Knowledge-Based Systems 29 (2012) 83–92

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Data mining techniques for the screening of age-related macular degeneration

Mohd Hanafi Ahmad Hijazi [a,c,*], Frans Coenen [a], Yalin Zheng [b]

[a] Department of Computer Science, University of Liverpool, Ashton Building, Ashton Street, Liverpool L69 3BX, UK
[b] Department of Eye and Vision Science, Institute of Ageing and Chronic Disease, University of Liverpool, UCD Building, Liverpool L69 3GA, UK
[c] School of Engineering and Information Technology, Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia

ARTICLE INFO

Article history:
Available online 23 August 2011

Keywords:
Image classification
Spatial-histograms
Image decomposition
Case based reasoning
Weighted frequent sub-graph mining

ABSTRACT

Age related macular degeneration (AMD) is the primary cause of adult blindness. Currently AMD cannot be cured, however early detection does allow the progress of the condition to be inhibited. One of the first symptoms of AMD is the presence of fatty deposits, called *drusen*, on the retina. The presence of drusen may be identified through the manual inspection/screening of retinal images. This task, however, requires recourse to domain experts and is therefore resource intensive. This paper proposes and compares two data mining techniques to support the automated screening for AMD. The first uses *spatial-histograms*, that maintain both image colour and spatial information, for the image representation; to which a case based reasoning (CBR) classification technique is applied. The second is founded on a hierarchical decomposition of the image set so that a tree representation is generated. A weighted frequent sub-graph mining technique is then applied to this representation to identify sub-trees that frequently occur across the data set. The identified sub-trees are then encoded in the form of feature vectors to which standard classification techniques can be applied.

1. Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in people over 50 years of age. It is caused by damage to the macula, a small area on the human retina that is responsible for seeing fine detail and colour [1]. Although there is no cure for AMD, the condition can be mitigated against in the event of early detection. One of the first symptoms of AMD is the presence of fatty deposits, called *drusen*, on the retina. These can be detected by inspection of retinal images routinely collected within screening programmes. This image inspection is usually conducted manually by trained clinicians. This paper describes two image classification mechanism designed to automate the identification of AMD in retinal images.

The main challenge of the retinal image AMD classification problem is that it is often difficult to distinguish drusen from background noise. This requires appropriate pre-processing of the image data. The need for appropriate image representations, to facilitate the application of data mining, has been identified as a generic challenge within the context of medical image classification in general [2,3]. In the context of AMD screening "standard" object segmentation techniques were found to be unsuitable as the shape and size of drusen varies significantly from image to image and tends to "blur" into the background. The motivation for the work described in this paper is the need for representations, that are entirely compatible with the application of data mining techniques, but which avoid the need for segmentation. In this paper two such techniques are proposed. The first is founded on spatial histograms, the second on a hierarchical decomposition of the "image space". The first is coupled with a case based reasoning (CBR) approach to classification, while the second uses a weighted frequent sub-graph mining technique to achieve the desired classification.

Spatial-histograms (first proposed in [4,5]) emphasise both colour and spatial information [6]. A *region* based approach is suggested in this paper whereby images are subdivided into "regions" and histograms are generated for each. The identified histograms were then conceptualised as time series where the *X*-axis represents the histogram "bin" number, and the *Y*-axis the bin size (number of pixels contained in each). Two different mechanisms were used to identify the desired regions. The first divided each image into 3×3 grid describing nine regions, while the second applied an angular partitioning describing eight regions. Both approaches produced better results then using colour histograms in isolation. The second proposed AMD screening techniques was founded on the concept of hierarchical decomposition whereby images are decomposed into successive sub-regions until either uniform regions or a maximal decomposition is arrived at. The decomposition was conducted using a novel, alternating, angular and circular decomposition. The resulting decomposition was stored in a sequence of tree structures, one per image, to which a

* Corresponding author at: Department of Computer Science, University of Liverpool, Ashton Building, Ashton Street, Liverpool L69 3BX, UK.
E-mail addresses: m.ahmad-hijazi@liverpool.ac.uk (M.H.A. Hijazi), coenen@liverpool.ac.uk (F. Coenen), yalin.zheng@liverpool.ac.uk (Y. Zheng).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.07.002

frequent sub-graph mining technique [7] was applied to identify frequently occurring sub-tree. These sub-trees were then used to form a set of feature vectors (one per image) to which established classification techniques could be applied.

The principal contributions of the work described in this paper are as follows:

(1) A novel approach to AMD screening.
(2) The use of a novel spatial histogram technique, founded on a sub-division of the image space into a small number of regions, to represent images.
(3) The use of a novel hierarchical decomposition technique, using interleaved angular and circular decomposition, to represent images.
(4) In relation to (2) the application of a CBR technique for classification using a time series analysis based mechanism to identify "similar cases".
(5) In relation to (3) the application of a frequent sub-graph mining technique, to generate feature vectors, that uses a weighting mechanism to reduce the search space.

The rest of this paper is organised as follows. Section 2 describes the application domain and Section 3 some relevant previous work. The necessary image preprocessing required for the proposed screening techniques to operate successfully are described in Section 4. The two proposed AMD screening techniques are then described in Sections 5 and 6 respectively. The evaluation of the proposed approaches is presented in Section 7, and some conclusions in Section 8.

2. Age-related macular degeneration

The work described in this paper is focused on the classification of retinal images, in particular the identification of age-related macular degeneration (AMD). As shown in Fig. 1(a), the macula is a small area approximately 5–5.5 mm in diameter centred on the fovea. The fovea is a concave central retinal depression approximately 1.5 mm in diameter which is thought to account for the most acute central and colour vision. The delicate cells of the macula become damaged and stop functioning properly in various conditions. Of these, AMD is the leading cause of irreversible vision loss in people aged 50 or over [1].

Early diagnosis of AMD is achieved by the identification of *drusen* [1,8], yellowish-white sub-retinal fatty deposits, by screening patient retinal images. The severity of AMD can be categorised into three classes: *early*, *intermediate*, and *advanced*. Early AMD is characterised by the existence of several small (63 μm in diameter) or a few medium (63–124 μm) sized drusen or retinal pigmentary abnormalities. The presence of at least one large (124 μm) and numerous medium sized drusen, or geographic atrophy, that does not extend to the centre of the macula, characterises intermediate AMD. AMD can be either *non-neovascular* or *neovascular* [8]. Advanced non-neovascular (dry) AMD exists once the drusen has reached the centre of the macula. *Choroidal neovascularisation* characterises advanced neovascular (wet) AMD. Drusen are often categorised as *hard* or *soft* drusen. Hard drusen have a well defined border, while soft drusen have boundaries that often blend into the retinal background. Fig. 1(a) shows an example of normal retinal image with the macula circled. A retina image that features drusen is given in Fig. 1(b) (drusen indicated by a white arrow). The classification of AMD images by means of drusen identification is not a straightforward process. Most of the previous works have focused on automatic drusen segmentation [9–13] as a necessary precursor prior to AMD classification. The work proposed in this paper, however, approaches the AMD screening problem without the need for the prior identification of the physical existence of drusen. The aim is to classify images as either "AMD" or "non-AMD".

3. Previous work

The earliest work reported in the literature concerning drusen detection is that of Sbeh et al. [14] who used mathematical morphology to identify "brightest points" and hence aid the detection of drusen. More recent work [9] used a wavelet analysis technique to extract drusen patterns, and multi-level classification (based on various criteria) for drusen categorisation. Other work on the identification of drusen in retina images has focuses on segmentation coupled with image enhancement approaches [11–13]. Rapantzikos et al. [13] adopted a multilevel histogram equalisation to enhance the image contrast followed by drusen segmentation, in which two types of threshold, global and local, were applied to retinal images. Köse et al. [11,12] proposed two approaches involving *inverse* drusen segmentation within the macular area. A region growing technique was used to identify "healthy" pixels by applying a threshold on the colour intensity levels [11]. Once this was done, the inverse of the segmented image was used to generate the segmentation of the drusen. A similar inverse segmentation approach, supported by statistical information, was adopted in [12]; where healthy *characteristic images* (CIs) were compared to new *Sample Images* (SIs) and a predetermined threshold applied to classify SI. In [10] another approach, based on a non-parametric technique for anomaly detection, was described that used a support vector data description (SVDD) to segment anomalous pixels.

There has been very little reported work on the application of image mining techniques for AMD screening. The existing work (see above) has been mostly focuses on the segmentation/identification of drusen. Of the reported work that the authors' are aware of, only two reports [9,10] extend drusen detection and segmentation to distinguish retinal images with and without AMD features.

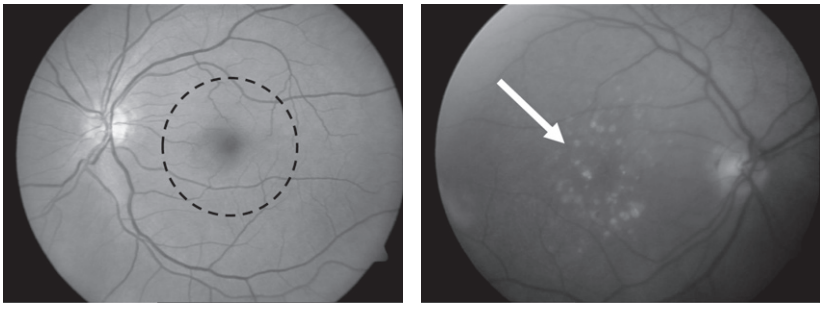

Fig. 1. Illustration of fundus images in grayscale: (a) normal and (b) AMD.