Automatic Screening and Classification using Machine Analysis Technique

Shrutika A. Patil
Department of Electronics and Telecommunications,
Shree L. R. Tiwari College of Engineering,
Mira Road (E), India.
email: shrutika.patil25@gmail.com

Mrs. Manjiri Gogate
Assistant Professor,
Department of Electronics and Telecommunications,
Shree L. R. Tiwari College of Engineering,
Mira Road (E), India.
email: manjiri_3980@rediffmail.com

Abstract - Diabetic Retinopathy (DR) is a microvascular complications caused by increase of insulin in blood, leading to blindness or vision loss because of changes in blood vessels of retina. DR is highly preventable with regular screening and timely intervention of lesions which can help ophthalmologists in detecting at an early stage. The background or non-proliferative DR contains four types of lesions, i.e. microaneurysms, hemorrhages, hard exudates and soft exudates. This paper presents a novel automatic approach for detecting DR in eye fundus images by employing image processing techniques. The proposed system consists of preprocessing, feature extraction, and classification stages. The classifier used are of two kinds categorized as low computational complexity such as k-nearest neighbor(kNN) and Gaussian Mixture Model(GMM) classifiers and as high computational complexity as Support Vector Machine(SVM). The proposed system uses genetic algorithm to evaluate and test publicly available retinal image database using performance parameters such as sensitivity, specificity and accuracy.

Keywords – Diabetic Retinopathy, Gaussian Mixture Model, Support Vector Machine, k-nearest neighbor

*****

I. INTRODUCTION

Diabetic Retinopathy is a microvascular complication caused by diabetes mellitus leading to damage in retina. It is a major cause of blindness in middle and advanced age group [1]. According to Diabetes Information Data, from total of 20.8 million people suffers from diabetes but only 14.6 million are diagnosed. DR is a progressive disease, i.e. the patient does not feel any change at early stage in his vision but when he feels it has already reached to its advance stage. Thus early detection of DR is necessary to prevent vision loss, which is done by regular screening and timely intervention of lesions that helps ophthalmologists in detection using eye fundus images.

DR contains four types of lesions namely microaneurysm (MA), Hemorrhages (HM), Hard Exudates (EX), and cotton wool spots(CWS) also called soft exudates. These are caused due to any changes in components of retina like blood vessels, fovea, macula and optic disc. Microaneurysms are small chambers like structure caused by local distension of capillary walls and are visible as small red dots [1]. Hemorrhages are huge blood clots that occurs into retina. MAs and HMs are termed as red lesions. While the bright lesions are EX and CWS that are deposited as bright yellowish patches of varying shapes and size. This is shown in Fig 1.

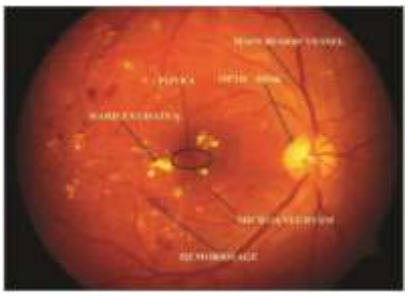

Fig.1. Retinal Abnormalities

A. Literature Survey

Several systems are designed for detection and diagnosis of DR that have been reported in literature. Ram *et al*.[4] presented a method for MA detection based on successive clutter rejection scheme. Their method first extracted large clutters with possible MAs and then based on some special features, it is rejected the false regions. In Alan d Fleming *et al*.[6] showed how image contrast normalization can improve the ability to differentiate MA and other dots on the retinal images. The watershed transform was applied to obtain better contrast normalization. In T. Walter *et al*.[14] candidate regions for exudates were extracted using morphological closing of luminance channel, local standard variation in sliding window, and watershed transform.

Priya R. *et al*.[9] applied fuzzy c-means clustering to segment the blood vessels in fundus images. In this, detection was done using Radial Basis Function Neural Network (RBFNN) method but didn't prove better in accuracy. Osareh *et al*.[13] used a detailed feature set

consisting of color, shape, size, and texture for exudate classification. They used a multilayered neural network classifier for this purpose.

Zhang *et al.* [7] has proposed a method that finds the correlation co-efficient between microaneurysms and Gaussian filter and thereby extracts the MAs. In this paper, an approach to the computer aided diagnosis (CAD) of DR is presented. C. Sinthaniyothin *et al*. in ordered to get exudates they considered the maximum variance to obtain the optic disk center and a region growing segmentation method. Gagnon *et al*.[11] tried to obtain disk localization from a template based on matching that utilizes the Hausdroff distance measure on the binary edge images by tracking the optic disk through a pyramidal decomposition. But these method failed because of presence of exudates with similar in brightness and size of the optic disk.

Sopharak *et al*.[12] have proposed a technique to detect a DR from fundus image automatically by using a set of optimally adjusted morphological operators and then by applying threshold operation. FCM- based method for optic disk and exudate detection was proposed by Yazid *et al*. [15]. their system used application of FCM clustering, edge detection, Otsu thresholding and inverse surface thresholding for this purpose.

In this paper we will propose a new system that will be able to detect all types of lesions with relatively different methods then those proposed earlier. In our system we will try to find the best possible regions using three classifiers and obtain accuracy that would compare all three classifiers. Further paper is organized in six sections. Section II describes in a brief about flow diagram of proposed system. Section III explains about the preprocessing stage with detailed description of all steps carried under preprocessing. Section IV lists about the features required for classification on lesion regions. Section V presents method for classification of DR and Non DR fundus images. Last section will give overview of expected results.

II. System Overview

System based on digital image processing and machine learning are playing vital role in biomedical research nowadays. Computer aided diagnostic systems have brought new horizons in detection and treatment of many common diseases [2]. The proposed system uses preprocessing, feature extraction, and classification as shown in Fig. 2.The first block in flow diagram is to acquire retinal fundus image from database available publicly.

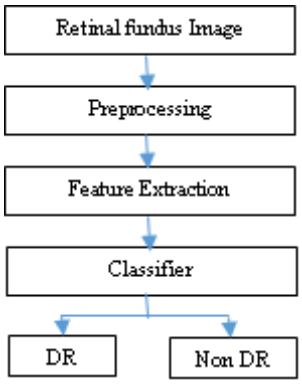

Fig.2. Flow Diagram of Proposed System

III. Preprocessing

Preprocessing is the initial step in image processing that includes steps such as background elimination, contrast enhancement, noise reduction, optic disk elimination. This steps are must because the available database are often noisy and poorly illuminated due to unknown noise and camera settings [3]. Thus accuracy and time consumption can be improved by cropping background and noisy regions from fundus images [2].

A. Background Elimination

In background elimination, the fundus image acquired from database is first scanned in column and rows pattern to obtain lesion region and eliminate the unwanted portion of image. Thus background elimination performs segmentation and edgedetection operations on image. Mean and variance based method is used for background elimination [2]. After the background is eliminated the image with region of interest are given to contrast enhancement step.

B. Contrast Enhancement

Contrast Enhancement is essential to distinguish DR regions from blood vessels, optic disk and other background of image. Bright lesions appear as bright spots while the red lesions appear as dark patterns in fundus images with highest contrast in green plane [14]. But in automated systems identification of these lesions is difficult due to different factors such as variability in image clarity, variation in image background texture, and confusion with optic disk pixels. An automated system should enhance the contrast of bright as well as red lesions with smoothing green channel image. Fig. 3 shows Green channel image and contrast enhanced bright regions.

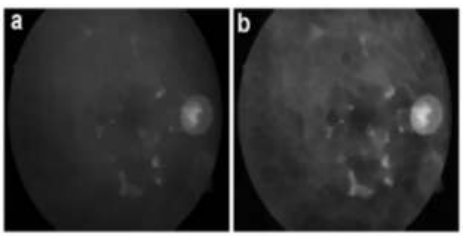

Fig.3. (a) Green channel image and (b) Contrast enhanced bright regions

C. Noise Reduction

Even after contrast enhancement there is possibility of presence of noise or error pixels in green channel image. Thus noise reduction is necessary which is done using median and Gaussian filtering. Filtering will suppress the error pixels alongside of lesion pixels. We will try to use 3*3 median filter to remove the poor illuminated pixels [3] such as red lesions. While bright regions can be detected using Gabor filter bank. The Gaussian kernel based Gabor filter can model wide range of shapes and sizes depending upon values of its parameters [16]. Gabor filter bank can be described using equation (1):

$$G_{FB} = \frac{1}{\sqrt{\pi r \sigma}} e^{-\frac{1}{2}[(\frac{d_1}{\sigma})^2 + (\frac{d_2}{\sigma})^2]} (d_1(cos\Omega + \iota sin\Omega)) \quad (1)$$

where σ = standard deviation of Gaussian, Ω = spatial frequency, and r = aspect ratio. θ is orientation of filter, d_1 = x cos θ + y sin θ and d_2 = -x sin θ + y cos θ. The contrast enhanced image $g$is convolved with Gabor filter G centered at location (s, t) to generate Gabor filter reponse γ for selected values of σ, Ω and θ, given in equation (2) [16]:

$$\gamma(\sigma, \Omega, \theta) = \sum_x \sum_y g(x,y) G_{FB}(s-x, t-y, \sigma, \Omega, \theta, r) \quad (2)$$

For considered frequency and scale values, computation of maximum Gabor filter bank response M_γ (σ, Ω) is done using equation (3) for spanning from 45° up to 180° at steps of 45°:

$$M_\gamma(\sigma, \Omega) = max|\gamma(\sigma, \Omega, \theta)| \quad (3)$$

The binary candidate regions are extracted from M_γ by applying low adaptive threshold value T[17]. The threshold value is calculated such that it should extract all possible candidate lesion regions. Fig. 4 shows output of Filter bank and adaptive thresholding technique.

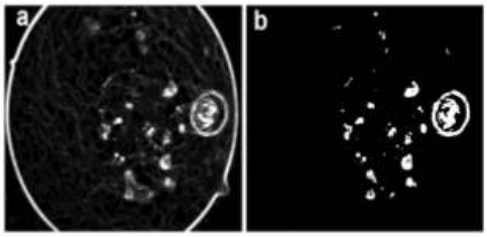

Fig.4. (a) Filtered Image and (b) Adaptive thresholding

D. Optic Disk Elimination

Optic disk appears as bright circular or elliptical region in fundus image and elimination of this region is essential in order to achieve robust DR detection. The regions segmented by thresholding of the filter-bank based enhanced image also contain an optic disk region and pixels due to their similarity with DR lesions. For accurate detection of lesions, these false and spurious pixels should be removed before classification stage [2]. In our system we will use morphological operation for elimination of Optic disk. Morphological operations are defined as closing or dilation and opening or erosion. Dilation is an operation that grows or thickens object while erosion operation shrinks or thins the objects in binary image [3]. Thickening and thinning process is controlled using a structuring element of defined shape and size. Here disk shape structuring element is considered. Fig. 5 segmentation and removal of optic disk region.

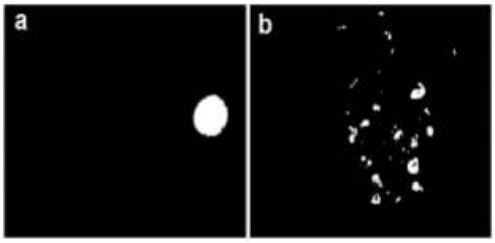

Fig.5. (a) segmented optic disk and (b) optic disk eliminated

IV. Feature Extraction

The preprocessing stage has extracted all possible regions that may be considered as DR lesions. Thus for classification of DR regions feature set needs to be formed that could represent each candidate region. Feature set includes following features described [2]:

*Area*is the count of number of pixels in candidate region and defined as $A = \Sigma_{vi}1$ sum of all the pixels in candidate region v_i.

Compactness is the measure of shape defined as $C = p^2/(4\pi A)$ where, p and A are perimeter length and area of candidate region, respectively.

Mean intensity is the mean intensity value of the contrast enhanced green channel for all pixels within the candidate region.

Mean hue, *Mean saturation* and *Mean value*for each candidate region is calculated in order to differentiate lesion region based on their color properties.

Mean gradient magnitude for edge pixels is computed to differentiate between strong and blurry edges.

Entropy is the value of all pixels in the square region including candidate region pixels and its neighboring pixels.

*Energy*is calculated by summing the intensity values of all pixels within the candidate region ad dividing it by total number of candidate region pixels.

V. Classification

The classifiers can be of two kinds categorized as low computational complexity classifiers and high computational complexity classifiers. Low computational complexity classifiers include Gaussian Mixture Model (GMM) and k-Nearest Neighbor (kNN) while high

computational complexity classifier is termed as Support Vector Machine (SVM). The spurious candidates can be removed using classifiers [2].

A. Gaussian Mixture Model (GMM)

In order to classify candidates as DR or Non DR region, a Bayesian Classifier using Gaussian functions is used called as Gaussian Mixture Model (GMM). Here two classes are defined as R1 = DR regions R2 = Non DR regions. This classification is carried out by simply dividing the data set into training and testing subsets randomly. The Bayes decision rule is used to obtain the decision rule based on estimates from training datasets as classifier is trained using training datasets which are labeled as lesions. The GMM characterizes as a classifier having halfway point between purely nonparametric and parametric models [2].

k – Nearest Neighbor (kNN)

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any technique we generally look at 3 important aspects:
1. Ease to interpret output
2. Calculation time
3. Predictive Power

KNN algorithm fairs across all parameters of considerations. It is commonly used for its ease of interpretation and low calculation time. In k- nearest neighbor classifier, the object is classified by a majority vote of its neighbors, with the object being assigned the most common class amongst its k-nearest neighbor.

Support Vector Machine (SVM)

SVM is a supervised learning process applied for analyzing the training data to find optimal way to classify DR images into respective classes: Mild, Normal, Severe.SVM models constructs a hyperplane for separating the given data linearly into separate classes as shown in fig 6(a).The classification parameters are formed according to calculated features using SVM algorithm.For nonlinear classification of given data, SVM uses nonlinear kernel function to map given data into high dimensional feature space where given data can be classified linearly shown in fig 6(b).Kernel function K(x,y) represents the inner product $<\phi(x), \phi(y)>$ in feature space [3]. Kernel function used is:

$$K(x, x') = \exp\left\{\frac{\|x-x'\|^2}{2\sigma^2}\right\} \quad (4)$$

where x and x' are training vectors; σ is parameter that controls width of Gaussian.

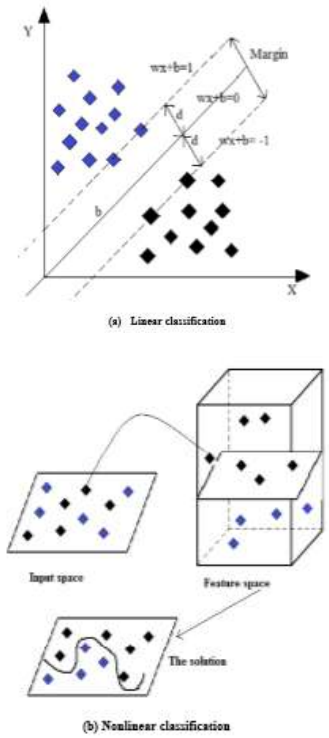

Fig.6. SVM Architecture

VI. Expected Results

The proposedsystem is expected to evaluate the performance parameters such as sensitivity, specificity and accuracy as figure of merit. Sensitivity is a true positive rate and specificity is a true negative rate. These parameters can be calculated using following equations:

$$\text{sensitivity} = \frac{T_P}{(T_P + F_N)}, \quad (5)$$

$$\text{specificity} = \frac{T_N}{(T_N + F_P)}, \quad (6)$$

$$\text{accuracy} = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)}, \quad (7)$$

REFERENCES

[1] ManojKumar S. B., Manjunath R. and H. S. Sheshadri, "Feature extraction from the fundus images for the diagnosis of Diabetic Retinopathy,"*2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, 2015, pp. 240-245.

[2] 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 Journal of Applied Optics*, 2012, Vol.51, No. 20, pp. 4858-4866.

[3] Aravind C., Ponnibala M., Vijayachitra S. "Automatic detection of microaneurysms and classification of diabetic retinopathy images using SVM technique," In: *IJCA Proceedings on international conference on innovations in intelligent instrumentation, optimization and electrical sciences (ICIIIOES)*,2013, Vol. 11, pp.18–22

[4] K. Ram, G. D. Joshi and J. Sivaswamy, "A Successive Clutter - Rejection-Based Approach for Early Detection of Diabetic Retinopathy,"*IEEE Transactions on Biomedical Engineering*, 2011, Vol. 58, No. 3, pp. 664-673.

[5] M. U. Akram, A. Tariq, S. A. Khan and S. A. Bazar, "Microaneurysm detection for early diagnosis of diabetic retinopathy,"*Electronics, Computer and Computation (ICECCO), 2013 International Conference on*, Ankara, 2013, pp. 21-24.

[6] Alan D. Fleming, Sam Philip, Keith A. Goatman, John A. Olson, Peter F. Sharp "Automated microaneurysm detection using local contrast normalization and local vessel detection,"*IEEE Trans. Med. Imag.*, 2006,Vol. 25,No.9, pp. 1223–1232.

[7] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recogn*., 2010, Vol. 43, No. 6, pp.2237– 2248.

[8] Prakash J., Sumanthi K."Detection and classification of microaneurysms for diabetic retinopathy,"*Int J Eng Res Appl*2013, Vol. 4, pp. 31–36.

[9] Priya R, Aruna P, Suriya R (2013) "Image analysis technique for detecting diabetic retinopathy,"*Int J Comput Appl*, Vol. 1, pp. 34–38.

[10] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher. "Automated detection of diabetic retinopathy on digital fundus images,"*Diabetic Medicine*, 2002,Vol. 19, pp. 105–112.

[11] L. Gagnon, M. Lalonde, M. Beaulieu, and M. C. Boucher, "Procedure to detect anatomical structures in optical fundus images," In *Proc.SPIE Medical Imaging: Image Processing*, 2001, pp. 1218–1225.

[12] Sopharak, A., Uyyanonvara, B., Barmanb, S., Willliamson, T. H., "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods,"*Comput. Med. Imaging Graph*, 2008, Vol. 32, No. 8, pp.720-727.

[13] A. Osareh, B. Shadgar, and R. Markham, "A computationalintelligence-based approach for detection of exudates in diabetic retinopathy images," *IEEE Trans. Inf. Technol. Biomed.*, 2009, Vol. 13, pp. 535–545.

[14] T. Walter, J. C. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathydetectionofexudatesincolorfundusimagesofthehuman retina," *IEEE Trans. Med. Imag.*,2002, Vol. 21, pp. 1236–1243.

[15] H. Yazid, H. Arof, and H. M. Isa, "Automated identification of exudates and optic disc based on inverse surface thresholding," *J. Med. Syst.*, 2012, Vol. 36, pp. 1997–2004.

[16] J. Sung, S. Y. Bang, and S. Choi, "A Bayesian network classifier and hierarchical Gabor featuresfor handwritten numeral recognition," *Pattern Recogn. Lett.* 27, 2006, pp. 66-75.

[17] R. C. Gonzalez and R. E. Woods, "Digital Image Processing,"*2*nd Edition, Prentice-Hall, 2002.