AN AUTOMATED DIABETIC RETINOPATHY CLASSIFICATION SYSTEM USING BAYESIAN LOGISTIC REGRESSION CLASSIFIER

Deepthi K Prasad
Full time research scholar, Department of CSE
B N M Institute of Technology
Bangalore, Karnataka, India

Vibha L
Professor, Department of CSE
B N M Institute of Technology
Bangalore, Karnataka, India

Venugopal K R
Principal, UVCE
Bangalore University
Bangalore, Karnataka, India

Abstract: Diabetic retinopathy (DR) is the damage to the retina due to elevated blood sugar levels that is the prime reason for loss of sight world-wide. The best solution to this problem is by controlling the blood sugar levels and also by regularly getting the eyes checked using an image capturing system to check for signs of any damage to the retina. This work proposes a method for identifying DR by the use of various image processing techniques. Textural attributes are extracted from the grey scale images and optimal features are selected. Bayesian Logistic Regression (BLR) classifier is used for classifying the images as DR-Present and DR-absent for the publicly available database. The results are compared with various performance metrics like accuracy, sensitivity and specificity. An accuracy of 98.4% is obtained in the proposed method.

Keywords: Bayesian Logistic Regression; Diabetic Retinopathy; Microaneurysms; Classification

INTRODUCTION

Diabetes mellitus is a malady that results when adequate amount of insulin is not produced by the pancreas. The progress of this disease affects the circulatory system and also damages the retina. Diabetic retinopathy (DR) is the most prevalent form of the diseases caused due to diabetes complication. It is caused by damage to the retinal blood vessels. According to an analysis of WHO [1], Diabetic retinopathy alone resulted in 1.9% of severe visual impairment across the world and 2.6% of blindness in 2010.The only solution to this problem is through the regular screening of the retina for damage and also by controlling the blood sugar levels. Figure 1 shows the normal image of the human retina.

.

Figure 1.Normal Retinal Image

DR causes damage to the blood vessels in the retina. Some of these blood vessels can leak fluids and other deposits, thereby causing blockage for the blood to pass through the vessels. Tiny reddish dots called Microaneurysms (MA) appear in the peripheral layers of the retina. They are the first vital signs which indicate the onset of DR. This stage is called Non Proliferative DR (NPDR). As it progresses, smaller blood vessels will be blocked and fragile capillaries begin to emerge. This stage is known as to as Proliferative DR (PDR).

Figure 2 depicts the sample retinal fundus image with DR. This stage may cause vision loss if it is not diagnosed and treated. Early detection of this disorder is the most critical need for avoiding blindness.
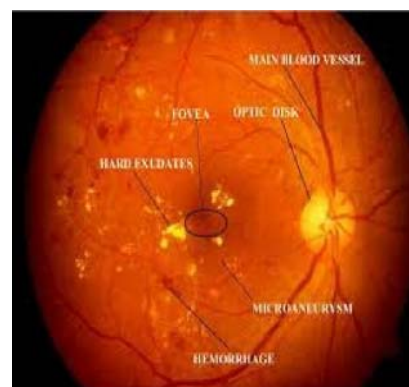

Figure 2. Sample Retinal Fundus Image with DR

Motivation: DR detection is complex task and has to be executed carefully. The automation of this system is highly beneficial especially in rural areas where there is scarcity of experts and expensive equipments.A computerised system that would initially scan for the presence or absence of DR can refer only those cases which need immediate attention of the

medical practitioner. It would be highly beneficial, economical and time-saving.

Contribution: The proposed work would classify a collection of input images in to DR-Present or DR-Absent using a Bayesian Logistic Regression Classifier with an accuracy of 98.4%.

The structure of this work is presented below.Section 2 describes the literature work in DR detection. The proposed system is presented in section 3. The next section focuses on the results and performance metrics.The last section concludes the work with prospective improvements.

LITERATURE SURVEY

The work done in the field of DR detection and classification is described below.

G. Gardner et al. [2] developed a neural network model for the identification of both hemorrhages and exudates from fundus images. This work resulted in a sensitivity of 88.4% and a specificity of 83.5%.

S. Abdelazeem [3] used a method which used recursive region growing to separate MA and blood vessels from input image. Neural networks were adopted to segment the blood vessels. MA are the residue that were left after the blood vessels were carefully segmented. The sensitivity of 77.5% and specificity of 88.7% was obtained using this methodology.

D.K Prasad et al. [4] propose the use of a retinal screening system to detect DR and classify them as normal or abnormal. One rule and back propagation neural network classifiers are used for publicly available dataset. The accuracy obtained is 98%.

James .A Hanely and Barbara J McNeil [5] developed a system to convert an area under a ROC curve obtained depending on the patients characteristic. The region would represent likelihood that an arbitrarily picked diseased patient is evaluated with more prioritized rank than a randomly selected patient without the disorder.

Grisan and Ruggeri [6] implemented a strategy to segment haemorrhages from the retinal images using the concept of local thresholding and the spatial thickness of the underlying pixels. The accuracy obtained was 94%.

Sergio and Daniel [7] developed a method for DR detection which used morphological operations. This model used back propagation neural networks and produced 88% sensitivity & 92% specificity.

Datta et. al [8] utilized contrast limited adaptive histogram equalization technique for the identification of changes that happen in retina when affected by DR. Various morphological functions are used for the segmentation of blood vessels and MA.

Somasundaram S.K & Alli P [9] proposed the use of "Machine Learning Bagging Ensemble Classifier (ML-BEC)" to detect DR. The method extracts attributes from the retinal images using a technique called t-distributed Stochastic Neighbor Embedding (t-SNE). They claimed that the classifier can achieve better classification accuracy when compared to individual classifiers.

Javeria et al. [10] developed an automated system for DR detection by detecting exudates from fundus images. Various statistical and geometric features are extracted from candidate

lesions. An accuracy of 98.5% is obtained for various publicly available databases.

PROPOSED SYSTEM

The architecture of the proposed system is shown in Figure 3. It has the following components
- Pre-Processing and Segmentation of the Input Image
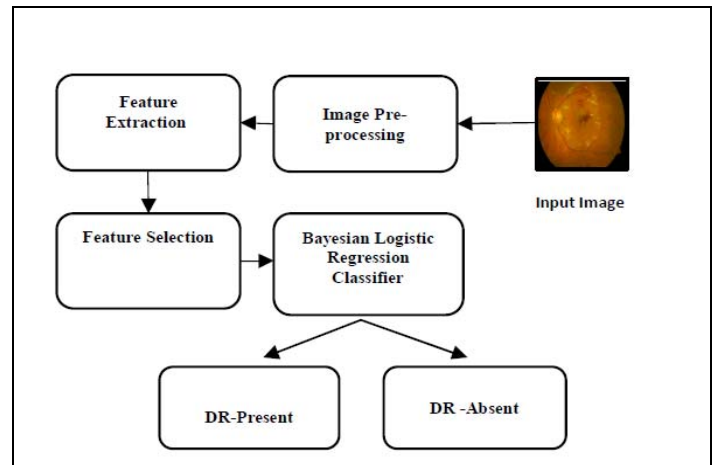- Extraction of Features
- Optimal Attribute Selection
- Classification

Figure 3. Architecture of the Proposed DR Detection System

A. *Image Pre-processing and Segmentation*

Image Pre-processing: The input data for the experiment is taken from a standard DR database DIARETDB0. The images are resized to 720 x 576 pixels but the aspect ratio is restored. All the three channels of the image are extracted and the extractions results are analyzed. The green index is chosen as MA exhibit increased contrast with the back ground. Contrast Limited adaptive histogram equalization technique is used [11] to highlight the MA and other lesions in a better way. For detecting the edges, Laplacian of Gaussian (LoG) [12] function is used to mark the edges as it highlights the regions where the intensity changes rapidly and as the image was already smoothened using image enhancing techniques.

Segmentation: The separation of blood vessels, MA and other retinal structures is a vital step in the DR detection. Morphological operations are used for segmenting the blood vessels and MA using the following steps mentioned in Table I.

Table I. Segmentation of Retinal Structures

Input : Edge detected Image.
Output: Images containing only MA, Blood vessels and Exudates.
Step 1: The circular border of the edge detected image is removed before filling up the enclosed area.
Step 2: The image containing MA and noise is got by subtracting away the edges image and removing the larger area.
Step 3: For obtaining the exudates, contrast limited adaptive histogram equalization is applied twice and threshold based segmentation is performed. This step will filter out only the exudates.
Step 4: The image containing exudates are compared with Step 2 outcome using AND logic to remove these exudates.

Step 5: To obtain blood vessels, the image is segmented with a given threshold. A clear view of blood vessels is obtained by using a median filter to remove the small area containing noise.

Step 6: The blood vessel image is compared using AND logic with the result from the Step 4 to remove the vessels. The resultant image containing only MA is obtained after removing the small noise and the optic disk area.

B. Feature extraction

The area of MA, blood vessels and exudates are calculated from the segmented image using the method given in [4].

The statistical features of the image like mean, third moment, entropy, standard deviation, homogeneity, gray level co-occurrence matrix (GLCM) and kurtosis are extracted [4].

C. Feature Selection

Significance test [13] is used for selecting the optimal features from the set of features extracted. Selecting the vital features among ten features which include three area estimation and various texture elements is done using significance test. Significance test calculates whether group of information are happened by chance, genuine occurrence or based on the level of genuine occurrence. The significance level is indicated by the term probability value (p-value). In the p-value is very less it shows that the set of information is statistically important compared to the one with higher value. Significance test selected only eight of these features and thus resulted in optimal feature selection.

D. Classification

The goal of classification is to separate the input data into distinct groups which do not overlap with each other.

Classification of the input images is performed using the Bayesian Logistic Regression (BLR) classifier [14] [20].

Logistic Regression is method to solve learning functions. It takes the form

$$f : X \rightarrow Y, \text{ or } P(Y|X) \qquad - (1)$$

Here Y is discrete valued and $X = hX1 \ldots Xni$ is a discrete vector which contains continuous or discrete variables. BLR assumes the distribution as a parametric form and directly estimates the parameters by considering the training set [19].

Problem Statement: For a given set of retinal images (1…M), where M value is more than 100, the objectives are :

(i) Accurate partitioning of images as DR-present or DR-Absent categories.
(ii) To improve the accuracy of classification and analyze the effectiveness of this approach.

RESULTS AND DISCUSSION

The implementation of the retinal structures segmentation and extraction and selection of features is carried out using MATLAB (MATrix LABoratory). The classification is performed using WEKA (Waikato Environment for Knowledge Analysis) software. The BLR implementation of

WEKA gives the user to select the required threshold value to to improve the classification accuracy.

E. Results of pre-processing and segmentation

The results obtained during pre-processing and segmentation process, which are explained in previous section are depicted in Figure 4.

F. Performance Analysis

Classification was performed using BLR mentioned in section III and the database used was DIARETDB0 [15] consisting of 130 images in which 110 have DR signs (DR-Present) and 20 are normal images (DR-Absent).
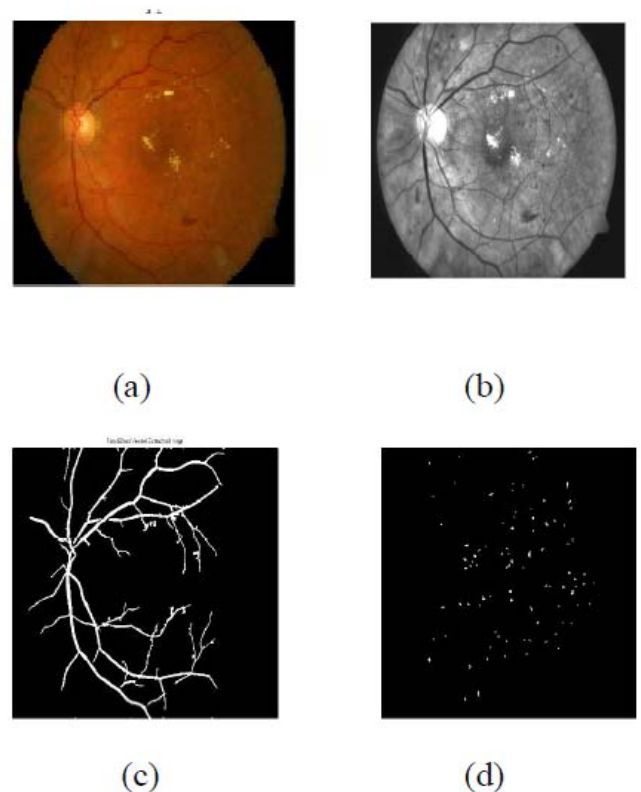

Figure 4. Segmentation Results (a) retinal image from DIARETDB0 (b) CLAHE image (c) blood vessel extracted (d) MA segmented

Confusion matrix:

Accuracy of classification specifies the amount of right predictions exhibited by the classifier to the total items in the test dataset. It can be computed using a confusion matrix. The general confusion matrix for a classification for binary classifier is given in Table II.

Table II. General Confusion Matrix for a Binary Class Classifier

Group		Assigned Classes	
		Yes	*No*
Outcome	*Yes*	**TP**	**FP**
	No	**FN**	**TN**

Here each item C_{pq} in the confusion matrix represents the number of items assigned by the classifier to a class C_p, but the actual class should be C_q

The meaning of the entries in the common confusion matrix is given below:

- TP (True Positive) - Input is DR-Present and detected as DR-Present
- TN (True Negative) – Input is DR-Absent and detected as DR- Absent
- FP (False Positive) – Input is DR-Absent and detected as DR- Present
- FN(False Negative) – Input is DR-Present and detected as DR-Absent

The input images are classified into two distinct classes namely DR-Present and DR-Absent. The confusion matrix obtained for the proposed work using BLR classifier is a 2X2 matrix as shown in   Table III.

Table III. Confusion matrix for BLR

Class	DR-Absent	DR-Present
DR-Absent	20	0
DR-Present	2	108

The average precision and recall values obtained were 0.98 and 0.9 respectively for the BLR classifier which indicates more stable results.

Accuracy: The accuracy of classification refers to the percentage of images correctly categorized as DR-Present or DR-Absent. The proposed method resulted in an accuracy of 98.4%.

Receiver Operating Characteristics (ROC): ROC curves assist in finding out the effectiveness of the classification. ROC curves are plotted for True Positive rate (Sensitivity) versus False Positive rate (Specificity) and are depicted in Figure 5. The maximum number of iterations chosen was hundred and the prior class chosen was Gaussian. Also the threshold value chosen was 0.4.

The proposed method outweighs the earlier methods [16], [17] & [18] in terms of increased accuracy for the DIARETDB0 database as depicted in Figure 6.
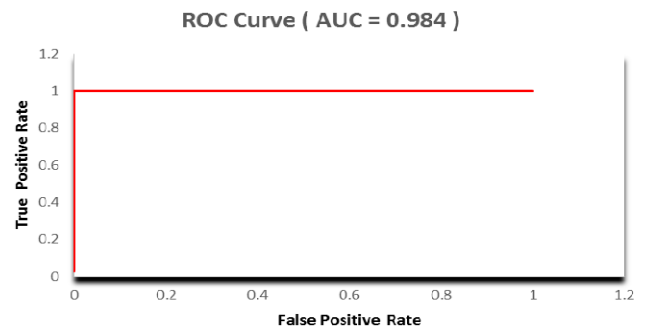

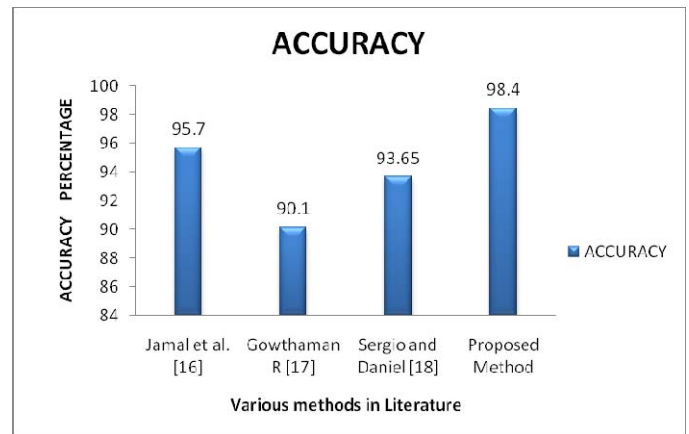Figure 5. ROC curve for the DIARETDB0 database using BLR

Figure 6. Performance Analysis of Various DR Detection Algorithms

CONCLUSION AND FUTURE WORK

Diabetic retinopathy is best detected at an earlier stage with the identification of microaneurysms. In the proposed work, the images from DIARETDB0 are pre-processed and segmented to separate microaneurysms using various morphological operations. Statistical features are extracted from the retinal images. The proposed work used Significance test to select optimal features. Classification is performed using the Bayesian Logistic Regression classifier.   The recorded results on DIARETDB0 database have proved to be better when compared to earlier methods in the literature. The accuracy of classification is 98.4% using Bayesian Logistic Regression classifier. As a future work, multistage classification of DR can be performed to detect the severity level of DR as mild, moderate or severe stages and with different image databases.

REFERENCES

[1] World Health Organization. Global report on diabetes. Geneva: World Health Organization; 2016

[2] G. Gardener, D. Keating, T. Williamson and A. Elliot, "Detection of diabetic retinopathy using neural network analysis of fundus images", British Journal of Ophthalmol., volume 80, number 11, pp. 937-948, 1996.

[3] S. Abdelazeem, "Microaneurysm detection using vessels removal and circular hough transform", In the Proceedings of the 19th National Radio Science Conference, pp. 421- 426, 2002.

[4] D. K. Prasad, L. Vibha and K. R. Venugopal, "Early detection of diabetic retinopathy from digital retinal fundus images," 2015 IEEE Recent Advances in Intelligent Computational Systems

(RAICS), Trivandrum, 2015,pp.240-245. doi: 10.1109/RAICS.2015.7488421.

[5] James A. Hanley and Barbara J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve", Diagnostic Radiology, volume 143, pp. 29-36, April 1982.

[6] E. Grisan and A. Ruggeri, "Segmentation of candidate dark lesions in fundus images",In the Proceedings of 29th IEEE EMBS, pp. 6735-6738, August 2007.

[7] Sergio Bortolin Junior and Daniel Welfer, "Automatic detection of microaneurysms and hemorrhages in color eye fundus images," International Journal of Computer Science and Information Technology, vol. 5, pp 21-37, October 2013.

[8] N. S. Datta, H. S. Dutta, M. De and S. Mondal, "An effective approach : image quality enhancement for microaneurysms detection of non-dilated retinal fundus image," Proceedings of International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013, vol. 10, pp. 731- 737, 2013.

[9] Somasundaram K, S. & P, Alli, " A machine learning ensemble classifier for early prediction of diabetic retinopathy",Journal of Medical Systems, December 2017

[10] Javeria Amin, Muhammad Sharif, Mussarat Yasmin, Hussam Ali, Steven Lawrence Fernandes, "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions",Article in Journal of Computational Science, March 2017.

[11] Zuiderveld, Karel. "Contrast limited adaptive histograph equalization." Graphic Gems IV. San Diego: Academic Press Professional, 1994. 474–485.

[12] M. Usman Akram, Anam Tariq, M. Almas Anjum, and M. Younus Javed, "Automated detection of exudates in colored retinal images for diagnosis of diabetic retinopathy," Appl. Opt. 51, 4858-4866 ,2012.

[13] Xu Q., Kamel M., Salama M.M.A.,"Significance test for feature subset selection on image recognition", In: Campilho A., Kamel M. (eds) Image analysis and recognition. ICIAR 2004. Lecture Notes in Computer Science, vol 3211. Springer, Berlin, Heidelberg, 2004

[14] J. G. Serra, P. Ruiz, R. Molina and A. K. Katsaggelos, "Bayesian logistic regression with sparse general representation prior for multispectral image classification," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 1893-1897.doi:10.1109/ICIP.2016.7532687

[15] Kauppi, T., Kalesnykiene, V., Kamarainen, J.-K., Lensu, L., Sorri, I., Uusitalo, H., Kälviäinen, H., Pietilä, J., "DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms", Technical report.

[16] I. Jamal, M. Akram, "Retinal image preprocessing: background and noise segmentation" , TELKOMNIKA Indonesian Journal on Electrical Engineering, Vol. 10, No. 3, pp. 537-544, 2012.

[17] Gowthaman R., "Automatic identification and classification of microaneurysms for detection of diabetic retinopathy," International Journal of Research in Engineering and Technology, vol. 3, pp. 464-473, February 2014

[18] Sergio Bortolin Junior and Daniel Welfer, "Automatic detection of microaneurysms and hemorrhages in color eye fundus images," International Journal of Computer Science and Information Technology, vol. 5, pp 21-37, October 2013

[19] Avali, Viji R., Gregory F. Cooper, and Vanathi Gopalakrishnan. "Application of bayesian logistic regression to mining biomedical data." AMIA Annual Symposium Proceedings 2014, pp. 266–273, 2014

[20] Alexander Genkin, David D. Lewis and David Madigan , "Large-scale bayesian logistic regression for text categorization", URL http://www.stat.rutgers.edu/~madigan/PAPERS/shortFat-v3a.pdf