IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 27-32

Formulation a Framework for Discovery of Knowledge Patterns in Diabetic Retinal Images Through Data Mining Techniques

Ankita Makode,Apeksha Meshram, Prof. Kalpna Malpe, Mayuri Gaikwad,Tejaswini Chavhan
Department of Computer Science and Engineering Guru Nanak Institute of Engineering and Technology,Dahegaon Nagpur,Maharashtra,India
Department of Computer Science and Engineering Guru Nanak Institute of Engineering and Technology,Dahegaon Nagpur,Maharashtra,India

Abstract-*The exploration of the existing research in automated retinal image analysis towards extraction of Describing, Structural and Disease patterns aiding Diabetic Retinopathy (DR) detection reveals several challenges which are stated as the need (i) to extract features that are very expressive (ii) to formulate a representative feature vector which helps in elevating the performance of data mining algorithms (iii) to devise data mining algorithms that could achieve improved accuracy as medical experts seek higher accuracy for automated medical systems and (iv) to develop hybrid methodologies that evolve efficient rules to handle complex scenarios. The stated barriers indicate the need for efficient image processing and data mining algorithms to mine significant patterns from retinal fundus images to detect Diabetic Retinopathy. This chapter presents definition of the problem undertaken and objectives of the research. It also highlights the research methodology adopted for undertaking this investigation. This research focuses on utilizing image processing and data mining methods to analyze and process the retinal fundus images for the detection of retinal image patterns namely Describing, Structural and Disease patterns through appropriate elicitation of the image features, targeting Diabetic Retinopathy detection. Therefore, deploying suitable image processing techniques to extract distinguishable features and application of relevant transformation and feature selection techniques to formulate a expressive feature vector and build efficient data mining (classification, clustering hybrid) models to predict the patterns in retinal images (Quality grading, Left or Right eye identification, macula and fovea, blood vessels, optic disc and exudates) aiding Diabetic Retinopathy detection, with a high accuracy rate. In view of this, the following objectives are articulated.*

I. Introduction

The emerging field of medical image mining emphasizes the extent to which image processing and data mining techniques are collectively helpful to medical experts in understanding and analysing medical images. Image mining focuses on methodologies for extracting useful knowledge from images, for which image processing and data mining have been utilised collectively. Image processing techniques deal with processing the image and making it suitable for further analysis (Bourne 2010). Then, data mining, the science of analyzing data from different perspectives, thereby finding new and interesting patterns, thus discovering meaningful correlations and trends from large data (Pal 2011), are adopted on the processed image to identify useful knowledge patterns. Hence, these techniques when employed together aids in efficient disease diagnosis. This thesis presents the research work undertaken to explore the utilization of image processing and data mining techniques in retinal image analysis, particularly Diabetic Retinopathy (DR) detection, on the publicly available benchmark data in the form of fundus images captured by fundus camera. This chapter presents an overview of image processing and data mining techniques utilized in the medical image analysis, followed by its application on retinal images and description of the datasets used for this purpose. Then, the organization of the remaining chapters of the thesis is provided.

1.1 Medical Image Mining

Medical images are obtained through medical imaging, a technique that involves different imaging modalities and processes to picture the human body and create visual representations of the interior of the body for clinical analysis, disease diagnosis and treatment. Therefore, medical images provide the information required to diagnose or treat or follow-up the severity of a disease. Hence, medical image analysis plays an important role in initiatives to improve public health. Manual analysis of these images for diagnosis and clinical judgement is a time and resource intensive task. Hence, automated medical image analysis is sought for this purpose. Out of the many techniques available, image processing and data mining have been extensively utilised for medical image analysis. Image processing techniques operate on image to make it suitable for extraction of

expressive features. Then, data mining techniques are adopted on these features to discover useful knowledge patterns. The efficiency of data mining techniques depends greatly on the representativeness of the features elicited through image processing techniques. Hence these techniques when employed together aids in efficient disease diagnosis.

1.2 Human Eye Anatomy

Automated medical image analysis has been widely used in the field of medicine and biology, out of which its application in the field of Ophthalmology is one of highly demanded. Ophthalmologists analysis the images of various parts of the eye to identify the health of the eye. The human eye, one of the most important sense organs responsible for the sensation of vision, constitutes a very complex structure and mechanism for visual processing.

Human eye is an intricate anatomical structure that enables the humans to see. Each part of the eye plays a distinct role in the process of vision. The ultimate goal of such an anatomy is to allow humans to focus images on the back of the retina. The anatomy of eye is shown in Figure 1.1.

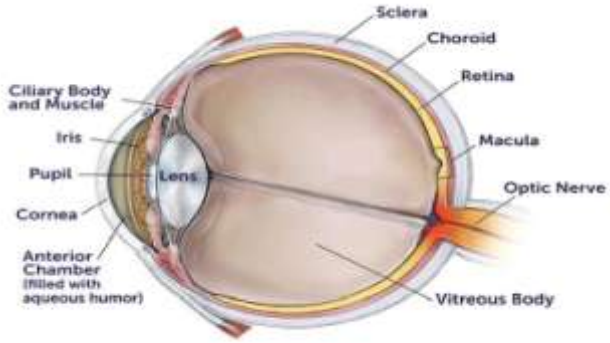
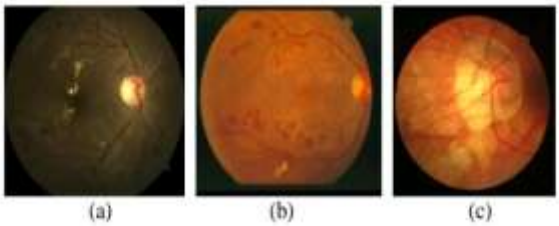


Figure 1.1 Components of the eye

1.3 Diabetic Retinopathy

Diabetic Retinopathy (DR), one of the primary causes for blindness in working age population, is a retinal disease caused due to prolonged high blood sugar levels owing to Diabetes Mellitus. Every subject diagnosed with diabetes has a high risk of developing DR particularly, when the blood sugar levels are not within limits. The vision impairment can be prevented if the disease is diagnosed and treated early, but the patient with DR do not realise any symptoms until the disease progresses to a later stage. During the onset of the disease, the blood vessels start to leak blood, lipids and proteins paving way for the appearance of small bright dots called exudates. They appear on the retina as bright, reflective white or cream coloured lesions. If this leakage occurs on the macula region, vision may be drastically affected. Examination of the retinal fundus image reveals these exudates, if the patient is affected by DR. Sample of retinal fundus images, affected by DR are shown in Figure 1.2.



1.2 Diabetic Retinopathy affected retinal fundus Image
(a)showing exudates (b)showing abnormal blood vessels (c)revealing abnormality in optic disc

1.4 Retinal Image Patterns

The knowledge patterns that can be revealed from retinal fundus images are grouped into three categories namely Describing, Structural and Disease patterns. Describing patterns report the characteristics of the image. Structural patterns reveal various retinal anatomical structures while disease patterns expose the diseased regions. These patterns are narrated here.

1.4.1 Describing Pattern

Describing patterns describe the image as a whole revealing its characteristics such as quality and left or right eye annotation. Retinal fundus images of left and right eye differ in the structural properties. Left or right eye (LR) identification is necessary for identifying nasal and temporal side of the OD. ISNT (Inferior, Superior, Nasal, Temporal) ratio, a measure computed to identify the health of the eye, varies in its calculation on whether is it computed on a left or right eye. Hence, LR identification is a pre-requisite for ISNT ratio calculation.

1.4.2 Structural Pattern

The structural patterns are the patterns that reveal the anatomical structures of the retina. Having mined the structural patterns, the result can be viewed as a binary image with the structural pattern identified as white and the others as black. In case of structural patterns, this binary image can then be post-processed through application of simple image processing techniques to establish the domain knowledge about the structure. The structural patterns include macula and fovea, blood vessels and optic disc (OD).

1.4.3 Disease Pattern

Disease patterns expose the diseased regions in the retinal fundus image. These patterns, similar to the structural patterns can be viewed as a binary image with the disease patterns as white and the other background regions as black. It is also post-processed to eliminate unwanted structures and establish domain knowledge.

1.5 Retinal Fundus Image Datasets

Retinal fundus images are available in public repositories. This encourages the researchers in this field to leverage their computational methods on such images and evaluate the performance of their algorithms. This provides a common platform for comparison of the proposed methods with the existing methods. A large variety of retinal image datasets are available in the publicly available repositories. Retinal datasets related to DR alone are taken up for investigation in this research. A few datasets also contain the benchmark ground truth of a few structural patterns annotated by ophthalmologists, which would help in evaluating the performance of the automated methods. The datasets namely HRF, DRIVE, STARE, DIARETDB0, DIARETDB1, HEI-MED and MESSIDOR datasets are investigated in this study.

II. Literature Review

This chapter elaborates on the existing literature pertaining to image processing and data mining methods towards retinal image analysis in detecting Diabetic Retinopathy (DR). Retinal image analysis involves identification of various Describing, Structural and Disease patterns. The earlier works attempted in extracting Describing (quality, left or right eye identification), Structural patterns (macula and fovea, blood vessels, optic disc) and disease pattern (exudates) are detailed here. To begin with, the review on various methodologies for extraction of describing patterns is presented in the following section.

2.1 Methods For Discovery Of Describing Patterns

Discovery of describing patterns include grading of quality in retinal images and identification of left or right eye in fundus images. The earlier works in this aspect are presented here. The sub-section below presents the review on quality grading of retinal fundus images.

2.1.1 Quality grading of retinal images

The earlier attempts on grading the quality of a retinal image are concisely presented here. The computational approaches developed towards quality assessment rely on histogram analysis, local and global generic image quality parameters, structural criteria, or both structural and generic image quality criteria. The approaches commonly involve segmentation of retinal anatomical structures, extracting properties of the anatomical structures, eliciting local and global generic features characterising the quality factors and categorisation of quality through application of classification techniques on extracted features from the retinal images.

2.1.2 Left or Right eye (LR) identification in retinal images

Left or Right eye (LR) identification plays a major role in identifying the nasal and temporal regions within the optic disc (OD) and also can aid in annotation of the image. The previous attempts in LR identification are presented in this section.

A method taking into account the position of optic nerve head with respect to the macula was investigated . The approach first identified the optic cup through its characteristic bright intensity. A ROI was formulated with identified pixel as centre. Then, the OD was segmented from the ROI through thresholding and

morphological processes. The sum of pixel intensities of the green channel in the left half of the disc was then compared with the right half.

2.2 Techniques For Segmentation Of Structural Patterns

The existing approaches towards extraction of structural patterns viz., macula and fovea, blood vessel and optic disc (OD) are discussed in this section. The following sub-section deals with the earlier attempts on segmentation of macula and localisation of fovea.

2.2.1 Macula segmentation and Fovea localisation

Macula is the dark region at the centre of the retina, responsible for detailed vision . The fovea forms the central part of the macula. Hence, identification of macula and fovea has been dealt together in the literature. The macula does not possess a well defined boundary, making it difficult to evaluate the extraction algorithms through segmentation accuracy.

2.2.2 Blood vessel segmentation

Blood vessels, one of the most complex retinal image patterns nourish the retina with blood supply. The properties of the blood vessel act as a diagnostic marker towards vision threatening retinal diseases, particularly DR. Hence, segmentation of blood vessel gains importance. Two benchmark datasets namely DRIVE (DRIVE:Digital Retinal Images for Vessel Extraction) and STARE have been widely used for the purpose of validation of algorithms towards vessel segmentation. Accuracy, sensitivity and specificity are used to quantify the performance of the algorithms. Retinal blood vessel segmentation is achieved through assigning each pixel as either a vessel pixel or non-vessel pixel.

2.2.3 Optic disc segmentation

Optic Disc (OD) is an important structural pattern appearing as a bright oval structure in the retinal fundus image. It is the entry point for blood vessels. The OD is composed of optic cup, optic rim and blood vessels emerging from it. The properties of OD and its constituents act as a diagnostic indicator for identifying retinal diseases primarily Glaucoma and DR. The literature reveals two levels of extraction with regards to OD: (i) The location of OD is identified (OD localisation or detection) and (ii) boundary information of OD is elicited (OD segmentation).

III. Implementation

3.1 Image Mining Framework To Discover Retinal Patterns For Diabetic Retinopathy Detection

An overall framework of this research is presented in this chapter. Pattern discovery from retinal images involves application of image processing techniques to formulate the feature vector and adoption of data mining techniques to find the various patterns viz. Describing, Structural and Disease patterns that aid Diabetic Retinopathy detection. The processes involved in the discovery of various patterns from retinal images.

3.1.1 Phases In Discovery Of Retinal Image Patterns

Pattern discovery from images requires the utilization of image processing and data mining techniques. For the data mining algorithms to work efficiently, a feature vector which represents the candidates in a distinguishable form is essential. To formulate such a feature vector, application of a few image processing steps is necessary. Once a representative feature vector is identified, the data mining procedures can be applied to extract the various patterns from the retinal images.

3.2 Tournament Selection Based Inter Hybrid Ranked Decision Forest For Quality Assessment Of Retinal Images

This chapter brings to light the supervised ensemble classification technique adopted to assess the quality of retinal fundus images. The existing approaches extracted quality related features and adopted classification techniques for quality categorisation. Exploration of various classification techniques in this regard was minimal in the existing research. Tournament Selection based Inter Hybrid Ranked Decision Forest is put forth for achieving this motive. The concept of Tournament selection adopted in Genetic Algorithms (GA) (Goldberg 1989; Man et al 1996) to select the fittest candidates for next generation is tailored to suit the data mining needs of the classification procedures in view to attain better classification accuracy. This chapter details the concept of the proposed Tournament Selection based Inter Hybrid Ranked Decision Forest followed by the description of the quality characteristics of the retinal images succeeded by the narration of the proposed methodology adopted for this task. Then, the improved results obtained through the proposed methodology are highlighted comparing it with the existing classification algorithms.

3.3 Ranked Ensemble Of Tournament Based Intra Hybrid Decision Trees For Left Or Right Eye (Lr) Identification

This chapter highlights the supervised classification techniques adopted to categorize left and right eye images. Existing works in this regard involve segmentation of anatomical structures such as macula, optic disc and blood vessels within the optic disc. As segmentation is computationally complex and error-prone task, a data mining based approach is attempted for LR identification. The proposed Tournament based Inter Hybrid Ranked Decision Forest (introduced in Chapter 5) achieves comparable performance with the existing methods. Hence, an improvised classification ensemble, Ranked Ensemble of Tournament based Intra Hybrid Decision Trees, is proposed for this purpose. This chapter describes the procedure of the proposed ranked ensemble of classification trees. Then, the significance of LR identification is projected followed by the description of the methodology adopted for this task. The results highlight the various attempts in choosing the feature set, color models, classification algorithms and the ensemble procedures to achieve the improved classification accuracy.

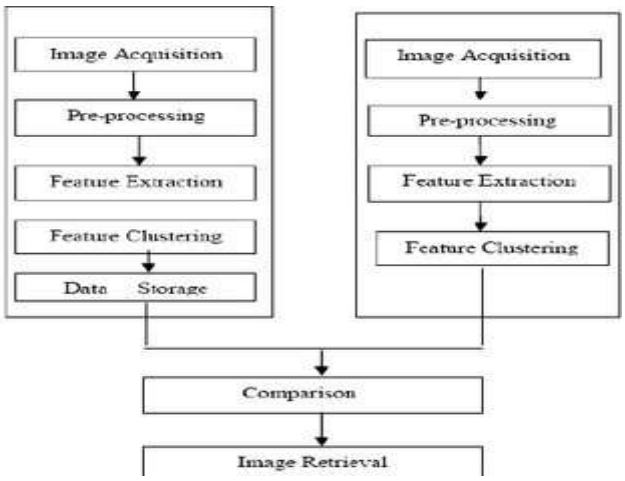
3.4 Heuristic Based Clustering For Macula Segmentation And Fovea Localisation

Macula is a significant structural pattern responsible for high resolution vision. Existing approaches towards extraction of macula and fovea mostly involve segmentation of optic disc and identifying macula with optic disc as reference. This is a computationally complex and error prone task. This work attempts to utilise data mining techniques for this purpose. Since macula lacks well defined border definition and no expert opinion is provided in this regard, supervised classification techniques cannot be adopted. Unsupervised clustering technique is sought for this purpose. This chapter introduces an unsupervised clustering algorithm for segmentation of macula. The algorithm targets at incorporating a heuristic based on measures indicating the statistical distribution of data for selecting the initial cluster centres that play a significant role in performance of the clustering algorithms. Initially, the clustering algorithm is explained. Then, the significance of extraction of macula region is quoted which is followed by the proposed methodology to serve the purpose. Then, the results are presented with respect to the literature. The following section deals with the heuristic based clustering algorithm.

3.5 Hybrid Model Using Naive Bayes And Ranked Ensemble Of Tournament Based Intra Hybrid Decision Trees For Diabetic Retinopathy Detection

This chapter introduces a hybrid model that utilises two classification algorithms in succession in order to achieve higher classification accuracy. The hybrid model is utilised to detect Diabetic Retinopathy (DR) from the entire image and/or segmented anatomical structures such as blood vessels and optic disc. DR detection is attempted through classification by individual and ensemble classifiers, but only less accuracy is reported. Hence, the decrease in accuracy is related to noise in data. Naive Bayes classifier, which incorporates Bayesian probability, is used for elimination of noise. Then, Ranked Ensemble of Tournament based Intra Hybrid Decision Trees is used for classification of the remaining data. To start with, the proposed hybrid model is explained. Then, the proposed methodology that uses the proposed hybrid model to detect DR through various sources is detailed. Followed by this, the results on DR detection are projected. The following section elaborates on the proposed hybrid classification model.

IV. System Architecture



V. Result

This chapter presents the conclusions of the research work undertaken highlighting some of the key points and research contributions on exploitation of image processing and data mining techniques towards discovery of retinal image patterns that aid in Diabetic Retinopathy (DR) detection.

VI. Conclusion

In view to detect Diabetic Retinopathy, three broad categories of patterns were formulated namely Describing patterns, Structural patterns and Disease patterns. The describing patterns describe the properties of the image such as (i) Quality and (ii) Left or right eye. The structural patterns are the anatomical structures in the retinal fundus image such as (i) macula and fovea (ii) blood vessels and (iii) optic disc (OD) while disease patterns refer to the diseased structures in the retinal fundus image, exudates in this work. The research methodology for discovery of patterns from retinal fundus images involved three phases namely image processing, data mining and image post processing phases. Exploitation of image processing and data mining methods revealed the importance and necessity for extraction of relevant features and categorisation methods that would yield high accuracy. This led to formulation of classification, clustering and hybrid models that could generate efficient rules for categorization

VII. Future Scope

The proposed methodologies for extraction of retinal image patterns could be utilised for automated retinal image analysis system to be used by the practitioners in the field of Ophthalmology for retinal image analysis and early diagnosis of sight threatening diseases particularly Diabetic Retinopathy. However, there still exist challenges that need to be addressed:

- The need for improved image filtering techniques to better expose the regions of interest and
- The need for new evaluation metric for assessing the distinguishability of split attributes.

Application

This software contains all the data related to the Diabetic Retinopathy disease. It detect the Diabetic Retinopathy disease of the human eye.

Reference

[1]. Eljkovic V, Bojic, M , Zhao, S , Tameze, C & Valev, V 2015, 'Exudates and Optic Disk Detection in Retinal Images of Diabetic Patients', Concurrency and Computation: Practice and Experience, vol. 27, no. 1, pp. 172-192.

[2]. Sunir. 2014. Diabetic Retinopathy. Available from: <http://www.msdmanuals.com/professional/eye-disorders/retinaldisorders/diabeticretinopathy> [29 January 2016].

[3]. Goldbaum M. 2013. Structured Analysis of the Retina. Available from : <http://www.ces.clemson.edu/~ahoover/stare/> [29 January 2016].

[4]. Gegundez-Arias, ME , Marin, D , Bravo, JM & Suero, A 2013, 'Locating the Fovea Center Position in Digital Fundus Images using Thresholding and Feature Extraction Techniques.', Computerized Medical Imaging and Graphics, vol. 37, no. 5, pp. 386393.

[5]. Fleming, AD , Philip, S , Goatman, KA , Sharp, PF & Olson, JA 2012, 'Automated Clarity Assessment of Retinal Images using Regionally based Structural and Statistical Measures', Medical Engineering & Physics, vol. 34, no. 7, pp. 849-859.

[6]. Paulus, J , Meier, J , Bock, R , Hornegger, J & Michelson, G 2010, 'Automated Quality Assessment of Retinal Fundus Photos ', International journal of computer assisted radiology and surgery, vol. 5, no. 6, pp. 557-564.