IMAGE & SIGNAL PROCESSING

A Novel Approach for Detection of Hard Exudates Using Random Forest Classifier

C. Pratheeba[1] · N. Nirmal Singh[2]

Abstract

Diabetic Retinopathy is the major cause of blindness for diabetics in which the retina is damaged. Regular screening system help in detecting the early symptoms like exudates, which are due to the leakage of blood pressure of vessels. The significant role of proposed system is detecting the hard exudates in prevention of visual loss and blindness. Many researchers studied and investigated about detecting the exudates region but not satisfied with their results. Fundamental medical image processing steps with different techniques are implemented by the proposed system. Random Forest is a novel classification which is applied on color retinal images able to classify cluster of data with high accuracy. The performance of the proposed system is obtained by analyzing the accuracy obtained from the Random Forest classifier. These images are obtained from Diabetic Retinopathy Database (DIARETDB) database. The simulation results are obtained with the help of MATLAB 2018. By applying novel classification techniques improves the automatic detection of hard exudates from color retinal images. The achieved accuracy is compared with existing classifiers since the proposed Random Forest classifier provides the accuracy of 99.89% applied on color retinal images.

Keywords Diabetic retinopathy · Classification · Random forest · MATLAB · Computer aided detection

Introduction

Diabetic Retinopathy (DR) is the impact of diabetes on the eye. It is the most important eye disease that causes blindness. Patients with diabetes are more eagerly to develop such disease as they grow older. Early intimations of DR are the emergence of microaneurysms, hemorrhages and hard exudates. DR begins when diabetes mellitus damages small retinal blood vessels causing microaneurysms (MA), the small swellings that form on the side of tiny blood vessels as shown in Fig. 1.

As the damage increases, it causes hemorrhages deep inside the retina. Retinal hemorrhage can cause severe vision problems. As the disease advances, the retina reacts by growing abnormally weak, leaky, fragile and misdirected blood vessels. The more these abnormal blood vessels grow, the more they bleed and pull on the retina. In addition, the leaked fluid produces sediments composed of lipid byproducts called exudates. Exudates are yellow and appear in different sizes and locations within the retina. There are no early symptoms for Diabetic Retinopathy as patients do not suffer from vision problems until the late stages of the disease when treatments could be inefficient. Therefore, an early diagnosis of DR is vital, and preventing the disease from developing is desirable. Furthermore, it is exceptionally noteworthy that diabetic patients experience a yearly eye exam [1–3]. The human eye can detect DR from retinal fundus images with the help of manual blood vessels segmentation and lesion extraction [20]. However, this is a very time-consuming process that requires training and could be affected by the observer's fatigue. Therefore, researchers have presented different solutions to detect DR automatically. The manifestations of the

---

✉ C. Pratheeba
   cpratheeba@yahoo.co.in

   N. Nirmal Singh
   nirmalsingh_n@hotmail.com

[1] Department of Electronics and Communication Engineering, Udaya School of Engineering, Vellamodi, India

[2] Department of Electronics and Communication Engineering, V.V College of Engineering, Thisayanvilai, Tirunelveli, India
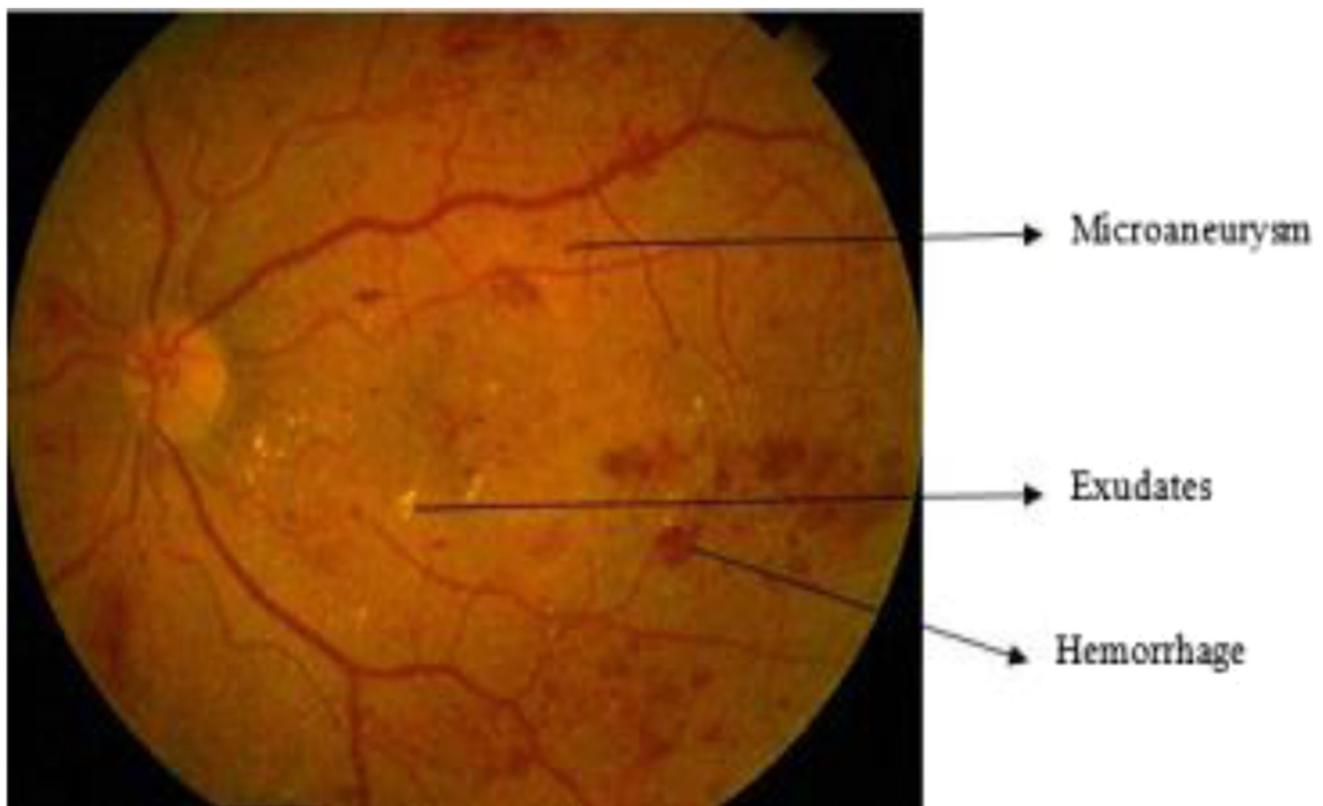
Fig. 1 Retinal image containing Non-Proliferative DR symptoms

abnormalities present in retinal OCT images obtained in DIARETDB database such as Hard exudates, microaneurysms, hemorrhages and soft exudates are discussed in which the area affected for Hard exudated and Hemorrhages is large whereas the area affected for Soft exudates and Microaneurysms is small. Several classification techniques have been proposed by the researchers and are implemented on classifying the retinal Optical Coherent Tomography (OCT) images are studied. Carrera et al. explored many detection techniques for classify the grade of non-proliferative Diabetic retinopathy for hard exudates detection system. The starting preprocessing step isolates hard exudates and micro aneurysms for feature extraction. The classifiers like Support vector Machine (SVM) and Decision tree is performed and it provides 92.5% of accuracy, 95% of sensitivity and predictive capacity of 94% [1].

Literature review

Gegundaz et al. studied various ways to reduce blindness of patient by finding the loss visual acuity at its early stages. A computer aided classification is applied on OCT fundus images of retina using feature extraction and supervised classification method so that the hard exudates region is detected from the OCT images. By using this method the system can

provides 90% of sensitivity and specificity of 70% from the screening of patient point of observation [2]. Jaya et al. proposed an approach to identify the Diabetic retinopathy at initial stage with the help of OCT images. This approach applies the Fuzzy support vector machine to detect the hard exudates from the color fundus images. This technique uses morphological operations to avoid false alarms. The achieved accuracy rate of this method is 96%, with a sensitivity of 94.1% and specificity of 90% [3]. Roy et al. identified the key concern which affects the accuracy of the classification techniques. Filter based extraction method and Fuzzy C means approaches are applied for exudates detection. Further SVM method is used for classifying the retinal fundus images into PDR and NPDR. This method can provides the accuracy of 96.23% which is higher than existing method [4]. Rajput and Preethi explored the challenging tasks in diabetic retinopathy identification and classification in the early stage of various medical images. This method classifies various pathologies and grades the exudates and non exudates. Hough transform approach is applied for Optic disc removal and k-means clustering method is used for hard exudates detection [5].

Ghazal et al. probed into the accurate requirements of classification techniques in the diagnosis of diabetic retinopathy. An improved CAD method is connected for detection and reviewing of NPDR. The novel strategy utilizes three stages for diagnosis beginning with integrated joint model

segmentation which joined the intensity and spatial information. Followed by some feature parameters such as thickness, curvature and reflectivity are calculated. At last step Deep Fusion Classification Network (DFCN) to classify the retinal fundus image. This method concludes that it provides 93% of accuracy, 91% of sensitivity and 97% of specificity [6]. Agurto et al. investigated the exudates present in retinal OCT images and introduced the feature extraction method followed by a classification technique. This method uses Partial Least Square (PLS) for classifying the images and achieved with high accuracy of 97% [7]. Akram et al. studied the various methods for detecting the Diabetic Retinopathy. A new hybrid classifier is applied for the detection of exudates region in the OCT image. This technique uses m-meriods based modeling approach and Gaussian mixture method as a hybrid classifier for better accuracy of classification [8]. Dutta et al. proposed an approach to identify the Diabetic Retinopathy. This method uses Deep learning model for classification which classifies the various kinds of pathologies in OCT images. Fuzzy C-means method is used to find the weights that present the harshness level of thresholds [9].

To improve the accuracy in classifying the retinal OCT images, the proposed system makes use of Random Forest classifier. This classification technique improves the accuracy by categorizing the exudates and non-exudates present in retinal OCTs efficiently and provide higher accuracy than others.

In this paper, the principle center is on classification of pixels between two classes, exudates and non-exudates from the retinal images. Existing methods do not provide sufficient accuracy in classifying the exudates and non-exudates images. The constraint in the conventional classifiers like Artificial Neural Networks, Deep Neural Networks and Convolutional Neural Networks do not provide efficient accuracy when applied on retinal OCT images [19]. The proposed classification technique Random Forest (RF) provides much better accuracy than the existing classification techniques attempted on retinal images.

Fig. 2 Block Diagram of proposed system
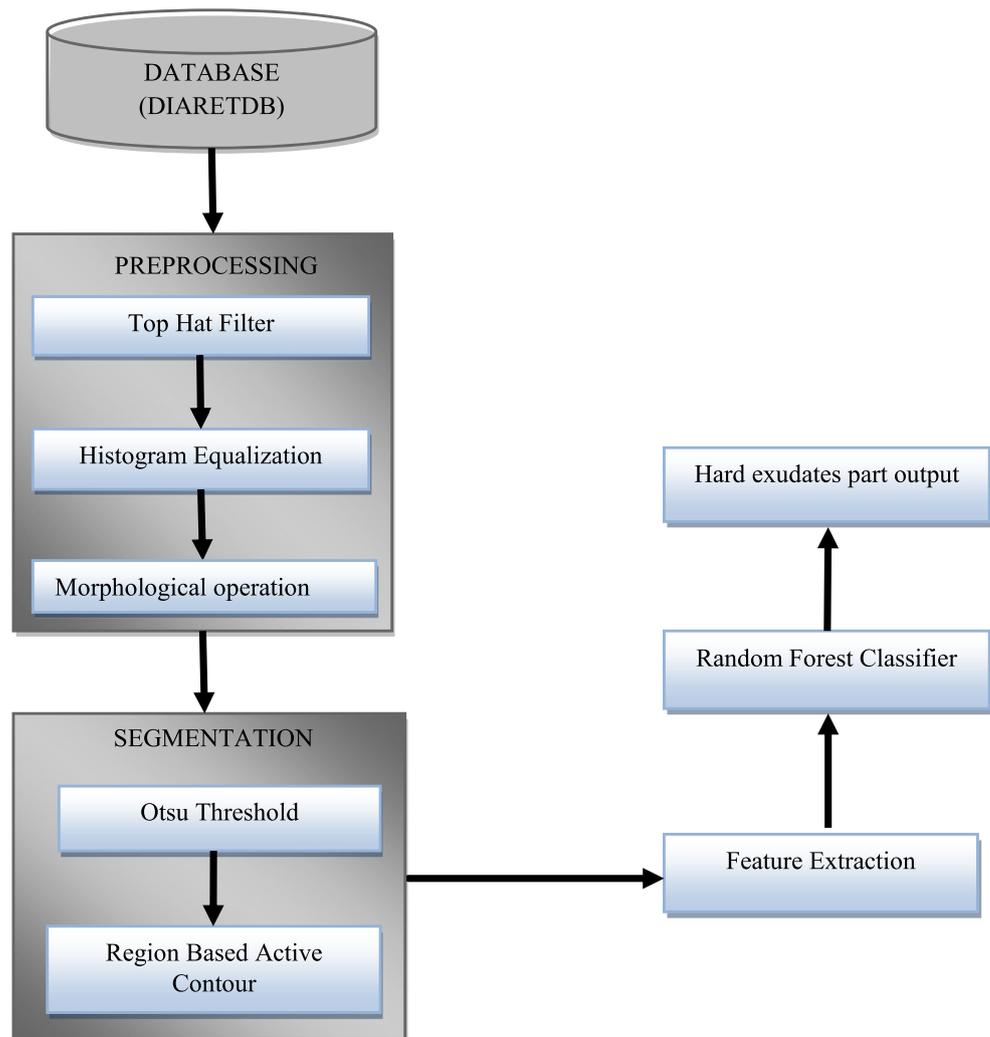
The main contribution of this paper is as follows;

1. Top Hat filter is not applied on any retinal OCT images whereas this filter is used by the proposed system to improve the quality and sharpens the edges of the exudates region in the OCT images [10].
2. With the help of Gray Level Co-occurrence Matrix (GLCM) feature extraction technique, the features which are beneficial for classifying the types of exudates [11].
3. The proposed Random Forest technique differentiates the exudates and non-exudates from the Optical Disk and the eye vessels. This technique gives better classification of retinal OCT images compared to other classification techniques applied on retinal OCT images.

The organization of the paper is as follows: Section I describes the introductory part of the proposed framework and the investigations of several researches papers are mentioned in Section II, section III clarifies the proposed framework, section IV states the obtained simulation results and their discussion and section V concludes the paper.

Proposed method

Problem statement

In recent times, Diabetic Retinopathy (DR), is considered as one of the major cause of blindness, would be help in slow down by early stage detection. To detect the symptoms like hard exudates region of the retinal images is one of the challenging tasks carried out by the ophthalmologists so far. There are several techniques proposed by the researchers to identify and classify the exudates and non-exudates region. But it does not provide better accuracy like classification technique used by the proposed system.

Overview

Hard exudates are yellowish intra-retinal stores, which are arranged in the back post of the fundus. The exudate is made out of serum lipoproteins that break from the unusually porous veins along the dividers of flawed microaneurysms. Hard exudates are frequently observed as either singular strips or groups of vast round rings encompassing the microaneurysms. They additionally show up close to the macula and are nearly connected with retinal thickening. Hard exudates are found in a few retinal vascular pathologies yet are an essential sign of diabetic macular edema. To be sure, diabetic macular edema is the central reason for visual weakness in diabetic patients. It requires to be analyzed at a beginning time. Laser treatment keeps loss of vision from macular edema [14].

The least complex and most productive approach to analyze macular edema is to identify hard exudates. In fact, the nearness of hard exudates inside 3000 μm of the focal point of the macula permits macular edema to be detected [15]. Exudates show up as bright patterns in retinal fundus images, and they have better complexity when contrasted with their surrounding background. Their shape and size differ essentially, and their borders are mainly irregular. Hard and soft exudates are separated based on their shading and the sharpness of their borders. Optic disc seems brighter than the background pixels in the pictures because of changes in light. The adjustment in lighting and the gray level changes because of vessels are the essential drivers for the high local contrast. In images with a couple of diminish hard exudates, one approach to enhance the last outcomes could be the utilization of alternative segmentation strategies that detect the dim hard exudates and eliminate with a high rate of noisy regions.

Figure 2 shows the fundamental image processing steps with the techniques used by the proposed CAD system. In image acquisition step, the images are acquired in DICOM format from the Diabetic Retinopathy Database (DIARETDB) dataset. It comprises of 200 images of different eye with low signal to noise ratio. These databases provide both normal and abnormal retinal images. The initial step is to apply preprocessing method to select the green channel from the fundus image and to find the boundaries of the eye from background and to improve the image to make it easier for further processing. In this step, the retinal images are preprocessed initially by histogram equalization to increase the contrast of images. Then the top hat filter is applied on the images to enhance the image in a dark background. The next step is to detect different segments of the eye such as OD, blood vessels, soft and hard exudates.

Since each of these objects are in different shapes and properties. The third step is to analyze the detected objects to extract features such as area and cumulative intensity from every input fundus image. The last step is by making use of the texture feature classification was performed to help medical doctors to better classify abnormality eyes from normal ones. The classification technique used by the proposed system is Random Forest which provides better accuracy of classifying the retinal images than any other classifiers.

Image acquisition

Image acquisition is the process of capturing the images or obtaining the images from the datasets. In the proposed system, the images are in DICOM format and it is obtained from Diabetic Retinopathy Database (DIARETDB) dataset. This dataset provides Optical Coherent Tomography (OCT) images of retina. This dataset comprises of diagnosis information of 200 images of different eye.

Image preprocessing

Image pre-processing is the process of enhancing the qualities of the images by changing the pixels' quality, contrast levels and grey values. The detection of abnormality in the fundus images with respect to the exudates classification, the first step involved is thepreprocessing. The main reason for preprocessing is to extract the boundaries of the eye and differentiate between eye and the background, to detect the OD and remove it from the fundus image, and to enhance the image. The distortion and noise of the input images are eradicated by the following preprocessing steps.

- **A. Top Hat Filter**

Filtering is the process of modifying and enhancing the images that eradicate the unnecessary distortions present in the images. It also improves some features of the images. One of the image enhancement filters is top hat filter that sharpens the edges of the image boundaries [10]. The image is altered by structuring element (SE) and is subtracted from the original image. With the help of this filter, the brightest spots of the images are highlighted.

- **B. Histogram Equalization**

The techniques used for enhancing the images are of two types namely frequency domain and spatial domain. It considerably improves the perception and interpretability of the image boundaries for human viewers. One of the frequency domain techniques is histogram equalization and it is used to improve the contrast of the images [12].

The cumulative density function (CDF) is applied in order to obtain transformation function effectively so that it produces the flat equalized histogram. The CDF of a digital image is obtained by averaging the probability distribution function (PDF) and is represented as.

$$S_0 = T(r_0) = \int_0^{r_0} P(r_0) dr_0; 0 \le r_0 \le 1 \tag{1}$$

Then differentiating the above equation

$$\frac{ds_0}{dr_0} = P(r_0) \tag{2}$$

Thus calculating the CDF should flattening the histogram equalization.

Histogram equalization is applied to filtered image where each pixel p is adapted using Eq. (3)

$$S(x,y) = \left[\frac{1}{h'^2} \sum_{q \in M_p} \partial (S_P - S_q) \right]^m \times 255 \tag{3}$$

where δ(x) is the delta function. The new enhancement method downgrades toStark's HE when m = 1.When m increases, the contrast between vessels and the background becomes larger, but at the same time, some noises on the background are also enhanced, and the boundaries of some vessels are distorted. Also, when the window size h' is small, the improved HE tends to use the local information to do the enhancement, which causes noisy results. On the other hand, if large value is chosen for h', the enhancement results become smoother, but detail information such as small vessels is smeared and the computational burden increases dramatically. As a result, based on the performance measure, the best enhancement performance is achieved when m = 8 and h = 81.

Morphological operation

In morphological operations, the objective is to change the structure or type of the objects utilizing a structuring component. These activities change the shape and size of the objects in the image. The two central morphological operations are dilation and erosion. Dilation enables objects to extend, along these lines possibly filling in little gaps and interfacing disjoint objects. Erosion shrinks objects by drawing without end (eroding) their limits. These operations are customized for an application by the correct determination of the structuring component, which decides precisely how the objects will be dilated or eroded. The dilation procedure is done by laying the structuring component on the image and sliding it over the image as like convolution [16]. The distinction is in the operation performed. It is best portrayed in an arrangement of steps as follows:

1. In the event that the root of the structuring component corresponds with a 'white' pixel in the image, there is no change; and the control is moved to the following pixel.
2. In the event that the root of the structuring component corresponds with a 'dark' in the image, all pixels secured by the structuring component are made as dark. The erosion procedure is like dilation, yet the pixels are swung to 'white', not 'dark'. As previously, the structuring component over the image is slided and after that the following steps are performed:

- If the origin of the structuring component harmonizes with a 'white' pixel in the image, there is no change; at that point control is moved to the following pixel.
- If the structuring component harmonizes with a 'dark' pixel in the image, and no less than one of the 'dark' pixels in the structuring component falls over a white pixel in the image, at that point the 'dark' pixel in the image are changed from 'dark' to a 'white'. These two central tasks, dilation and erosion, are joined into more complex

sequences. Opening comprises of a erosion followed by a dilation and used to take out all pixels in regions that are too little to contain the structuring component. Closing has dilation taken after by erosion and used to fill in openings and little holes. Closing and opening will produce different outcomes despite the fact that both comprise of erosion and dilation.

Image segmentation

Several methods depend on segmentation results in order to obtain description and recognition of the images. Medical image segmentation divides the original image into its constituent regions or objects. It is also known as image partition method. In medical imaging systems, segmentation plays significant role in extracting the features from the images and these features are useful for classifying the images into anatomical parts like tumors, cells, tissues, blood vessels and so on.

Hard exudates segmentation

DR classification is done by estimating the presence of hard exudates. The segmentation of hard exudates is done first. As brightness characteristics of OD and hard exudates are similar, the OD region should be masked [17, 18]. The segmentation of hard exudates is done with the following three steps:

1. OD and hard exudate detection
2. Segmentation of detected OD and hard exudates
3. Removal of OD region from the segmented image

Step 1:   *OD and Hard Exudate detection*

After the removal of blood vessels, the OD and hard exudates region are detected by rearranging each pixel in the sliding neighborhood of the image into a column of a temporary matrix.

Step 2:   *Segmentation of detected OD and hard exudates*

The image after OD and hard exudates detection is converted back to binary with a threshold value to segment out the region of OD and hard exudates.

Step 3:   *Optic Disc Removal* from the segmented image

Since detected hard exudates are required for DR classification, the unwanted OD is masked in the segmented image. The OD (Optic Disc) is the leaving point of retinal nerve fiber and entering point for retinal blood vessels. The shape of the OD is round and it is the brighter region in the retina. In this method optic disc removal is the important step and it is the significant source of false positives in hard exudates detection [13].

Preprocessed image is the initial stage and remain with that use the entropy based approach to guess the location of the optic disc center. The OD is situated in a high intensity region now the vessels have maximal directory entropy. It consist of multi scale ring shaped top hat filter here the OD center of radius is equal to the third of the region of interest. To reduce radius and position, top hat filter is used so that it eliminates the bright lesions in OD.

Algorithm for OD elimination:

- *Step 1: The segmented image is treated as input image*
- *Step 2: The spatial value of brightest point is used for extracting the cluster corresponding to bright color. This extracted cluster includes bright lesions along with OD.*
- *Step 3: As the size of lesions is small compared to OD, image opening is used to eliminate lesions with structuring element of B disk type with radius 5.*
- *Step 4: As removing the lesions leaves OD alone in the image, subtracting this image from the image obtained in Step 2 removes OD and leaves only bright lesions.*

Thresholding

Thresholding is considered as one of the most foremost segmentation instrument by exchange the original pixels with black value pixels. i.e. changing the gray into binary image. An edge is chosen empirically by investigating the whole image set to get just the exudates. While contrasting with the threshold, if the pixel's intensity is higher, at that point the pixel is set to white else, it is set to black. As the exudates and optic disc have comparative brightness values, the output image after use of the threshold will have both exudates and optic disc.

One of the binarization algorithms is Otsu's threshold method. This strategy finds the ideal threshold values where the expansion of foreground and background pixels spread must be least [21, 28–31].

Region based active contour

Region Based Active Contour is one of the well known segmentation procedures that successfully segment the images with inadequately characterized boundaries. This procedure is also known as Chan-Vese algorithm. The fundamental reason for existing is to recognize the hard exudates region on the retinal image with fitting points of interest of their correct shape. By applying region based active contour model the boundary of extracted exudates region will be accomplished [18].

Region Based Active Contour has numerous improvements over the Edge based segmentation techniques. Firstly, it doesn't rely upon the image gradient and utilize the image statistical data inside and outside the contour to control the development. Furthermore, the results of segmentation will not rely upon the initial contour area.

The Chan-Vese segmentation algorithm is proposed to segment objects without clearly distinct limits. This algorithm relies upon level sets which are iteratively advanced to decrease the energy that is controlled by weighted qualities equivalent to the summation of intensity contrasts from the average value inside the segmented region. This segmentation technique relies upon the boundary length of the segmented region. Chan and Vese stated an energy function $F(c_0, c_1, C)$ which is characterized as

$$F(c_0, c_1, C) = \mu.length(C) + v.area(inside(C))$$
$$+ \lambda_0 \int |u_0(x, y) - c_0|^2 dxdy$$
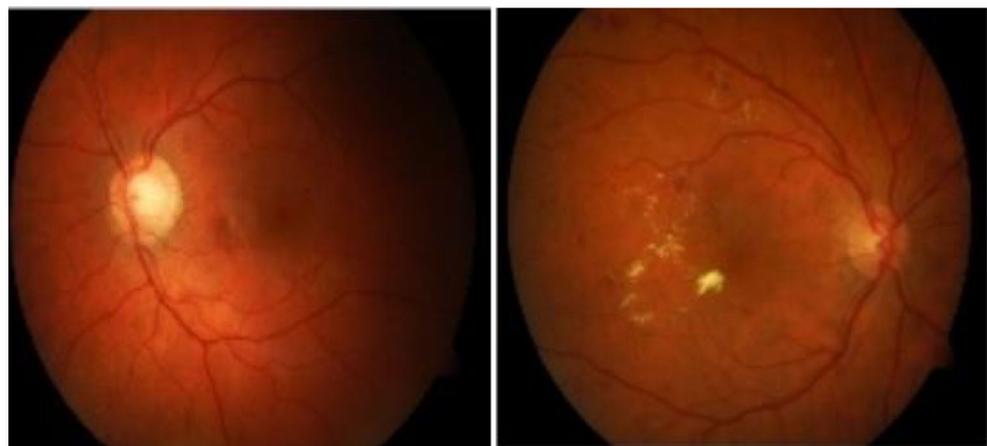$$+ \lambda_1 |u_0(x, y) - c_1|^2 dxdy \qquad (4)$$

Here C is the evolving curve,

c_0 and c_1	are the values of u inside and outside of C,
$\mu \geq 0, v \geq 0, \lambda_0 \lambda_1 > 0$	are the values used as constants,
u_0	is an input image.

By applying the level set method on the images, the minimization problem is resolved and is explained by the following expression

$$\underset{c_0, c_1, C}{inf} F(c_0, c_1, C) \qquad (5)$$

Implicitly the curve, C is defined by Lipschitz's zero level set function, $\phi(x, y, t)$ and is given as

$$C(t) = \{(x, y) | \phi(x, y) = 0\} \qquad (6)$$

where "t" is an artificial time.

The denomination of initial contour for $t = 0$, is $(x, y, 0) = \phi_1(x, y)$. The gradient flow Chan-Vese is defined as

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \left[\mu \, div \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - v - \lambda_0 (u_0 - c_0)^2 + \lambda_1 (u_0 - c_1)^2 \right]$$
$$= 0 \text{ in } \Omega$$

$$\qquad (7)$$

$$\frac{\delta_\epsilon(\phi)}{|\nabla \phi|} \frac{\partial \phi}{\overrightarrow{n}} = 0 \text{ in } \partial\Omega \qquad (8)$$

Feature extraction

A feature is a piece of information that describes a specific item in an image. The visual features of medical images are analyzed with the help of texture feature of images. For example, the texture of hemorrhage or the shape of exudates is considered as a feature of a DR image. Therefore, a larger number of extracted features will lead to better image classification. However, as the number of features increases, the computing time increases too. There are numerous statistical methods available for analyzing the gray level spatial distribution of images. One of the popular feature extraction techniques is Gray Level Co-occurrence Matrix (GLCM) which is applied on the images to extract the texture features. It evaluates the pair of pixels with the values obtained from the images. GLCM texture features like contrast, correlation, energy, homogeneity, mean, standard deviation,

Fig. 3 Color fundus input image **a** Normal Image **b** Abnormal Image

(a)   (b)

Fig. 4 Grayscale image **a** Normal Image **b** Abnormal Image

(a)

(b)
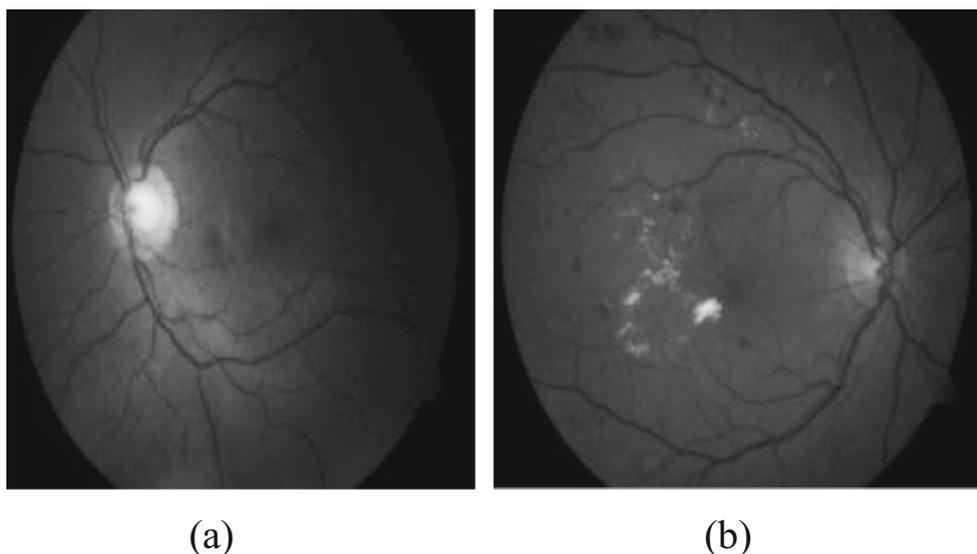
skewness, kurtosis and entropy are extracted from the desired region in the image.

Therefore, it is important to select only the most appropriate and useful features.In this research, initially 31 features have been extracted; however, only the best ten features were selected. The initial 31 features were extracted from Green, Red and Blue channels of the fundus images, as well as the gray scale image. By analyzing these features, the Green channel has more details compared to the rest of the channels. Therefore, to have a faster training and avoid under-fitting problem in the training process, only used the features extracted from the Green channel. In addition, only a selection of the 135 images in DIARETDB was used. For each image, the proposed algorithm computed and extracted the ten selected features using seven images. The features were selected to ensure that they bring meaningful information to the classification process. The total area of exudates (hard and soft) and hemorrhages, their cumulative intensities, and the percentage of their cumulative intensity with respect to the cumulative intensity of the whole fundus image are detected.

- **GLCM:**

The GLCM is a static geometric device for removing second order texture data from the MR images. A GLCM is a matrix here the number of rows and columns is equal to the number of individual gray [11].

- **Energy:**

Energy is calculated by

$$En = \sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f^2(x,y)} \tag{9}$$

Fig. 5 Histogram Equalization image **a** Normal Image **b** Abnormal Image
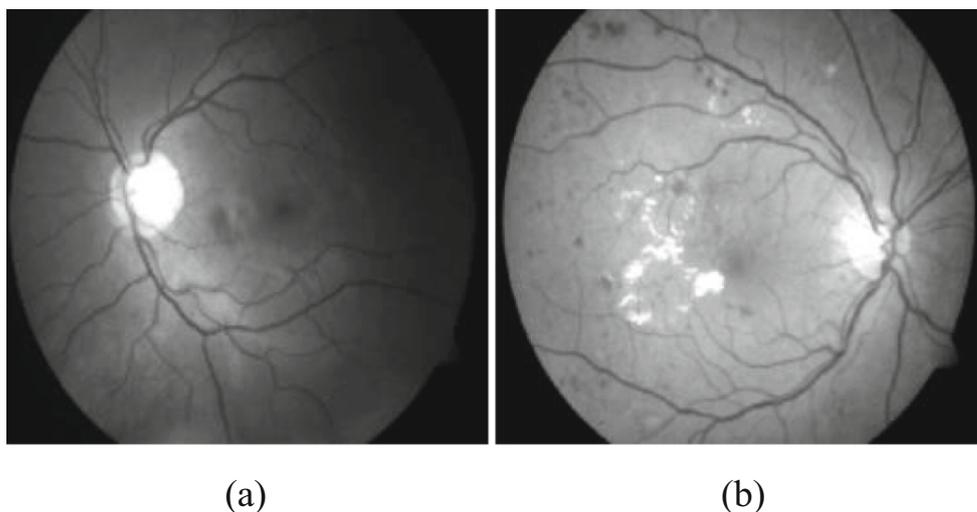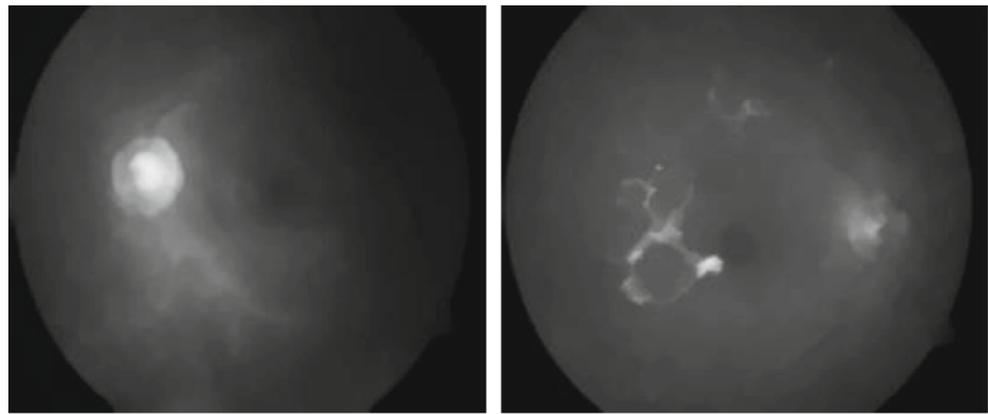

(a)

(b)

Fig. 6 Morphological Operation **a** Normal Image **b** Abnormal Image

(a)

(b)

- **Contrast:**

   Contrast is calculated by

$$C_{on} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (x-y)^2 f(x,y) \qquad (10)$$

- **Correlation:**

   Correlation feature is calculated by

$$C_{orr} = \frac{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (x,y) f(x,y) - M_x M_y}{\sigma_x \sigma_y} \qquad (11)$$

- **Homogeneity:**

   Homogeneity is calculated by

$$IDM = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \frac{1}{1 + (x-y)^2} f(x,y) \qquad (12)$$

- **Mean:**

   Mean is calculated by.

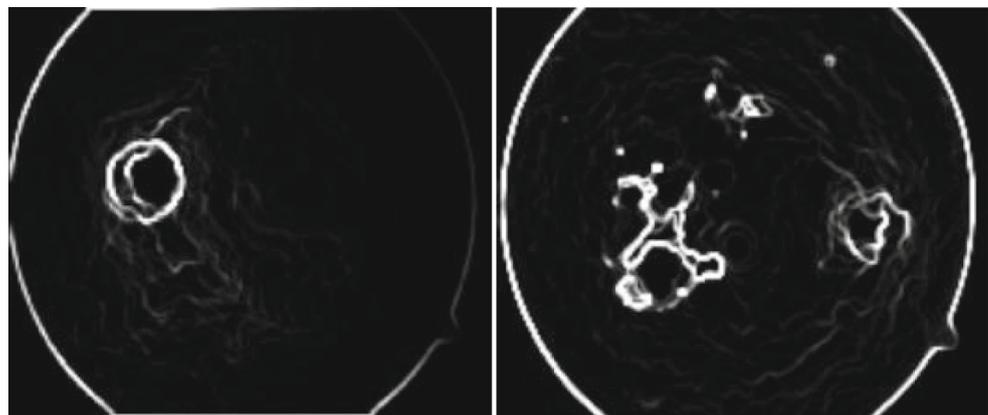$$\text{Mean } \mu_i = \frac{1}{N} \sum_{j-1}^{N} f_{ij} \qquad (13)$$

Fig. 7 OD and Hard exudates detection **a** Normal Image **b** Abnormal Image

(a)

(b)

Fig. 8 OD and Hard exudates segmentation **a** Normal Image **b** Abnormal Image

(a)   (b)
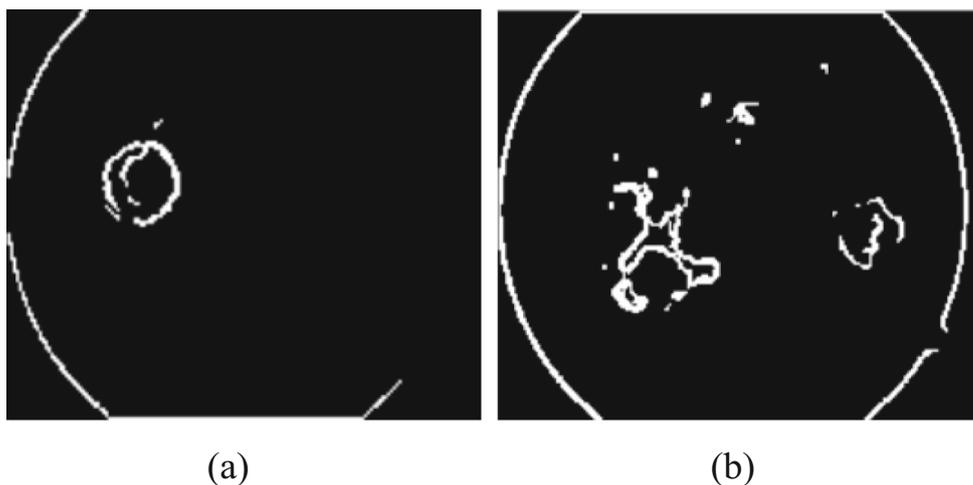
- ***Standard Deviation:***

  Standard deviation is calculated by.

$$SD = \sigma_i = \left(\frac{1}{N}\Sigma_{j-1}^{N}\left(f_{ij}-\mu_i\right)^2\right)^{\frac{1}{2}} \tag{14}$$

- ***Skewness:***

  Skewness is calculated by

$$S_k(X) = \left(\frac{1}{m \times n}\right)\frac{\Sigma\left(f(x,y)-M^3\right)}{SD^3} \tag{15}$$

- ***Kurtosis:***

  Kurtosis is calculated by

$$K_{urt}(X) = \left(\frac{1}{m \times n}\right)\frac{\Sigma\left(f(x,y)-M^4\right)}{SD^4} \tag{16}$$

- ***Entropy:***

  Entropy is calculated by

$$E = -\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f(x,y)\log_2 f(x,y) \tag{17}$$

  Where, N is the total number of pixels in the image.

Fig. 9 OD Elimination **a** Normal Image **b** Abnormal Image

(a)   (b)

Table 1   Accuracy of optic disk detection
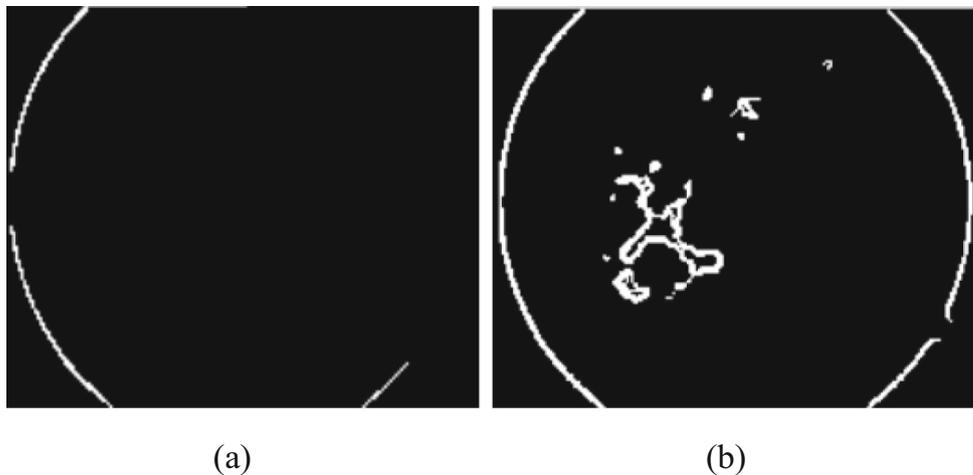
Threshold value	Number of images	Percentage (%)
3	120	92.78
4	118	91.0
6	111	86.34
8	104	80.35
12	89	68.9
15	87	66.8
17	83	64.98
20	79	60.10

Classification

Image classification is the process of grouping images into different classes. Imageclassification is an important process in the field of medical image processing. The outcome of the feature extraction process has a significant effect on the results of image classification. Moreover, image classification is based on the similarity of image items. To illustrate, the classification system consists of a set of predefined images of known features which are used to compare with an image of unknown class. However, the classifier requires training in order to be able to perform its function. Therefore, extracting the appropriate features is an essential step in image classification.

Random Forest

The Random Forests algorithm is one of the novel classification techniques able to classify large amounts of data with accuracy. It can use both for classification and the regression type of troubles. It is supervised classification approach which produces the forest

Table 3   Dataset used for Classification

Condition	Number of images	Training image	Testing image
Normal	100	70	25
Abnormal	100	60	35
Total	200	130	60

with a amount of tress [9]. In common, the number of trees in the forest the additional robust the forest looks likes. It could be also said that the greater the number of trees in the forest gives the better accuracy results. There are many advantages of random forest algorithms. The classifier can hold the missing values. It can also model the random forest classifier for categorical values [10]. The over fitting problem will not come in any classification problem. Most significantly it can be used for feature engineering which means identifying the most important feature out of the available feature from the training dataset.

Results and discussion

In this section, the obtained result for diabetic retinopathy classification is discussed. Random Forest applied on retinal OCT images to detect and classify the hard exudates images based on the severity. So far researches have not implemented Random Forest on retinal OCT images. Henceforth a novel classification technique, Random Forest is applied on retinal OCT images which provide 99.89% of accuracy in classify the images. This helps in enhancing the classification technique implemented on the CAD system to diagnose the diabetic retinopathy OCT images.

Table 2   Extracted features from retinal image

Image. No	contrast	correlation	energy	homogeneity	Mean	Standard Deviation	Skewness	Kurtosis	Entropy
sample 1	0.012420635	0.91048191	0.848983637	0.993789683	0.5	0.023087036	5.16E-15	1	1
sample5	0.021618185	0.878375428	0.801103954	0.989190908	0.5	0.162210296	0	1	1
sample13	0.04427676	0.845709397	0.670713766	0.97786162	0.5	0.370311821	0	1	1
sample21	0.038366435	0.038366435	0.721953112	0.980816782	0.5	0.453821132	0	1	1
sample11	0.03931199	0.885895682	0.617706692	0.980344005	0.5	0.159664711	0	1	1
sample 31	0.019876764	0.728026089	0.907434973	0.990061618	0.5	0.199404112	0	1	1
sample 46	0.041679154	0.857949326	0.666647532	0.979160423	0.5	0.011596551	0	1	1
sample 28	0.017188092	0.760345411	0.911387068	0.991405954	0.5	0.014778532	−7.98E-15	1	1
sample 43	0.027594949	0.882738166	0.737838906	0.986202525	0.5	0.02761252	0	1	1
sample 35	0.032301587	0.853233104	0.748654114	0.983849206	0.5	0.425218663	0	1	1
sample 54	0.022602634	0.926160853	0.671801857	0.988698683	0.5	0.012445079	9.48E-15	1	1
sample 57	0.023174603	0.790126475	0.866940703	0.988412698	0.5	0.130178358	−9.04E-16	1	1

Table 4 Performance of the Random Forest algorithm in detecting exudates

Image	TP	FP	FN	TN	Sensitivity%	Specificity%	Precision%	Accuracy%
1	321	0	49	431,630	86.76	100.00	100.00	99.99
3	3680	0	1012	427,308	78.43	100.00	100.00	99.77
5	542	37	42	431,379	92.81	99.99	93.61	99.98
6	51	0	36	431,913	58.62	100.00	100.00	99.99
8	270	318	30	431,382	90.00	99.93	45.92	99.92
12	4090	255	452	427,203	90.05	99.94	94.13	99.84
13	54	0	0	431,946	100.00	100.00	100.00	99.77
17	811	278	137	430,774	85.55	99.94	74.47	99.90
18	1166	250	49	430,545	95.97	99.94	82.34	99.93
22	207	95	73	431,625	73.93	99.98	68.54	99.96
24	785	52	110	431,053	87.71	99.99	98.00	99.87
26	188	223	148	431,441	55.95	99.95	45.74	99.91
27	1053	0	0	430,947	100.00	100.00	100.00	100.00
29	53	216	0	431,731	100.00	100.00	19.70	99.95
31	427	795	77	430,701	84.72	99.82	34.94	99.80
33	136	647	88	431,129	60.71	99.97	17.37	99.83
35	372	35	106	431,487	77.82	99.00	99.40	99.97
41	349	0	0	431,651	100.00	100.00	100.00	100.00

Implementation environment

The proposed CAD system is implemented with the help of MATLAB 2018 software. The MATLAB code is embedded with statistical and image processing tools. The Retinal OCT images of 200 patients obtained from DIARETDB database are analyzed and examined to obtain the classification results of the proposed system.

Simulation results

The following steps show the results obtained by giving the original retinal OCT image from the database. Preprocessing techniques are applied on the sample images to remove the noise and enhance the quality of images. This step is followed by classifying the exudates and non exudates region from the images. The resultant extracted features from the sample images are given in the table.

Table 5 Performance of Random Forest algorithm in detecting hard exudates

Image	TP	FP	FN	TN	Sensitivity%	Specificity%	Precision%	Accuracy%
1	321	0	49	431,630	86.76	100.00	100.00	99.99
4	3680	0	1012	427,308	78.43	100.00	100.00	99.77
5	542	37	42	431,379	92.81	100.00	100.00	99.98
6	51	0	36	431,913	58.62	100.00	100.00	99.99
11	270	318	30	431,382	90.00	100.00	100.00	99.92
12	4090	255	452	427,203	90.05	99.94	100.00	99.84
15	54	0	0	431,946	100.00	100.00	100.00	100.00
19	811	278	137	430,774	85.55	100.00	100.00	99.90
21	1166	250	49	430,545	95.97	100.00	100.00	99.93
26	207	95	73	431,625	73.93	99.98	95.90	99.96
24	785	52	110	431,053	87.71	100.00	100.00	99.87
26	188	223	148	431,441	55.95	100.00	100.00	99.91
27	1053	0	0	430,947	100.00	100.00	100.00	100.00
29	53	216	0	431,731	100.00	100.00	100.00	99.95
31	427	795	77	430,701	84.72	100.00	100.00	99.80
33	136	647	88	431,129	60.71	100.00	100.00	99.83
36	372	35	106	431,487	77.82	100.00	100.00	99.97
41	349	0	0	431,651	100.00	100.00	100.00	100.00

Table 6 Average performance result in detecting exudates and hard exudates

Measurement	Detecting exudates	Detecting hard exudates
Sensitivity	81.76	75.54
Specificity	100.00	100.00
Precision	81.10	99.81
Accuracy	99.85	99.89

Step 1: Image acquisition Take an image from the database: Images are obtained from DIARETDB database. The color fundus images are taken as input with a size of 576×720 pixels for exudates detection and severity classification. Figure 3 shows the sample input images used for analyze and are given into the CAD system.

Step 2: Preprocessing An image is given as input to pre-processing techniques to remove the noise and improve the quality. This involves RGB to gray scale conversion, image filtering, histogram equalization and morphological operation. Figure 4 illustrates the RGB to Gray scale conversion of (a) normal and (b) abnormal image. Here the abnormalities are more visible in gray scale image.

The gray scale image is further preprocessed to correct the problem of illumination variation that occurs in when pictures are taken. Other problems corrected by this process include the enhancement of the contrast between the exudates and vein network and the background to aid in segmentation and detection of the abnormalities. Therefore, Top Hat filtering and Histogram equalization is done. Top Hat Filter sharpens the edges of the retinal OCT images which consequently enhances the sharp edges of hard exudates region. Histogram equalization is applied on these preprocessed images to increase the pixel quality of the images. (Fig. 5).

To remove the retinal blood vessels, the morphological closing operation is performed over the enhanced image. Figure 6 shows the result of morphological closing operation.

Step 3: Segmentation After preprocessing, the preprocessed is given to the segmentation techniques to isolate the hard exudates region from the retinal OCT images.

(i) *OD and Hard Exudate detection*

The detected OD region and hard exudates region in the retinal image is shown in Fig. 7

(ii) *Segmentation of detected OD and hard exudates*

Figure 8 shows the segmented optic disc and hard exudates portion.

(iii) *Optic Disc Removal* from the segmented image

Figure 9 shows the removal of optic disc region from the segmented image. The optic disk detection was carried out on the images in the database changing the threshold value as illustrated in Table 1.

Step 4: Feature extraction It is significant to extract the features from the segmented images. The main functionality of this is to define the shapes of hard exudates region. Statistical parameters are calculated by GLCM feature extraction is given in the Table 2. Here the parameters such as contrast, correlation, energy, homogeneity, mean, standard deviation, skewness, kurtosis and entropy are mentioned respectively.

Step 5: Classification Random Forest classifier classifies the retinal OCT images and provides better accuracy. The achieved accuracy is 99.89%. The database used for retinal image classification is given in Table 3.

Performance evaluation

Following performance measures are analyzed for retinal OCT images. The main factors like Accuracy, Sensitivity and Sensitivity are considered and the obtained values are 99.89%.

Table 7 Comparison of classification obtained by Different approaches

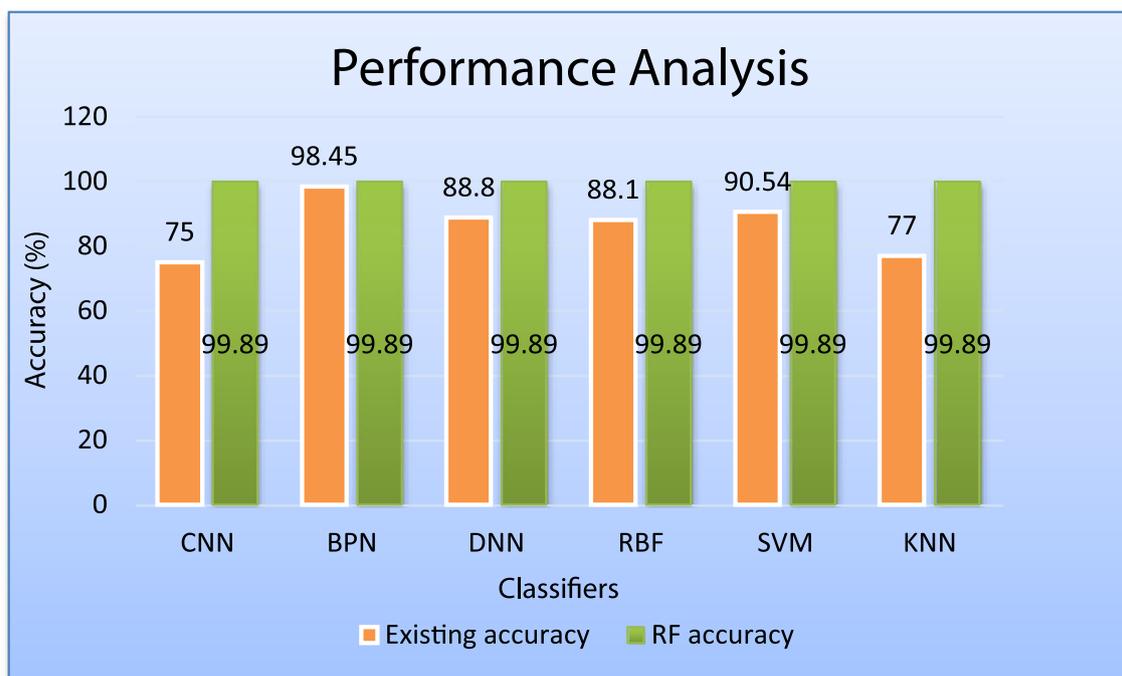| Approaches (applied on retinal OCT images)	Obtained Accuracy (%)
Convolutional Neural Network (CNN) [26]	75
Back Propagation Neural network (BPN) [24]	98.45
Deep Neural Network with Conventional feature textures (DNN) [25]	88. 8
Radial Basis Function (RBF) [27]	88.1
Support Vector Machine (SVM) [22]	90.54
KNN Classifier [23]	77
Random Forest	99.75

Fig. 10 Performance analysis

Accuracy Accuracy is reliability of the proposed method to separate the hard exudates regions from retinal OCT images. It is defined by

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ positive + True\ Negative + False\ Negative} \times 100\%$$

Sensitivity It is the percentage of the actual hard exudates region that has been truly detected by the proposed system.

$$Sensitivity = \frac{True\ Positive}{False\ Negative + True\ Positive} \times 100\%$$

Specificity It is the percentage of actual background of the hard exudates region that has been truly detected by the proposed system.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ positive} \times 100\%$$

Tables 4 and 5 shows the Random Forest algorithm resulted in a specificity of 100%. Table 6 illustrated to obtain an average sensitivity and specificity of 81.76% and 100%, respectively in detecting exudates. It is proved that the proposed methodology managed to detect hard exudates since Diabetic Retinopathy with an accuracy of 99.89% in all the images. Table 7 shows the accuracy obtained by different approaches and it is compared to the proposed technique. The proposed classification technique achieved the accuracy of 99.89% that is higher than the other existing classification techniques applied on retinal OCT images.

Figure 10 shows the graphical representation of comparison of different classification techniques applied on retinal OCT images. X-axis represents the various classification techniques compared to Random Forest and Y-axis represents the achieved accuracy obtained by the classification techniques.

Conclusion

This research proposes a novel technique to identify exudates using morphological methods and classify these exudates using the opinion of Random Forest algorithm. The strength of this approach is the ability to determine whether each exudates is hard exudates or not, separately. The pre-processing for Diabetic Retinopathy OCT images was done by top hat filter

gives better filtering effect and was enhanced by Histogram equalization. By seeing the results of segmentation technique it concludes that Region Based Active Contour technique gives the satisfactory results than other existing techniques. The feature extraction from retinal OCT images has been done by GLCM provides the better values. Finally, the classification technique is based on Random Forest implemented to check whether the images are exudates or non-exudates. The performance of Random Forest is compared to SVM classifier, CNN, DNN, KNN, RBF and BPN. The achieved accuracy of proposed system is 99.89% which is higher than existing classification techniques for retinal OCT images. Moreover this technique provides less consumption time to classify the images.

References

1. Carrera, E., Andres, G., and Carrera, R., Automated detection of Diabetic Retinopathy using SVM. In Electronics, Electrical Engineering and Computing, IEEE, pp. 1-4, 2017.
2. Marin, D., Gegundaz, A., Garrido, F., Vasello, J., Ponte, V., and Bravo, M., An exudates detection method for diagnosis risk of diabetic macular edema in retinal images using Feature based and Supervised classification. Medical and biological Engineering and Computing, Springer, pp.1-4, 2018.
3. Dheeba, J., and Singh, N., Detection of Hard exudates in color fundus images using fuzzy support vector machine based expert system. Journal of Digital imaging, Springer, pp.761-768, 2015.
4. Roy, A., Dutta, D., Bhattacharya, P., and Choudhury, S., Filter and fuzzy C means based feature extraction and classification of diabetic retinopathy using support vector machine. Interference Conference on Communication and Signal Processing, IEEE, pp.6-8, 2017.
5. Rajput, G. and Preethi, N., Detection and classification of exudates using K-means clustering in color retinal images. International Conference on Signals and Image Processing, IEEE, pp.126-130, 2014.
6. ElTanboly, A., Ghazal, M., Khalil, A., Shalaby, A., Mahmoud, A., El-Azab, M., Schaal, S., and Baz, E., An integrated framework for automatic clinical assessment of diabetic retinopathy grade using spectral domain OCT images. In Biomedical imaging, IEEE, pp.1431-1435, 2018.
7. Agurto, C., Yu, H., Wigdahl, J., Pattichis, M., Nemeth, S., Barrigo, S., and Soliz, P., A multiscale optimization approach to detect exudates in the macula. IEEE Journal of Biomedical Healty Informatics 18(4):1328–1336, 2014.
8. Akram, U., Khalid, S., Tariq, A., Khan, S., and Azam, F., Detection and classification of retinal lesions for grading of diabetic retinopathy. Computers in Biology and Medicine, Elsevier, pp.161-171, 2014.
9. Dutta, S., Manideep, B., Basha, S., Caytiles, D., and Iyenagar, N., Classification of diabetic retinopathy images by using deep learning models. International Journal of Grid and Distributed Computing, Springer 11(1):89–106, 2018.
10. Yavuz, Z., and Kose, C., Blood vessel segmentation from retinal images based on enhancement methods. Signal processing and Communications applications, IEEE, pp. 907-910, 2014.
11. Chand, R., and Dheeba, J., Automatic Detection of Exudates in color fundus retinopathy images. Indian J. Sci. Technol., 8(26):1–6, 2015.
12. Meshram, S. P., and Pawar, M. S., Extraction of retinal blood vessels from diabetic retinopathy imagery using contrast limited adaptive histogram equalization. Int J Adv Comput Theory Eng, 2(3): 143–147, 2013.
13. Akiyama, S. D., Kishi, S., Matsumoto, H., Observation of neovascularization of the disc associated with proliferative diabetic retinopathy using OCT angiography", Springer, pp. 1-6, 2018.
14. Walter, T., Klein, J., Massin, P., and Erginay, A., 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:1236–1243, 2002.
15. Giancardo, M., Karnowski, G. S., Tobin, K., and Chaum, E., Exudates based diabetic macular edema detection in fundus images. Med. Image Anal.:216–226, 2012.
16. Basha, S. S., Prasad, K. S., and Automatic detection of hard exudates in diabetic retinopathy using morphological segmentation and fuzzy logic, pp. 211–218, 2008.
17. Kussakunniran, W., Rithipravat, P., and Zhang J., Hard exudates segmentation based on learned initial seeds and iterative graph. Computer methods and programs in biomedicine, Elsevier, pp.173-183, 2018.
18. Bhargavi, R., and Rajesh, V., Exudate detection and feature extraction using active contour model and SIFT in color fundus images. J. Eng. Appl. Sci.:2362–2365, 2015.
19. Adem, K., Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks. Elsevier, pp. 289-295, 2018.
20. Amin, Y., Sarif, M., Ali, H., and Fernades, S. L., "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions", Journal of Computational science, Elsevier, pp. 153-164, (2017).
21. Kayal, D., and Banerjee, S., Detection of hard exudates using 2D Otsu algorithm in digital retinal fundus image. CSI Transactions on ICT, 5(1):53–57, 2017.
22. Tjandrasa, H., Putra, R., Wijaya, A., and Arieshanti, I., Classification of non-proliferative diabetic retinopathy based on hard exudates using soft margin SVM. IEEE Conference on control system, computing and Engineering, 2013.
23. Poonam, R., Manza, R., and Jonatham, P., Computer aided hard exudates detection on digital fundus images using morphology and multi-resolution analysis. International Journal of advanced computer technology. pp. 20-27.
24. Asha, G., Nasiha, A., Jayaram, M. A., and Manjunath, A., Exudates detection in retinal images using Back Propagation Neural network. Int. J. Comput. Appl. 25, 2011.
25. Al-Bander, B., Al-Nuaimy, W., Al-Taee, M. A., Williams, B. M., & Zheng, Y., Diabetic macular edema grading based on Deep neural network. Iowa Research Opthalmic Medical Image Analysis:121–128, 2016.
26. Pratt, H., Coenen, F., Broadbent, D., Harding, S., and Zheng, Y., Convolutional Neural Networks for diabetic retinopathy. International Conference on Medical Imaging:200–205, 2016.
27. Garcia, M., Sanchez, C. I, Poza, J., Lopez, M. I., and Hornero, R., Detection of hard exudates in retinal images using a Radial Basis Function Classifier. Ann. Biomed. Eng. 37(7):1448–1463, 2009.
28. Amel, F., Mohammed, M., & Abdelhafid B., Improvement of the Hard Exudates Detection Method Used For Computer-Aided Diagnosis of Diabetic Retinopathy. International Journal of Image, Graphics & Signal Processing, 4(4) 2012.

29. Joshi, S., & Karule, P. T., A review on exudates detection methods for diabetic retinopathy. Biomedicine & Pharmacotherapy, 97, 1454-1460, (2018).

30. Kumar, B. V., Janani, K., & Priya, N. M. , January). A survey on automatic detection of hard exudates in diabetic retinopathy. In 2017 InternationalConference on Inventive Systems and Control (ICISC) (pp. 1-11). IEEE., (2017)

31. Ravindraiah, R. and Reddy, S. C. M., Exudates detection in diabetic retinopathy images using possibilities C-means clustering algorithm with induced spatial constraint.In Artificial Intelligence and Evolutionary Computations in Engineering Systems (pp. 455-463). Springer, Singapore., 2018.