Texture Analysis for Glaucoma Classification

Suraya Mohammad *†, D.T.Morris*
* School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester,UK
†Universiti Kuala Lumpur, British Malaysian Institute, Gombak, Selangor, Malaysia

***Abstract*—In this paper, we present our ongoing work on glaucoma classification using fundus images. The approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF). This texture measurement is chosen because it can address the illumination issues of the retinal images and has a lower degree of computational complexity than most of the existing texture measurement methods currently used in the literature. Contrary to other approaches, the texture measures are extracted from the whole retina image without targeting any specific region. The method was tested on a set of 196 images composed of 110 healthy retina images and 86 glaucomatous images and achieved an area under curve (AUC) of 84%. A comparison performance with other texture measurements is also included, which shows our method to be superior.**

***Keywords*—BRIEF, Glaucoma, Texture**

I. Introduction

Glaucoma is the second leading cause of blindness worldwide. A large majority of those affected from glaucoma are from developing countries such as Asia and sub-Saharan Africa. Glaucoma damage is irreversible, so early detection can prevent severe vision loss. Unfortunately it has been shown that only half of the prevalent cases are identified. This may be due the fact that first, glaucoma is an asymptomatic disease that patients do not notice until they experienced vision loss and secondly, there is no single test to identify people with glaucoma [1]. These have severely hampered the establishment of screening based programs to detect the disease.

The screening program is where a grader will examine a retinal image and decide whether referral for further clinical examination by an ophthalmologist is necessary. With this screening program it is hoped that more people with glaucoma can be detected at an earlier stage so that timely treatment can be provided to maintain the patient's quality of life and to reduce the cost involved in surgical treatments such as laser treatment and drainage implant. With current techniques, screening for glaucoma is not cost effective. Improved methods of glaucoma classification could reduce the costs of screening programs and make them cost effective.

There are essentially three tests used in the diagnosis of glaucoma; tonometry, optic disc/nerve layer examination and visual field testing. Tonometry is a process of measuring intraocular pressure (IOP). IOP alone has limited effectiveness as a population based screening tool. Many studies have shown that there is no cut off IOP value that discriminates between normal eyes and those with glaucoma [1].

Visual field testing has been shown to have relatively high sensitivity and specificity when used as a screening test [2]. However, the test is time consuming and requires sophisticated equipment and trained, well-motivated operators to help and guide the subject. As it is a subjective examination, it assumes that patients understand the testing instructions, cooperate and complete the test. Many people have found a visual field testing a difficult task to perform. It was reported that more than 90% of over 40-year-olds were able to complete the screening test, but for older patients (over 70 years of age) and those with visual impairment the proportion drops to 71% [1].

Optic disc assessment can be the method of choice for glaucoma screening. It involves examination of the optic disc for signs of glaucoma either directly or through several imaging instrument such as fundus imaging and 3D imaging instruments. The 3D imaging instruments have the advantages over fundus imaging with regards to glaucoma assessment as they allow the acquisition and analysis of 3D measures of the optic disc.

With a 3D image, the optic cup depth, one of the main indicators for glaucoma can be measured and thus provide a more accurate assessment of the disease. However they are expensive instruments and thus not available in many primary care centres. For large scale screening a cheaper option such as the colour fundus image is preferred. In addition trained personnel are normally required to operate the equipment to ensure that the images captured are at the standard required for the glaucoma classification to work correctly. Using a fundus camera on the other hand, does not require a skilled operator and the time taken to acquire the images is also reasonably short.

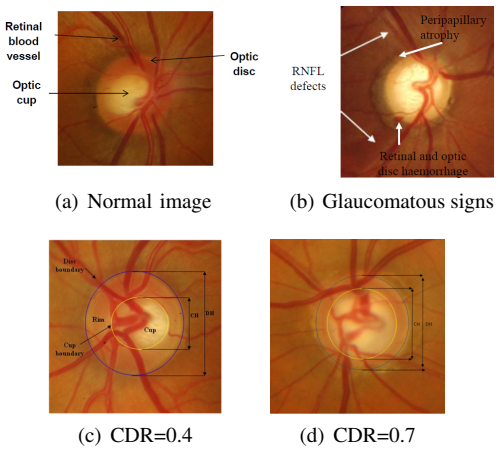
In this paper, we present the latest results of our ongoing work on glaucoma classification using fundus images. The approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF) [3]. The texture measures are extracted from whole retina image without targeting on any specific region.

The remaining part of the paper is as follows. The literature review is presented in Section II. Section III and Section IV, described the proposed approach and the result of the study respectively. Finally in Section V, we present our conclusion.

II. Literature Review

Previously published work on glaucoma detection follows two distinct paths; based on segmentation of retinal objects and based on classification using image features. The first approach requires segmentation of the main retinal structures and based on the segmented structures, some parameters are calculated and used to label the image as normal or glaucomatous.

Examples of retina images (normal and glaucomatous) are shown in Figure 1. The image in Figure 1(a), shows an example of normal retina images. The main features of the retina image are the optic disc, the optic cup and the retinal blood vessels.



(a) Normal image (b) Glaucomatous signs

(c) CDR=0.4 (d) CDR=0.7

Fig. 1. First ine: Anatomy of a retina image: (a) normal retinal image showing all the major features and (b) glaucomatous sign often found in the eye effected by glaucoma. Second line: CDR value for normal (c) and glaucomatous (d) image.

Contrast this with an appearance often found in glacomatous retina image as shown in Figure 1(b). The signs of glaucoma appeared here are enlargement of the optic cup (cupping), the presence of peripapillary atrophy, retinal and optic disc haemorrhages and the presence of Retinal Nerve Fiber Layer (RNFL) defects.

Measurements based on cup to disc ratio (CDR), blood vessel area and retinal nerve fibre layer (RNFL) defect detection [4][5] have been proposed for glaucoma detection. Out of those, CDR is the most common; it is defined as the ratio between the vertical diameter of the cup to the vertical diameter of the disc (Refer to Figure 1(b) and Figure 1(d). To measure it, segmentation of optic disc and optic cup are needed.

One of the advantages of using CDR for glaucoma assessment is that it is quite sensitive to glaucomatous changes in the optic disc since it encodes local cup deformation [6]. Enlargement of the cup size as often found in the glaucomatous retina will result in a bigger vertical cup diameter and hence a higher value of CDR. However some studies found out that CDR alone is inconsistent in explaining optic disc damage caused by glaucoma [7]. CDR, like IOP, is also distributed as a continuous variable with no sharp cut off value [1]. In addition, there are cases where the CDR measured is small but the patient has a significant visual field loss and vice versa. This is because the cup size is determined in part by disc diameter so that a large disc normally will have large cup and small disc will normally have a small cup.

Methods based on CDR also require an accurate segmentation of both the optic disc and optic cup. Any errors in segmentation may lead to significant change in the CDR measurements and thus the misdiagnosis of the disease. The reported accuracy of object segmentation for glaucoma assessment using fundus images is yet to achieve the required specificity and sensitivity required for large scale screening of glaucoma. One of the reasons is because a correct segmentation of the optic disc and optic cup itself is difficult to achieve. Variation in shape and size of the optic disc are known to hamper the performance of many automatic optic disc and cup segmentation system. Low contrast near the optic cup/disc boundary and the interleaving retinal blood vessels add to the difficulty in performing the segmentation.

To avoid the difficulties associated with optic disc and optic cup segmentation and relying solely on CDR measurement, an alternative approach is to extract image features from the retina image and use them for glaucoma classification. These features can be extracted over the entire image or on a specific region of interest namely the optic disc region. Several features have been used, for example work in [8] uses image intensities, Discrete Fourier Transform (DFT) coefficients and B-splines coefficients, work in [9][10][11][12] use textural features and a combination of textural and structural features is used in [6].

The advantage of this approach is it does not require segmentation as it performs a statistical data mining technique on image patterns themselves. With careful design, this approach is capable of achieving robustness against inter and intra image variations [6].

III. Method

This section presents the description of the retina image classification approach to identify normal and glaucomatous retinal image. Our approach uses the second technique described earlier that is based on classification using image features. It follows the standard machine learning pipeline, and basically it consists of two procedures, feature extraction and classification. This is illustrated in Figure 2.

A. Features extraction

To capture the image information for glaucoma classification, we use BRIEF features. The BRIEF features are extracted from the green channel of the retinal image and then represented as a histogram and used as image representation in a similar manner as local binary patterns (LBP). We choose the green channel because it provides better contrast compared to the blue and red channels.

Formerly BRIEF was used as a feature for image matching and recently it has been used as a texture measurement for optic disc segmentation [13]. This texture measurement is chosen because it can address the illumination issues of the retinal images and has a lower degree of computational complexity than most of the existing texture measurement methods currently used in the literature.

BRIEF texture measurement uses a binary string to encode the appearance of image patches. It form a descriptor by taking image patches and computing the result of a predefined binary test of n pixel pairs. The intensities of pairs of pixels are compared. If one pixel is greater than the other by a certain *Threshold* then the binary test is set to '1' otherwise it is set to '0'. The final output of the BRIEF operation is defined as binary string of length n.

The formal definition of BRIEF as a texture measure used in this paper is as follows:

A test τ defined on patch p of size SxS as:

$$\tau(p; \underline{x}, \underline{y}) = \begin{cases} 1 & \text{if (p(}\underline{x}\text{) - p(}\underline{y}\text{)) >Threshold} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

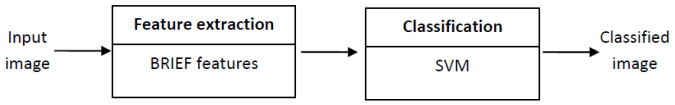


Fig. 2. Processing pipeline for the glaucoma classification.

where $p(\underline{x})$ and $p(\underline{y})$ are the pixel intensities at location $\underline{x}$ and $\underline{y}$. The BRIEF descriptor is then defined as the n bit vector:

$$f_n = \sum_{1 \le i \le n} 2^{i-1} \tau(p; \underline{x}_i, \underline{y}_i) \quad (2)$$

In the current implementation, the descriptor is calculated using 27x27 image patches and is 16 bits long, similar to value used in [13]. As for the threshold value, it is based on the estimated noise level of the image, which is provided by the noise standard deviation (σ). In the literature, various methods for the evaluation of σ, have been proposed. The two common approaches are either block based or filtering. We used a method which is based on filtering.

Briefly, it involved the following steps: first we blurred the image using a Gaussian filter to suppress image structure. Then the difference image is computed by subtracting the filtered image from the original. The noise level (i.e. noise standard deviation or σ) is estimated using the difference image which is assumed to contain the noise signal only.

We assumed that the noise affecting the image is Gaussian. For a Gaussian distribution of noise, around 68% of the sample distribution will have a value within 1σ from the mean, around 95% will have a value within 2σ from the mean and around 99% of have a value within 3σ from the mean. In other words, around 1% of the peaks which are more than 3σ brighter will be due to noise and the other 99 % are real values. Thus, in our work, we set the threshold value to 3σ, this will ensure that around 99% of the difference is real.

In handling retina images, pre-processing is normally performed before other processing can take place. During this stage, methods to correct uneven illumination, a problem in retinal images, and remove other image structures that may interfere with the classification (e.g. blood vessels) are employed. With respect to glaucomatous classification, in many automated glaucoma detection system, the blood vessels are often removed as their presence may lead to emphasizing the blood vessels and not the signs related to glaucoma [8]. In addition, each retina image is normally normalized to a standard size, to ensure robustness to optic disc size variations [6].

We, on the other hand, did not pre-process the image for illumination correction or for the blood vessel removal. This is because the BRIEF features are already invariant to image illumination, thus correcting the illumination is not needed. As for retaining the blood vessels, the reason is because a vascular change is also one of glaucoma's indicators. In glaucoma, the optic cup is enlarged, and the blood vessels within the optic disc are often displaced with changes to their tortuosity. This difference may be captured in the feature extraction step (i.e. BRIEF features calculation step), and may help in distinguishing between normal and glaucomatous cases.

Normalization with respect to optic disc was also not conducted. This is because our dataset contains optic discs of variable sizes. Although it is believed that normalization of optic disc size can lead to robustness against optic disc size variations, the process may, as an unwanted side effect, disturb the natural textures of the retina images, which can lead to bias during classification. Thus, in this investigation, normalization was not used.

B. Classification

In this step, a classification and the associated class label (glaucoma or normal) is computed based on the given features. A Support Vector Machine (SVM) is used as the classifier. SVM is a linear classifier and determines a maximum-margin and soft hyper plane that best separates the considered classes. We use SVM with the LIBSVM package [14].

C. Comparison performance

Other than evaluating the BRIEF's performance in classifying glaucomatous and normal retina images, we also perform comparison performance between BRIEF and other well-known texture measurements: Gray Level Co-Occurrence Matrix (GLCM) [15], Grey Level Difference (GLD) [16], Local Binary Pattern (LBP) [17], Rank transform [18] and Completed Modelling of LBP (CLBP) [11]. These particular texture measurements were chosen because, like BRIEF, they also characterise an image through the probability of occurrence of a pattern associated to a neighbourhood of a given size and in the case of CLBP, it was used for glaucoma classification in [12].

GLCM measures the joint probability of grey-levels of two pixels standing in a predefined relative position. In our implementation these parameters were used: window size is 3x3, displacement is 1 pixel and the directions were 0, 45, 90 and 135 degrees. A GLCM matrix is normally converted into Haralick features before being used for classification; however for this experiment we used the GLCM features directly, similar to the implementation used in [19]. GLD is quite similar to GLCM, but rather than measuring the joint probability of the two pixels, it measures their differences. In the implementation of GLD, similar parameters were used as for GLCM.

LBP forms a descriptor by thresholding the values in the periphery of the neighbourhood of the centre value. The output is the binary string which is then mapped to its decimal representation and collected into a histogram which is used as image description. In this experiment we used the original LBP which is calculated in 3x3 image patches. Rank transform measures the number of pixels in the periphery of 3x3 regions whose intensity is less than the intensity of the central pixel. The result will be a value ranging from 0 to 8, i.e. a rank transform will give 9 sets of possible patterns.

CLBP is an extension of LBP. In CLBP, a local region is represented by its centre pixels and a local difference. The centre pixels represent the image gray level and are converted into binary code by global thresholding (CLBP_C).

The local difference is then decomposed into sign (CLBP_S) and magnitude (CLBP_M) components. The sign component is equivalent to normal LBP. The CLBP feature is the combination of the CLBM_C, CLBP_M and CLBP_S. In the implementation of CLBP, similar parameters were used as for LBP.

IV. Result and Discussion

We tested the approach using 196 retina images, 86 are glaucomatous and 110 are normal images. The images were provided by the Manchester Royal Eye Hospital UK. During the classification experiment, we employed 10 fold cross validation, where one tenth of the images are used for testing and the remaining for training in each fold. This is to ensure that the result of the glaucoma assessment is not biased due to the training set used.

A. Classification result

The receiver operating characteristic (ROC) curve and balance accuracy (Bac) were used as the performance measures. The ROC shows classifier performance for different decision thresholds. It can provide information on how to tune the decision threshold in order to achieve the best trade-off between sensitivity and specificity. As for balance accuracy, we used it because our dataset has imbalanced normal and glaucomatous classes. Balance accuracy is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on either class. In the case of balance classes, this performance parameter corresponds to the classification accuracy. The result is shown in Table I. Based on the table, we obtained a balance accuracy of 78% and an Area under curve (AUC) of 84%.

Samples of images which have been classified correctly and incorrectly are shown in Figure 3. Figures 3(a)-3(e) show correctly classified normal images (true negatives). The correctly classified images mostly consist of images with good contrast and are healthy looking. This corresponds with typical appearances usually found in normal images. In contrast, Figures 3(f)-3(j) show images which are correctly classified as glaucoma (true positives). The images shown are characterised by bigger cups and in some cases the presence of atrophy (the pale regions outside the disk) and these characteristics again correspond to common glaucoma patterns.

The difficulties in performing automatic glaucoma classification can be appreciated by looking at the misclassified images, shown in Figure 3(k)-3(t). The images shown in Figure 3(k)-3(o) are misclassified normal images (false negatives). It can be seen that the false negatives are images which show similar characteristics to the glaucomatous retina such as a bigger cup area. One example is the normal image in Figure 3(k) which visually quite similar to the glaucomatous image in Figure 3(s). Similar observation can be made for misclassified glaucoma images (false positives), most of the false positives are images which have characteristics that resemble normal images.

B. Comparison performance result

An ROC of BRIEF features in comparison with GLCM, GLD, LBP, CLBP and Rank transform is shown in Figure 4. As shown in the figure, BRIEF features achieved the highest AUC with 84%, compared to the other textural features evaluated. LBP achieved the second highest AUC with 77% and CLBP in the third place with 76%.

TABLE I.
Classification results. Bac = Balance accuracy, AUC = Area under curve

Bac(%)	AUC (%)
78	84

Although the combination of CLBP is found to be effective in classifying glaucoma in [12], the same performance is not replicated with our dataset. Two reasons may contribute to its lower performance. First, in the original paper the dataset used to test the approach consisted of 41 retina images, 13 are glaucomatous images and 28 are non glaucomatous images. Given the small size of the dataset, it is understandable if the result is better than the result achieved here. The second reason is in their work, the image resolution is reduced to 225 x 225, however no justification is given for why this is necessary. It can be argued that the scaling process can provide bias to the final classification.

Two of the published works in glaucoma classification utilising image features and tested with a larger dataset are [8], [6]. Work in [8] proposed a two stage glaucoma classification system using features based on intensities, DFT coefficients and B-spline coefficients. The features were further processed using PCA to reduce the feature dimensions. Pre-processing included non-uniform illumination correction, vessel removal, optic disc region extraction and optic disc region normalization. It was tested with 239 glaucomatous and 336 normal images and achieved an AUC of 88%.

The other work [6] used a hybrid approach to glaucoma classification. The approach combined image features derived from the optic disc region, based on: CDR, cup to disc area ratio (CDA), RNFL defect detection probabilities and atrophy defect detection probabilities. The approach was evaluated using 1962 images, 808 images were used for training and 1154 for testing and used five fold cross validation. The AUC obtained were 78%, 64% using only CDR, 61% using only CDA, 68% using only features derived from optic disc regions and 73% using a combination of CDR, CDA and defect detection probabilities from RFNL and atrophy.

For the purpose of this comparison, we will only consider the result obtained using only image features. Altogether, [6] extracted 84 image features (both structural and statistical) from the optic disc for the classification. Pre-processing included feature normalization with respect to optic disc size and optic disc region extraction. The AUC achieved using only image features was 68%.

Based on the obtained AUC, the results in [8] are better than ours. However, their approach involved down sampling of larger optic disc sizes during the normalization step. As stated by the author himself, this step can bias the pure glaucomatous variation and may affect the classification. To account for this effect, they only evaluated their approach using a dataset with limited optic disc size variation. This restriction was not

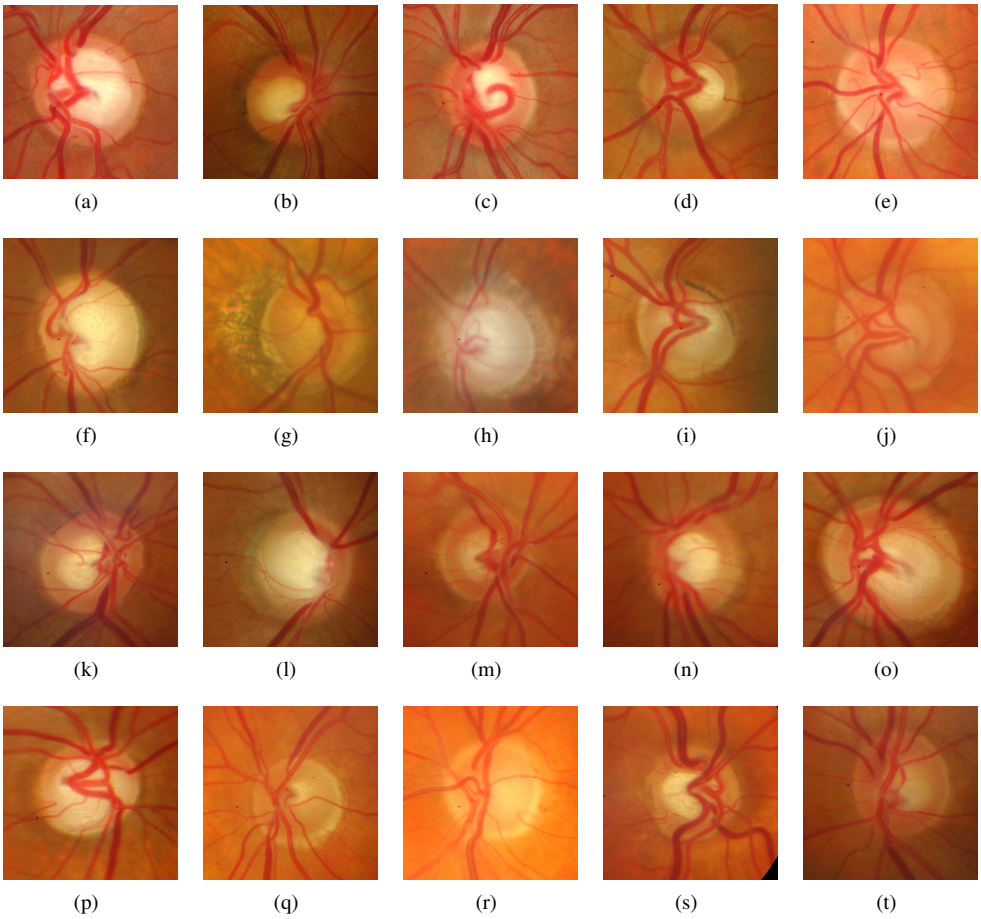


Fig. 3. Samples of retina images who has been correctly and incorrectly classified using kNN classifier. First row: True negatives, Second row: True positives, Third row: False Negatives and Fourth row: False positives

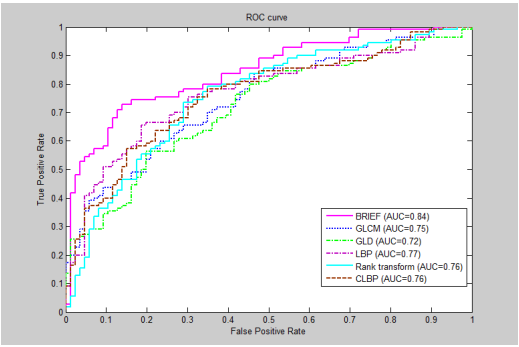


Fig. 4. The Receiver operating characteristic (ROC) curve of BRIEF in comparison with GLCM, GLD, LBP, CLBP, combination of LBP and CLBP and Rank Transform.

applied in our dataset and in [6]. For the record our dataset has an average vertical optic disc diameter of 444 ± 53 pixels. Further evaluation by [6] revealed that performance by [8] drops to AUC of 61% when tested using a wider variation of optic disc sizes.

V. Conclusion and Future Work

In this paper, we described an approach for a glaucoma classification system using textural feature description based on BRIEF. Test evaluation performed using the approach showed that the method was able to achieve an AUC of 85% for glaucoma classification. Comparison of performance between BRIEF and other texture measurements (GLCM, CLD LBP, CLBP and Rank Transform) was also conducted and based on the result, we showed that the BRIEF features performed better.

The key differences of our approach compared to others are: firstly, the use of a textural feature for the classification. In most of the methods more types of features are normally used, which means some sort of feature selection needs to be applied to find the most relevant features to drive the classification. Second, in our approach the image features are derived from whole image rather than concentrated on the optic disc region. This way, we take account of both deformation found within the optic disc and from regions in the retina background. The method is also reasonably fast and simple to implement, which makes it appropriate for glaucoma screening.

In the present work we do not pre-process the image for blood vessel removal and we do not consider the effect of disc size on the classification result. Therefore for future work

we intend to investigate the effect of including them into consideration. An evaluation using a larger dataset will also be conducted to fully evaluate the feasibility of the approach for the screening program.

REFERENCES

[1] A. Rotchford, "What is practical in glaucoma management?" *Eye*, vol. 19, no. 10, 2005.

[2] J. Katz, A. Sommer, D. E. Gaasterland, and D. R. Anderson, "Comparison of analytic algorithms for detecting glaucomatous visual field loss," *Archives of ophthalmology*, vol. 109, no. 12, p. 1684, 1991.

[3] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 778–792.

[4] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Yamamoto, and H. Fujita, "Automated determination of cup-to-disc ratio for classification of glaucomatous and normal eyes on stereo retinal fundus images," *Journal of biomedical optics*, vol. 16, no. 9, pp. 096 009–096 009, 2011.

[5] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, "Automated diagnosis of glaucoma using digital fundus images," *Journal of medical systems*, vol. 33, no. 5, pp. 337–346, 2009.

[6] G. D. Joshi, "Automatic retinal image analysis for the detection of glaucoma," Ph.D. dissertation, International Institute of Information Technology, Hyderabad, 2014.

[7] J. S. Lai, C. C. Tham, and J. C. Chan, "The clinical outcomes of cataract extraction by phacoemulsification in eyes with primary angle-closure glaucoma (pacg) and co-existing cataract: a prospective case series," *Journal of glaucoma*, vol. 15, no. 1, pp. 47–52, 2006.

[8] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson, "Glaucoma risk index: Automated glaucoma detection from color fundus images," *Medical image analysis*, vol. 14, no. 3, pp. 471–481, 2010.

[9] U. R. Acharya, S. Dua, X. Du, S. Vinitha Sree, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, no. 3, pp. 449–455, 2011.

[10] M. R. K. Mookiah, U. Rajendra Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," *Knowledge-Based Systems*, vol. 33, pp. 73–82, 2012.

[11] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1657–1663, 2010.

[12] M. Alsheh Ali, T. Hurtut, T. Faucon, and F. Cheriet, "Glaucoma detection based on local binary patterns in fundus photographs," *Proc. SPIE, Medical Imaging, Computer Aided Diagnosis*, vol. 9035, pp. 903 531–903 531–7, 2014.

[13] S. Mohammad, D. T. Morris, and N. A. Thacker, "Segmentation of optic disc in retina images using texture," in *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 1, Lisbon, Portugal, 5-8 January, 2014*, 2014, pp. 293–300.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[15] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 6, pp. 610–621, 1973.

[16] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 6, no. 4, pp. 269–285, 1976.

[17] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 582–585.

[18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer VisionECCV'94*. Springer, 1994, pp. 151–158.

[19] P. Paclík, S. Verzakov, and R. P. Duin, "Improving the maximum-likelihood co-occurrence classifier: a study on classification of inhomogeneous rock images," in *Image Analysis*. Springer, 2005, pp. 998–1008.