Earlier Diabetic Retinopathy Detection Using Advanced Pre-Processing Methods and SVM Classification

Sherin Stanly, Dr.R.Kavitha Jaba Malar,

Mphil Scholar, Nanjil Catholic College of Arts and Science, Kaliakkavilai, Manonmaniam Sundaranar university, tirunelveli 627 012. 
Assistance Professor, Department of Computer Science, Nanjil Catholic College of Arts and Science, Kaliakkavilai. sherinstan.mphil@gmail.com

Abstract— Diabetic retinopathy (DR) is a vital cause that leads to blindness and it is one of the most important complications in diabetes. In this, the early diagnosis can prevent the patient from loss of sight/vision. Here, the paper is all about the detection of diabetic retinopathy in prior through a computer-aided diagnosis on which, the digital image processing of the retinal image is generated. Initially, an image processing stage isolates the Haemorrhages, Microaneurysms (MA), Exudates and Cotton Wool Spots to extract features which are used to support vector machine to identify the retinopathy grade of each retinal image. Also, the performance of this proposed system is assessed by the use of 400 retinal fundus images in the Kaggle Diabetic Retinopathy Detection database. The experiment is carried through various stages which primly consist of classification work on all stages of the image processing after it has been analyzed and studied in detail. Indeed it proves that this method of diagnosis is highly efficient with high rate of success in the detection of diabetic retinopathy from the retinal fundus image.

Keywords— Diabetic retinopathy, machine learning, exudate detection, SVM

I. INTRODUCTION

Diabetes is a vital disease which is caused due to the inadequate or absolute lack of insulin discharge. This leads to the impediment related to insulin discharge which is the excess of sugar retaining in blood, indeed result in accumulation of glucose in high rate in blood within the body. This accumulation or occurrence affects the entire human body which in turn results in diabetes. Diabetes affects various organs of the body such as kidney, heart and eyes. Among this, diabetes which affects or damages the blood vessels in the eye is referred to as diabetes retinopathy (DR). It is one of the major public health problems around the globe which leads to blindness. Based on the survey done by the World health organization, it is stated that the global prevalence of diabetes estimated in 2000 is 2.8% but in 2030 the diabetes rate will increase to 4.4%. In current practice, the DR stages are screened or diagnosed manually by the ophthalmologist through image generated by the non-fundus camera but the DR stages are classified into two major sectors as NPDR (Non-proliferative diabetic retinopathy) and PDR (Proliferative diabetic retinopathy). This typology is functioned/ performed through various types, numbers and locations of any micro vascular lesion or DR feature in the image, Yet the complete screening process is time-consuming, tedious, and prone to error and it also requires the expertise of trained personnel. In order to reduce such difficulties, a computer-aided diagnosis has been exploded in the past to diagnose and mitigate diagnostic between manual readers.

![Fig. 1 Stages of Diabetic Retinopathy](image-url)
Fig.1 shows the different types of diabetic retinopathy. Diabetic retinopathy is divided commonly into two stages.
1. Non-proliferative diabetic retinopathy.
2. Proliferative diabetic retinopathy.

In the non-proliferative diabetic retinopathy Microaneurysms, Haemorrhages, Exudates and Cotton Wool Spots are present. The proliferative stage is categorized based on the severity of diabetic retinopathy. They are classified as mild, moderate and severe. In this stage, abnormal blood vessels are formed. Its result is mostly a loss of vision. The primary aim of this research is to identify the diabetic retinopathy before it attains the PDR stage to develop.

The rest of the paper is organized as follows. Section II explains the literature survey. Section III illustrates the block diagram of the proposed methodology. Section IV explains the results and discussion, Section IV discusses the conclusion of the paper.

II. LITERATURE REVIEW

Shailesh Kumar and Basant Kumar [1] described Diabetic Retinopathy Detection by Extracting Area and Number of Microaneurysm from Colour Fundus Image. This method consists of three modules: pre-processing techniques, feature extraction and classification techniques. Adaptive histogram equalization is used for contrast enhancement. Median filter and morphological process are used to extract the exudates from color Fundus images. In this method SVM classification is used for DR detection which classifies the image into two classes such as DR eye and healthy eye.

Arisha Roy, Debasmita Dutta and Pratyusha Bhattacharya have proposed Filter and Fuzzy C Means Based Feature Extraction and Classification of Diabetic Retinopathy using Support Vector Machines [2]. This method consists of four main methods: Pre-processing, Selecting a specific region from the Histogram, Use of Convex Hull on blob measurements and Optical Disk Measurement. This method select exudates from fundus images using Fuzzy C means technique with the removal of Optical Disk. Proposed method also isolates Neovascularization from retinal images. Finally the classification is done by using support vector machine.

Moriun Akter and Mohammad Shorif Uddin presented Morphology-Based Exudates Detection from Color Fundus Images in Diabetic Retinopathy [3]. This Exudates detection method includes: Histogram equalized grayscale image, Binary image by thresholding, Eroded image, Dilated image, Erosion to remove exudates and gives only optic disk, Dilation of optic disk, Distance transform of the image and Exudates image by Watershed transform. The author claimed that their proposed method attains 99% accuracy.

Navoneel Chakrabarty developed a Deep Learning Method for the detection of Diabetic Retinopathy [4]. Important steps used in this method: Data Pre-Processing and the CNN layers. Pre-processing step includes: Conversion to Weighted gray scale, resizing and Pixel Rescaling. The automatic diabetic retinopathy system is developed by using Python’s and Scipy libraries.

Shuang and Yogesan Kanagasingam described Exudate Detection for Diabetic Retinopathy With Convolutional Neural Networks [5]. Process flow of: Removal of Optic Disc Detection, Removal of Retinal Vessels, Ultimate Opening and Convolutional Neural Networks. In this method exudate candidates are first extracted with morphological ultimate opening techniques and then the candidate points are passed to the trained CNN deep networks for classification.

III. PROPOSED METHODOLOGY

The objective of our proposed method is an earlier diabetic retinopathy detection and classification. There are four main modules in it viz. pre-processing, Segmentation, feature extraction and classification. In pre-processing we use the latest developed pre-processing methods. In feature extraction, texture and color features are extracted. For classification, SVM classifier is used. Fig 2. shows our proposed architecture.

A. Input Image

The images necessary for our research have been downloaded from the kaggle website. In this repository varieties of fundus images taken from different types of lighting conditions are available. In our dataset 250 images are used for training and 244 for testing, of these images used for training 150 images are that of the retina of the people affected by different types of retinopathy. The remaining are the images of the unaffected retina. There are five training labels (0,1,2,3&4). These labels are named Normal, Microaneurysms, Haemorrhages, Soft Exudates, and Cotton Wool Spots respectively. In this image dataset two
types of images are available one is of the right eye and other is of the left eye. The image of the right eye is named right and that of the left is named left. The following table has detail of the training data set.

Table 1 training data set

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Number of Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>71</td>
</tr>
<tr>
<td>1</td>
<td>Microaneurysms</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>Haemorrhages</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Soft Exudates</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>Cotton Wool Spots</td>
<td>18</td>
</tr>
</tbody>
</table>

**B. Image Pre-Processing**

Image pre-processing is an important aspect of diabetic retinopathy detection. In our proposed method pre-processing is a set of methods which includes green channel extraction, CLAHE, filtering and binarization. This is useful in visualizing the obscure parts of images and in normalizing the over-bright parts of the images taken in different types of lighting conditions.
**Green Channel Extraction**

In the pre-processing stage the green channel extraction is an indispensable aspect [6]. The color retinal images are normally low contrast images. Its blood vessels and affected regions are dim. After the green channel is extracted red and blue channels are used in the later stage to extract color features. Fig 3.1 shows the green channel extracted image.

**Contrast Limited Adaptive Histogram Equalization**

The AHE transforms all the pixels into grey scale image [7]. The transformation function is derived from the neighbourhood region each and every pixel is transformed based on the histogram if the sequence surrounding the pixel. In our new method we apply the CLAHE contrast enhancement method. All the important functions of this method have been derived from the AHE method. The CLAHE method separate the input image we provide into 3 parts viz, red green and blue respectively. Fig 3.2 shows the contrast enhanced image.

**Color Normalization**

Color normalization is an inevitable aspect in the fundus images pre-processing. This is used to obtain a standardized color range. In this research we used Affine color Normalization method [2].

**Optic Disc Removal**

It is another important step in retinopathy detection [8]. Unless the optic disc in the retinal image is removed probably the ratio of the false positive well increase so, this is an indispensable step. In this method we apply the entropy based approach and estimates the location of the OD. Normally this OD location would be high intensity region. In the forthcoming steps its right position and shape can be found out. Fig 3.1 shows the OD removed image.
C. Segmentation
In our proposed method four types of segmentation are done. First the exudates area is segmented. Secondly Haemorrhage area is segmented. Microaneurysms area is segmented thirdly and the cotton wool spots are segmented finally. Mathematical morphological operation and Fuzzy C-means clustering [9] method are used for the same. Fig 4 shows the segmented image.

![Segmented Image](image)

**Fig 4 segmented image.**

D. Feature Extraction
Feature extraction is an important aspect in machine learning. In our proposed method we classify the diabetic retinopathy into four types [10]. The following features are extracted for the classification. The extracted features are stored as feature set. In this work fourteen features are used. Of them eleven features are texture based features and others are disease based features. For SVM classification texture features of Haemorrhages area, Microaneurysms area, cotton wool spots area and exudates area features are used. In our method for the feature extraction Grey-Level Co-Occurrence Matrix (GLCM) is used. The features are explained below.

**Mean**
Mean of pixels in image.

\[
\mu_n = \sum_{i,j=0}^{L-1} (Z_{i,j} - m)^n p(Z_{i,j})
\]

(1)

\(m\) is the mean value of \(z\).

**Standard Deviation**
Standard deviation is the deviation of \(z\).

\[
m = \sum_{i,j=0}^{L-1} Z_{i,j} p(Z_{i,j})
\]

(2)

**Entropy**
Entropy of the image.

\[H(z) = - \int p(z) \log p(z) dz\]

is the entropy of the image.

**Contrast**
Contrast is the different between the color and brightness of the given image.

\[
\text{Contrast} = \sum_{i,j} |i - j|^2 p(i,j)
\]

(5)
Correlation
Correlation is calculated by using the following formula.

\[ d_{rs} = 1 - \frac{(x_r - x'_r)(x_s - x'_s)^T}{(x_r - x'_r)^T (x_r - x'_r)^T} \]  

\[ (x_s - x'_s)(x_s - x'_s)^T \]  

Energy
Energy is calculated by using the following formula. It provides the sum of squared elements in the GLCM.

\[ \text{Energy} = \sum_{i,j} p(i,j)^2 \]  

Homogeneity
Homogeneity illustrated in formula 9.

\[ \text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1 + |i-j|} \]  

ASM is the calculation of local homogeneity and the opposite of Entropy.

\[ \text{ASM} = \sum_{i,j} (p(i,j))^2 \]  

Inverse differential moment
It is explain in the following equation. It is a direct calculation of the local homogeneity of a fundas image.

\[ \text{IDM} = \sum_{i} \sum_{j} \left\{ \frac{p(i,j)}{1 + (i-j)^2} \right\} \]  

Skewness
It is used for calculate the Skewness of the histogram. It is calculated by using the following formula.

\[ \mu_3 = \sum_{i,j=0}^{l-1} (Z_i - m)^3 p(Z_i) \]  

Disease based features
It provides the information about area of the disease spread in the retina image.

\[ \rho(i,j) = \frac{1}{M-N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} b_i (i - m, j - n) \]  

E. Classification using SVM
Fundus images are classified by SVM classifier as normal or abnormal [11]. The SVM classifier is a supervised learning algorithm its decision boundary is segmented by a hyper plane. Mainly three types of SVM are used linear SVM, non-linear SVM and binary SVM. We use the linear SVM for our research. The formula of the linear SVM is given bellow.
To evaluate our proposed method a computer with intel core i7 processor and 4GB RAM, 400 GB hard-disc space is used. To develop this image-processing system Matlab 2016 software and windows 10 operating system are used.

In medical diagnosis usually medical input data are classified into two types, whether the disease is present or not. Its classification accuracy is calculated based on specificity, precision, accuracy and F-measure. For experimental analysis of the proposed method 400 retinopathy affected and non-affected images are used totally. In this method, training database 350 images are used for training purpose of these 300 are normal images and 50 are abnormal. The training dataset is trained by SVM classifier. In the training time the important feature like blood vessel area, microaneurysm area and exudates areas are extracted and stored for training purpose. Our testing data set comprises 50 retinopathy images of which 11 are abnormal and 39 normal. The images are tested with SVM classifier. In our proposed performance measurement is shown in fig 1. In this the already existing methods also have been compared. Our experimental results prove that the proposed method is better than the already existing methods. Table2 explain the comparison details with existing methods. Fig 5 shows the graphical representation of the performances with the existing methods.

1. Sensitivity – Implies that abnormality of retina is classified correctly, which is defined as.

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

2. Specificity - Implies the normality of retina is classified correctly, which is defined as.

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

The following parameters are used for the sensitivity and specificity calculations.

- True Positive (TP) – It implies that the abnormality is predicted correctly as abnormality.
- True Negative (TN) - It implies that the normality is predicted correctly as normality.
- False Positive (FP) – It implies that the normality is predicted wrongly as abnormality.
- False Negative (FN) – It implies that the abnormality is predicted wrongly as normality.

For the performance analysis, we compare our results with already developed retinopathy methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Method</td>
<td>96.8</td>
<td>95.1</td>
<td>94.7</td>
</tr>
<tr>
<td>[1]</td>
<td>95.7</td>
<td>94.3</td>
<td>93.1</td>
</tr>
<tr>
<td>[2]</td>
<td>94.1</td>
<td>90.2</td>
<td>93.5</td>
</tr>
<tr>
<td>[3]</td>
<td>95.3</td>
<td>92.1</td>
<td>92.7</td>
</tr>
<tr>
<td>[4]</td>
<td>93.2</td>
<td>90.19</td>
<td>91.6</td>
</tr>
<tr>
<td>[5]</td>
<td>92.2</td>
<td>89.9</td>
<td>92.40</td>
</tr>
</tbody>
</table>

Table 2 Performance comparison
In our method green channel extraction, contrast enhancement and filtering are done for preprocessing. Morphological operation and fuzzy C-means clustering are used for segmentation. Segmentation divided into four parts. They are blood Haemorrhage segmentation, Microaneurysms segmentation, Exudates segmentation and Cotton Wool Spots Segmentation. Feature extraction carried out by using Gray level Co-occurrence matrix. Classification is done by using SVM classifier. In this research we have proposed effective retinopathy detection method. We have used strong preprocessing method. It is very useful to detect the diabetic retinopathy earlier and is highly useful for the ophthalmologists and the patients. In our work four different retinal abnormalities are detected. The experimental result shows that new method is much better than the already prevailing methods.

**REFERENCES**


