



DETECTION OF RETINAL ABNORMALITIES USING MACHINE LEARNING METHODOLOGIES

*R. Saha**, *A. Roy Chowdhury**, *S. Banerjee†*, *T. Chatterjee‡*

Abstract: This paper presents an algorithm for the design of a computer aided diagnosis system to detect, quantify and classify the lesions of non-proliferative diabetic retinopathy as well as dry age related macular degeneration from the fundus retina images. Symptoms of non-proliferative diabetic retinopathy in images consist of bright lesions like hard exudates, cotton wool spots and dark lesions like microaneurysms, hemorrhages. Dry age related macular degeneration is manifested as a bright lesion called drusen. The proposed system consists of two parts: image processing, where preprocessed gray scale images are segmented to extract candidate lesions using a combination of Gaussian filtering and multilevel thresholding followed by classification of the different lesions in non-proliferative diabetic retinopathy and age related macular degeneration using perceptron, support vector machine and naïve Bayes classifier. From the comparative performance analysis of the classification techniques, it is observed that comparable results are obtained from single layer perceptron and support vector machine and they both outperform naïve Bayes classifier. The classification accuracy of support vector machine classifier for dark lesion class is 97.13 % and the classification accuracy of single layer perceptron for bright lesion class is 95.13 % with optimal feature set.

Key words: *diabetic retinopathy, age related macular degeneration, multilevel thresholding, perceptron, support vector machine*

Received: January 5, 2018

DOI: 10.14311/NNW.2018.28.025

Revised and accepted: August 30, 2018

1. Introduction

Diabetic retinopathy (DR) is a retina related disease occurring due to complications of diabetes. Computer aided diagnosis (CAD) is becoming increasingly popular in the early diagnosis of the disease. In this paper, we present an effort to

*Rituparna Saha; Amrita Roy Chowdhury; Computer Science and Engineering Department, Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India, E-mail: rsaharitu.saha@gmail.com, amrita.me.cse@gmail.com

†Sreeparna Banerjee – Corresponding author; Department of Natural Science and IEM, Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India, E-mail: bsreep96@hotmail.com

‡Tamojit Chatterjee; Regional Institute of Ophthalmology, Calcutta Medical College, Kolkata, West Bengal, India, E-mail: drtamojit@gmail.com

develop a CAD system for the detection of abnormalities of the retina like diabetic retinopathy and age related macular degeneration (AMD). DR is broadly classified as non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR), which denote the early stage and later stage of DR, respectively. CAD system adopts a pixel wise identification approach [9] which could be helpful in detecting small lesions and thus aid ophthalmologists in diagnosis and therapy planning. The clinical symptoms visible for DR are mainly two types of lesions: dark lesions and bright lesions. Dark lesions are caused by the leakage of blood from the blood vessels. Two types of such lesions are hemorrhages (HEM) and microaneurysms (MA). Bright lesions in DR consist of hard exudates (HEX) caused by leakage of lipid and proteins and cotton wool spots (CWS) caused by damage to nerve fibers. Due to age, a DR patient may also have some other symptoms in the retina like drusen, which is a manifestation of dry AMD and which is similar in appearance to DR bright lesions. Therefore the differentiation of the drusen with the DR bright lesions is another important task of early stage DR detection. CAD system adopts a pixel wise identification approach which could be helpful in detecting small lesions and thus aid ophthalmologists in diagnosis and therapy planning.

In this paper, the CAD system is designed for the detection and classification of four major type NPDR lesions: MA, HEX, CWS, HEM along with dry AMD lesion called drusen. The proposed system first extracts all the affected candidate regions from the retina images using a combination of Gaussian filtering and multilevel thresholding. These candidate regions are then used for feature extraction. Clinical symptoms that show up in retina fundus images are summarized in Tab. I.

This paper extends our previous works [18–20, 27] where only DR related abnormalities were detected using Fuzzy c means clustering approach. In this paper, a new way is presented to detect both types of DR lesions (dark and bright) and also dry AMD using multilevel thresholding [24]. The proposed technique examines three different machine learning approaches: single layer perceptron (SLP), support vector machine (SVM) and naïve Bayes (NB) approaches for classification after performing optic disc (OD) localization and elimination using an edge detection technique based on universal law of gravity following [25] and also the detection and elimination of blood vessel regions [10] using morphological operators.

The related work is described in Section 2. Section 3 represents retina image database collection used for the current research work. The proposed methodology is described in Section 4. Results and discussion are reported in Section 5. Finally, conclusions and future work are stated in Section 6.

2. Related Works

As vision loss due to DR is a major concern worldwide, computer based techniques for its detection specially, hemorrhages and neovascularization occurring during the later stages [4], has recently generated a lot of interest. In addition, AMD, especially dry AMD, can cause additional complications [5] in the elderly population. Hence, computer assisted diagnosis will greatly aid ophthalmologists in the early diagnosis and subsequent treatment of the patient and mitigate vision loss.

Lesion names	Clinical symptoms	Clinical etiology
MA	Small, dark red dots	First clinical sign of NPDR, occurs in the inner layer of retina [1]
HEM	Dark red, varies in size and shape	Resulted from ruptured microaneurysms [1]
HEX	Bright yellowish and varies in size	Occurred due to accumulation of lipoproteins derived from leakage of abnormal vessels [1]
CWS	White or yellowish white lesions with blurry edges	Nerve fiber layer infarctions from occlusion of precapillary arterioles [1]
Drusen	Discrete, small, yellow deposits	Form as deposition of membranous material between the plasma membrane and the basement membrane of the retinal pigment epithelium (RPE) on Bruch's membrane [5]

Tab. I *Characteristic features of retinopathy lesions.*

Efforts to extract bright lesions [2, 4, 6, 13–17, 19, 28] and dark lesions [3, 6–8, 11, 28] have been reported in the literature. These methods performed candidate lesion extraction using segmentation techniques as well as the use of morphological operations and wavelets, to name a few. These were done after image preprocessing from RGB to green channel image, and noise removal, as required. This was followed by classification using k-nearest neighbor [3], rule based supervised learning [8], support vector machine [12, 20] and naïve Bayes machine learning [19] classifiers. In [13], a combination of Gaussian mixture model and k-nearest neighbor classifier has been used to extract the lesions using a reduced number of features. An Amplitude Modulation-Frequency Modulation technique [11] has been proposed to investigate DR and AMD. Experiments have been performed in the frequency domain with high and medium pass filters to extract smaller structures like MA, HEX and dot-blot HEM. Low pass filters were used to extract larger structures like vessels in this technique.

In all the methods mentioned above except Akram et. al. [6], lesion identification was done manually. In the work of Akram et. al. [6], automatic detection of lesions was done after detection of blood vessel tree and OD. Enhancement of blood vessels was achieved by inverted green channel followed by Gabor wavelet. Binary mask for blood vessel tree segmentation was created by multilayer thresholding and adaptive thresholding. In the detection of OD, Hough transform preceded by Canny edge detection was used to identify the circular shape of OD.

3. Database

The proposed system uses ninety images out of a total of one hundred and thirty images from DIARETDB0 [21] and seventy five images out of a total of eighty nine images from DIARETDB1 [22,23] databases consisting of color fundus images of patients with DR symptoms: HEX, CWS, MA, HEM. Twenty five dry AMD images of patients from the well-known city eye hospital have also been utilized to develop a composite database for the experiment.

4. Proposed Method

The proposed method consists of image processing to extract candidate lesions, described below followed by classification using SVM, SLP and NB classifiers.

4.1 Image processing

Image processing procedures are described below.

4.1.1 Image preprocessing

Preprocessing consisted of color normalization using histogram specification. Gray scale image conversion was performed using green channel extraction. Contrast limited adaptive histogram equalization (CLAHE) [14] was used for contrast enhancement without enhancing the noise content of the image. The extraction of dark and bright lesions was done using a combination of two techniques, namely segmentation using Gaussian filtering, using a standard deviation $\sigma = 32$ for best results, and multilevel thresholding.

In this paper, the smoothing operation was succeeded by the subtraction operation between the preprocessed image without smoothing and the resultant smoothed image to derive an image suitable for selecting the threshold values to segment the different color lesions.

Multilevel thresholding was done to segment mainly four regions: black corner portion around the retina, a region with dark red color denoting the blood vessel tree, dark lesions such as HEM, MA, a region with bright yellow color denoting the OD, bright lesions such as HEX, CWS or drusen and another region denoting the surface or base of the retina. For this reason, three threshold values T_C, T_D and T_B with $T_C < T_D < T_B$ were used to find out two binary images of a fundus retina image where one binary image denotes the extracted dark lesions and another denotes the binary image with extracted bright lesions of that fundus image.

4.1.2 Blood Vessels Detection and Elimination

The resultant binary image of the extracted dark lesion region containing MA, HEM and some blood vessel areas are shown in Fig. 1. In this paper we followed our preceding work [19] of blood vessels detection algorithm based on morphological operation, with some modifications. The morphological closing operation was applied on the preprocessed gray scale image, shown in Fig. 1(a). The output image of closing operation was then binarized using thresholding as well as to reduce the

noise portions from the binarized image. The connected components having Solidity value less than 2 were only considered as blood vessels and finally, blood vessel trees are extracted as shown in Fig. 1(b). Solidity is defined by Eq. (1) where the term Area denotes the number of pixels in the connected component and the term ConvexArea denotes the number of pixels in the convex image of the component

$$\text{Solidity} = \frac{\text{Area}}{\text{ConvexArea}}. \quad (1)$$

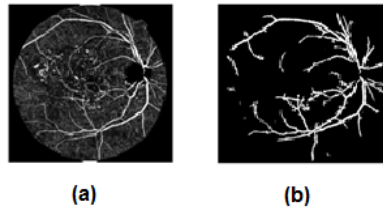


Fig. 1 Blood vessels extraction: (a) Morphological closing operation result, (b) Final extracted blood vessels.

4.1.3 Optic Disk Detection and Elimination

The extracted image for bright lesions contained both lesion and non-lesion portions like HEX, CWS, drusen and OD. To eliminate OD, an edge detection technique using the theory of universal gravitational law [25] was applied followed by cropping out the OD portion and binarization using thresholding and region filling. The resultant image gave the detected OD region as shown in Fig. 2. Based on the detected OD size and position, a binary mask was generated as shown in Fig. 2(d) for performing logical operation to eliminate the OD from the extracted bright lesion region. This approach was able to retain the full shape of the OD. Finally, elimination of OD was done by applying logical operations between the generated binary mask and the extracted bright lesion area with OD region.

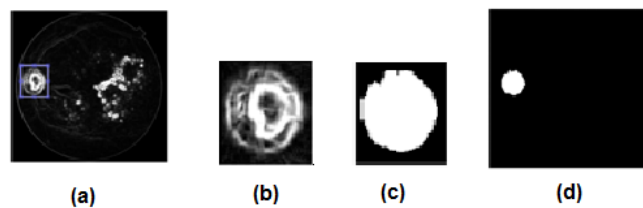


Fig. 2 Optic disc detection: (a) Edge detected image, (b) Cropped OD portion, (c) Filled image, (d) Generated binary mask.

4.1.4 Feature Selection

The most crucial step of the proposed CAD system was feature extraction. From the resultant binary dark and bright lesion images, some useful region based features were selected for classification purposes. In the resultant output image of extracted bright and dark lesions, each of the connected components were considered as a region as shown in Fig. 3(b). In this proposed work, fourteen types of region based features were selected. Of these, a combination of seven types of basic geometrical shape based features and seven types of texture based features were considered. In the following Tab. II and III, the selected features are listed.

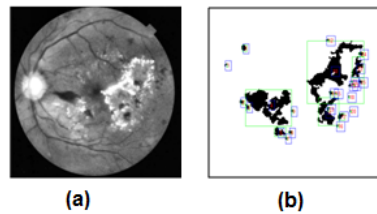


Fig. 3 (a) Preprocessed image, (b) Extracted bright lesion image with marked regions.

Feature	Feature names	Explanation
F1	Area of connected region	Number of pixels in the region
F2	Perimeter of region	Distance between each adjoining pair of pixels around the border of the region
F3	Convex area of region	Specifies number of pixels in convex image of region
F4	Solidity of region	$\text{Solidity} = \frac{\text{Area}}{\text{ConvexArea}}$
F5	Circularity of region	$\text{Circularity} = \frac{\text{Perimeter}^2}{4 \times \pi \times \text{Area}}$
F6	Compactness of region	$\text{Compactness} = \frac{\text{Area}}{\text{Perimeter}^2}$
F7	Eccentricity of region	Ratio of the distance between the foci of the ellipse and its major axis length for the ellipse that has the same second-moments as the region

Tab. II Selected features based on shape.

4.2 Classification

The proposed system experiments with perceptron classifier with hidden layer zero (which is equivalent to SLP) and then compares the performance of perceptron with SVM classifier and NB classifier using Weka data mining tool [29].

Feature	Feature names	Explanation
F8	Mean (μ) of region	$\mu = \sum_{i=0}^{L-1} ip(i)$, where $p(i)$ denotes the first order histogram estimation for gray level i , ($0 \leq i \leq L - 1$ where $L - 1$ denotes the maximum intensity level).
F9	Entropy of region	Entropy = $-\sum_{i=0}^{L-1} p(i) \times \log_2 p(i)$ where μ denotes simple mean intensity value of region, i denotes the different gray levels ($0 \leq i \leq L - 1$ where $L - 1$ denotes the maximum intensity level).
F10	Energy or Uniformity of region	Energy = $\sum_{i=0}^{L-1} p^2(i)$ where μ denotes simple mean intensity value of region, i denotes the different gray levels ($0 \leq i \leq L - 1$ where $L - 1$ denotes the maximum intensity level).
F11	Standard Deviation (σ) of region	$\sigma = \sqrt{\frac{1}{L-1} \sum_{i=0}^{L-1} (i - \mu)^2}$ where μ denotes simple mean intensity value of region, i denotes the different gray levels ($0 \leq i \leq L - 1$ where $L - 1$ denotes the maximum intensity level).
F12	Smoothness (R) of region	$R = 1 - \frac{1}{1+\sigma^2}$ where σ denotes the Standard deviation
F13	Color	Gray value of the pixel
F14	Sharpness of edge	Nature of the edge of abnormal object

Tab. III Selected features based on texture.

4.2.1 Perceptron algorithm

The Perceptron algorithm, originally proposed by Rosenblatt [26], is a supervised learning technique for binary classification. It is the simplest form of neural network for classification of linearly separable sets of patterns. In two dimensional space, a straight line is used to separate the linearly separable patterns, for three dimensions, a plane and for the dimensions greater than three, a hyper plane is used to separate the -1 pattern from $+1$ pattern.

In our work, we used multilayer perceptron (MLP) provided by Weka data mining tool [29] with zero hidden layer. A network of MLP contains input layer, hidden layer/layers and output layer with weighted connection. For each node, a weighted sum of the inputs to that node is evaluated and generates the results depending on a thresholding generally using a sigmoid function. The weights are

initially learned from the training set and then the change of weight is computed using Eq. (2).

$$\text{Weight}_{\text{next}} = \text{Weight}_{\text{current}} + \Delta\text{Weight}, \quad (2)$$

where $\Delta\text{Weight} = (-LG) + (M\Delta\text{Weight}_{\text{prev}})$, L is learning rate, G is gradient determined using back propagation algorithm, M is momentum and $\Delta\text{Weight}_{\text{prev}}$ is previous change in weight.

In Weka data mining tool, for perceptron classifier, stratified cross validation is used by wrapper method for optimization. A 10 fold cross validation technique was performed using all the data for both training and validation. Each dataset was checked for validation exactly once. In the 10 fold cross validation technique, the recorded dataset was randomly split into 10 folds of equal size. In each iteration, one fold was used as a testing set and other 9 folds are used for training the classifier. Then cross validation estimate of the accuracy was calculated by averaging all the test results over folds. Stratified cross validation maintains the correct ratio of each class value.

An MLP with zero hidden layer is equivalent to SLP which is suitable when the data is linearly separable. In the current research work, the dataset contains feature values obtained from retina fundus images with different abnormalities like HEX, CWS, MA, HEM and drusen. The bright lesion classes (HEX, CSW and drusen) are separable from the dark lesion classes (MA and HEM) with respect to color feature.

Among the bright lesion classes, HEX and CWS differs in color feature as HEX are bright yellow and CWS are whitish. The sharp edges of HEX distinguishes it from CWS which has blurry edges.

Drusen are circular in shape where HEX in most of the cases are irregular in shape. Drusen are more clustered in nature than HEX. So the features circularity and compactness separate HEX from drusen.

Among dark lesion classes, MA are very small in size and thus are separable from HEM which are medium or large in size. MA appears as the first signs of DR and are distinguished from HEM which appears at later stages of DR. As the dataset of the current research work is linearly separable, SLP is selected.

4.2.2 Support Vector Machine

Support Vector Machine classifier, a supervised learning technique is used to classify a feature vector \mathbf{x} into two classes c_1 and c_2 using a linear discriminant function given by Eq. (3).

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (3)$$

where \mathbf{w} is a weight vector, perpendicular to the hyperplane. This denotes the orientation of the hyperplane and b is the bias, denoting the position of hyperplane. In the classification of feature vector \mathbf{x} , if $g(\mathbf{x}) > 0$ then $\mathbf{x} \in c_1$ class. In case of $g(\mathbf{x}) < 0$, $\mathbf{x} \in c_2$ class. Otherwise $g(\mathbf{x}) = 0$ denotes that \mathbf{x} is on the hyperplane. SVM basically tries to maximize the distance of the hyperplane between two classes by maximizing the distance of the plane from each of the vectors. In Weka [29], sequential minimal optimization algorithm is used for the training purpose of the support vector classifier. The default kernel is polynomial kernel with exponent

parameter. As the retina abnormalities dataset of this research work are linearly separable, the value of exponent is set to 1 for linearity.

4.2.3 Naive Bayes Classifier

Naive Bayes (NB) classifier is a supervised, probabilistic learning technique based on Bayes' Theorem. This technique classifies a set of patterns with an assumption of independence of predictors that describes instances. Bayes' theorem is shown in Eq. (4).

$$P(y|x_1, x_2, \dots, x_n) = \frac{P(y)P(x_1, x_2, \dots, x_n|y)}{P(x_1, x_2, \dots, x_n)}, \quad (4)$$

where $P(y|x_1, x_2, \dots, x_n)$ is the posterior probability of classification, $P(x_1, x_2, \dots, x_n|y)$ is the probability of predictors, $P(y)$ denotes the prior probability and $P(x_1, x_2, \dots, x_n)$ is the prior probability of predictors for a given class variable y and a dependent feature vector (x_1, x_2, \dots, x_n) .

5. Results and Discussion

The calculated statistical feature vectors of the extracted candidate regions were fed into the classifier. Two datasets were designed, combining the feature vectors of the true positive lesions separately for both bright and dark lesions. Classification was performed on the set of feature vectors that helps to classify bright lesions into three classes class 1, class 2, class 3 denoting HEX, CWS, drusen respectively and dark lesions into two classes class 1, class 2 denoting MA, HEM respectively. A total of 638 abnormal regions were detected using image processing from the collection of one hundred and ninety retina fundus images. Tab. IV describes the details of the abnormal objects detected from the retina images.

Dataset	No of lesion type	Extracted no. of lesions
1st dataset	Dark lesions = 297	MA = 106 HEM = 191
2nd dataset	Bright lesions = 341	HEX = 96 CWS = 48 Drusen = 197

Tab. IV Dataset information of extracted lesions.

Performance measures are defined by Eq. (5), (6) and (7)

$$\text{sensitivity}_i = \frac{TP_i}{TP_i + FN_i}, \quad (5)$$

$$\text{specificity}_i = \frac{TN_i}{TN_i + FP_i}, \quad (6)$$

$$\text{accuracy}_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}. \quad (7)$$

True Positive rate (TP_i) denotes number of lesions correctly classified in i -th class. False Negative rate (FN_i) denotes number of lesions of the particular class that are incorrectly classified. False Positive rate (FP_i) denotes number of lesions of other classes incorrectly classified. True Negative rate (TN_i) denotes number of lesions of other classes correctly classified. Tab. V and VI shows the comparative performance analysis of bright lesions and dark lesions respectively using different classification techniques.

Methods	Sensitivity [%]	Specificity [%]	Accuracy [%]
NB	76.23	80.20	77.32
SVM	86.65	93.47	95.06
SLP	87.66	93.34	94.89

Tab. V Comparative performance analysis of bright lesions.

Methods	Sensitivity [%]	Specificity [%]	Accuracy [%]
NB	87.20	80.50	85.19
SVM	90.11	87.66	97.13
SLP	91.35	88.58	97.04

Tab. VI Comparative performance analysis of dark lesions.

From the above comparisons, it is clear that SLP gives comparable results with SVM and outperforms NB. Tab. VII shows the sensitivity, specificity, accuracy rate and area under Receiver Operating Characteristic (ROC) curve for each lesion type for SLP.

Lesion type	Sensitivity	Specificity	Accuracy	Area under curve
HEM	90.40 %	87.94 %	96.86 %	0.93
MA	92.30 %	89.23 %	97.23 %	0.97
HEX	83.19 %	92.60 %	93.27 %	0.86
CWS	95.56 %	97.13 %	98.90 %	0.99
Drusen	84.23 %	90.29 %	92.50 %	0.82

Tab. VII Performance rate of each lesion type for SLP classification.

To improve the accuracy level of the classification as well as narrow down the search, a feature extraction process is carried out, based on an adoption of a decision tree approach. The level of optimization is achieved by selecting optimal features from the set of fourteen features F1, F2, ..., F14.

The J48 classifier uses the C4.5 algorithm to build the decision tree. The training data set $\mathbf{S} = (s_1, s_2, \dots, s_n)$ contains samples where each sample s_i is again a m -dimensional vector $(r_{1,i}, r_{2,i}, \dots, r_{m,i})$. Here, r_j denotes the feature value of s_i along with its class in which s_i falls.

At each node, the algorithm selects the attribute which has the highest normalized information gain. In this case, among the texture based feature set, mean, uniformity of region, color and sharpness of edge are selected as optimal features and among shape based feature set, area, solidity and eccentricity of region are selected as optimal features which help to classify the bright lesions. This set of optimal features score 95.13% accuracy using SLP which is an improvement over other methods. For dark lesion classification, all of the fourteen features are selected as optimal features. Tab. VIII describes the accuracy achieved by SLP, SVM and NB with optimal feature set for bright lesion classification.

Feature extraction		Accuracy		
Feature name	Feature no.	NB [%]	SVM [%]	SLP [%]
Area, Solidity, Eccentricity, Mean, Uniformity, Color, Sharpness of edge	F1, F4, F7, F8 F10, F13, F14	79.36	95.08	95.13

Tab. VIII Feature selection of bright lesions.

The performance of the proposed system was compared with the other existing techniques. Tab. IX shows a comparison table of bright lesions i.e. HEX, CWS and drusen classification results with other methods. Tab. X depicts a comparison table for dark lesions i.e. MA, HEM classification, with other proposed methods.

Method	Selected Lesions	Accuracy [%]
Niemeijer et. al. [2]	HEX, CWS & Drusen	95.00 (area under ROC curve)
Rocha et.al. [15]	HEX only	95.30 (area under ROC curve)
Akram et. al. [6]	HEX only	97.56
Proposed method (using SLP with optimal features)	HEX, CWS & Drusen	95.13

Tab. IX Comparison table for bright lesion classification with other systems.

Method	Selected Lesions	Accuracy [%]
Niemeijer et al. [3]	MA, HEM	100.00 (sensitivity) and 87.00 (specifiicty)
Rocha et. al. [15]	MA, HEM	93.30 (area under ROC curve)
Akram et al. [6]	MA, HEM	98.12
Proposed method (using SVM)	MA, HEM	97.13

Tab. X Comparison table for dark lesion classification with other systems.

6. Conclusion

The proposed work is successfully able to design a CAD system for the diagnosis of the early stages of DR, specifically NPDR as well as dry AMD. With the help of the features selected in this technique, we are able to detect the different types of bright lesions as well as dark lesions. In comparison with other research works [3, 6, 15] the important aspects of the proposed algorithm are discussed in the following.

The novelty of this paper lies in designing a CAD system with the combination of Gaussian filtering and multilevel thresholding approach to segment the candidate lesions, which is more efficient than the segmentation using Fuzzy c means clustering technique. Another strong side of this technique is that it is able to retain the full contour information of the candidate lesions. The same single procedure is also able to extract the most challenging NPDR dark lesion symptom which is, the first manifestation of DR called MA. In addition to that it also considers the most popular dry AMD symptom, drusen with the other bright lesion symptoms of NPDR. The proposed method is able to distinguish the different types of bright lesions like CWS, drusen and HEX where other methods focus on HEX only. Hemorrhages are symptoms associated with the advancement of NPDR. The proposed technique is also able to identify such lesions. The proposed system is able to handle the detection and elimination of the non-lesion regions namely blood vessels and OD using strong and effective approaches. Here the proposed OD detection technique is able to extract the full shape of the OD.

Another interesting aspect of this approach is that the preprocessing part of the proposed system does not need to separate the black background region from the main retina portion in the fundus retina image. For feature extraction, a combination of strong features such as geometrical shape based and texture based features are extracted to get the detailed information of the lesions. Though SVM is considered as an efficient classification technique over SLP in terms of time complexity and accuracy, SLP classification technique is appreciated as its accuracy for the detection of bright lesion classes is more than SVM classifier on extracted feature set. The linearly separable dataset used for this research work is well suited to SLP.

The performance accuracy of SVM and SLP for bright lesion classification using fourteen features are 95.06% and 94.89%, respectively. Using optimal feature set, SLP achieves an accuracy of 95.13% which outperforms SVM with 95.08% of accuracy. In the detection of dark lesions, SVM and SLP achieve 97.13% and 97.04% of classification accuracy, respectively. Hence, SVM and SLP are comparable in accuracy for this research work.

The proposed technique is robust to the image resolution, retinal color intensity variation and the position, size of the lesions. Beside these salient aspects, there are still some places for improvement of the proposed approach. One such limiting factor is that the size of database used for training and testing purpose, may be extended to modify the accuracy level of the proposed system. It may be mentioned here that, this paper proposes a methodology for both identifying and distinguishing different types of dark and bright lesions, and is thus able to perform better than other methods, and with comparable accuracy. The proposed system shows a reliable and quick diagnosis approach for the early stage of DR called NPDR.

Acknowledgement

The authors would like to acknowledge a grant from TEQIP and Department of Biotechnology, Government of India (No. BT/PR4256/BID/7/393/2012 dated 02.08.2012) for supporting this research. Authors would like to acknowledge Medical College, Kolkata for supplying images for this project and for their great support.

References

- [1] Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. *Ophthalmology*, 1991, 98(5), pp. 786–806, doi: [10.1016/S0161-6420\(13\)38012-9](https://doi.org/10.1016/S0161-6420(13)38012-9).
- [2] NIEMEIJER M., GINNEKEN B.V., RUSSELL S.R., SUTTORP-SCHULTEN M.S., ABRAMOFF M.D. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative Ophthalmology and Visual Science*, 2007, 48(5), pp. 2260–2267, doi: [10.1167/iovs.06-0996](https://doi.org/10.1167/iovs.06-0996).
- [3] NIEMEIJER M., GINNEKEN B.V., STAAL J., SUTTORP-SCHULTEN M.S., ABRAMOFF M.D. Automatic Detection of Red Lesions in Digital Color Fundus Photographs. *IEEE Transaction on Medical Imaging*, 2005, 24(5), pp. 584–592, doi: [10.1109/TMI.2005.843738](https://doi.org/10.1109/TMI.2005.843738).
- [4] ANITHA J., HEMANTH D.J. An Efficient Kohonen-Fuzzy Neural Network Based Abnormal Retinal Image Classification System. *Neural Network World*, 2013, 2, pp. 149–167, doi: [10.14311/NNW.2013.23.011](https://doi.org/10.14311/NNW.2013.23.011).
- [5] CRABB J.W., MIYAGI M., GU X., SHADRACH K., WEST K.A., SAKAGUCHI H., KAMEI M., HASAN A., YAN L., RAYBORN M.E., SALOMON R.G., HOLLYFIELD J.G. Drusen proteome analysis : An approach to the etiology of age-related macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(23), pp. 14682–14687, doi: [10.1073/pnas.222551899](https://doi.org/10.1073/pnas.222551899).
- [6] AKRAM M.U., KHALID S., TARIQ A., KHAN S.A., AZAM F. Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine*, 2014, 45, pp.161–171, doi: [10.1016/j.combiomed.2013.11.014](https://doi.org/10.1016/j.combiomed.2013.11.014).
- [7] SPENCER T., OLSON J.A., MCHARDY K.C., SHARP P.F., FORRESTER J.V. An Image-processing Strategy for the Segmentation and Quantification of Microaneurysms in Fluorescein Angiograms of the Ocular Fundus. *Computers and Biomedical Research*, 1996, 29(4), pp. 284–302, doi: [10.1006/cbmr.1996.0021](https://doi.org/10.1006/cbmr.1996.0021).
- [8] FRAME A.J., UNDRILL P.E., CREE M.J., OLSON J.A., MCHARDY K.C., SHARP P.F., FORRESTER J.V. A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms. *Computers in Biology and Medicine*, 1998, 28(3), pp. 225–238, doi: [10.1016/S0010-4825\(98\)00010-0](https://doi.org/10.1016/S0010-4825(98)00010-0).
- [9] STAAL J., ABRAMOFF M.D., NIEMEIJER M., VIERGEVER M.A., GINNEKEN B.V. Ridge- Based Vessel Segmentation in Color Images of the Retina. *IEEE Transaction on Medical Imaging*, 2004, 23(4), pp. 501–509, doi: [10.1109/TMI.2004.825627](https://doi.org/10.1109/TMI.2004.825627).
- [10] NIEMEIJER M., ABRAMOFF M.D., GINNEKEN B.V. Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Transaction on Medical Imaging*, 2009, 28(5), pp. 775–785, doi: [10.1109/TMI.2008.2012029](https://doi.org/10.1109/TMI.2008.2012029).
- [11] AGURTO C., MURRAY V., BARRIGA E., MURILLO S., PATTICHIS M., DAVIS H., RUSSELL S., ABRAMOFF M., SOLIZ P. Multiscale AM-FM methods for diabetic retinopathy lesion detection. *IEEE Transaction on Medical Imaging*, 2010, 29(2), pp. 502–512, doi: [10.1109/TMI.2009.2037146](https://doi.org/10.1109/TMI.2009.2037146).

- [12] WISAENG K., HIRANSAKOLWONG N., POTHIRUK E. Automatic Detection of Retinal Exudates using a Support Vector Machine. *Applied Medical Informatics*, 2013, 32(1), pp. 33–42, available at <http://ami.info.umfcluj.ro/index.php/AMI/article/view/406/pdf>.
- [13] ROYCHOWDHURY S., DARA D.K., PARHI K.K. DREAM: Diabetic retinopathy analysis using machine learning. *IEEE Journal of Biomedical and Health Informatics*, 2014, 18(5), pp. 1717–1728, doi: [10.1109/JBHI.2013.2294635](https://doi.org/10.1109/JBHI.2013.2294635).
- [14] SOPHARAK A., UYYANONVARA B., BARMAN S., WILLIAMSON T.H. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 2008, 32(8), pp. 720–727, doi: [10.1016/j.compmedimag.2008.08.009](https://doi.org/10.1016/j.compmedimag.2008.08.009).
- [15] ROCHA A., CARVALHO T., JELINEK H.F., GOLDENSTEIN S., WAINER J. Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Transactions on Biomedical Engineering*, 2012, 59(8), pp. 2244–2253, doi: [10.1109/TBME.2012.2201717](https://doi.org/10.1109/TBME.2012.2201717).
- [16] OSAREH A., MIRMEHDI M., THOMAS B., MARKHAM R. Automated identification of diabetic retinal exudates in digital colour images. *British Journal of Ophthalmology*, 2003, 87(10), pp. 1220–1223, doi: [10.1136/bjo.87.10.1220](https://doi.org/10.1136/bjo.87.10.1220).
- [17] BANERJEE S., KAYAL D. Detection of hard exudates using mean shift and normalized cut method. *Biocybernetics and Biomedical Engineering*, 2016, 36(4), pp. 679–685, doi: [10.1016/j.bbe.2016.07.001](https://doi.org/10.1016/j.bbe.2016.07.001).
- [18] DASGUPTA M., BANERJEE S. Case Based Reasoning in the Detection of Abnormalities in Retina Images: A Survey. *International Journal of Research in Electronics and Computer Engineering*, 2014, 2(2), pp. 93–99, available at <http://nebula.wsimg.com/b3ca8575be5f3030ecf39657506853c7?AccessKeyId=DFB1BA3CED7E7997D5B1&disposition=0&alloworigin=1>.
- [19] SAHA R., ROYCHOWDHURY A., BANERJEE S. Diabetic Retinopathy Related Lesions Detection and Classification using Machine Learning Technology. *Springer International Publishing Switzerland, Artificial Intelligence and Soft Computing (ICAISC)*, 2016, 9693, pp. 734–745, doi: [10.1007/978-3-319-39384-1_65](https://doi.org/10.1007/978-3-319-39384-1_65).
- [20] ROYCHOWDHURY A., BANERJEE S. Detection of Abnormalities of Retina Due to Diabetic Retinopathy and Age Related Macular Degeneration Using SVM. *Science Journal of Circuits, Systems and Signal Processing*, 2016, 5(1), pp. 1–7, doi: [10.11648/j.cssp.20160501.11](https://doi.org/10.11648/j.cssp.20160501.11).
- [21] KAUPPI T., KALESNYKIENE V., et. al. DIARETDB0-Standard Diabetic Retinopathy Database, Calibration level 0. IMAGERET project 2007, <http://www.it.lut.fi/project/imageret/diaretdb0>.
- [22] KAUPPI T., KALESNYKIENE V., et. al. DIARETDB1-Standard Diabetic Retinopathy Database, Calibration level 1. IMAGERET project 2007, <http://www.it.lut.fi/project/imageret/diaretdb1>.
- [23] KAUPPI T., KALESNYKIENE V., et. al. DiaRetDB1 V2.1 - Diabetic Retinopathy Database and Evaluation Protocol. IMAGERET project 2007, http://www.it.lut.fi/project/imageret/diaretdb1_v2_1.
- [24] ARORA S., ACHARYA J., VERMA A., PANIGRAHI P.K. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. *Pattern Recognition Letters*, 2008, 29(2), pp. 119–125, doi: [10.1016/j.patrec.2007.09.005](https://doi.org/10.1016/j.patrec.2007.09.005).
- [25] ALSHAYEJI M., AL-ROOMI S.A., ABED S.E. Optic disc detection in retinal fundus images using gravitational law-based edge detection. *Medical & Biological Engineering & Computing*, 2017, 55(6), pp. 935–948, doi: [10.1007/s11517-016-1563-0](https://doi.org/10.1007/s11517-016-1563-0).
- [26] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 1958, 65(6), pp. 386–408, doi: [10.1037/h0042519](https://doi.org/10.1037/h0042519).
- [27] BANERJEE S., ROYCHOWDHURY A. Case based reasoning in the detection of retinal abnormalities using decision trees. *Procedia Computer Science*, 2015, 46, pp. 402–408, doi: [10.1016/j.procs.2015.02.037](https://doi.org/10.1016/j.procs.2015.02.037).

- [28] AGURTO C., BARRIGA E.S., MURRAY V., NEMETH S., CRAMMER R., BAUMAN W., ZAMORA G., PATTICHIS M.S., SOLIZ P. Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. *Investigative Ophthalmology & Visual Science*, 2011, 52(8), pp. 5862–5871, doi: [10.1167/iovs.10-7075](https://doi.org/10.1167/iovs.10-7075).
- [29] WITTEN I., FRANK E., HALL M. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edition, Morgan Kaufmann, 2011.