An Approach for Image Data Mining using Image Processing Techniques

Amruta V. Anasane[1]
[1]*Department of Computer Science & Engineering, Prof. Ram Meghe College of Engineering & Technology, Amravati, Maharashtra, India*

Abstract- The automatic telemedicine system for computer-aided screening and grading of diabetic retinopathy depends on consistent detection of retinal lesions in fundus images. The automatic detection of both microaneurysms and hemorrhages in color fundus images is described and validated in this paper. Classifiers such, k-nearest neighbor (KNN), support vector machine (SVM), and AdaBoost are analyzed for classifying retinopathy lesions from non-lesions. SVM classifiers are found to be the classifiers for bright and red lesion classification. This lesion categorization problem deals with unbalanced data sets and SVM or combination classifiers derived from SVM using more classification error due to the data inequality. Diabetic retinopathy (DR) is a complication of diabetes that can lead to impairment of vision and even blindness. It is the most common cause of blindness in the operational age population. DR can be managed using available treatments, which are successful if diagnose early. Since DR is asymptomatic until late in the disease method regular eye fundus assessment is necessary to monitor any changes in the retina.
Keywords – Fundus images, SVM, Diabetic retinopathy (DR), micro aneurysms (MA) and hemorrhages (HE) etc.

I. INTRODUCTION

Diabetic retinopathy (DR) is a problem of diabetes that can lead to mutilation of vision and even blindness. It is the most general cause of blindness in the operational age population. One out of three diabetic person presents signs of DR and one out of ten suffers from its most severe and threatening forms. Since DR is asymptomatic until late in the disease process, regular eye fundus examination is necessary to check any changes in the retina. As the number of ophthalmologists available are less, there is an urgent need for automatic detection of DR to cover the large diabetic population while reducing the clinical burden on retinal specialists. This process can be carried out in two steps i.e. firstly detecting the DR cases and then analyzing and grading them. Indeed, the identification of the harshness level, through DR grade, allows more suitable and consistent referral to treatment centers. A computer screening and grading system relies on the automatic recognition of lesions. Fundus images with DR display red lesions, such as micro aneurysms (MA) and hemorrhages (HE), and light lesions, such as exudates and cotton wool spots. In this paper, we are focusing only on MA and HE which are initial symptoms of DR. MAs are the first signs of DR, HEs are also important for DR screening and useful for grading. In fact, retinal HEs are the result of MA. When it starts to leak into the retinal layers, a more severe level of DR is indicated. According to the most common DR strictness scale, their existence and number indicate either a moderate or a severe no proliferative DR. When blood is leaking into the nerve fiber layer, it is flame HE. Blood leaking deeper in the retinal layer is an indication of blot HE. It appears larger than a dot HE, and its borders are irregular, leading to various shapes. In fact, none of the patterns are similar nor any pattern matches to all the other possible sizes and shapes of HEs. A common line of attack adopted in the literature for combined MA and HE detection consists in identifying all dark-colored structures in the image, mainly through a thresholding, combined with adapted preprocessing, and then in removing the vessels from the resulting set of candidates. In this paper, we have proposed a method for the detection of both MAs and HEs that does not require preceding vessel segmentation. A supervised organization scheme is considered to discriminate between lesions and other structures like vessel segments and background noise. Candidate regions are recognized only after the pre-processing step. To sort out the required candidates, features are extracted and then are to be used. The condition of the human system is an important problem factor in a large number of medical conditions like atherosclerosis or diabetes. An organ that is particularly sensitive to vascular system pathologies is the eye. A breakdown of blood vessels in the retina has severe impact on the quality of vision. In modern, the most common cause of such anomalies is diabetes, which according to American Diabetes continues to increase. As a result, diabetic retinopathy affects over a sector of adults with diabetes, and is currently the most common cause of sightlessness in the Western world. In pattern recognition terms, detection of blood vessels is a segmentation task, where the objective is to separate the arrangement of interest from the background. From the viewpoint of machine learning, it is a problem of binary classification: allot every image pixel to the positive or negative decision class. The decision on pixel’s class can be made based on its neighborhood, defined as a centered on the pixel to be classified. The content of a patch is feed into a classifier, which is to return the decision on the middle pixel.

II. BACKGPROUND

The retinal angiography currently is a key characteristic for confirming the activities of lesions in the management of a wide range of retinal diseases, such as diabetic maculopathy and malarial retinopathy. The Y. Zhao et al. [1] has proposed automated retinal image analysis of images, especially for leak detection, is relatively unfamiliar. The contrast to the large number of studies on detecting various retinal lesions in color fundus photograph, relatively few methods have been proposed on automated detection or quantification of leakage. The proposed method is used to detect three types of leakage on images from eyes with MR. This method can add up the number of leakage sites and measures their sizes and has a reasonable performance over only 05 images of MR. However, it only uses the intensity information to generate the saliency map for the detection, which may undergo when some non-leakage areas also have high intensities. Rabbani et al. [2] has proposed a method to detect leakage in FA images of subjects with diabetic. The employed can active contour segmentation model to detect the boundaries of leaking areas. This method is designed to detect areas of leakage in a circular region centered at the fovea on 24 images. Martinez-Costa et al. [3] suggested that any pixels with statistically high increment in gray level along the FA sequence close to the fovea center could be segmented as leakage, and applied this criterion to detect the leakage in the macula due to retinal layer occlusion. However, this method requires manual detection of the fovea center. Phillips et al. [4] calculated the gradient of fluorescence intensity, and then threshold the gradient values only to determine leakage regions in DR images. However, the supervised methods are limited by their dependence on training datasets derived from manual annotation. The performance of the classifier will be inherently dependent on the quality of this annotation. Yuan et al. [5] proposed a saliency based sore detection method from the wireless capsule endoscopy (WCE) images. It uses a multi-level super pixel representation as the preprocessing step for saliency detection, and the saliency map is generated from different levels by integrating all obtained saliency maps according to the color and texture features. This method is capable of accurately representing the contours of the regions, and these regions are located through an image feature encoding and recognition method. The limitation of this method is that neither its efficiency nor its potential is well demonstrated, because the dataset used for validation is too small. Mahapatra and Sun et. al. [6] used the saliency and gradient information in a Markov random field for non-rigid registration of dynamic MR images. This approach attempts to address the problem that most no rigid registration algorithms fail to give satisfactory results in the presence of intensity changes. Although the saliency provides high quality contrast-enhanced images, the incline information can still be influenced by noise.

This paper introduces the methods for development of an automatic telemedicine system for computer-aided screening and grading of diabetic retinopathy depends on reliable detection of retinal lesions in fundus images. **Section I** Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on images. **Section VI** proposed method and outcome result possible. Finally, **section VIII** concludes the paper.

III. PREVIOUS WORK DONE

In research literature, an image is taken as input which is fundus in nature along with the binary mask of its region of interest (ROI). The ROI is the area mostly round, surrounded by some black background. It produces a probability color map for red lesion detection. To support different image resolutions spatial calibrations are applied. The input image is preprocessed via smoothing and normalization. The optic disc is automatically detected and discarded thus for red lesions detection. Based on their intensity and contrast, the candidate regions corresponding to potential lesions, are identified in the preprocessed image. DSF together with color features are extracted for each candidate. The candidates are classified according to their probability of being actual red lesions. For Spatial Calibration, Y. Zhao et. al. [1] proposed images cannot be resized. Rather, the diameter of ROI is taken as a size invariant. This proposition is reasonable since most of the images for DR screening are acquired with a field of view. The H. Rabbani et. al [2] proposed light Equalization method, in which lesions may be hardly visible in areas of poor contrast and/or low brightness. Moreover, in a telemedicine context, images are variable in terms of color and quality. Consequently, preprocessing steps are required to address these issues. To overcome the effect, the clarification equalization method resulting color image is subtracted from the original one to correct for potential variations. The average intensity of the original channel is added to keep the same color range as in the original image. The A. M. Mendonca et. al [7] has proposed optics disk removal method. The Optic Disk is a significant source of false positives in red lesion detection. Starting from the preprocessed image, we first use an entropy-based approach to estimate the location of the Optic Disk's center. Basically, high intensity is the location of the Optic Disk, where the vessels have maximal directional entropy. Optic Disk's radius and its refined position is estimated using an optimization. This consists in a ring shaped matched filter to the image in a sub-ROI centered on the first estimation of the Optic Disk center, of radius equal to a third of the ROI radius. The radius and position of the matched filter that minimizes the difficulty are selected as the OD's final radius and center position. The author L.

Breiman et. al [8] proposed the Random Forest classifier to distinguish between lesions and non-lesions. This worked a success and due to its numerous advantages was widely used in computer vision. It is works well for non-lesions classification with high-dimensional and noisy data. It is robust against outliers and over-fitting. A Random Forest is a combination of decision trees trained independently using bootstrap samples drawn with replacement from the training set. N. Beucher et. al [9] has proposed Dynamic Shape Features. There are several regions that corresponds to non-lesions, such as vessel segments and remaining noise in the retinal background and thus to distinguish between these, the dynamic shape features based on shape features was proposed. In a topographic representation, each candidate corresponds to a water source. Morphological flooding is applied to starting from the lowest water source and ending when the retinal background is reached. It is indeed hypothesizing that when the flooding reaches the retinal background intensity, the catchment and no longer contextually represent a red lesion.

IV. EXISTING METHODOLOGIES

The author has studied various different method to detect novel method for automatic detection of both micro aneurysms and hemorrhages in color fundus images is described and validated. The main contribution is a new set of outline features, called Dynamic Shape Features, that do not require precise segmentation of the regions to be secret. These features represent the evolution of the shape during image flooding and allow to distinguish between lesions and vessel segment. It proves to be vigorous with respect to variability in image resolution, quality and gaining system.

A. Optic Disc Localization Technique –

The proposed method takes as input a color fundus image together with the binary mask of its region of interest. The ROI is the circular area surrounded by a black background. The description of the lesions was provided, method was evaluated on a per lesion basis, meaning we analyzed its performance in detecting every single lesion. The number of lesions as well as their location and type are crucial to determine DR severity level, therefore per-lesion performance must be as good. This method was evaluated on a per-image basis thus providing the report of diagnosis for each image. The method's evaluation is done on performance in discriminating images with / without signs of DR. The ROC arc as a global detection gain. Three databases were used for which a per-image annotation was provided. It should be noted that because no manual lesion segmentations were provided with these databases, author could not train lesion classifier on these datasets. Instead, early built classifiers were used to identify the lesions in the new images. Although this might be seen as a disadvantage, which able to assess the performance of our lesion detection under sensible conditions and its robustness in the face of data changes.

B. Convolution Neural Network –

The convolution neural network is a composite of multiple elementary processing units, each feature several weighted inputs and one output, performing convolution of input signals with weights and transform the outcome with some form of nonlinearity.

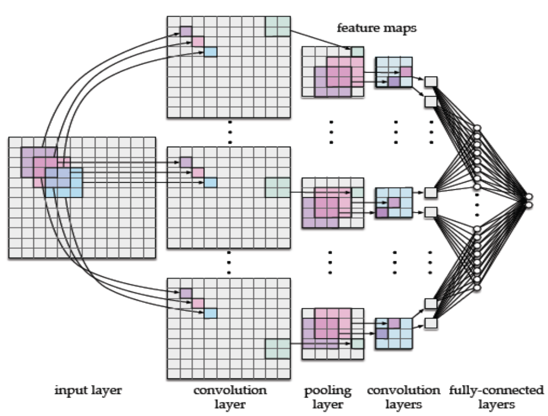

Figure 1. Convolution Neural Network Architecture

The units are arranged in rectangular layers, and their locations in a layer write to pixels in an input image as shown in Figure1. The arrangement of units is the primary characteristics that makes CNNs suitable for processing visual information. The other features are local connectivity, parameter sharing and pooling of hidden units.

V. ANALYSIS & DISCUSSION

Diabetic retinopathy is a severity disease which causes blindness among working age people. This research work presents a retinal vessel segmentation technique, which can be used in computer based retinal analysis. This proposed method could be used as a system for the early detection of diabetic retinopathy. The algorithm implemented in this work can be effectively used for detection and analysis of structures in retinal images. The retinal blood vessel morphology helps to classify the severity and identify the consecutive stages of a number of diseases. The changes in retinal vessel diameter are one of the symptoms for diseases based on pathology. Back propagation is a systematic method for training multi-layer artificial neural network. The back propagation provides an efficient technique to change the weights in a forward network, with differentiable activation function units, to learn a training set of input output examples. If the number of layers in the network is represented as n, and if Li represents the number of neurons in the ith layer,

$$N = \sum_{i=1}^{n-1} L_i L_{i+1} \quad \text{where } i = 1, 2, \ldots, n$$

The learning capacity is an important parameter which determine the ability of a system to learn. The data input given to the neural network are the three primary color components of the image, i.e., red, green and blue. The neural network applied is a three-layer neural network, which consists of one input layer, three hidden layers and one output layer. The different topologies of hidden layers with different numbers of neurons and found three hidden layers has been tested by authors; each containing neurons presents the finer results the output layer contains a neuron. The pixel samples are taken from vessel and non-vessel regions to train the neural network. The author has used the samples collected from the retinal images of original dataset which contains around 300 samples for training and testing the neural network.

The A. M. Mendonca et. al [2] has the proposed method obtained an AUC of 0.85 and a sensitivity of 93.9% at a specificity of 50%. The CARA1006 databases, the best performance was also achieved using R Fcara with an AUC of 0. 9666.The high-resolution Erlangen images, at a set point of 0.7, the method obtained a sensitivity and specificity both of 93.333%.

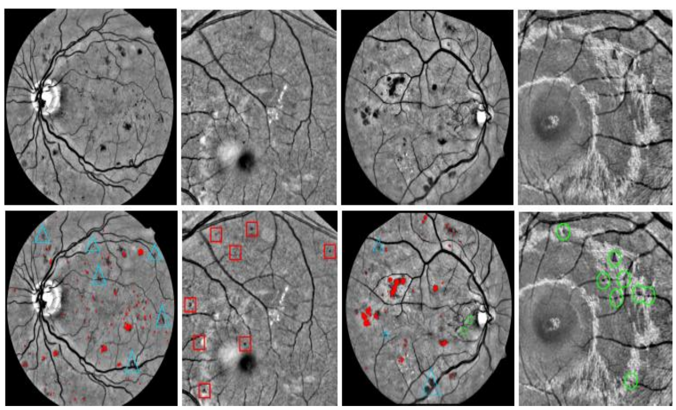

Figure 2. Lession Detection Example.

The only image with DR that was missed corresponds to a mild non-proliferative DR with only two Mas i.e. one is at the edge of the ROI and the second is directly connected to a large vessel. The two false positive images are glaucoma cases.

Table -2 Comparision between Algorithms

Optics Disk Localization	Convolution Neural Network	Proposed Algorithm
OD is located in a high intensity region where the vessels have maximal directional entropy	The convolution neural network (CNNs) is a composite of multiple elementary processing units, each featuring several weighted inputs and one output	It detects ME and HE where fundus image having maximum threshold at vessels and optic disk.
Multiscale ring-shaped matched filter to the image in a sub-ROI centered on the first estimation of the OD's center	The units are arranged in rectangular layers (grids), and their locations in a layer correspond to pixels	After the retinal image preprocessing , SVM Feature Extraction and Classification happens for retinopathy detection
It is convenient for non-linear classification with high-dimensional and noisy data	Suitable for processing visual information	It is more convenient for microanyermus and homorrages detection

In this paper a new diagnostic scheme has been presented which automatically detects and segments the blood vessels in retinal images, without any user intervention. Computerized segmentation of blood vessels from fundus images provides the means for automated examination and assessment by ophthalmologists. The number of vessels in a retinal image is more, or when large number of images is captured, manual detection of the characteristics of vessels becomes impossible, and computerized segmentation is the only possibility in this situation. This retinal vessel segmentation technique gives the knowledge about the location of vessels which paves a way for the screening of diabetic retinopathy. The main advantage of our method is the ability to identify and classify the image pixels as vessels or non-vessels, automatically.

Table -1 Experiment Result

	Original Lena Image (PSNR)	Watermarked Lena Image (PSNR)
BJUT Watermark Image	**33.1224**	**41.9946**
Bobbol Watermark Image	**33.1224**	**47.5911**
DDNT Watermark Image	**33.1224**	**45.8103**

Table 1 show the peak signal to noise ratio of performance of our proposed method of watermarked image and original image with various watermark image, where our watermarked images peak signal to noise ratio has a better performance than others.

VI .PROPOSED METHODOLOGY

The proposed algorithm could be used as a pre-screening system for the early detection of diabetic retinopathy. The algorithm implemented in this work can be effectively used for detection and analysis of vascular structures in retinal images. The retinal blood vessel helps to classify the harshness and identify the successive stages of a number of diseases. The changes in retinal vessel diameter are one of the symptoms for diseases based on pathology. Training dataset is used for fundus image data mining, the SVM classifier is used for fundus image features extraction. After feature extraction the images are classified into hemorrhages or microaneurysm identified images. The pre-processed input fundus image is also classified into HE and ME using data mining techniques.

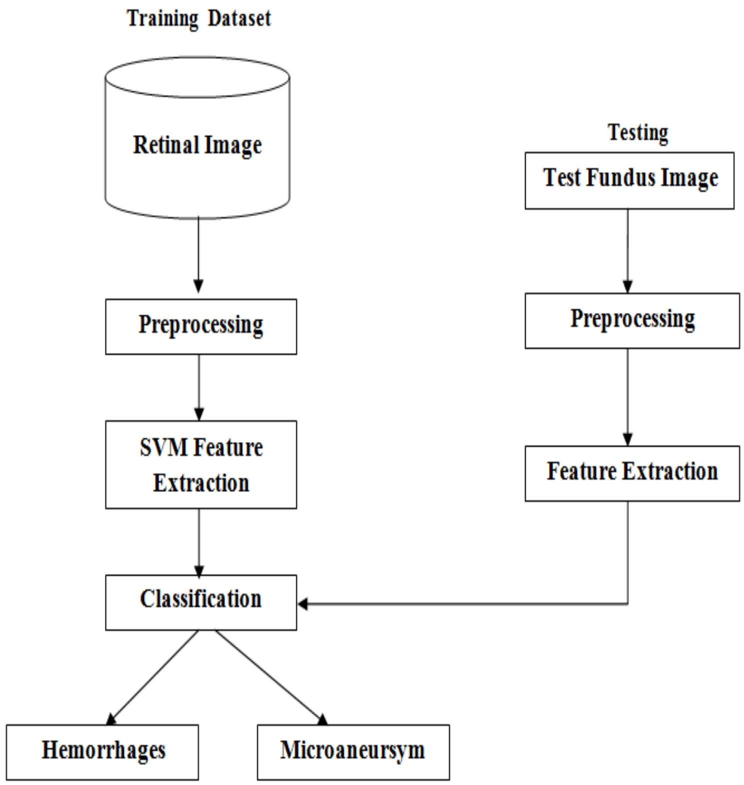

Figure 3. Diabetic Retinopathy Detection using Image Processing Technique.

The MA and HE detection consists in identifying all dark-colored structures in the image, mainly through a thresholding, combined with adapted preprocessing. MAs and HEs that does not require prior vessel segmentation. The author has considered a supervised classification scheme to discriminate between lesions and other structures like vessel segments and background noise through tracking.

Algorithm :- db_detection(image)

Step 1: Select the fundus image from dataset

Step2: Detect for Homorrages with mildDr/SevierDr

Step 3: If Image with homorrages detected calculate elapsed time.

Else

goto Step 6 for NoDr.

Step 4: Detect for Microanyermus with mildDr/SevierDr

Step 5: If Image with microanyermus detected calculate elapsed time.

Else

goto Step 6 for NoME.

Step 6: Exit

Algorithm- HE & ME Detection Algorithm.

The dB detection(image) algorithm gets fundus image as an input from FD Dataset. The algorithm classifies it into microaneuryms or hemorrhages detection with mildDr and SevierDr. If NoDr and NoME detected then exited and print the message. If hemorrhages detected then processed image is displayed onto the axes, and if microaneuryms detected along the vessels then the small dots with processed image is detected.

VII. RESULT & ANALYSIS

The blood vessels of the fundus image from an operator and the accuracy of the detection of blood vessels is found to be high. The algorithm for the automatic segmentation of blood vessels from the retinal images is tested with the images of Fundus Image dataset. The Elapsed time of existing methodology called Optic Disk Localization and proposed methodology is calculated along different fundus images as shown in following graph. This technique comes under the basic classification of pixel processing-based approach. An important feature of this method is its adaptability to image intensity properties, as most of the other algorithms are solely based on threshold values computed from the image information. The proposed method assigns an inveterate classification result as vessel or non-vessel to each pixel, while compared with other vessel segmentation algorithms.

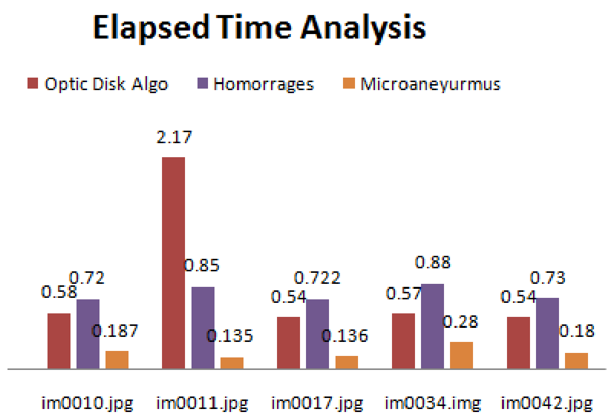

Figure 4. Elapsed Time Analysis of Algorithms.

The detection of some of the other features of diabetic retinopathy such as haemorrhages and exudates can be further evaluated by the characteristics of graph given below for optic disk algorithm and proposed algorithm required time.

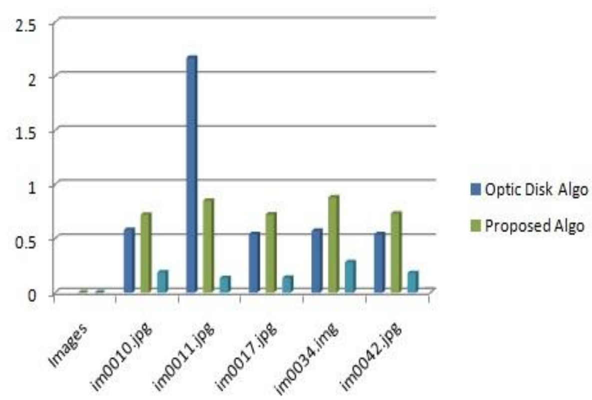

Figure 5. Comparision betwwen optic disc and proposed algorithm for time calculation.

VIII. CONCLUSION

This paper focused on the study of diagnostic scheme has been presented which automatically detects and segments the blood vessels in retinal images, without any user intervention. The Computerized segmentation of blood vessels from fundus images provides the means for automated examination and assessment by ophthalmologists. The number of vessels in a retinal image is more, or when large number of images is captured, manual detection of the characteristics of vessels becomes impossible, and computerized segmentation.

IX. FUTURE SCOPE

The demonstrated effectiveness together with its simplicity makes this computerized ME and HE detection method a suitable diagnostic tool for the complete pre-screening system for early diabetic retinopathy detection. In Future, more suitable algorithms are detected for future prediction of diabetic retinopathy precisely and correctly.

X. REFERENCES

[1] Y. Zhao, I. J. MacCormick, D. Parry, S. Leach, N. Beare, S. P. Harding, and Y. Zheng, “Automated detection of leakage in fluorescein angiography images with application to malarial retinopathy,” Sci. Rep., vol. 5, p. e10425, 2015.

[2] H. Rabbani, M. Allingham, M. P., S. Cousins, and S. Farsiu, “Fully automatic segmentation of fluorescein leakage in subjects with diabetic macular edema,” Invest. Ophthalmol. Vis. Sci., vol. 56, no. 3, pp. 1482– 1492, 2015.

[3] L. Martinez-Costa, P. Marco, G. Ayala, E. D. Ves, J. Domingo, and A. Simo, “Macular edema computeraided evaluation in ocular vein occlusions,” Comput. Biomed. Res., vol. 31, pp. 374–384, 1998.

[4] R. Phillips, T. Spencer, P. Ross, P. Sharp, and J. Forrester, “Quantification of diabetic maculopathy by digital imaging of the fundus,” Eye, vol. 5, pp. 130–137, 1991.

[5] Y. Yuan, J. Wang, B. Li, and M. Meng, “Saliency based ulcer detection for wireless capsule endoscopy diagnosis,” IEEE Trans. Med. Imag., Apr. 2015.

[6] D. Mahapatra and Y. Sun, “Mrf-based intensity invariant elastic registration of cardiac perfusion images using saliency information,” IEEE Trans. Biomed Eng., vol. 58, no. 4, pp. 991–1000, Apr. 2011.

[7] A. M. Mendonca, A. Sousa, L. Mendonc¸a, and A. Campilho, “Automatic localization of the optic disc by combining vascular and intensity information.” Computerized medical imaging and graphics, vol. 37, no. 5-6, pp. 409–17, 2013.

[8] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001.

[9] N. Beucher et al., “Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs.” IEEE transactions on medical imaging, vol. 29, no. 1, pp. 185–95, 2011.

[10] N. Cheung, P. Mitchell, and T. Y. Wong, “Diabetic retinopathy.” Lancet,vol. 376, no. 9735, pp. 124–36, 2010.A. A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme," Pattern Recognition Letters, vol. 26, pp. 1019-1027, 2005.