence is characterized by increased levels of glucose in the blood [1]. Diabetic patients are easily affected by visual loss and blindness due to the problem of diabetic retinopathy. In order to prevent this serious eye disease, various effective treatments are available to diagnose the progression of disease in early stages. Accordingly, the contrast levels of the retinal images are captured and analyzed by trained individuals [2]. Due to large diabetic population, this manual assessment is very time consuming and costly. So, the automatic detection system is introduced by various researchers for diagnosing the diabetic retinopathy. The diabetic retinopathy disease leads to damage the retinal blood vessels, which cause the outflow of blood and liquefied on the retina and form features such as microaneurysms, haemorrhages, exudates, and venousloops. With sequence, the blockages and injure to blood vessels remove areas of the retina with their blood supply. Importantly, the recognition of diabetic retinopathy is performed by the detection of micro aneurysms(MAs) in the fundus image of the eye based on their grey level and high contrast level [5]. However, the detection of diabetic retinopathy is one of the challenging tasks due to the variation of brightness and contrast level among different patients in different photographs.

For the early stage detection of retinal disease, the blood vessels based segmentation is one of the major tasks in preprocessing step. However, the manual segmentation is very difficult due to multifarious nature of the vascular network. So, various automatic retinal vessel segmentation methods are proposed by researches based on two types of classification techniques named as

supervised and unsupervised classification. The supervised methods require a feature vector for each pixel for classifying the abnormal retinal images. However, most of the extracted features are completely redundant to the classified target. In order to maintain the classification accuracy, the proper selections of features are used in the classification. In this paper, we have proposed support vector machine based classifier for diagnosing the diabetic retinopathy problems. Here, the diabetic retinopathy detection process is performed based on the steps like preprocessing, segmentation, feature extraction followed by obtaining the result through SVM classifier. Initially, the unwanted information from the original images are removed by Gaussian filter and the preprocessed data are applied to segmentation process based on optical disk, blood vessel and damage area using curvelet transform [10]. Then, the segmented data are further processed to extract various features like mean, variance, perimeter, diameter, damage area value, maximum intensity and minimum intensity. Finally, these extracted features are fed to the proposed SVM classifier for determining the result whether the patient is diabetic or not.

The main contribution of this paper is the classification of diabetic detection using SVM based classification. In addition, features extracted from the retinal image include mean, variance, area, perimeter, diameter, maximum intensity, minimum intensity, damage area value and entropy. Finally, the classification accuracy is calculated to determine the effectiveness of the proposed SVM classifier. Then, the organization of this paper is discussed as follows: section 2 discusses about various diabetic related works and problem definition is described in section 3. Section 4 provides the detailed description of the proposed technique. Then, the experimental analysis and discussions about the proposed method is discussed in section 5. Finally, section 6 gives the conclusion of this paper.

2. Literature Review

There has been several works in the literature associated with diabetic detection from retinal images, and we discuss some of those works in this section. M. Kavitha *et al.* [1] have proposed a technique for automated screening of diabetes in a person from the retinal images using blood vessel, optical disk and damage area-based features. Furthermore, they have extracted various features like mean, variance, area, minimum intensity, maximum intensity, perimeter and diameter from the segmented results. Finally, they have used Levenberg–Marquardt algorithm-based neural network to determine whether the patient is diabetic or not. Finally, they have compared the proposed neural network with various existing classifiers such as K-NN, ANFIS, Kohonen network. Amol Prataprao Bhatkar *et al.* [13] have developed MLP classifier for detecting the Diabetic Retinopathy in Retinal Images. In this paper the Multi Layer Perception Neural Network (MLPNN) classifier was presented to classify whether the retinal images as normal and abnormal. Different features of retinal images such as 64-point discrete cosine transform along with nine statistical parameters were extracted from retinal images and used as inputs to the classifier. Keerthi Ram *et al.* [12] have proposed a successive clutter-rejection-based approach for detecting the Diabetic Retinopathy in early stage. In this paper, they have formulated MA detection as target detection in clutter problem, where the probability of occurrence of target is considerably small compared to the clutter. The processing stages were designed to reject specific classes of clutter while passing majority of true MAs, using a set of specialized features.

3. Problem Definition

This section shows the classification problems while detecting the diabetics from retinal images. In [1], they have employed Levenberg–Marquardt algorithm-based neural network to determine whether the patient is diabetic or not. However, the large neural network based classification requires high processing time during the diabetic detection process. In addition, neural network modeling requires greater computation resources for performing the operation. Furthermore, the learning algorithm of LM has the disadvantage that if the value of damping factor is large, the designed Hessian matrix is not used at all.

4. Proposed Methodology

This section shows the detailed description of the proposed Technique for detecting the Diabetic through Retinal Image automatically. Proposed technique makes use of SVM classifier to accomplish the automatic detection task of diabetic with better classification accuracy. Fig. 1 shows the block diagram representation of the proposed technique.

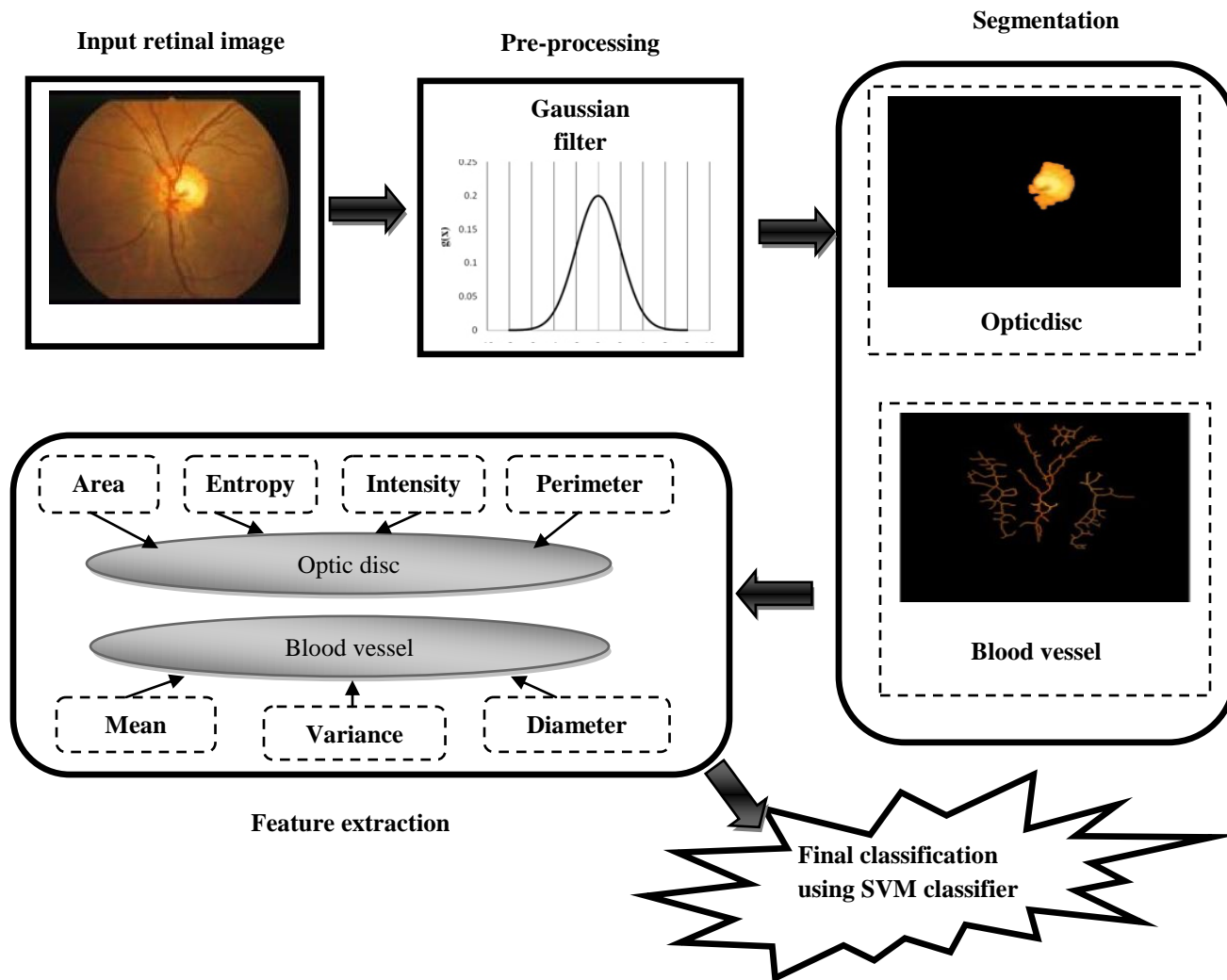


Fig.1. Block diagram representation of the proposed SVM classification technique for diabetic detection from retinal image

4.1 Pre-Processing of an Input Retinal Image Using Gaussian Filter

This section shows the pre-processing phase of an input image, in which various errors from the images are corrected before applying the segmentation process. Basically, the retinal images are affected by noise, errors and imperfections in various manners. So, these input images are cannot fed directly for performing the segmentation process. In order to overcome these drawbacks, we have used one of the efficient filtering techniques called Gaussian filter. The image noises are reduced by Gaussian filter that is used for improving the quality of an image for further processing. Then, the mathematical representation of an one dimensional Gaussian filtering function is represented as follows:

$$G(i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{i^2}{2\sigma^2}}$$

Where, σ is defined as the standard deviation of the distribution, which is defined as the ratio of sample rate and cutoff frequency. Similarly, the impulse response of the Gaussian filter can be represented as follows:

$$G(i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sigma^2 n^2}{2}}$$

From the results of Gaussian distribution function, the convolution matrix is developed. Then, the generated convolution matrix is applied to an original input retinal image. Then, the average weights of the every pixel are evaluated using original pixel and neighboring pixels based on the evaluation of highest weight and lowest weight respectively. Finally, the quality of an input image is improved by using the Gaussian filtering process with the help of developed noise free images.

4.2 Segmentation

Once the pre-processing stage is completed, the segmentation process is carried out for dividing the image into segments [3]. The image segmentation is mainly based on the predetermined features of an image pixel. In this paper, the segmentation process is carried out based on both optic disc and blood vessel that can be discussed in detailed manner as follows.

4.2.1 Optic Disc Based Segmentation

Basically, the blood vessels and optic nerves are enters into the retina of human eye through the Optic disc region that is the brightest part of the normal fundus images. In this paper, we have used curvelet transform for optic disc segmentation and the segmentation process is improved by morphological operation. The sequential processing of optic disc segmentation is carried out in following ways. Initially, the input image is converted to grey scale image and the curve like structure are efficiently detected by curvelet transform [10]. After performing the segmentation process, the morphological operation is carried out for an efficient segmentation.

a) RGB to grey scale image

Initially, the color input image is converted to grey scale image which can be varies from black to white based on the intensity value. Then, the curvelet transform is applied based on the following rule that is $W=L^2$. Where, W and L denote the superposition functions of the various length and width.

b) Curvelet transforms

In curvelt transform, the grey scale image is degraded into sequence of disjoint scales, in which the analysis is carried out using FFT based Ridgelet transform. The Ridgelet transform is used to represent objects with curved edges using an appropriate multi-scale localization. An important thing in curvelt transform is to provide better information in spatial domains and to reduce the scale redundancy by restoring the scarcity. In this paper, the curvelet transform is used to represent the curvelet structure of the optic disc and also producing the noise free image. Finally, the optic disc segments are collected by curvelet transform.

c) Morphological operation:

Once, the segmentation process is completed, the morphological operation is carried out for improving the segmentation results, which is defined by image processing operations that process image based in its shapes. Here, the segmented image pixel is compared with the corresponding input image pixel with its nearest value. Then, the values of each pixel are applied to the morphological operation and the objects present in an image are added with the pixel value. Finally, the features are extracted from the residual image for further processing.

4.2.2 Blood Vessel Based Segmentation

In order to detect the eye diseases in early stage, the blood vessel based segmentation is carried out. The steps involved in the blood vessel segmentation are grey scale image conversion, motion filtering, image adjustment, binary conversion, morphological operation and thinning. Initially, the input RGB image is converted to grey scale image. Then, the converted grey scale image is applied to motion filtering that is used to approximating the linear motion of the image. Then, the difference between the original image and filtered image are calculated as follows:

$$D = OP - MF$$

Where, OP is the original processed image and MF denotes the motion filtered image. After performing the motion filtering operation, the image adjustment process is carried out. Here, the image adjustment is performed based on the following parameters such as image intensity levels, color balance and brightness. Here, the color casts are removed by adjusting the image intensity levels and the brightness adjustment provides the better image appearance. Then, the color imperfections in the image are corrected by color balance adjustment. After performing the image adjustment, the binary conversion process is applied to convert the original preprocessed image into binary image. Then, the binary image is applied to morphological image for selecting the processing pixels based on threshold value. Finally, the thinning algorithm is applied to the selected pixels areas and removes the pixels which have been present inside the shape to shrink it without breaking it apart. Then, the feature vectors are extracted from the segmented results.

4.3 Feature Extraction

Once the optic disc and blood vessel based segmentation process is over, feature extraction is carried out from the segmented result [9]. Finally, extracted features like area, mean, intensity, entropy variance perimeter and diameter are combined to produce the feature vector.

4.3.1 Optic Disc Features

From the optic disc segmentation, area, perimeter, entropy and intensity features are extracted. Here, the total numbers of pixels that present in the corresponding segments denotes the features based on area. Then, the perimeter feature is extracted from the boundary pixels. Suppose $P_1, P_2, \dots, P_B, \dots, P_N$ be the pixels in the concerned segment the extracted features representation based on area and perimeter is represented as follows:

$$\text{Area} = N ; \text{Perimeter} = B$$

Then, the entropy of pixel is calculated using following formula:

$$\text{Entropy} = - \sum_x \text{Prob}_x \log_2 \text{Prob}_x$$

Where, the differences between the two adjacent pixels are equal to x that is denoted as the probability Prob_x . Then, the intensity based features are extracted based on the minimum and maximum value.

$$\text{Maximum Intensity} = \text{Max}(I_1, I_2, \dots, I_m)$$

$$\text{Minimum intensity} = \text{Min}(I_1, I_2, \dots, I_m)$$

Where, I_1, I_2, \dots, I_N be the intensity of the pixels. Then, $\text{Max}(I_1, I_2, \dots, I_m)$ and $\text{Min}(I_1, I_2, \dots, I_m)$ denotes the maximum and minimum intensity of pixels respectively.

4.3.2 Blood Vessel Based Features

In blood vessel based feature extraction, three features are extracted such as mean, variance and diameter.

$$\text{Mean} = \sum_{x=1}^N \frac{I_x}{n}$$

Where, I_x is the intensity of the image pixels and N is the total number of pixels present in the image. Then, the image feature based on variance is calculated as follows:

$$\text{Variance} = \sum_i \sum_j i^m j^n I(i, j)$$

Where, $I(i, j)$ denotes the intensity of the pixels of an image. Then, the image features are extracted based on the diameter that can be represented as follows:

$$\text{Diameter} = \frac{2\sqrt{\text{Area}}}{\pi}$$

Finally, the extracted features are applied to the proposed SVM classifier for classifying the diabetic detected pixels.

4.4 Diabetic Detection Using the Proposed SVM Classifier

SVM classification is one of the supervised learning methods used in medical diagnosis for the purpose of classification and regression, which is shown in fig. 2. Here, the proposed classification technique is used to overcome drawbacks of the neural network based classification [4,6]. When, leading with large datasets, the support vectors [7, 8] are specifically used to identify the separation based on the simple solution. Basically, the empirical classification error can be minimized by the SVM classifier, so it is called as maximum margin classifiers. In this paper, we have used SVM classifier to find the optimal way to classify the input images into their respective classes like diabetic pixels or not. In this proposed classifier, the hyperplane model is used for separating the classes of an object. Here, the learning algorithm of SVM classifier is used to produce the classification parameter based on the extracted features like mean, perimeter, diameter, maximum intensity, minimum intensity and area. Then, the diabetic images are classified using the derived classification parameters. In addition, SVM classifier is used to map the inputs into two dimensional feature spaces in implicit manner using kernel trick. In order to fix the non linear curves to the data, the SVM classifier makes use of a kernel function which is used to map the data into different space using hyperplane. Then, the inner product of the image can be represented in future space using kernel function.

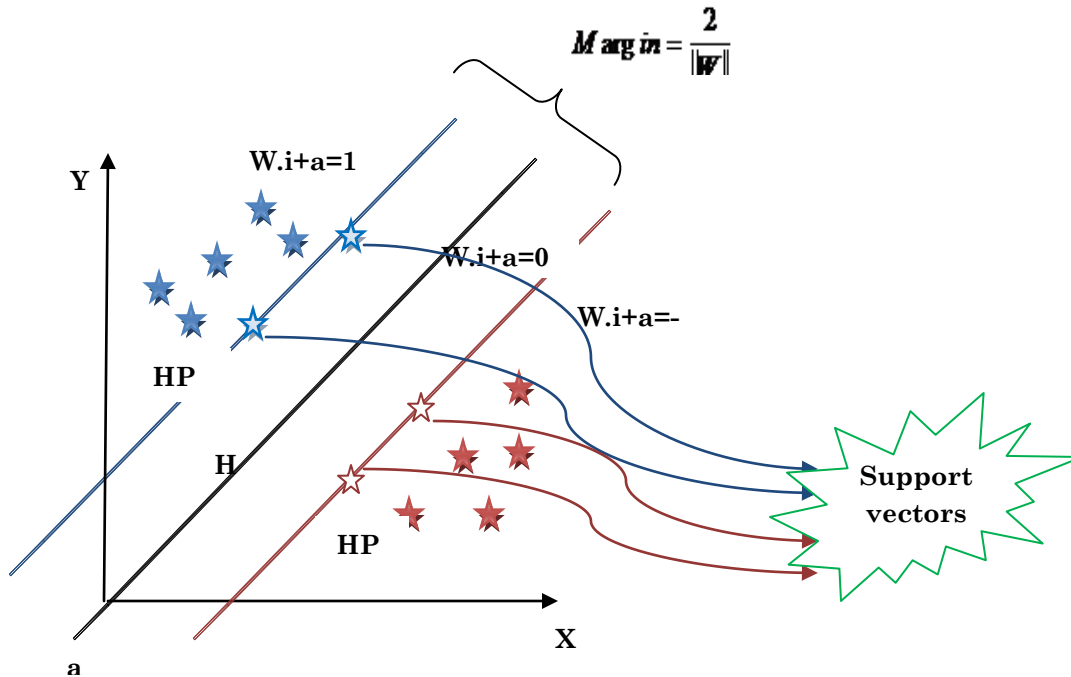


Fig. 2. SVM classification model

In SVM classifier, the optimal separating hyperplane is used to minimize the risk of misclassifying examples of the test dataset and maximize the separation between the two parallel hyperplanes. For example, the labeled training data points can be represented in the following form.

$$T = \{(i_1, j_1), (i_2, j_2), \dots, (i_n, j_n)\}$$

Where, (i, j) is the data pint of the training data samples and $j_n = \frac{1}{-1}$ is a constant which is used to denote the class belongs to the data point of the p -dimensional real vector denoted as i_n . Then, the number of data samples present in the training data is defined as n . Initially, the SVM vector is used to map the input vectors into a decision value and the classification process is carried out using the threshold value. Then, the training data can be viewed by hyperplane separation model that can be represented as follows:

$$W \cdot i + a = 0$$

Where, p -dimensional weight vector can be denoted as W and a is an offset parameter which is used to increase the margin. Then, the hyperplane is used to separate the training points based on the margin value. In order to minimize the weight vector, we need to ensure that for all x either. $w \cdot i_x + a \geq 1$ or $w \cdot i_x + b \leq -1$

In order to analyses the high dimensional data, the radial basis kernel function is used as the classifier. The mathematical representation of the radial based kernel function can be represented as follows:

$$K(i_x, j_y) = \exp\left(-\gamma \|i_x - j_y\|^2\right) ; \gamma > 0$$

Where, the kernel parameter can be denoted as γ and the training vector can be denoted as i_x . Here, the selection of large kernel parameter leads to give the smoother decision surface with standard resolution boundary. Then, the best parameter set is applied to the training dataset for the classification process. Finally, the proposed SVM classifier is used to classify the testing dataset with the better classification accuracy.

5. Result and Discussion

This section shows the experimental results of the proposed SVM classification based Diabetic Detection from Retinal Images. The proposed method is implemented using Matlab. In this paper, we have collected the images from the Stare database [1].

Evaluation metrics: In this paper, the proposed SVM classifier is analyzed using the evaluation metrics of classification accuracy. Basically, the accuracy of classification is defined as the percentage of correctly classified normal images and abnormal images, which can be represented as follows:

$$\text{classification accuracy} = \frac{TP + TN}{TP + TN + FN + FP}$$

Where, TP denotes the true positive value and TN denotes the true negative value. Then, false positive and false negative value can be represented as FP and FN respectively.

5.1 Experimental Results

In, this section fig. 3 shows the experimental results of the proposed diabetic retinopathy process. Here, the input image with diabetic problem is shown in fig. 3 (a). Then, the pre-processing phase is applied for getting the enhanced image which is shown in fig. 3 (b). Then, the segmentation process is applied for both optic disc and blood vessel features, which have been shown in fig. 3 (c) and 3 (d).

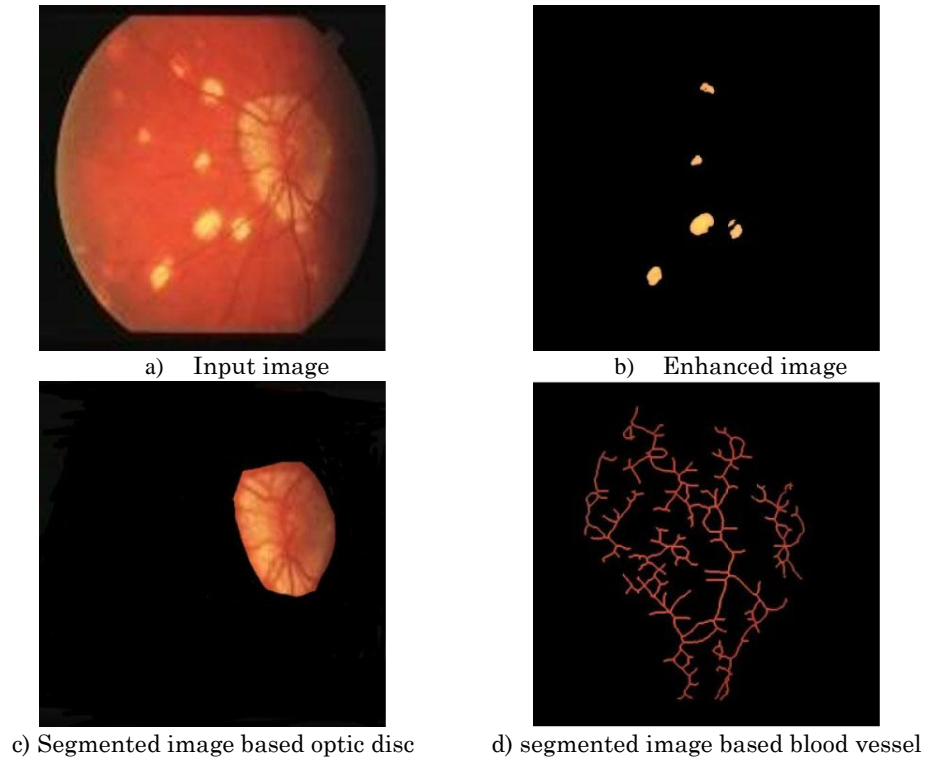


Fig. 3. Experimental result of abnormal retinal image

5.2 Comparative Analysis

This section shows the comparative analysis of the proposed technique with existing classifiers such as K-nearest neighbor's method [11], neural network, and Naive bayes classifier [14]. Fig. 4 shows the comparative results of the various existing classifiers with the proposed SVM classifier based on the classification accuracy. Here, the proposed classification accuracy is analyzed based on the number of training data samples. Here, four sets of data samples are considered for the evaluation process. Initially, 50 number of training data samples are taken for the analysis. Then, the classification accuracy is measured using various classifiers. Accordingly, the maximum classification accuracy is obtained by the proposed SVM classifier as 85% and the minimum classification accuracy is obtained by the naive bayes classifier as 72%. Further increasing the data sample as 100, the classification accuracy of the SVM, NN, K-NN and NB classifier are obtained as 89%, 82%, 80% and 75% respectively. By analyzing the fig. 4, we can conclude that, the proposed SVM classifier has achieved high-accuracy diabetic detection value when compared to the accuracy of the existing classifiers.

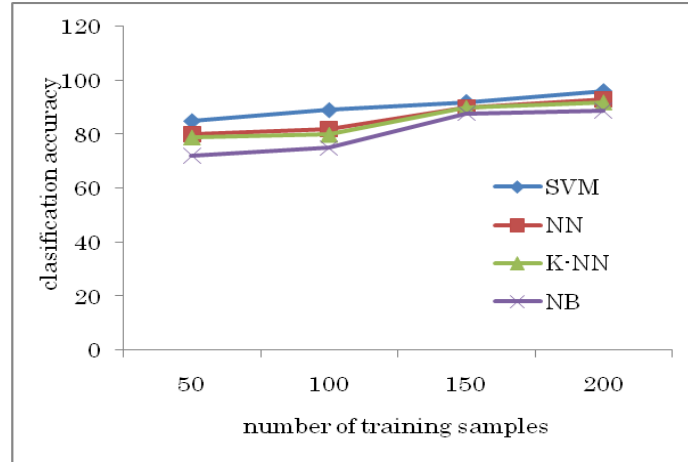


Fig. 4. Comparison results of our proposed classification technique with other existing classification methods

6. Conclusion

In this paper, we have proposed an efficient classification technique for detecting the disease of diabetic retinopathy which causes cataracts, glaucoma and, most importantly, damage to blood vessels inside the eye. Here, the proposed diabetic detection process has been performed in four phases namely pre-processing, segmentation, feature extraction and SVM classification. Here, the segmentation process has been done in two ways such as optic disc segmentation and blood vessel segmentation. In order to analyze the proposed technique, the images are collected from the stare database. Then, various features like mean, area, diameter, minimum intensity and maximum intensity were extracted in both optic disc and blood vessel segmented case. Finally, the performance parameter of the classification accuracy is used to evaluate the proposed technique with other existing classification models. From the results, we can conclude that our proposed technique has achieved the better classification accuracy of 96%.

Reference

- [1] M. Kavitha, S. Palani, "Blood Vessel, Optical Disk and Damage Area-Based Features for Diabetic Detection from Retinal Images", *Science and Engineering*, vol. 39, pp. 7059–7071, 2014.
- [2] Syna Sreng, Noppadol Maneerat, Don Isarakorn, Jun-ichi Takada, "Feature Extraction from Retinal Fundus Image for Early Detection of Diabetic Retinopathy", pp. 26-29, 2013.
- [3] Cemal Kose, Ugur Sevik, Cevat Ikibas, Hidayet Erdol, "Simple methods for segmentation and measurement of diabetic retinopathy lesions in retinal fundus images", *computer methods and programs in biomedicine*, vol. 107, pp. 274–293, 2012.
- [4] R.A. Welikalaa, M.M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, S.A. Barman, "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy", *Computerized Medical Imaging and Graphics*, vol. 43, pp. 64–77, 2015.
- [5] Balint Antal, Andras Hajdu, "An Ensemble-Based System for Microaneurysm Detection and Diabetic Retinopathy Grading", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, 2012.
- [6] Madhavi Pradhan, Ketki Kohale, Parag Naikade, Ajinkya Pachore, Eknath Palwe, "Design of Classifier for Detection of Diabetes using Neural Network and Fuzzy k-Nearest Neighbor Algorithm", *Computational Engineering Research*, vol. 2, no. 5, 2012.
- [7] V. Anuja Kumari, R. Chitra, "Classification Of Diabetes Disease Using Support Vector Machine", *Engineering Research and Applications*, vol. 3, no. 2, 2013.
- [8] P.R. Thorat, K.V. Patil, "Identification of Diabetic Retinopathy from Segmented Retinal Fundus Images by using Support Vector Machine", *Engineering and Advanced Technology*, vol. 5, no. 4, 2016.
- [9] Li, H.; Chutatape, O., "Fundus image features extraction", *IEEE Transactions on Engineering Med. Biol. Soc.* vol. 1, pp. 3071–3073, 2000.

- [10] Choi, Myungjin, Rae Young Kim, and Moon-Gyu Kim; The curvelet transform for image fusion. In: International Society for Photogrammetry and Remote Sensing, vol. 35, pp. 1–5, 2004.
- [11] G. G. Rajput, Preethi N. Patil, "Detection and classification of exudates using k-means clustering in color retinal images", In proceedings of international conference on signal and image processing, 2014.
- [12] Keerthi Ram, Gopal Datt Joshi, Jayanthi Sivaswamy, "A Successive Clutter-Rejection-Based Approach for Early Detection of Diabetic Retinopathy", IEEE Transactions on Biomedical Engineering, vol. 58, no. 3, 2011.
- [13] Amol Prataprao Bhatkar, Dr. G.U.Kharat, "Detection of Diabetic Retinopathy in Retinal Images using MLPclassifier", Nanoelectronic and Information Systems, 2015.
- [14] M. L. Raymer; L. A. Kuhn; W. F. Punch, "Knowledge discovery in biological data sets using a hybrid Bayes classifier/evolutionary algorithm", In Proceedings of international symposium on Bioinformatics and Bioengineering Conference, pp. 236 - 245, 2001.