Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 402 – 408

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Case Based Reasoning in the Detection of Retinal Abnormalities using Decision Trees

Sreeparna Banerjee[a,]*, Amrita Roy Chowdhury[a]

[a]*West Bengal University of Technology, BF 142, Salt Lake, Kolkata and 700064, India*

Abstract

The most common abnormalities in retina images occur due to age related macular degeneration and diabetic retinopathy. In this paper a decision support system is proposed to classify these abnormalities. The process involves the combination of contextual information with images obtained from a database. Decisions trees are proposed for this purpose as they can combine contextual information with images. Images are pre- processed and segmented to obtain the regions of interest for the individual manifestations in each of these diseases. Matching of candidate segmented images with prototype segmented images is performed along with the matching of the associated contextual information.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: CBR; similarity based retrieval; decision Trees; retinal Images;

1. Introduction

Over the past couple of decades, Case Based Reasoning (CBR) has evolved as a popular paradigm for decision making in real world problems. CBR hinges on the fact that similar cases have similar solutions. This methodology has been adapted from the physicians' approach to diagnosis and therapy planning. The knowledge acquired by medical experts is a combination of text book knowledge and knowledge acquired from clinical experience. The basic idea of CBR [1] is to retrieve cases from a case database and establish the relevance between candidate and prototype cases through a similarity measure. If the first case history involves the analysis and classification of sets of longitudinal series of multimedia image sets, automatic indexing using digital content, referred to as Content

* Corresponding author. Tel.: +91-9830496271
E-mail address: sreeparnab@hotmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.037

Based Image Retrieval (CBIR) [2] is a possible solution for defining similarity measures. Thus, retrieval of images from a database based on similarity based criteria is an important procedure in CBR [3, 4, 5]. When the retrieval process involves heterogeneous information like images and contextual information, the CBR system encounters problems in aggregating these variables and dealing with missing information. Decision trees [3] are well suited for retrieval to process heterogeneous as well as incomplete information. This note proposes an approach to classify abnormalities in fundus images of the retina using decision trees. The abnormalities of the retina that have been considered for the present study are those arising from age related abnormalities like Age Related Macular Degeneration(AMD) and those arising from Diabetic Retinopathy (DR) like micro aneurysm(MA), hard exudates (HE), cotton wool spots (CWS) and hemorrhages. Table 1 lists the manifestations of the abnormalities as seen in retina fundus images, obtained in public databases. To the best of our knowledge, there is no existing work combining AMD and DR. This is a pioneer attempt in integrating contextual information with images.

Table 1.Properties of different retinal abnormalities

Name of disease	Types of abnormalities	Characteristic	Colour	Shape	Frequency	Area affected
AMD	Dry	retinal pigment epithelial layer	yellow	circular	many	large/small
AMD	Wet	bleeding, leaking from blood vessels	dark red	not defined	few	large
DR	Micro aneurysms	occur secondary to capillary wall out pouching due to pericyte loss	red	small circular	few/many	small
DR	Haemorrhages	occur as microaneurysms rupture in the deeper layers of the retina	bright red	not defined	few	large
DR	Hard exudates or waxy exudates	breakdown of the blood-retina barrier, allowing leakage of serum proteins, lipids, and protein from the vessels	yellow	circular	many	large
DR	Soft exudates or cotton wool exudates	nerve fiber layer infarctions from occlusion of precapillary arterioles	white or yellowish white	not defined	few	small

2. Previous Work

Segmentation of retina images to extract abnormalities, and in particular, exudates, using fuzzy c-means have been used successfully in earlier work. Mathematical morphology has also been used to extract exudates and micro aneurysms [9, 13]. Hue and intensity characteristics of regions of interest have been utilized in exudates extraction [10]. Naive Bayes and SVM classification has been used to classify exudates [14]. Micro aneurysms have been detected using Gabor filters [15]. Decision Support systems have been used to classify diabetic retinopathy images [12]. Machine learning has been used for exudates extraction [14]. Decision trees along with association rule mining has been used for general medical image classification [17]. A comparative study of Case Based Reasoning and data mining has been carried out [16]. These approaches have been confined to DR cases. In our proposed algorithm, the Decision Tree approach has been employed to integrate AMD cases with DR. The advantage is in the use of contextual information to differentiate between AMD and DR.

3. Image Processing

Diabetic Retinopathy (DR) and Age related Macular Degeneration (AMD) are the most common causes for loss of vision. In DR, retina contains blood clots as well as lipids and proteins. In AMD, severe bleeding from blood vessels of retina may cause blurred vision leading to loss of sight.

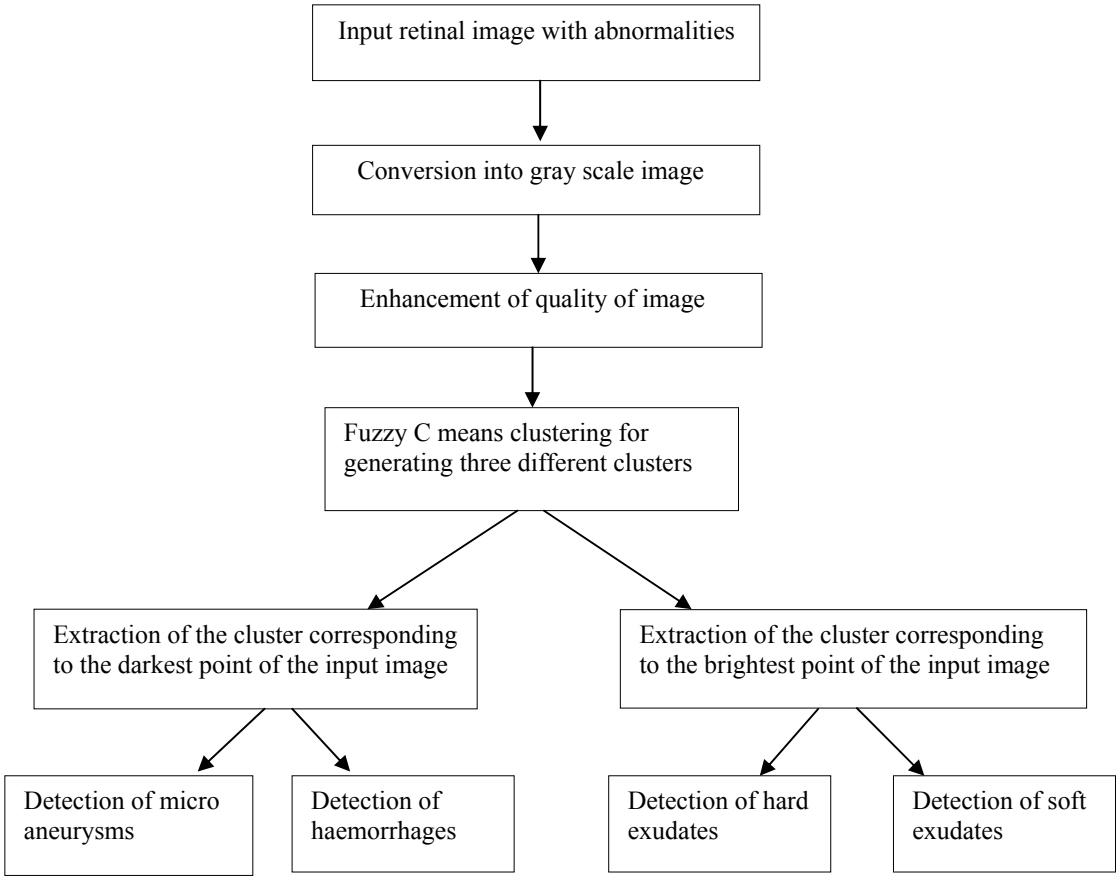

Fig. 1. Flowchart of image pre-processing

Retinal images having abnormalities in public database are considered as input images for the training sets. Normally, retinal images are colored images. So we convert the input images to gray scale images. Then preprocessing on images were performed for enhancing the quality of the images. After that Fuzzy C means (FCM) clustering technique is applied to the images. FCM is based on minimization of the function

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left[x_i - c_j\right]^2 \tag{1}$$

where m > = 1, u_{ij} is the degree of membership of x_i in the cluster j, c_j is center of j^{th} cluster and $[*]$ is used to express similarity between any measured data and center.

Any diseased retinal image has three different color groups. Blood vessel tree and haemorrhages are dark red; Optic Disk, exudates in Diabetic retinopathy and drusens in AMD are bright yellow; mucus and membrane of the retina are of yellowish red color. So we select the number of clusters as three. After three different clusters are

generated, then, depending on the abnormality type, an appropriate cluster is selected. If micro aneurysms or haemorrhages are to be detected from Diabetic Retinopathy images, then the cluster containing the dark red portion should be extracted from the clustered image and the region of interest should be selected from that. If the exudate is a region of interest, then the cluster containing bright yellow color is selected. The flow of the algorithm is depicted in Fig. 1. Images from the public databases were divided into two groups, one consisting the training set comprising of 10 images of each abnormality (60 images) and the other the test set (79 images) out of a total of 139 images.

4. Decision tree

Decision Tree is a decision support methodology which uses a set of rules to divide a population of cases into homogeneous groups and has a flow chart like structure. Each internal node denotes a test on an attribute, each edge denotes the outcome of the test and leaf nodes contain the class label. Each rule comprises of tests on some attributes of a group. These attributes could be images or contextual information.

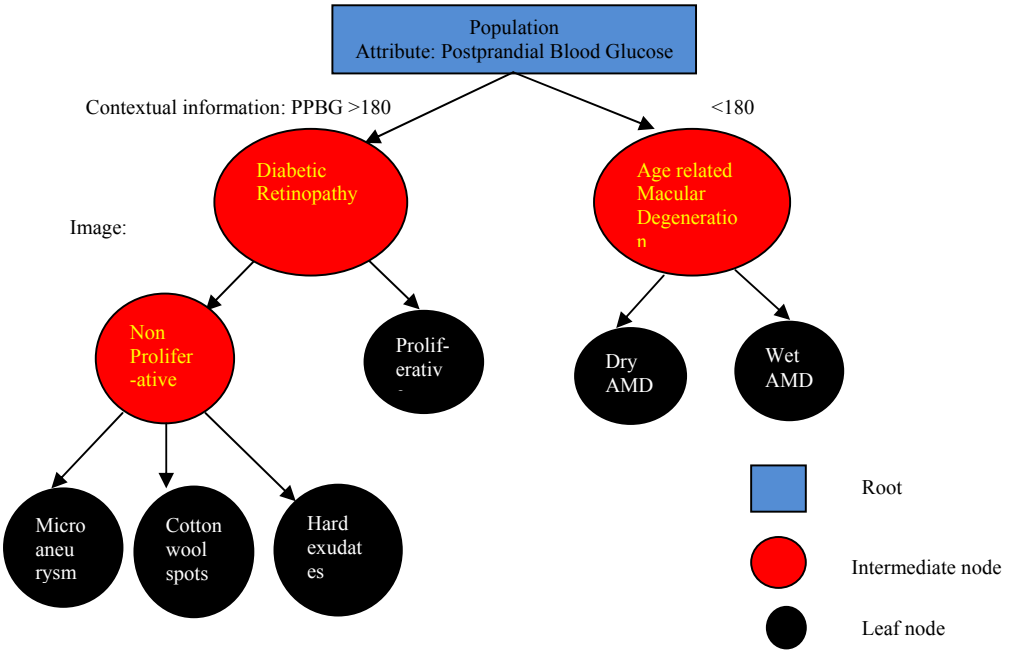

Fig. 2. Decision Tree

Following [3, 4] the tree is made up of a single node containing the whole case population, at the onset of the learning process. After the Decision Tree is constructed, each case i is assigned to each leaf j where j=1...N with a weight w_{ij}. This weight w_{ij} is 0 or 1 if every test attribute is known for i. At the beginning, the whole population forms the root of the tree. For generating different branches of the tree, the discriminating features are selected from the population characteristics. These features are called tests which generate new child nodes. The power of discrimination can be measured by Shannon entropy gain G [3].

$$G = \left(\sum_{n=1}^{N} I^n \right) - I^0$$
$$I^i = \sum_{c=1}^{C} p_c \log p_c, i = 0,1,..N \quad (2)$$

where p_c is the percentage of cases with label $c\,(c = 1,\ldots,C)$ in a node, I^0 is the entropy in the parent node

(before dividing it) and $I^i\left(i=1,...,N\right)$ is the entropy in the i[th] child node. A similarity measure[3] is defined between prototype and candidate cases as follows. At the end of learning each prototype example is assigned to each leaf $\left(j=1,...,N\right)$ with weight w_{ij} equal to zero or 1 if every attribute is known or an intermediate value between 0 and 1 otherwise. When a candidate case (c) is compared to a prototype case (i) which falls in some leaf j, the similarity measure (initially set to zero) is increased by $w_{cj}.w_{ij}$ [3]. The final similarity measure is a sum of these values over all leaves N. Ordering of the prototypes are by decreasing order of this similarity measure. The accuracy is the percentage of retrieved cases whose label matches the query. The heterogeneous variables include age, gender, postprandial blood glucose (PPBG) levels and family history of the disease (e.g. diabetes). The attribute chosen to separate the population into two homogeneous groups is PPBG, which is less than 180 for normal population. Subsequently, image features are analyzed for classification as shown in Fig. 2. For AMD, drusen (yellowish) and haemorrhages are considered. For DR, reddish spots (MA) or yellowish patches (HE), fluffy white (CWS) or haemorrhages.

5. Results

Images of abnormalities of retina images obtained from the prototype database and the regions of interest (ROI) extracted are depicted in Fig. 3 to Fig. 8. In addition to Fuzzy C- Means, morphological operations were performed to remove the optic disc for hard exudates extraction as shown in Fig. 3. For candidate images, the same image processing procedures are followed to obtain ROI. Fig. 3 to Fig. 6 contains images of DR while Fig. 7 and F8 depict images of AMD. One of the earliest manifestations of DR is MA which appears as red spots as seen in Fig. 6. Blood vessels also appear in Fig. 6 which can be removed using morphological operations. In the next stage, there is a leakage of lipid proteins from the blood vessels leading to hard exudates which are round and yellowish in colour (see Fig. 3). The optic disk is also a yellow disk which is removed through morphological operations. Soft exudates or cotton wool spots appear as fluffy white deposits (Fig. 4) also occur during non proliferative DR. Haemorrhages or blood leakage (Fig. 5) occur during the advanced stages which become very severe during the proliferative phase of DR. AMD is initially manifested as the more prevalent form of dry AMD when yellowish deposits know as drusen (Fig. 8) are seen near the macula. Additional blood clots/haemorrhaging is observed in the more severe wet type of AMD (Fig. 7). Prototype images with the above mentioned manifestation are

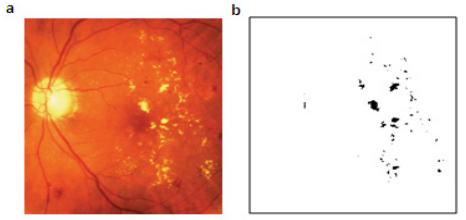

Fig. 3. (a)DR image with hard exudates; (b) hard exudates detected.

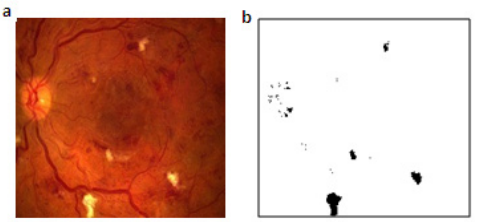

Fig. 4. (a) DR image with cotton wool spots; (b) cotton wool spots detected.

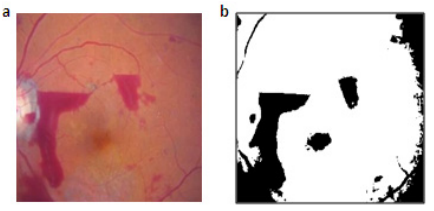

Fig. 5. (a) DR image with haemorrhages; (b) haemorrhages detected

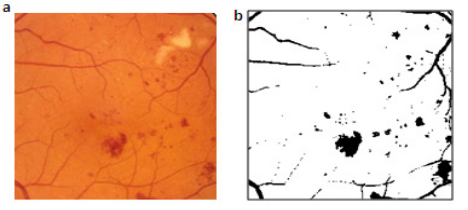

Fig. 6. (a) DR image with micro aneurysms ; (b) micro aneurysms detected

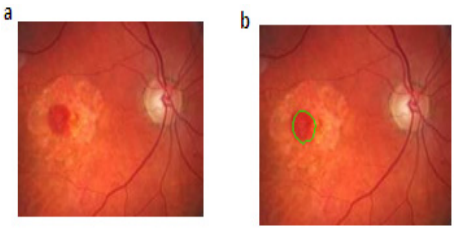

Fig. 7. (a) wet AMD image; (b) blood clot detected

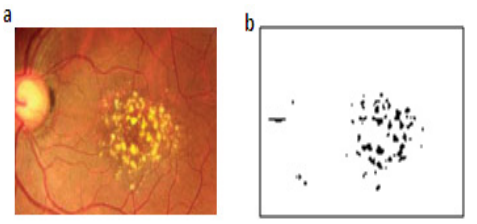

Fig. 8. (a)dry AMD image; (b) dry drusen detected

6. Conclusion and future work

The diseased retinal images are processed and the abnormalities are detected using the proposed algorithm. Several images are collected from public database and used for this purpose. Abnormalities of retina due to Diabetic Retinopathy (DR) and Age related Macular Degeneration (AMD) is considered here and treated separately. The DT algorithm which is in the developmental stages can be improved to automatically detect the type of the disease in retina i.e. whether DR or AMD.

Future work consists of developing a more comprehensive DT framework by including more attributes and values. Also the existing framework needs to be tested and tried out on extensively on many cases to standardize the procedure. The proposed algorithm will be tried on many more test images to be obtained from medical centers, in order to standardize the technique.

Acknowledgment

The authors would like to acknowledge a grant from Department of Biotechnology, Government of India (No. BT/PR4256/BID/7/393/2012 dated 02.08.2012) for supporting this research.

References

1. A. Aamodt. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, vol. 7, no. 1, p. 39–59, March 1994.
2. A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain. Content- based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no.12, p.1349–1380, December 2000.
3. Gwenole Quellec, Mathieu Lamard, Lynda Bekri, Guy Cazuguel, B´eatrice Cochener, Christian Roux. Multimedia medical case retrieval using decision trees. 29th Annual International Conference of the IEEE EMBS , Lyon, France, August 23-26, 2007.
4. M. Das Gupta, S. Banerjee. Similarity Based Retrieval In Case Based Reasoning For Analysis Of Medical Images. Proceeding ICIPACV 2014, Istanbul, Turkey.
5. Gwenole Quellec, Mathieu Lamard, Guy Cazuguel, Christian Roux, B´eatrice Cochener. Case Retrieval in Medical Databases by Fusing Heterogeneous Information. IEEE Transaction on medical imaging 2010, p.1-11.
6. A. Osareh, M. Mirmehdi, B. Thomas R. Markham. Automated Recognition of exudate maculopathy using Fuzzy c-means clustering and neural networks. Procedia MIUA, 2001 BMVC UK, Claridge E., Bamber J. eds. p.49- 52.
7. A. Osareh, M. Mirmehdi, B. Thomas, R. Markham. Comparative exudate classification using support vector machine and neural networks. Procedia MICCAI 2002 (LNCS 2489) p. 413-420
8. A. Sopharak, B. Uyyanonovara, S. Barman. Automatic Exudates Detection from non-dilated diabetic retinopathy retinal images using fuzzy c- means clustering. Sensors, p. 2148-2161 (2009).
9. M. Ponni Bala, S. Vijaychitra. Computerised Retinal Image Analysis to Detect and Quantify Exudates Associated with Diabetic Retinopathy. Intl. J. Computer Applications, 54(2), p. 7-12 (2012).
10. Asha Gowda Karegowda, A. Nasiha, M.A. Jayaram, A.S. Majunath. Exudates Detection in Retinal Images using Back Propagation Neural Network. Intl. J. Computer Applications, 25(3), p. 25-31 (2011).
11. T. Akila, G. Kavitha. Detection and Classification of Hard Exudates in Human Retinal Fundus Images Using Clustering and Random Forest Methods. IJETAE 4(2), p. 25-29 (2014)
12. M. Madheswaran, S. Jerald Jeba Kumar. An Improved Medical Decision Support System To Grading The Diabetic Retinopathy Using Fundus Images. ICTACT Journal On Image and Video Processing, November 2012, Volume: 03, Issue: 02, p. 502-510 (2012).
13. C. Aravind, M. Ponni Bala, S. Vijaychitra. Automatic Detection of Microaneurysms and Classification of Diabetic Retinopathy Images using SVM Technique. Intl. J. Computer Applications, ICIIIOSP-2013, p.18-22 (2013)

14. A Sopharak, Akara, Dailey, Matthew N., Uyyanonvara, Bunyarit, Barman, Sarah, Williamson, Tom, Nwe, Khine Thet and Moe, Yin Aye. Machine learning approach to automatic exudate detection in retinal images from diabetic Patients. Journal of Modern Optics, 57(2), p.124-135 (2010).
15. R. Gowthaman. Automatic Identification and Classification of Microaneurysma For Detection Of Diabetic Retinopathy. IJRET 3(2), p. 464-473 (2014).
16. Vimala Balakrishnan, Mohammad R. Shakouri, Hooman Hoodeh, Loo, Huck-Soo. Predictions Using Data Mining and Case-based Reasoning: A Case Study for Retinopathy. World Academy of Science, Engineering and Technology, International Journal of Computer, Information, Systems and Control Engineering 6(3), p. 20-23 (2012).
17. P. Rajendran, M.Madheswaran. Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm. Journal of Computing, 2(1), p.127-136 (2010).