Computers in Biology and Medicine 45 (2014) 161–171

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Detection and classification of retinal lesions for grading of diabetic retinopathy

M. Usman Akram [a,*], Shehzad Khalid [b], Anam Tariq [a], Shoab A. Khan [a], Farooque Azam [a]

[a] Department of Computer Engineering, College of Electrical & Mechanical Engineering, National University of Sciences & Technology, Islamabad, Pakistan
[b] Department of Computer & Software Engineering, Bahria University, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 23 July 2013
Accepted 18 November 2013

Keywords:
Diabetic retinopathy
NPDR
Microaneurysms
Haemorrhage
Hard exudates
Cotton wool spots
m-Mediods
Classification

ABSTRACT

Diabetic Retinopathy (DR) is an eye abnormality in which the human retina is affected due to an increasing amount of insulin in blood. The early detection and diagnosis of DR is vital to save the vision of diabetes patients. The early signs of DR which appear on the surface of the retina are microaneurysms, haemorrhages, and exudates. In this paper, we propose a system consisting of a novel hybrid classifier for the detection of retinal lesions. The proposed system consists of preprocessing, extraction of candidate lesions, feature set formulation, and classification. In preprocessing, the system eliminates background pixels and extracts the blood vessels and optic disc from the digital retinal image. The candidate lesion detection phase extracts, using filter banks, all regions which may possibly have any type of lesion. A feature set based on different descriptors, such as shape, intensity, and statistics, is formulated for each possible candidate region: this further helps in classifying that region. This paper presents an extension of the *m*-Mediods based modeling approach, and combines it with a Gaussian Mixture Model in an ensemble to form a hybrid classifier to improve the accuracy of the classification. The proposed system is assessed using standard fundus image databases with the help of performance parameters, such as, sensitivity, specificity, accuracy, and the Receiver Operating Characteristics curves for statistical analysis.

1. Introduction

Diabetic Retinopathy (DR) is an eye abnormality caused by long term diabetes and it is one of the causes of vision impairment. DR is the most common cause of blindness before the age of 50 years [1, 2]. DR is a progressive disease but the main issue with the disease is that a patient with DR has almost no signs of vision impairment at the initial stages of the disease. The severity of DR is determined by the number and types of lesions present on the surface of the retina.

The human retina consists of different components, such as blood vessels, the fovea, the macula, and the optic disc (OD). DR is broadly divided into two stages: non-proliferative DR (NPDR) and proliferative DR (PDR). NPDR occurs when the blood vessels get damaged inside the retina and leak fluid onto the retina [3], causing the retina to become wet and swollen. In NPDR, different signs of retinopathy can exist, such as microaneurysms (MAs), haemorrhages (HMs), exudates (hard and soft) (EXs), and inter-retinal microvascular abnormalities (IRMA) [4]. PDR is an advanced stage of DR in which new abnormal blood vessels start growing in different regions of the retina and may lead to total blindness. In this paper, we mainly consider only those NPDR lesions which are MAs, HMs, or EXs.

MAs are the first sign of DR to be visible to an ophthalmologist; they occur due to leakage from tiny blood vessels of the retina. They are of smaller size, are circular in shape, and are red in color. HMs occur when the walls of MAs get ruptured. Dot haemorrhages are like bright red dots, and blot haemorrhages are larger red lesions [2]. When the leakage of blood contains lipids and proteins, it creates yellow spots on the retina known as EXs. They cause complete blindness if the accumulation of the lipid is near or on the macula. MAs and HMs are referred to as dark lesions and EXs as bright lesions [4]. The ophthalmologists normally grade NPDR into three categories: i.e., mild, moderate and severe, depending on the location and occurrence of the lesions [4]. Fig. 1 shows a healthy retina along with its main components. It also shows examples from different categories of NPDR.

There are many recent methods in the literature for the accurate detection of MAs, HMs and EXs by considering them individually and in a collective way. Ref. [6] presented a method based on successive clutter rejection in which a feature based system is formulated which passes only true MAs while rejecting false classes of clutter. For the accurate diagnosis of DR, the University of Iowa introduced the Retinopathy Online Challenge [7]. The results of the first international competition were reported

* Corresponding author. Tel.: +92 3336913921.
E-mail addresses: usmakram@gmail.com (M. Usman Akram), shehzad_khalid@hotmail.com (S. Khalid), anam.tariq86@gmail.com (A. Tariq), shoab@carepvtltd.com (S.A. Khan), farooque.azam@gmail.com (F. Azam).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.11.014

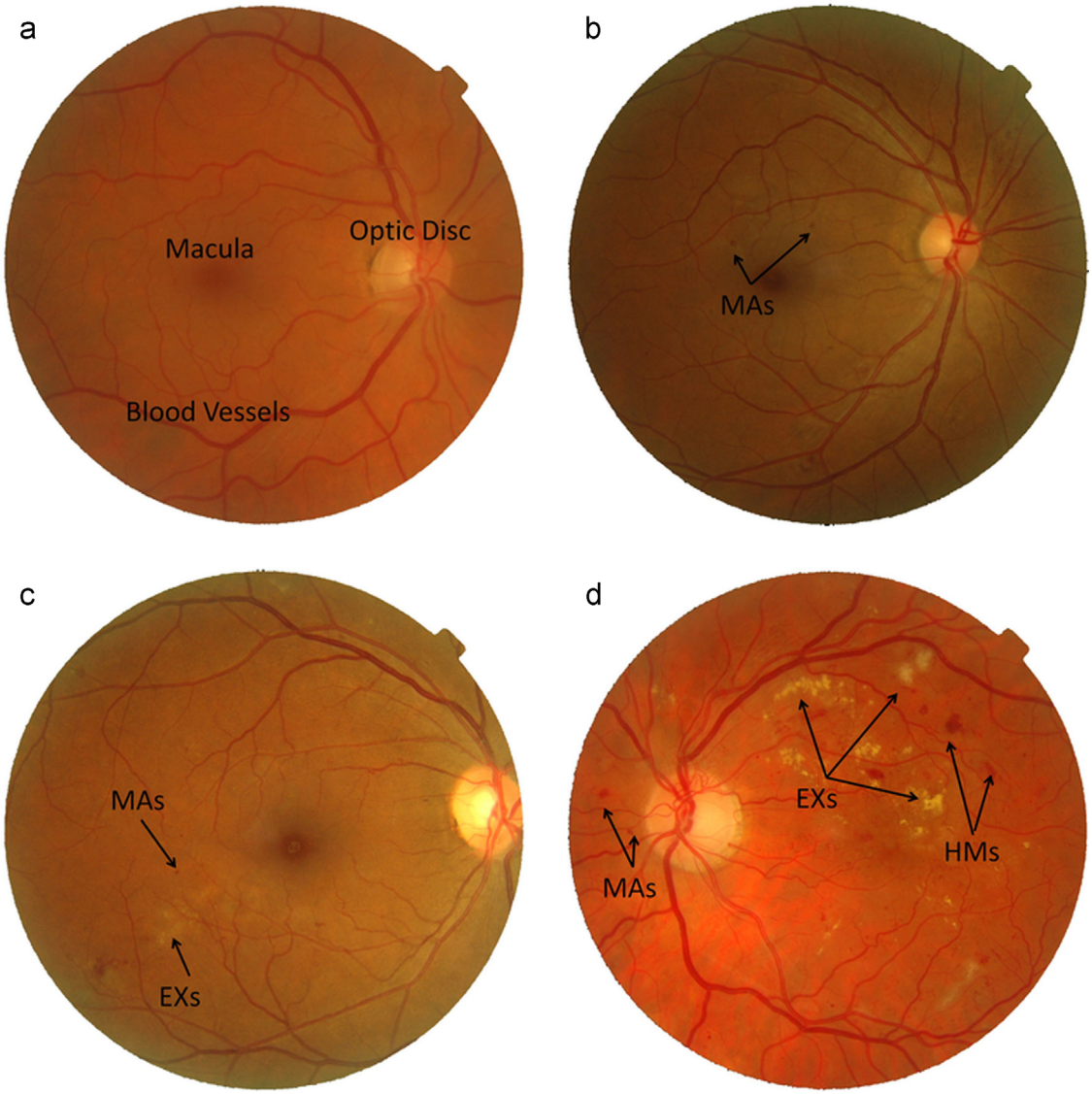

Fig. 1. Human retina and NPDR stages: (a) normal retina along with main components, (b) mild NPDR, (c) moderate NPDR, and (d) severe NPDR.

in [8]. Wavelet transform based matching using a Gaussian template is proposed in [9]. The technique was tested on 120 fundus images. Some automated diagnostic systems for DR considered haemorrhage and microaneurysm (HMA) as a common class [10] and used a moat operator for the detection. They have used only 30 fundus images for the classification of HMA and EXs. A similar study was done in [11] by using retinal images from 1273 patients. Hatanaka et al. [12] carried out the detection of HMs using the hue, saturation and value (HSV) model and the Mahalanobis distance, and tested the algorithm on 125 fundus images.

A multilayer neural network classifier was used to classify the segmented EXs in [13]. They used color, shape, size, and texture as the features and applied a genetic algorithm to choose a suitable subset of those features. The reported sensitivity and specificity are 96% and 94.6%, respectively. A generic contextual information based computer aided system was described in [14]. They described the spatial relation between the lesions and the identified exudates in two dimensional retinal images. A hybrid fuzzy neural network based classifier is used in [15] for dark and bright lesion detection. In [17], candidate regions for EXs were extracted using morphological closing of the luminance channel, local standard variation, and the watershed transform. Acharya et al. [18] proposed a system for the detection of NPDR and PDR by feeding higher order spectra based features to SVM. Their system graded the retinal images into different stages of NPDR with an average accuracy of 82%. Another system for automated detection of MAs, HMs, and EXs was presented in [19]. The proposed method achieved 82.6% and 88.3% accuracies for HMA and EXs, respectively.

This paper describes a system for the detection and classification of different types of NPDR lesions. The proposed technique extracts potential candidates for different signs of NPDR, i.e., MAs, HMs, and EXs. It then formulates a features set for each lesion depending on their properties. The true lesions are selected and classified using a hybrid classifier which is a weighted combination of multivariate *m*-Mediods and a Gaussian Mixture Model (GMM). The system uses the classification results. Then, based on the types, number, and location of the lesions, it grades the input retinal image into different categories of NPDR. The novelty of the proposed system lies in modeling of *m*-Mediods based classifier for grading NPDR.

This paper contains eight sections. Section 2 gives an overview of the complete proposed system and its different phases. Sections 3–5, respectively, present in detail the proposed techniques for preprocessing, detection of candidate lesions, and descriptions of all the features which we extract for all candidate lesions.

The modeling of the proposed hybrid classifier and grading criteria are described in Section 6, followed by the experimental results and comparisons with previous techniques in Section 7. The last section summarizes the paper.

2. System overview

DR is a progressive disease and its detection at an early stage is very crucial for saving a patient's vision: this requires regular screening. An automated screening system for DR can help in reducing the chances of complete blindness due to DR along with lowering the work load on ophthalmologists. A computer aided diagnostic system for screening for DR should differentiate between a normal retina and a retina with possible DR. Moreover, it should grade the affected retina into the different categories of DR [20]. This paper presents a method for classifying the lesions present in a retinal image, which is then used to grade the affected retina according to the various categories of NPDR. Fig. 2 shows a complete flow diagram for the proposed system. The system is divided into three phases: preprocessing, retinal image analysis for the detection of possible lesion regions, and the classification of the regions into different lesions. The first phase performs a background estimation and extracts the blood vessels and optic disc to facilitate further processing. The second phase applies a filter bank to extract all possible regions of lesions, and represents each region by a feature vector. The last phase classifies the lesions into three classes: MAs, HMs, and EXs, using a hybrid classifier. Finally, the input image is graded as normal or as belonging to any one of the three stages of NPDR.

3. Preprocessing and component extraction

Digital fundus images normally contain a main region at the center of the image, surrounded by dark background pixels. In the automatic diagnosis of DR, as only the main retinal pixels are required for processing, we therefore separate the background pixels from the foreground in the preprocessing step. Our preprocessing algorithm separates the background using a mean- and variance-based method and it also removes the noise from the image using hue, saturation, and intensity channels. The detailed work for the background separation used in this system is given in [21]. After the background separation, the system extracts the main components, such as the optic disc and the blood vessels, from the retina, something which will help eliminate any spurious and false regions caused by their

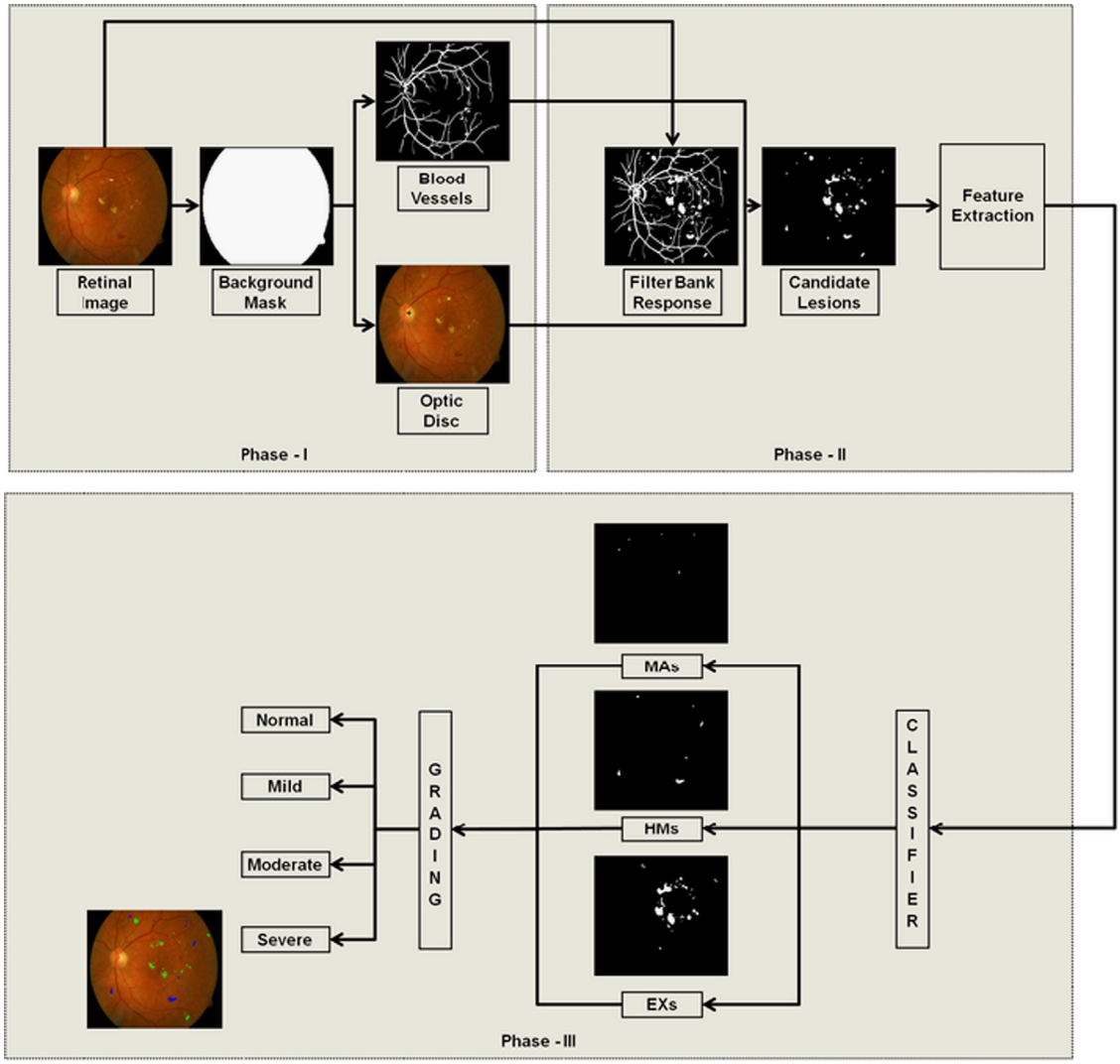

Fig. 2. Flow diagram for the proposed NPDR classification system.

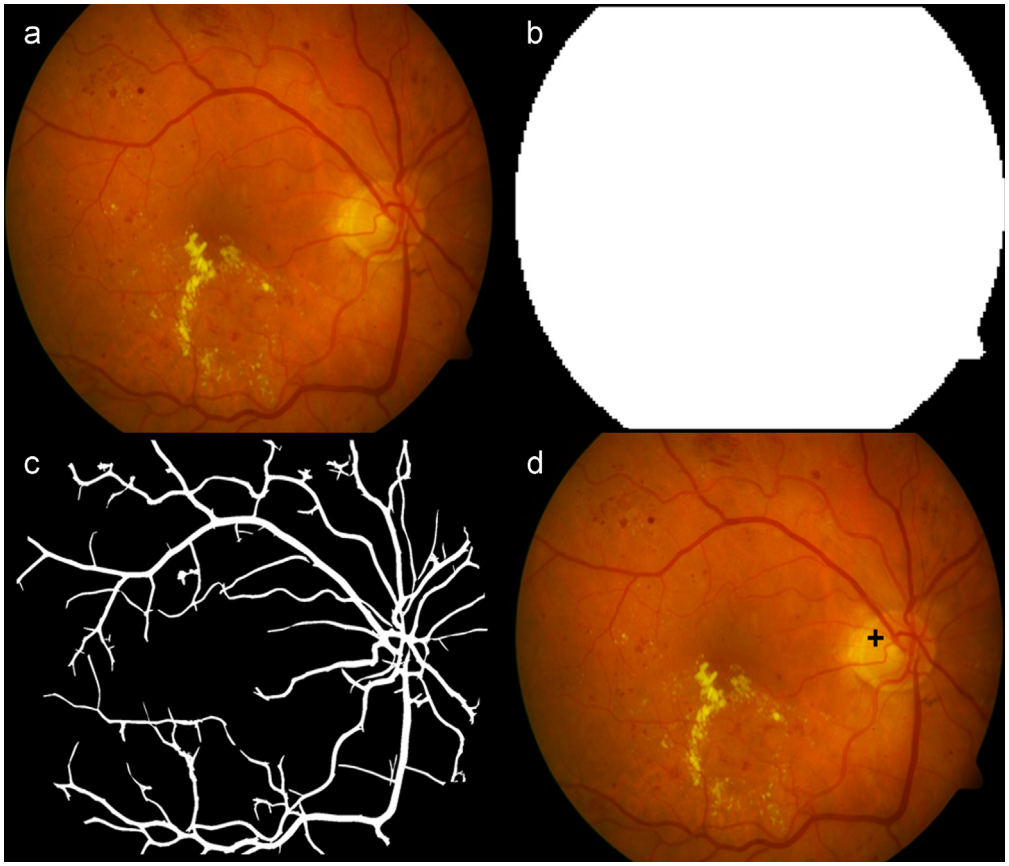

Fig. 3. Preprocessing and component extraction: (a) original retinal image; (b) background mask; (c) extracted vascular pattern; (d) localized optic disc.

similarities with bright and dark lesions, respectively. The system uses a Gabor wavelet and multilayered thresholding based method for blood vessel segmentation. Gabor wavelets are used due to their directional enhancement capabilities. Multilayered thresholding helps segment the large vessels along with thin capillaries. The optic disc is localized using an averaging filter and finding the circular region with the maximum intensity values. The algorithms for blood vessel segmentation and optic disc localization are given in [24] and [25], respectively. Fig. 3 shows the outputs of the preprocessing and component extraction phase along with the original retinal image.

4. Candidate lesion detection

NPDR lesions appear as different spots on the surface of the retina. A computer based diagnostic system should detect all the lesions present in the input fundus image for classification of the different signs of DR. The MAs and HMs appear as red or dark regions in color or gray image, whereas EXs appear as yellowish or bright spots and these lesions have the highest contrast in the green plane of the color image [17]. We use morphological opening and closing operations on the acquired retinal image to preserve the dark and bright regions, respectively. These give us fundus regions ϕ_f and ζ_f containing dark and bright regions, respectively, but they need contrast enhancement. The objective of the contrast enhancement is to make the detection of the dark and bright regions easy, by using an adaptive contrast enhancement technique [22] specified by

$$g = 255\frac{[\Phi_w(\phi_f) - \Phi_w(\phi_{fmin})]}{[\Phi_w(\phi_{fmax}) - \Phi_w(\phi_{fmin})]} \tag{1}$$

where Φ_w is the sigmoid function for a window defined by

$$\Phi_w(\phi_f) = \left[1 + \exp\left(\frac{m_w - f}{\sigma_w}\right)\right]^{-1} \tag{2}$$

and ϕ_{fmax}, ϕ_{fmin} are, respectively, the maximum and minimum intensity values of the opened green channel image. Here, m_w and σ_w are the mean and variance of the intensity values within the window, g represents the contrast enhanced image which is given as an input to the Gabor filter banks for enhancement. A Gaussian kernel function is used to represent a Gabor filter and it can model a variety of shapes by tuning its parameters [23], which makes it suitable for lesion enhancement. The Gabor filter bank is defined as follows:

$$G(x, y, \sigma, \Omega, \theta, r) = \frac{1}{\sqrt{\pi r}\sigma} e^{-(1/2)[(d_1/\sigma)^2 + (d_2/\sigma)^2]}(d_1(\cos\ \Omega + \imath\ \sin\ \Omega)) \tag{3}$$

where σ, Ω and r are the standard deviations of the Gaussian, the spatial frequency, and the aspect ratio, respectively, θ is the orientation of the filter, $d_1 = x\ \cos\ \theta + y\ \sin\ \theta$, and $d_2 = -x\ \sin\ \theta + y\ \cos\ \theta$. The contrast enhanced image g is convolved with the Gabor filter G centered at location (s,t) to generate the Gabor filter response γ for selected values of σ, Ω, and θ as [23]

$$\gamma(s, t, \sigma, \Omega, \theta, r) = \sum_x \sum_y g(x, y) G(s - x, t - y, \sigma, \Omega, \theta, r) \tag{4}$$

For the considered frequency and scale values, the maximum Gabor filter bank response $M_\gamma(\sigma, \Omega)$ is computed using Eq. (6) for θ spanning from 45° up to 180° at steps of 45°.

$$M_\gamma(\sigma, \Omega) = \max|\gamma(\sigma, \Omega, \theta)|. \tag{5}$$

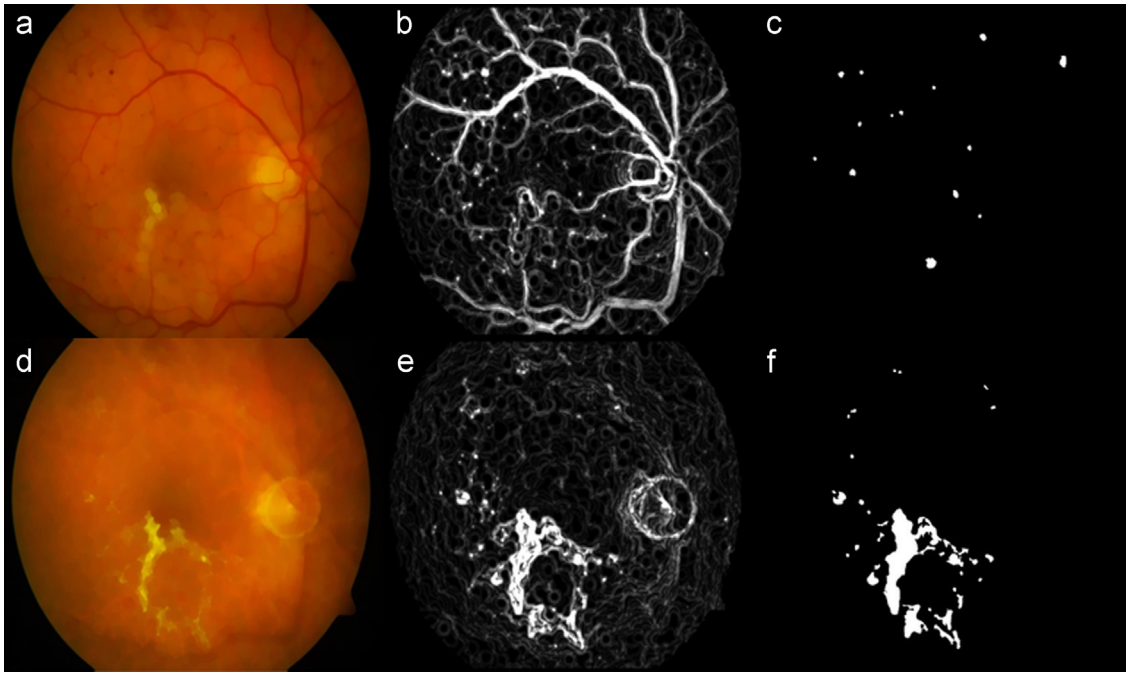

Fig. 4. Candidate lesion detection: (a,d) morphological opening and closing; (b,e) Gabor filter bank responses for dark and bright regions; (c,f) binary maps for candidate regions containing possible MAs, HMs and EXs.

Table 1
Description of features for classification of NPDR lesions.

Notation	Name	Description
f_1	Area	The sum of pixels in possible candidate region.
f_2	Eccentricity	The distance between foci of ellipse divided by its major axis.
f_3	Perimeter	The count of boundary pixels.
f_4	Mean intensity	The mean value of all green channel pixels within candidate region.
f_5	Aspect ratio	The ratio of major axis to minor axis of candidate region.
f_6	Compactness	Tells about the circularity of candidate region
f_7–f_9	Mean HSV	Mean values of hue, saturation and value for all pixels inside candidate region.
f_{10}	Mean enhanced intensity	The mean value of contrast enhanced green channel pixels within candidate region.
f_{11}	Mean gradient magnitude	The mean value for boundary pixels of candidate lesion.
f_{12}	Mean box gradient	The mean value of neighborhood pixels in a square region outside the candidate lesion.
f_{13}	Third moment	The skewness of histogram for all pixels in bounding square around candidate region.
f_{14}	Entropy	Calculated by including all pixels in bounding square around candidate region.
f_{15}	Mean range filter	The mean value of a local range of fundus image using a 5×5 neighborhood where range is the difference between maximum and minimum values in the specified neighborhood.
f_{16}	Energy	The sum of squares of all pixel intensities within candidate region.

The binary candidate regions for MAs, HMs and EXs are extracted from M_γ by applying a low threshold value T. A little bit of postprocessing is performed by applying morphological operations to fill the holes inside a connected region. These candidate regions also contain non-lesional pixels, such as blood vessels and OD, and for a reliable detection of the lesions, it is important to remove all such regions. The spurious regions are eliminated by removing the vessels and OD pixels, which were extracted in the preprocessing stage. Fig. 4 shows the outputs of the different phases of detection of candidate lesions. It shows the binary maps for MAs, HMs and EXs after the removal of blood vessels and OD.

5. Feature extraction

MAs, HMs and EXs appear with different properties, such as color, size, and shape. MAs are small in size and they appear in a dark red color with a circular shape; HMs are medium size dark red dots; and EXs are of yellow and whitish colors. For an automated system to distinguish between these lesions, a feature set is formed for each candidate lesion using four main descriptors: shape, color, intensity, and statistical features. In our feature set, we have not added Gabor features since they have already been used in the lesion enhancement and detection of the candidate regions. Table 1 contains the description of the features which we use in the proposed system.

6. Classification and grading

In this section, we model the lesion classes using labeled training samples, building on the feature space representation of the lesions outlined in the previous section. We present an extension of the m-Mediods based modeling approach, as presented in [26, 27], to cater for the multimodal distributions that

are expected in different lesion classes and we combine this modified *m*-Mediods based classifier with a GMM to form a hybrid classifier. The reason for choosing GMM is to cater for overlapping lesion patterns where the modeling of the distribution gives a good clue whereas *m*-Mediods caters for multivariate distributions of the samples within the pattern without neglecting patterns with small membership counts.

6.1. Gaussian Mixture Model (GMM)

In order to classify the candidate regions as dark and bright lesions based on their distributions, we use a two class Bayesian classifier using Gaussian functions, known as a Gaussian Mixture Model (GMM). The parameters for GMM are optimized using Expectation Maximization (EM) which is an iterative method that chooses the optimal parameters by finding the local maximum value of the GMM distributions for the training data. The details of the GMM which we have used here are given in [16].

6.2. m-Mediods based classifier

Given the feature space representation of the lesions, we generate model of normality for the different lesion classes. Our modeling approach, referred to as multivariate *m*-Mediods modeling, models the class containing *n* members with *m*-mediods known *a priori*. Let $DB^{(i)}$ be the classified training samples associated to lesion class *i*, and let *W* be the weight vector representation of the output nodes. The algorithm to model the lesion class is as follows:

1. Initialize the Semi-Fuzzy Self-Organizing Map (SFSOM) network with m_{init} neurons. We empirically set $m_{init} = 3{*}m$.
2. Initialize W_i (where $1 \le i \le m_{init}$) using the PDF $N(\mu, \Sigma)$ approximated using the training lesion samples in $DB^{(i)}$.
3. Determine the *k* Nearest Weights (*k*-NW) to the input feature vector *F* using

$$k-NW(F,\mathbf{W},k) = \{P \in \mathbf{W} \mid \forall R \in P, S \in \mathbf{W} - C, \|F,R\| \le \|F,S\| \wedge |P| = k\} \quad (6)$$

   where **W** is the set of weight vector representations of all output nodes, *P* is the set of *k* closest weight vectors, $\|.,.\|$ is the Euclidean distance function, and $k = \delta(t)$ where $\delta(t)$ is a function that determines the number of nearest neighbors to be impacted by the current learning iteration. The number of neighbors to be updated in a give training iteration decreases with the passage of time to move from coarse learning to fine learning of the network.
4. Train the SFSOM network by updating the *k* nearest weights (*P*) using

$$W_c(t+1) = W_c(t) + \alpha(t)\zeta(j)(F - W_c(t)) \quad \forall W_c \in P \quad (7)$$

   where W_c is the weight vector associated to the output neuron *c*, *j* is the order of closeness of W_c to *F* ($1 \le j \le k$), $\alpha(t)$ is the learning rate of the SFSOM in the current training cycle *t*, and $\zeta(j,k) = \exp(-(j-1)^2/2k^2)$ is a membership function which has an initial value of 1 and decreases with increasing values of *j*.
5. Reduce the learning rate $\alpha(t)$ and neighborhood size $\delta(t)$ as follows:

$$\alpha(t) = 1 - e^{2(t - t_{max})/t_{max}} \quad (8)$$

$$\delta(t) = \lceil \delta_{init}(1 - e^{2(t - t_{max})/t_{max}}) \rceil \quad (9)$$

   where t_{max} is the total number of training iterations and δ_{init} is the number of neighbors that were affected in the initial training iteration.
6. Iterate steps 3–5 until the maximum number of training iterations is reached.
7. Filter the output nodes with no training samples allocated to them.
8. Identify the closest pair of weight vectors, indexed as (*a*,*b*), given by the condition

$$(a,b) = \arg\min_{(i,j)} \|W_i, W_j\| \times \sqrt{|W_i| + |W_j|} \quad \forall i,\ j \wedge i \ne j \quad (10)$$

9. Merge the weight vectors W_a and W_b using

$$W_{ab} = \frac{|W_a| \times W_a + |W_b| \times W_b}{|W_a| + |W_b|} \quad (11)$$

   where $|.|$ is the function that returns the number of training lesion samples associated to a given weight vector.
10. Iterate steps 8–9 until the number of output nodes is equal to *m*. Append the weight vector **W** to the list of mediods $\mathbf{M}^{(i)}$ modeling the pattern *i*.
11. Approximate the density of the local distribution around each mediod by computing the mean of the distance of the mediod from its *k* nearest mediods. Append the average distance to that $\mathbf{D}^{(i)}$ which is in correspondence to a given mediod in the mediods list $\mathbf{M}^{(i)}$.

Once the multivariate *m*-Mediods based models of all the lesion classes have been generated, the classification of unseen lesion samples is achieved by computing the distance of the feature vector representation of the unseen lesion sample from the models of the different lesion classes, and is assigned to the class with the minimum distance. The algorithm for classification of unseen samples using *m*-Mediods based model comprises the following steps:

1. Select the *k* nearest mediods of test lesion *Q*, from the *m*-Mediods model of each class using

$$k-NM\{i\}(Q,\mathbf{M},k) = \{P \in \mathbf{M}^{(i)} \mid \forall R \in P,\ S \in \mathbf{M}^{(\mathbf{i})} - P \ \|Q,R\| \le \|(Q,S\| \wedge |P| = k\} \quad \forall i \quad (12)$$

   where $\mathbf{M}^{(\mathbf{i})}$ is the set of mediods from modeling class *i* and **P** is the ordered set of *k* closest mediods arranged in a ascending order in terms of their distance from the query sample *Q*.
2. Compute the membership $\Im\{i\}$ of the unseen lesion sample in the class *i* as

$$\Im\{i\} = 1 - \frac{\frac{\sum_{j=1}^{k} \|Q, k-NM\{i\}_j\|}{k}}{\overline{\mathbf{D}^{(i)}}} \quad (13)$$

   where $\overline{\mathbf{D}^{(i)}}$ is the average of the mean distances corresponding to mediods in **P** as identified in Eq. (12). The mean distance corresponding to a given mediod is precomputed and stored in $D^{(i)}$ as specified in step 11 of the modeling algorithm. Assign the sample to any of the lesion classes having the highest probability $\Im$.

The proposed multimodal *m*-Mediods classifier is further integrated with a GMM classifier using a weighted probabilistic ensemble to form a hybrid classifier. The selection of GMM for the ensemble is based on the fact that GMM can handle overlapping class distributions such as are expected in the problem at hand. The integration with GMM will enhance the capacity of the proposed multivariate *m*-Mediods classifier especially when we have a high degree of overlap amongst the class distributions.

Given a test lesion sample *Q*, the classification using the proposed classification framework, based on a probabilistic measure of the evidence from different classifiers, is carried out as

follows:

$$class(Q) = \arg \max_{\forall class_i} \left(\sum_{i=1}^{c} a_k * P_{\mathbf{C_k}}(y = class_i|Q) \right) \quad (14)$$

where $P_{\mathbf{C_k}}(y = class_i|Q)$ is the probability of $class_i$ given a test lesion sample using classifier k which is combined in a weighted ensemble framework by applying weights a_k to the probabilistic prediction of the class $\mathbf{C_k}$. A genetic algorithm is applied to search for an optimal combination of weights for each individual classifier in which the classifier ensemble is expected to give the best performance. We have employed k-fold cross-validation to evolve and evaluate the weight sets generated using the genetic algorithm.

6.3. Grading of NPDR

The classification stage labels each region in a candidate binary map as an MA, an HM, or an EX. Once the classification is made, the next stage is the grading of the input retinal image based on the type and count of the lesions extracted from it. We have used the following rules to grade the input image into different categories.

1. Normal: No lesion found.
2. Mild: Any MAs found satisfy $0 < \text{MAs} \leq 5$.
3. Moderate: Few MAs, HMs, or EXs are found such that $5 < \text{MAs} \leq 10$ and $0 < \text{HMs, EXs} \leq 5$.
4. Severe: MAs, HMs and EXs are found in greater quantity than moderate.

7. Experiments

7.1. Material

Some benchmark databases are publicly available for the evaluation of algorithms for the automated screening and diagnosis of DR. The purpose of these databases is to check the validity of automatic screening of DR and to compare the results with existing techniques. We have used four publicly available databases: DRIVE, STARE, DIARETDB, and MESSIDOR. DRIVE (Digital Retinal Images for Vessel Extraction) is a database which has been designed in the Netherlands to evaluate and compare different algorithms for vessel segmentation [29]. The images were captured by screening 400 diabetic patients between 25–90 years of age and 40 of them have been selected randomly. The images were captured using a Canon CR5 Non-Mydriatic retinal camera with a Field of View (FOV) of 45° and a resolution of 768×584. One of the oldest and most used retinal image databases is the STARE (STructured Analysis of REtina) dataset, which was designed for the structured analysis of the retina [28]. The database contains RBG images with 8 bits per channel and of size 700×605. DIARETDB (DIAbetic RETinopathy DataBase) is a database which was designed to evaluate automated lesion detection algorithms [30]. It contains 89 retinal images with different retinal abnormalities and is the best means of checking the accuracy of lesion detection. The images were captured with a 50° FOV and a resolution of 1500×1152. The MESSIDOR is another database which has been established to facilitate computer aided DR lesion detection [31]. The database was collected using a TopCon TRC NW6 Non-Mydriatic fundus camera with 45° FOV and resolutions of 1440×960, 2240×1488 and 2304×1536 with 8 bits per color plane.

The descriptions of the databases are gathered at two levels: image and lesion, with the help of ophthalmologists. The image level description contains a single description of the complete image, viz., normal or with DR. This means that if a fundus image contains any type of lesion, it is considered as affected. The affected images are further categorized into three categories depending on the type and number of lesions present on the surface of the retina: mild, moderate, and severe. Table 2 shows the image level descriptions for all databases.

Table 2
Image level database description.

Database	Images	Normal	With DR	Mild	Moderate	Severe
DRIVE	40	33	7	2	5	0
STARE	81	30	51	11	24	16
DIARETDB	89	5	84	31	34	19
MESSIDOR	1200	540	660	153	247	260
Total	1410	608	802	197	310	295

Table 3
Lesion level dataset specification.

Dataset	Total lesions	MAs	HMs	EXs
Set A	1546	492	366	688
Set B	1334	440	274	620
Set C	3480	1132	840	1508

The second level of description is at the lesional level, and provides a detailed description by giving the exact position, count, and type of lesions present in the fundus image. For the lesion level description, we need manually annotated ground truths created by medical experts. A binary map for each image in the databases is created using the information given in the ground truth by medical experts for the evaluation of the lesion level. We have generated three datasets using the above-mentioned four databases in such a way that Set-A contains lesions from DRIVE and STARE images, Set-B contains lesions from DIARETDB, and Set-C contains lesions extracted from MESSIDOR by medical experts. The datasets contain a total of 6360 lesions from 1410 retinal images with 3544 red lesions and 2816 bright lesions. The detailed specifications of the different types of lesions are given in Table 3.

7.2. Results

The results of a detailed quantitative and comparative analysis of the proposed system are given in this section. Detailed feature vectors for lesions present in the ground truth are used to generate models for GMM and m-Mediods. The patterns for m-Mediods are modeled using 100 mediods per pattern. Modeling of the patterns and the classification of unseen samples using the GMM are based on the approach described in [16]. We have randomly selected p% of the dataset, where $p = 50$, and treated it as the training set; the rest of the dataset is treated as the test set. We have used 7-fold cross validation to learn the weights of our classifier ensemble using the proposed weight learning approach. The experiment was repeated 50 times and the accuracies have been averaged to remove any bias caused by the different selections of the training and test sets.

The performance of the proposed system was measured using sensitivity (Sen), specificity (Spec), accuracy (Acc), and the Area Under the Receiver Operating Characteristics (ROC) Curve (AUC) as the figure of merit. Sensitivity, specificity, and accuracy are defined by Eqs. (15)–(17).

$$Sen = \frac{TP}{(TP + FN)} \quad (15)$$

$$Spec = \frac{TN}{(TN+FP)} \tag{16}$$

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{17}$$

where

- TP, True Positives: Lesion regions which are correctly classified by the classifier.
- FP, False Positives: Non-lesion regions which are wrongly classified as lesion regions by the classifier.
- TN, True Negatives: Non-lesion regions which are correctly classified by the classifier.
- FN, False Negatives: Lesion regions which are wrongly classified as non-lesion regions by the classifier.

The performance of proposed system in terms of sensitivity, specificity and accuracy for grading of DR at lesion and image level for all four databases are given in Table 4.

The statistical analysis of proposed system is done with the help of Receiver Operating Characteristics (ROC) curves which are plots of true positive rate (sensitivity) versus false positive rate (1-specificity). This analysis is done for performance evaluation of proposed Hybrid Classifier (HC). The experiments are conducted for lesion level evaluation. Fig. 5 shows the averaged ROC curves for GMM, *m*-Mediods and HC for three classes (MAs, HMs and EXs) using all databases. The overall area under the curve (AUC) for proposed system is 0.981.

To further evaluate the performance of proposed method, we compare the results of proposed system with existing techniques at lesion level. Tables 5 and 6 show the comparison of different methods with proposed method (PM) for dark (MAs and HMs) and bright (EXs) lesions respectively. It is important to note here that not all these systems have used same databases but the purpose to include these comparisons is to show that proposed system performed well in terms of sensitivity, specificity, accuracy and AUC.

To compare our system with the existing well-established classifier ensemble methods such as ADABOOST, Bagging and random subspace methods (RSM), we have tested these algorithms on our datasets and their comparison with proposed hybrid classifier is given in Table 7.

Once the classifier has detected all the lesions present in an input image, the system grades the image as normal, mild,

Table 4
Performance evaluation of proposed system for grading of DR.

Data Bases	Image Level			Lesion Level		
	Sen	Spec	Acc	Sen	Spec	Acc
DRIVE	95.23	94.73	95	94.26	94.74	94.03
STARE	100	91.30	97.5	98.79	97.43	97.89
DIARETDB	96.42	95.02	95.50	93.08	92.76	92.96
MESSIDOR	99.50	97.73	98.90	97.61	97.03	97.59
Overall	99.17	97.07	98.52	96.41	93.86	94.98

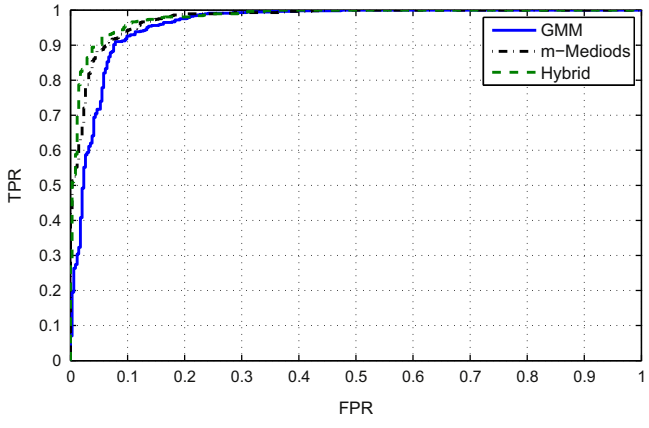

Fig. 5. ROC curves: lesion level performance of GMM, *m*-Mediods and HC for MAs, HMs and EXs.

Table 5
Comparison results for red lesions (MAs and HMs) detection.

Methods	Sen	Spec	Acc	AUC
Niemeijer et al. [8]	100	87	–	0.881
Quellec et al. [9]	89.62	–	92.9	0.92
Keerthi et al. [6]	88.46	–	–	0.85
Kahai et al. [33]	100	67	–	–
Larsen et al. [34]	71.4	96.7	82.6	0.903
Sinthanayothin et al. [10]	77.5	88.7	–	–
Proposed method	**97.83**	**98.36**	**98.12**	**0.974**

Table 6
Comparison results for EXs detection.

Methods	Sen	Spec	Acc	AUC
Sinthanayothin et al. [10]	88.5	99.7	–	–
clara et al. [32]	90.2	90	–	–
Niemeijer et al. [35]	95.0	86.0	–	0.95
Osareh et al. [5]	93	94.1	93.4	–
Walter et al. [17]	92.74	100	96.7	–
Proposed method	**97.39**	**98.02**	**97.56**	**0.97**

Table 7
Comparison of hybrid classifier with existing ensemble methods.

Methods	Acc
Adaboost	92.35
Bagging	96.78
RSM	96.23
Proposed method	**98.52**

Table 8
Confusion matrix, showing the NPDR grading at image level for all databases. The grading levels normal, mild, moderate, and severe are represented by I, II, III, and IV, respectively.

	Assigned Label						Assigned Label			
DRIVE	True Label	I	II	III	IV	STARE	True Label	I	II	III	IV
	I	32	1	0	0		I	28	2	0	0
	II	1	1	0	0		II	0	11	0	0
	III	0	0	5	0		III	0	1	23	0
	IV	0	0	0	0		IV	0	0	0	16
	Assigned Label						Assigned Label			
DiaretDB	True Label	I	II	III	IV	MESSIDOR	True Label	I	II	III	IV
	I	4	1	0	0		I	531	9	0	0
	II	1	29	1	0		II	4	149	0	0
	III	0	1	33	0		III	0	3	243	1
	IV	0	0	0	19		IV	0	1	2	257

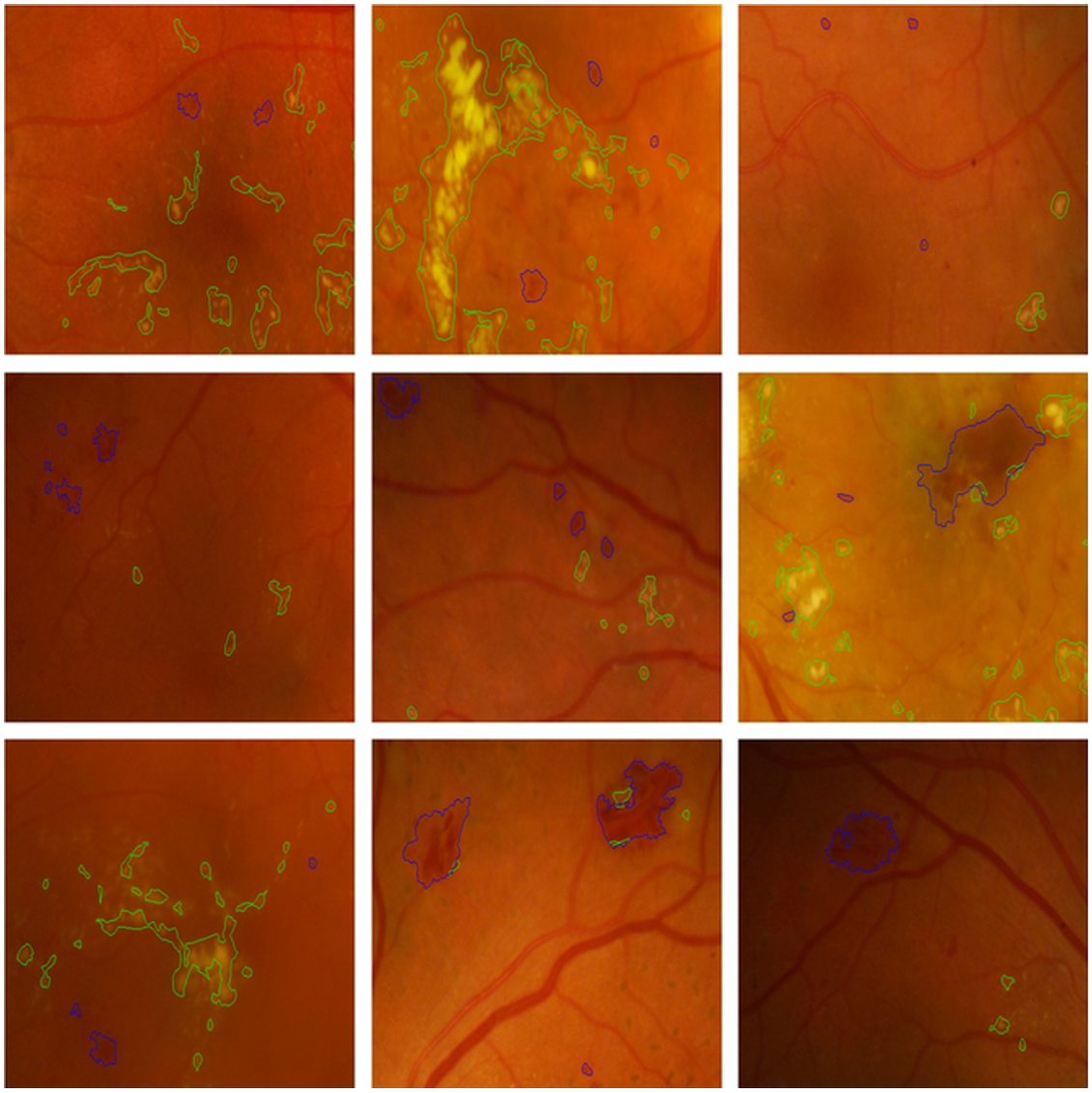

Fig. 6. Detection of dark and bright lesions. MAs and HMs are highlighted with a blue boundary and EXs are highlighted with a green boundary. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

moderate, or severe, based on the count, type, and location on the surface of retina of the lesions. Table 8 shows the total number of images which are graded correctly/incorrectly in the different severity levels.

Fig. 6 shows examples of the detection of both dark lesions and bright lesions. The lesions are highlighted with colored boundaries. Fig. 7 shows some examples of retinal image grading. MAs and HMs are marked with blue and EXs are marked with green.

8. Discussion and conclusion

In this paper, a system for the reliable grading of colored fundus images in different stages of non-proliferative Diabetic Retinopathy (NPDR) has been presented. A three stage model, comprising preprocessing, retinal image analysis, and classification, has been proposed. The preprocessing phase extracts background pixels to enable the processing of the further stages on the foreground pixels only. The main components, such as the vascular pattern and optic disc, are also extracted in the first phase to facilitate the later steps. The second phase detects all possible dark regions and bright regions which might include any type of Diabetic Retinopathy (DR) lesion, and then represents each region by a number of features for accurate classification.

In order to improve the classification of DR lesions, a new hybrid classifier has been proposed: it is an ensemble comprising both Gaussian Mixture Model (GMM) based classifier and *m*-Mediods based classifier. The purpose of using an ensemble was to ensure the involvement of the best properties of these two classifiers. This hybrid classifier (HC) assigns different weights to the classification probabilities to improve the overall accuracy, where the weights are learned using a genetic algorithm to enhance the classifier's performance. The testing of the proposed system as performed at the image and the lesional levels for a detailed evaluation. We have used sensitivity, specificity, accuracy, and the area under the Receiver Operating Characteristics (ROC) curves for its evaluation. The comparison between the classifiers using their ROC curves has shown the validity of the proposed HC classifier. It achieved a value of 0.981 as its area under the ROC curve, as against 0.977 for *m*-Mediods and 0.963 for GMM.

The results have shown that the proposed system detects all types of NPDR lesions and correctly grades the retinal images with high accuracy. The proposed system focused on the reliable

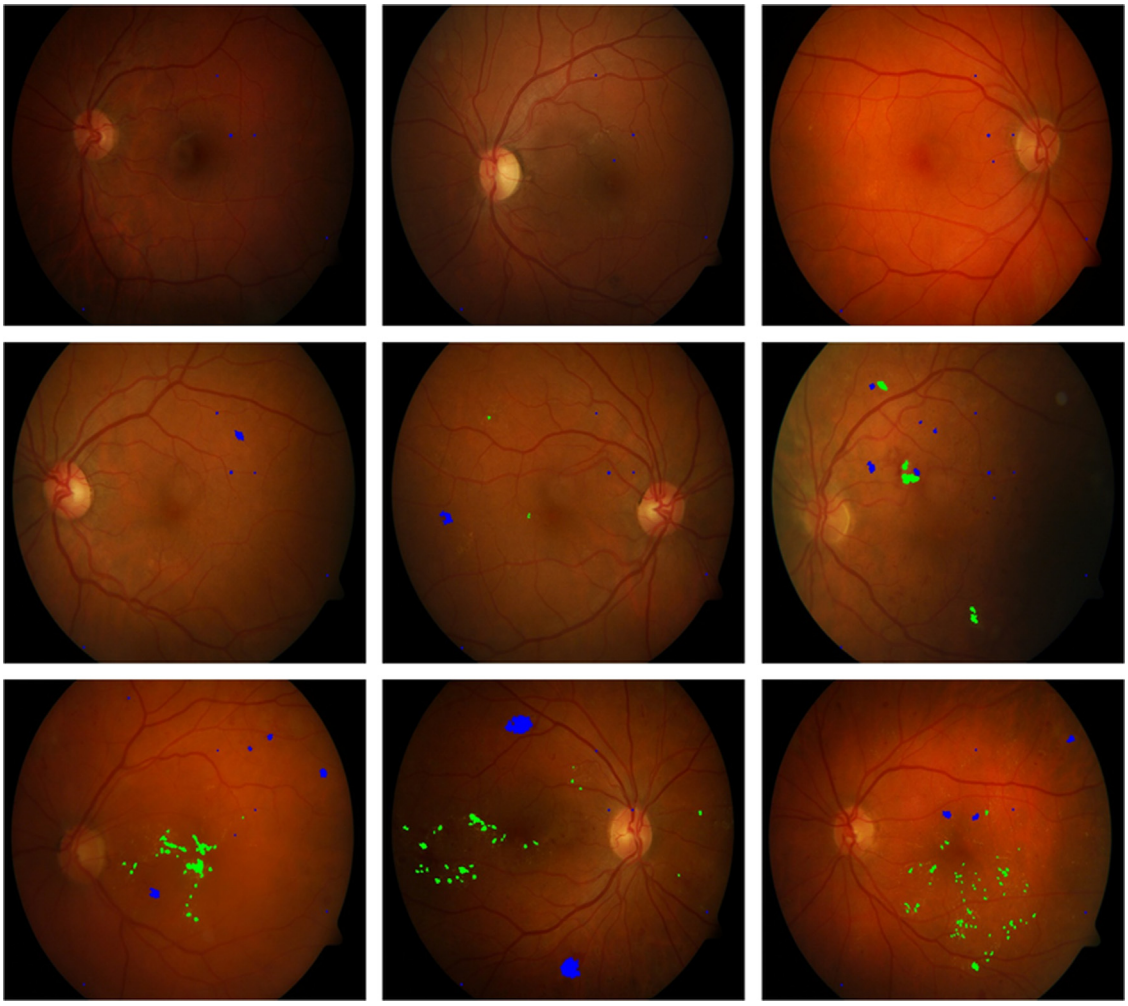

Fig. 7. Grading of NPDR. First row: mild NPDR, second row: moderate NPDR, third row: severe NPDR. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

detection of retinal abnormalities in order to develop an automated screening system for DR, and the results have shown the validity of the proposed system. Such systems can be used by non-experts for the initial examination of patients, so that they can then refer only those cases to ophthalmologists which are graded as having DR.

Conflict of interest statement

None declared.

Acknowledgments

This research was funded by the National ICT R&D Fund, Pakistan.

References

[1] R. Klein, B.E.K. Klein, S.E. Moss, Visual impairment in diabetes, Ophthalmology 91 (1984) 1.
[2] A.K. Sjolie, J. Stephenson, S. Aldington, E. Kohner, H. Janka, L. Stevens, J. Fuller, EURODIAB Complications Study Group, Retinopathy and vision loss in insulin-dependent diabetes in Europe, Ophthalmology 104 (1997) 252.
[3] P.C. Ronald, T.K. Peng, A Textbook of Clinical Ophthalmology: A Practical Guide to Disorders of the Eyes and Their Management, 3rd edition, World Scientific Publishing Company, Singapore, 2003.
[4] R.N. Frank, Diabetic retinopathy, Prog. Retin. Eye Res. 14 (2) (1995) 361.
[5] A. Osareh, M. Mirmehdi, B. Thomas, R. Markham, Automated identification of diabetic retinal exudates in digital colour images, Br. J. Ophthalmol. 87 (10) (2003) 1220.
[6] K. Ram, G.D. Joshi, J. Sivaswamy, A successive clutter-rejection-based approach for early detection of diabetic retinopathy, IEEE Trans. Biomed. Eng. 58 (3) (2011) 664.
[7] Retinopathy Online Challenge Website. The University of Iowa and the ROC organizers [Online]. Available ⟨http://roc.healthcare.uiowa.edu/results.php⟩.
[8] M. Niemeijer, B. van Ginneken, M. Cree, A. Mizutani, G. Quellec, C.I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A.M. Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, M. Abramoff, Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs, IEEE Trans. Med. Imaging 1 (29) (2010) 185.
[9] G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, C. Roux, Optimal wavelet transform for the detection of microaneurysms in retina photographs, IEEE Trans. Med. Imaging 27 (9) (2008) 1230.
[10] C. Sinthanayothin, J.F. Boyce, T.H. Williamson, H.K. Cook, E. Mensah, S. Lal, D. Usher, Automated detection of diabetic retinopathy on digital fundus images, Diabet. Med. 19 (2) (2002) 105.
[11] D. Usher, M. Dumsky, M. Himaga, T. Williamson, S. Nussey, J. Boyce, Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening, Diabet. Med. 21 (2004) 84.
[12] Y. Hatanaka, T. Nakagawa, Y. Hayashi, M. Kakogawa, A. Sawada, K. Kawase, T. Hara, H. Fujita, Improvement of automatic hemorrhages detection methods using brightness correction on fundus images, SPIE Med. Imaging 6915 (2008) 5429.
[13] O. Alireza, B. Shadgar, R. Markham, A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images, IEEE Trans. Inf. Technol. Biomed. 13 (4) (2009) 535.
[14] I.S. Clara, M. Niemeijer, I. Ivana, D. Alina, S.A.S. Maria, M.D. Abramoff, B. van Ginneken, Contextual computer-aided detection: improving bright lesion detection in retinal images and coronary calcification identification in CT scans, Med. Image Anal. 16 (1) (2012) 50.

[15] M.U. Akram, S.A. Khan, Automated detection of dark and bright lesions in retinal images for early detection of diabetic retinopathy, J. Med. Syst. 36 (5) (2012) 3151.
[16] M.U. Akram, A. Tariq, M.A. Anjum, M.Y. Javed, Automated detection of exudates in colored retinal images for diagnosis of diabetic retinopathy, OSA J. Appl. Opt. 51 (20) (2012) 4858.
[17] T. Walter, J.C. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy—detection of exudates in color fundus images of the human retina, IEEE Trans. Med. Imaging 21 (10) (2002) 1236.
[18] U.R. Acharya, K.C. Chua, E.Y.K. Ng, W. Wei, C. Chee, Application of higher order spectra for the identification of diabetes retinopathy stages, J. Med. Syst. 32 (6) (2008) 431.
[19] S.C. Lee, E.T. Lee, Y. Wang, R. Klein, R.M. Kingsley, A. Warn, Computer classification of nonproliferative diabetic retinopathy, Arch. Ophthalmol. 123 (6) (2005) 759.
[20] H.N. Iyer, A. Can, B. Roysam, C.V. Stewart, H.L. Tanenbau, A. Majerovics, H. Singh, Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy, IEEE Trans. Biomed. Eng. 53 (6) (2006) 1084.
[21] A. Tariq, M.U. Akram, An automated system for colored retinal image background and noise segmentation, in: Proceedings of IEEE Symposium on Industrial Electronics and Applications, vol. 405, 2010.
[22] C. Sinthanayothin, J.F. Boyce, H.L. Cook, T.H. Williamson, Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images, Br. J. Ophthalmol. 83 (1999) 902.
[23] J. Sung, S.Y. Bang, S. Choi, A Bayesian network classifier and hierarchical gabor features for handwritten numeral recognition, Pattern Recognit. Lett. 27 (1) (2006) 66.
[24] M.U. Akram, S.A. Khan, Multilayered thresholding-based blood vessel segmentation for screening of diabetic retinopathy, Eng. Comput. 29 (2013) 165.
[25] M.U. Akram, A. Khan, K. Iqbal, W.H. Butt, Retinal image: optic disk localization and detection, in: International Conference on Image Analysis and Recognition, Lecture Notes in Computer Sciences, vol. 6112, 2010, p. 40.
[26] S. Khalid, Motion based behavior learning, profiling and classification in the presence on anomalies, Pattern Recognit. 43 (1) (2010) 173.
[27] S. Khalid, Activity classification and anomaly detection using *m*-Mediods based modeling of motion patterns, Pattern Recognit. 43 (10) (2010) 3636.
[28] A. Hoover, M. Goldbaum, Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels, IEEE Trans. Med. Imaging 22 (8) (2003) 951.
[29] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501.
[30] T. Kauppi, V. Kalesnykiene, J.K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, J. Pietilä, DIARETDB1 Diabetic Retinopathy Database and Evaluation Protocol, Technical Report, 2006.
[31] MESSIDOR ⟨http://messidor.crihan.fr/index-en.php⟩.
[32] I.S. Clara, M. García, A. Mayoc, M.I. López, R. Hornero, Retinal image analysis based on mixture models to detect hard exudates, Med. Image Anal. 13 (4) (2009) 650.
[33] P. Kahai, K.R. Namuduri, H. Thompson, A decision support framework for automated screening of diabetic retinopathy, Int. J. Biomed. Imaging (2006) 1.
[34] M. Larsen, J. Godt, N. Larsen, H. Lund-Andersen, A.K. Sjolie, E. Agardh, H. Kalm, M. Grunkin, D.R. Owens, Automated detection of fundus photographic red lesions in diabetic retinopathy, Invest. Ophthalmol. Vis. Sci. 44 (2) (2003) 761.
[35] M. Niemeijer, B. van Ginneken, S.R. Russell, S.A. Maria, S. Schulten, M.D. Abramoff, Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis, Invest. Ophthalmol. Vis. Sci. 48 (5) (2007) 2260.