

## Data Mining for AMD Screening: A Classification Based Approach

Mohd Hanafi Ahmad Hijazi  
Applied Computing Group  
Faculty of Computing and Informatics  
Universiti Malaysia Sabah, Malaysia  
e-mail: hanafi@ums.edu.my

Frans Coenen  
*Department of Computer Science*  
University of Liverpool  
Liverpool, UK  
e-mail: coenen@liv.ac.uk

Yalin Zheng  
*Department of Eye and Vision Science*  
University of Liverpool  
Liverpool, UK  
e-mail: Yalin.Zheng@liv.ac.uk

**Abstract**—This paper investigates the use of three alternative approaches to classifying retinal images. The novelty of these approaches is that they are not founded on individual lesion segmentation for feature generation, instead use encodings focused on the entire image. Three different mechanisms for encoding retinal image data were considered: (i) time series, (ii) tabular and (iii) tree based representations. For the evaluation two publically available, retinal fundus image data sets were used. The evaluation was conducted in the context of Age-related Macular Degeneration (AMD) screening and according to statistical significance tests. Excellent results were produced: Sensitivity, specificity and accuracy rates of 99% and over were recorded, while the tree based approach has the best performance with a sensitivity of 99.5%. Further evaluation indicated that the results were statistically significant. The excellent results indicated that these classification systems are ideally suited to large scale AMD screening processes.

**Keywords** - age-related macular degeneration, data mining, decision support techniques, classification, retinal image

### I. INTRODUCTION

In this paper the authors propose and compare three different mechanisms for representing retina images in such a way that classification techniques can be applied to these images so as to support retina screening activities. The idea is to avoid or minimize the use of segmentation techniques, the technology on which retina image analysis normally relies, see for example [1], [2], [3], [4], [5], but instead use alternative whole image representation techniques that do not rely on segmentation. The first technique represents images in terms of histograms which are in turn conceived of as time series curves. The curves associated with a pre-label training set of images are then stored in a Case Base to which a Case Based Reasoning (CBR) tool is applied. The second representation comprises a purely statistical approach whereby a collection of statistics are extracted from an image training set and stored in a tabular (feature vector) format to which established classification techniques can be applied. The third techniques uses a hierarchical decomposition mechanism to generate a set of trees, one per image in the training set, which are then processed using a frequent subgraph mining technique to produce a feature representation (to which established classification techniques can again be applied).

The distinctions between the techniques described in this paper and those found in the literature is that: (i) lesion feature identification is not required in order to perform screening, (ii) novel forms of retinal image representations (histograms and trees) are employed, and (iii) image mining approaches are applied so as to allow for the discovery of patterns (or knowledge) that indicate if AMD is featured within given retinal images.

For evaluation purposes the proposed mechanisms were applied to the detection of Age-related Macular Degeneration (AMD). AMD is a condition where the delicate cells of the macula become damaged (and stop functioning properly) in the later stages of life. AMD is the leading cause of adult blindness in the UK, typically affecting people who are aged 50 years and over. In the UK, it is estimated that in 2020 this age group will comprise a population of 25 million people and more than 7% of these are expected to be affected [6]. AMD is currently incurable and causes permanent total blindness. However, there are new treatments that may stem the onset of advanced AMD if detected at a sufficiently early stage [7]. The diagnosis of AMD is typically undertaken through the careful inspection of retinal images by trained clinicians. Fig. 1 shows some example images. Fig. 1(a) presents a normal retina image,

Fig. 2(b) a retina that displays signs of early stage AMD and (c) a retina that features neovascular AMD.

The rest of the paper is organized as follows. Section 2 presents the background of the work described in this paper. Data preparation is presented in Section 3. The three proposed image classification approaches for AMD screening are described in Sections 4, 5 and 6 respectively.

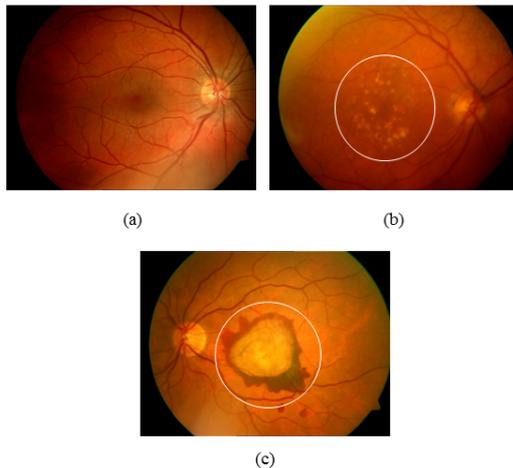


Figure 1. Examples of retinal images: (a) normal retina, (b) retina displaying features (drusen) of early AMD, and (c) retina featuring advanced neovascular AMD.

Section 7 presents a comparison between the proposed approaches and other approaches found in the literature. Some conclusions are provided in Section 8.

## II. LITERATURE REVIEW

There has been much reported work on image classification of all kinds. Typical applications include the classification of photo banks and satellite imagery. Image classification has also been applied to many medical applications. A good example is work on functional Magnetic Resonance Imaging [8]. The challenge of image classification (as also demonstrated in this paper), is not the classification techniques themselves, these are well understood; but the representation of the images in such a way that classification techniques can be applied. This processing typically includes many elements such as deblurring, colour and intensity equalisation, image enhancement of all kinds, noise removal and so on. Much existing work on automated AMD detection using retina images has not been directed at classification, but at the identification of features in retina images which can then be used for prediction purposes. This feature identification is often founded on some form of segmentation, a subject of much continuing investigation and research.

The diagnosis of AMD is typically undertaken by manual inspection of retinal images by trained clinicians. In most cases, an early indicator of AMD is the presence of

drusen, yellowish-white subretinal deposits, on the macula as shown in Fig. 1(b). The presence of large and numerous drusen indicate an early sign of AMD. Drusen can be categorised into hard and soft drusen. Hard drusen have well-defined borders, while soft drusen tends to blend into the retinal background.

The earliest work reported in the literature concerning the automated or semi-automated diagnosis of AMD is that of [5] who used mathematical morphology to detect drusen. Other work on the identification of drusen in retina images has focuses on segmentation coupled with image enhancement approaches [3], [4], [9]. The work described in [4] adopted a multilevel histogram equalisation technique to enhance the image contrast followed by drusen segmentation using both global and local thresholds. A different concept, founded on the use of histograms for AMD screening is also proposed in this paper. In [3], [9] a two phased approach was proposed involving inverse drusen segmentation within the macular area. In [10] a signal based approach called AM-FM was proposed to generate multi-scale features to represent drusen signatures. Images were partitioned into sub-regions; features were then extracted from each sub-region. A wavelet analysis technique to extract drusen patterns, and a multilevel classification for drusen categorisation were described in [1]. A set of rules were used to identify potential drusen pixels. In [11], a content-based image retrieval technique was employed to get a probability of the presence of a particular pathology. Segmentation of objects was first conducted; features were then extracted from the identified objects. A recent work in [12] used greyscale features extracted from the fundus images, which includes fractal dimensions, Gabor wavelet and entropy.

Of the reported work found in the literature that the authors are aware of, only five reports [1], [10], [11], [13], [12] extend drusen detection and segmentation to distinguish retinal images according to whether they exhibit AMD or not. However, most of the work (unlike the work described in this paper) first required the identification (segmentation) of AMD pathologies (drusen) using image processing and content based image retrieval techniques. The distinctions between the techniques proposed in this paper and the above four methods are: (i) that drusen identification (segmentation) is not required in order to perform AMD screening, and (ii) that image mining techniques are utilised to allow the discovery of patterns that indicate if AMD exists or not. The work presented in this paper is similar to [12] where drusen identification is not required. The differences are colour and spatial information are considered in this paper to support the classification.

## III. DATA PREPARATION

To evaluate the proposed approaches described above, two publicly available retinal image datasets were used: (i)

the ARIA<sup>1</sup> and (ii) the STructured Analysis of the Retina<sup>2</sup> (STARE) datasets. Both data sets featured normal retinæ, retinæ that showed signs of AMD and retinæ that feature Diabetic Retinopathy (DR). DR is another retina condition that leads to blindness which is typically identified through screening. Thus the data sets could be used for binary classification purposes (AMD vs. non-AMD) or multi-class classification purposes. ARIA is an online retinal image archive produced as part of a joint research project between St. Paul's Eye Unit at the Royal Liverpool University Hospital (RLUH) and the Department of Eye and Vision Science (previously part of School of Clinical Sciences) at the University of Liverpool. ARIA has a total of 220 manually labeled images. Of these, 101 featured AMD, 59 featured DR and 60 were normal. The STARE dataset was generated as part of a joint project between the Shiley Eye Center at the University of California and the Veterans Administration Medical Center (both located in San Diego, USA). A total of 174 STARE images were acquired for the work described, of these, 64 featured AMD, 72 DR and 38 normal. Both image sets were acquired using similar fundus camera equipment. Thus, with respect to the evaluation described below, the datasets were combined to produce a single large dataset comprising 394 images, of which 165 featured AMD, 131 DR and 98 neither AMD nor DR. The pre-processing that was applied to the image data sets is considered, in this section under two headings: image enhancement and noise removal (see below).

#### A. Image Enhancement

The image enhancement process applied to the collected retina data comprised four components.

1. Region Of Interest (ROI) identification.
2. Colour normalisation.
3. Illumination normalisation.
4. Contrast enhancement.

Before any enhancement could be applied to the digital retinal images the Region Of Interest (ROI) had to be delimited. The reason for this was that enhancement should only be applied to the retina (the ROI) and not to the dark background introduced as part of the image acquisition process. The retina comprises mostly coloured pixels, while the surrounding background comprises mostly black (or dark coloured) pixels (see Figure 1). ROI identification was achieved by applying an image mask to the retinal images so as to isolate and remove the dark background pixels from the original coloured retinal images.

Once the ROI had been identified, the next step was to normalise the colour variations. The aim was to standardize the colours across the set of retinal images. Colour normalization was achieved using the Histogram Specification (HS) approach described in [14]. This approach operates by mapping the colour histograms of

each image to a reference image colour histograms [14], [15]. The task thus commenced with the selection of a reference image that represented the best colour distribution determined through visual inspection on the set of retinal images by a trained clinician. Next, the RGB channel histograms of the reference image were generated. Finally, the RGB histograms of other images were extracted and each of these histograms was tuned to match the reference image's RGB histograms.

Colour normalisation does not eliminate illumination variation. In most of the acquired retinal images, the region at the centre of the retina tends to be brighter than those that are closer to the retina periphery. Illumination variation is of less importance for AMD screening than DR screening as drusen tends to appear in the macula region (centre of the retina); however, luminosity normalisations will enhance the detection of retinal structures such as blood vessels. Illumination normalisation was conducted using an approach, originally proposed in [16], that estimates luminosity (and contrast) variations according to the retinal image colours.

The final stage of the image enhancement pre-processing was contrast enhancement. To this end a Histogram Equalisation (HE) method, called Contrast Limited Adaptive Histogram Equalisation (CLAHE) [17], [18], was applied. HE is a common technique used to enhance contrast. The idea is to distribute colour intensities by spreading out the most frequent intensity values so as to produce a better colour distribution of an image. It improves the contrast globally but unfortunately it may cause bright parts of the image to be further brightened and consequently cause edges to become less distinct. Thus the CLAHE method, that locally equalises the colour histograms, was adopted.

#### B. Noise Removal

Common retinal anatomical structures often serve to "confound" any desired retina image analysis. In the context of the work described here retinal blood vessels were considered to fall into this category. Blood vessel removal commenced with the segmentation of the blood vessels. Various techniques have been proposed for retinal blood vessel segmentation, for the purpose of the work described here an approach that used wavelet features and a supervised classification technique, as suggested in [19], [20], were employed.

Another common retinal structure that could be removed from retinal images is the Optic Disc (OD). However, it is difficult to achieve high OD localisation accuracy in the case of retina images that feature severely damaged retinæ or images of low appearance quality. Thus, the routine localisation and removal of the OD was omitted from the image pre-processing task as standard. However, as will be noted later in this paper, one of the proposed approaches does adopt OD removal under certain conditions.

<sup>1</sup> [http://www.eyecharity.com/aria\\_online](http://www.eyecharity.com/aria_online)

<sup>2</sup> <http://www.ces.clemson.edu/~ahoover/stare>

#### IV. TIME SERIES APPROACH

This first proposed retina image classification method, founded on a time series based representation derived from colour histograms, is presented in this section. A CBR approach [21] was employed to achieve the desired classification. Three histogram based time series generation process were considered: (i) Colour Histograms (CH), (ii) Colour with Optic disc removed Histograms (COH) (ii) and Spatial Colour Histograms (SCH). These were coupled with two CBR approaches

(CBR1 and CBRplus); the first utilised a single Case Base (CB), while the second used two CBs. An overview of the time series approach is given in the following four sub-sections, readers interested in a much more detailed description of the proposed time series base approach are referred to [8].

##### A. Histogram Generation

The histograms used with respect to the time series based approach were generated using the RGB colour model. As stated above, three different categories of histogram were considered: (i) CH, (ii) COH and (iii) SCH. All channels in the RGB colour model were considered as this was found to produces better results than when using individual colour channels [22].

The first strategy extracted CH directly, conceptualised them as a time series (one per image) and stored them in a CB together with their class labels. Previous work had indicated that the removal of irrelevant objects that are common across an image set may improve classification performance. Earlier findings [22], [23] also indicated that, with respect to the retinal images, the OD can obscure the presence of features such as drusen. Hence the second strategy, COH, removed the OD pixels prior to histogram generation. This required identification and segmentation of the OD. There is a significant amount of reported work that has been conducted on OD identification. The adopted approach with respect to the work described in this paper was to localise the OD, by projecting the 2-Dimensional (2-D) retinal image onto two 1-Dimensional (1-D) signals (representing the horizontal and vertical axis of the retinal image) in a similar manner to that proposed in [3], [24].

Given two different images it may still be possible to generate two identical colour histograms. Thus, using colour information alone may not be sufficient for image classification. The third strategy, SCH, adopted a spatial-colour histogram [25], [26] based approach, a technique that features the ability to maintain spatial information between groups of pixels. A region-based approach was employed, whereby the images were subdivided into regions and histograms generated for each region. Feature selection was also applied to the SCH so as to eliminate less discriminative regions and reduce the overall number of SCH to be considered.

Whatever the case, all the generated histograms were conceptualized as time series where the X-axis represents

the histogram bin number, and the Y-axis the size of the bins (number of pixels contained in each). Note that the histograms were normalised so as to avoid the misinterpretation of the distances between the points on two time series caused by different offsets in the Y-Axis.

##### B. Case Based Generation

To facilitate classification, a CBR approach was adopted whereby a collection of labelled cases (examples) was stored in a CB. A new case to be classified (labelled) could thus be compared with the cases contained in the CB and the label associated with the most similar case selected. Two CBR approaches were considered. The first approach (CBR1) used a single CB for classification. The second approach (CBRplus) used two CBs, a primary CB and a secondary CB. The idea here was that the secondary CB acts as an additional source for classification to be used if the primary CB does not produce a sufficiently confident result. For the work described in this section the CH representation was used for the primary CB, while the COH representation was used for the secondary CB. The intuition here was that one drawback of histograms that exclude the OD pixels is that it may result in the removal of pixels representing significant features; especially where the features are close to, or superimposed over, the OD. To reduce the effect of such errors on the classification performance, the utilisation of COH was thus limited to the secondary CB only.

##### C. Case Retrieval

The fundamental idea of CBR is that we resolve a new case according to previously experienced cases contained in a CB. In the classification analogy we wish to classify a new case (image) according to previously classified cases (images) contained in the CB. To achieve this, the new case (described by a time series) needs to be compared with the time series associated with the previous cases and the most similar case or cases identified. The label associated with the most similar case can then be used to categorise the new case. A similarity checking mechanism is therefore required. To this end Dynamic Time Warping (DTW) [27], [28] was adopted because it has been shown to be an effective time series comparison technique [29] and has been successfully applied in a wide range of applications [30], [31]. Further details concerning case retrieval using DTW, as advocated in this paper, can be found in [32].

##### D. Initial Experiments

A sequence of experiments was conducted to: (i) identify the ideal number of histogram bins ( $W$ ) to represent images, (ii) compare the performance of the proposed representation using the CBR1 and CBRplus approaches, (iii) to evaluate the overall use of the proposed histogram based feature selection process. Evaluation was conducted using both a binary (AMD vs. non-AMD) data set and a multi-class data set. Some results of initial experiments have been reported in [32], [33]; the distinctions with the experiments presented in this paper are that in this earlier work: (i) a smaller

number of retinal images was used and/or (ii) evaluation was conducted on binary classification problems only.

The results demonstrated that the histograms extracted from all three RGB channels, combined and quantised to  $W$  colours, produced slightly better overall results than using histograms generated from the green channel alone. Previous work [4], [34] has suggested that the green channel is the most informative channel, but this view was not supported by the experiments conducted by the authors, no particular channel consistently produced a better performance than any other. The suggested explanation for this results is that histograms representing all channels are more informative, and therefore more discriminative (in the context of retina image classification), than the green channel histogram alone. With respect to the  $W$  parameter the results clearly indicate that the  $W \leq 64$  performed better on all evaluation metrics used (note that the maximum number of colours produced by the RGB colour model is 16 777 216). With respect to the performance of individual histogram generation processes, CH produced the best performance. The CBRplus approach tended to produce a comparable performance to the CBR1 approach. However, the CBRplus approach incurred higher computational cost. The results also indicated that by applying feature selection to SCH improved the classification performances, particularly in the multi-class setting.

## V. TABULAR FEATURES APPROACH

In this section the second proposed image classification approach is presented. The approach is founded on a tabular representation that utilises the basic 2-D array image format. The work described in the foregoing section demonstrated that the combination of colour and spatial information (spatial colour histograms) tends to produced better classification performance than when using colour information alone. Therefore, the proposed tabular representation presented in this section utilised both colour and spatial information to identify image features (defined them in terms of statistical parameters) which can be extracted either directly or indirectly from the representation. Two parameter extraction strategies were considered: (i) global extraction where the entire image is taken into consideration, and (ii) local extraction by partitioning the image down to some level of decomposition ( $D_{max}$ ) and extracting parameters on a region by region basis. We refer to the first strategy as S1 and the second as S2. In both cases a feature selection process was applied where the top  $K$  features were selected, partly so that the most discriminating parameters are used for the classification and partly so that the overall number of parameters to be considered was reduced. The rest of this section is arranged as follows. Section 5.1 considers the adopted features, 5.2 the feature selection process and 5.3 the results and conclusions from some preliminary experiments.

### A. Feature Extraction

The most common statistical image parameters are those that can be derived from colour, texture or shape information. With respect to the work described in this paper only colour and texture information were considered as we are interested in the composition of the entire image and not individual shapes within it. A total of fifteen features were used in the proposed tabular based image representation categorised as follows:

- **Features generated directly from the pixel colour information contained in the image (six features).** The six colour features extracted were the average values of each of the RGB colour channels (red, green and blue) and the HSI components (hue, saturation and intensity). These values were computed directly from a 2-D array colour representation of each image.
- **Features generated from a colour histogram representing the colour information contained in the image (two features).** The two histogram based features were: (i) histogram spread and (ii) histogram skewness. In this case only the green channel colour histogram was used as this has been demonstrated to be more informative than the other channels in the context of retina image analysis [4], [34], although this was not fully supported by our own experiments (see above). Once extracted, each histogram was normalised with respect to the total number of pixels of the ROI in the image. The histogram spread (also known as variance),  $h_{spread}$ , and skewness,  $h_{skew}$ , were computed as follow:

$$h_{spread} = \frac{1}{|h|} \sum_{i=1}^{|h|} (\hat{h}_i - \bar{h})^2 \quad (1)$$

$$h_{skew} = \frac{1}{|h|} \sum_{i=1}^{|h|} (\hat{h}_i - \bar{h})^3 \quad (2)$$

where  $|h|$  is the number of histogram bins,  $\hat{h}$  is the normalised histogram and  $\bar{h}$  is the histogram mean.

- **Features generated from the co-occurrence matrices representing the image (three features).** A co-occurrence matrix is a matrix that represents image texture information in the form of the number of occurrences of immediately adjacent intensity values that appear in a given direction  $P$  [14], [35]. Fig. 2 shows an example of a 6 x 6 image  $I$  and its corresponding co-occurrence matrix,  $L$ . To construct

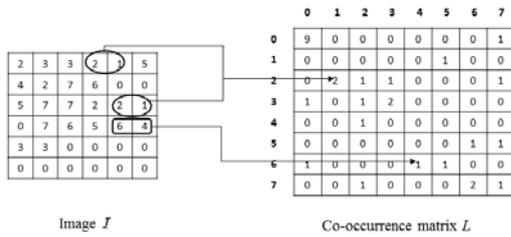


Figure 2. An example of image and its corresponding co-occurrence matrix ( $P = 0^\circ$ ).

$L$ , a position operator,  $P$ , has to be defined. Four possible different directions can be used to define  $P$ :  $0^\circ$ ;  $45^\circ$ ;  $90^\circ$  or  $135^\circ$  (see Fig. 3 where  $X$  is the pixel of interest). With reference to the co-occurrence matrix,  $L$ , in Fig. 2, the number of different intensity values is in the range of 0 to 7, thus a matrix of size  $8 \times 8$  is produced.  $P$  is defined as  $0^\circ$ , which means that the neighbour of a pixel is the adjacent pixel to its right. As shown in Fig. 2, the position (2, 1) contains a value of 2 as there are two occurrences of pixels with an intensity value of 1 positioned immediately to the right of a pixel with an intensity value of 2 in  $I$  (as indicated by oval shapes in Fig. 2). The same applies to the element (6, 4) of  $L$  that holds a value of 1 as there is only one pixel with an intensity value of 6 with a pixel with an intensity value of 4 immediately to its right in  $I$ , and so on. With respect to the approach described in this section, four co-occurrence matrices (one for each  $P$  direction) were generated for each image. Three textural features were then extracted from each matrix: (i) correlation, (ii) energy and (iii) entropy.

- **Features generated using a wavelet transform (four features).** A single level 2-D Discrete Wavelet Transform (DWT) was employed to generate the four wavelet based features used. The features were extracted by computing the average of four types of DWT coefficient:

- 1)  $W_{\varphi}^j(j_0, m, n)$
- 2)  $W_{\psi}^j(j, m, n)$
- 3)  $W_{\psi}^V(j, m, n)$
- 4)  $W_{\psi}^D(j, m, n)$
- 5) The first wavelet based feature is a scale based DWT; while the remaining features correspond to the wavelet response to intensity variations in three different directions: horizontal, vertical and diagonal respectively. To generate  $W_{\varphi}^j(j_0, m, n)$  and  $W_{\psi}^j(j, m, n)$  ( $t \in \{H, V, D\}$ ), assume an image  $f(x, y)$  of size  $row \times col$ . The DWTs were then computed as follows [14]:

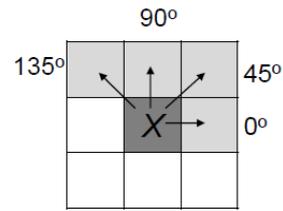


Figure 3. Position operator values.

$$W_{\varphi}^j(j_0, m, n) = \alpha \sum_{x=0}^{2^j-1} \sum_{y=0}^{2^j-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (3)$$

$$W_{\psi}^j(j, m, n) = \alpha \sum_{x=0}^{2^j-1} \sum_{y=0}^{2^j-1} f(x, y) \psi_{j, m, n}(x, y) \quad (4)$$

where  $j = 0, 1, \dots, J - 1$  is the scaling value,  $m = n = 0, 1, \dots, 2^j - 1$  are the translation parameters,  $\varphi_{j_0, m, n}$  and  $\psi_{j, m, n}$  are the scaling and translation basis functions respectively and  $\alpha = \frac{1}{\sqrt{row \times col}}$ .

Both functions are defined as [14]:

$$\varphi_{j_0, m, n} = 2^{j/2} \varphi(2^j x - m, 2^j y - n) \quad (5)$$

$$\psi_{j, m, n} = 2^{j/2} \psi(2^j x - m, 2^j y - n) \quad (6)$$

The 2-D scaling,  $\varphi(x, y)$ , and translation,  $\psi^t(x, y)$ , functions were derived from their corresponding 1-D functions as follow [14]:

$$\varphi(x, y) = \varphi(x)\varphi(y) \quad (7)$$

$$\psi^H(x, y) = \psi(x)\psi(y) \quad (8)$$

$$\psi^V(x, y) = \psi(x)\psi(y) \quad (9)$$

$$\psi^D(x, y) = \psi(x)\psi(y) \quad (10)$$

The values of  $\varphi(t)$  and  $\psi(t)$  were determined by the types of wavelet filters used. In this chapter, the most common *Haar* wavelet filter was employed and defined as:

$$\varphi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 0.5 \\ -1, & 0.5 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

The overall feature generation process is illustrated in Fig. 4. Each of the pre-processed colour images was first

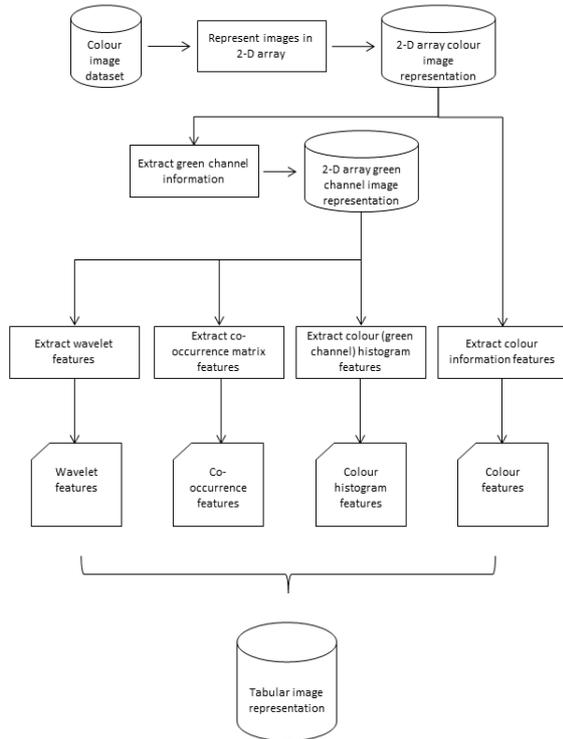


Figure 4. Block diagram of features extraction steps.

represented in a 2-D array form. The size of the array was equivalent to the size of the image it represented; each element of the array contained a pixel intensity value. The colour-based features were extracted directly from this array. The other categories of feature (histograms, co-occurrence matrix and wavelet) were extracted from the green channel representation of the images. Thus, a 2-D array of the green channel image representation was generated from each image. The colour (green channel) histogram, co-occurrence matrix and wavelet-based features were then extracted from this array. The resulting features were kept in a tabular form where each column represents a feature, and each row an image.

The first strategy (S1) was to extract these features with respect to the entire image. The second strategy (S2) was to first partition each image into  $R$  sub-regions using a quad-tree image decomposition technique, the features of interest were then generated from each sub-region. Note that in the context of the work presented in this paper, the decomposition of an image was conducted until some predefined maximum depth,  $D_{max}$ , was reached. Since quad-trees are more suited to square images, the image size was first expanded so that both the height and width of the images were identical. In the context of the work described in this thesis, the dimensions of each retinal image were fixed to  $768 \times 768$  pixels. This was achieved by expanding

the images with zero valued pixels. The extracted feature vectors were then arranged according to the order of the

0	1	4	5
2	3	6	7
8	9	12	13
10	11	14	15

Figure 5. Ordering of sub-regions produced using a quad-tree image decomposition ( $D_{max} = 2$ ).

sub-regions that they represent in an ascending manner, such that the features of the first sub-region formed the first 15 features, the second sub-region formed the next 15 features, while the  $R^{th}$  sub-region formed the last 15 features. Fig. 5 shows the sub-region ordering of an image using a quad-tree of depth,  $D_{max} = 2$ . The value of  $R$  is thus determined by the value of  $D_{max}$  such that  $R = 4^{D_{max}}$ .

*B. Feature Selection*

The next step was to reduce the number of extracted features; the aim being to prune the feature space so as to increase the classification efficiency (through removal of redundant or insignificant features) while at the same time maximising the classification accuracy. The adopted feature selection process comprised a feature ranking strategy based on the discriminatory power of each feature and selection of the best top  $K$  performing features. By doing this, only the most appropriate features were selected for the classification task and consequently a better classification result could be produced. The feature ranking mechanism employed used Support Vector Machine (SVM) weights to rank features [36]. The main advantage of this approach was its implementational simplicity and effectiveness in identifying relevant features.

*C. Preliminary Experiments*

The final stage in the tabular representation process was the classification stage. The nature of the tabular feature space representation permitted the application of many different classification algorithms. In this section, three classification algorithms were used:  $k$ -NN, Naïve Bayes (NB) and SVM. A number of preliminary experiments were conducted to: (i) identify the most appropriate strategy, S1 or S2 (partitioning or non-partitioning) and (ii) determine the most appropriate value for the  $K$  parameter. With respect to S1, using all features produced a better performance, with regards to accuracy and AUC, than when using a reduced number of features. However, with respect to strategy S2, using a reduced number of features ( $50 \leq K \leq 400$ ) produced the best overall performance with respect to all the considered evaluation metrics. Thus, overall, strategy S2 performed better than S1, irrespective of the classification

algorithms used. The conjectured reason for this is that the localized features extracted using S2 are likely to be more informative. With respect to  $D_{max}$  values, the classification tends to performed the best when  $D_{max} = 3$  and  $D_{max} = 4$ .

VI. TREE BASED APPROACH

This section presents the third image classification approach considered in this paper. The approach is founded on the idea of representing retina images using a quad-tree based approach. This representation was deemed appropriate for retinal images as the utilisation of spatial based features tends to produce better classification performances (as illustrated by the two previously proposed techniques described in Sections 4 and 5 above). A similar idea has also been used with some success in the context of analysing MRI brain scan data [8].

The proposed approach comprised three steps: (i) image decomposition, (ii) weighted frequent sub-graph mining and (iii) feature selection and classification. The generation of hierarchical trees to represent each image was done by decomposing the image into regions (a similar idea was adopted in some cases with respect to the tabular technique described above) that satisfied some condition, which then resulted in a collection of tree represented images (one tree per image). Next, a weighted frequent sub-graph (sub-tree) mining algorithm was applied to the tree represented image data in order to identify a collection of weighted sub-trees that frequently occur across the image dataset (an idea suggested by the work presented in [37], [38]). The identified frequent sub-trees were then defined the elements of a feature space that may be used to encode the individual input images in the form of feature vectors itemising the frequent sub-trees that occur in each image. A feature selection strategy was applied to the identified set of frequent sub-trees so as to reduce the size of the feature space. The pruned feature space was then used to define the image input dataset in terms of a set of feature vectors, one per image. Once the feature vectors were generated, any one of a number of established classification techniques could be applied. With respect to the work described in this paper two classification algorithms were used: NB and SVM. Each stage is described in further detail in the following sub-sections.

A. Image Decomposition

A number of image decomposition techniques have been proposed in the literature. The mechanism proposed by the authors, and first suggested in [39], proceeds in a recursive manner as commonly used by other established image decomposition techniques. The novelty of the proposed approach was that a circular and angular interleaved partitioning was used. In the angular partitioning the decomposition was defined by two radii (spokes) and an angle describing an arc on the circumference of the image “disc”. The circular decomposition was defined by a set of concentric circles with different radii radiating out from the

centre of the retinal disc. Individual regions identified during the decomposition were thus delimited by a tuple comprising a pair of radii and a pair of arcs. The technique is illustrated in Fig. 6 which, from left to right, shows four iterations of the decomposition process together with the tree structure produced. The main advantage of the technique (with regard to retinal images) is that it allows for the capture of different levels of detail. Dense detail from the central part of the retina disc image (where the most relevant image information can be found) and sparse detail from the periphery; consequently contributing to the production of a better classifier.

From Fig. 6 the decomposition commences with an angular decomposition to divide the image into four equal sectors. If the pixels making up a sector have approximately uniform colour intensity no further decomposition is undertaken. All further decomposition is then undertaken in a binary form by alternating between circular and angular decomposition. In the example, sectors that are to be decomposed further are each divided into two regions by applying a circular decomposition. The decomposition continues in this manner by alternatively applying angular and circular partitioning until uniform sub-regions are arrived at, or a desired maximum level of decomposition,  $D_{max}$ , is reached. Fig. 7 shows the tree generated from Fig. 6.

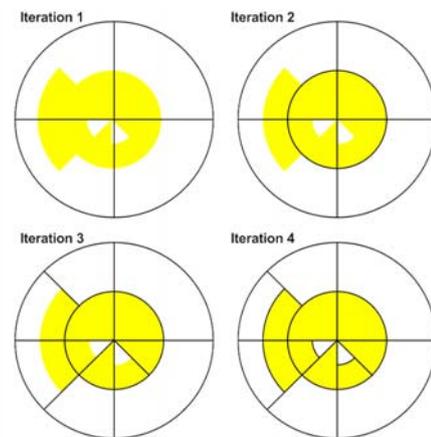


Figure 6. Angular and circular retina image decomposition, iterations 1 to 4 (from top right to bottom left)

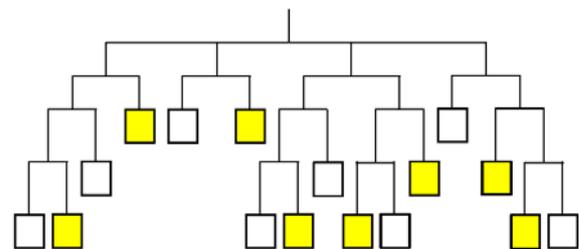


Figure 7. Tree data structure generated from the example hierarchical decomposition shown in Fig. 6

Before the partitioning is commenced, the centre of the retina disc has to be defined and the background and blood vessel pixels removed. This was achieved using a “mask” as described previously in Section 3.1. The image background,  $imbg$ , is defined as:

$$imbg = M \cap RV \quad (13)$$

where  $M(x)$  is 1 if  $x$  is a retinal pixel and 0 otherwise; and  $RV(x)$  is 0 if  $x$  is a blood vessel pixel and 1 otherwise. Using the mask the image ROI was identified and the partitioning commenced. Throughout the process the tree data structure was continuously updated such that each identified region was represented as a “node” in the tree, whilst the relationship between each node and its parent node was represented by edges. The average intensity value of the region was stored at the associated node. The RGB (red, green and blue) colour model was used to extract the pixel intensity values, thus each pixel had three intensity values (red, green, blue) associated with it, hence three trees were generated initially and then merged at the end of the process.

The nature of the termination criterion is important in any image decomposition technique. For the work described here a similar termination criterion as described in [40] was adopted. The homogeneity of a parent region,  $\omega$ , was defined according to how well a parent region represents its child regions’ intensity values. If the intensity value, which is derived from the average intensity values of all pixels in a particular region, of a parent is similar (less than a predefined homogeneity threshold,  $\tau$ ) to all of its child regions, the parent region is regarded as being homogeneous and is not decomposed further. Otherwise, it will be further partitioned. Calculation of the  $\omega$  value for a child region  $i$  of a parent region  $p$  was formulated as:

$$\omega_i = \frac{|\mu_p - \mu_i|}{\mu_p} \quad (14)$$

where  $\mu_p$  is the average intensity value for the parent region and  $\mu_i$  is the average intensity value for child region  $i$ . Note that a lower  $\tau$  value will make the decomposition process more sensitive to colour intensity variations in the image, and will produce a larger tree as more nodes will be generated (but limited to some maximum number of nodes that can be produced by a predefined maximum level of decomposition,  $D_{max}$ ). The decomposition process is performed iteratively until  $D_{max}$  is reached or all sub-regions are homogeneous. Further details of the image decomposition process can be found in [32].

On completion the intensity values stored at nodes in a tree using a label set  $\{equal, high, low\}$ , while the edges were labelled according to the set  $\{nw, sw, ne, se, inner, outer\}$ . If the original intensity values were used as node labels very few frequently occurring sub-graphs would have been found (see below).

### B. Weighted Frequent Sub-Graph Mining

A Weighted Frequent Sub-graph Mining (WFSM) algorithm was applied across the tree dataset. Frequent Sub-graph Mining (FSM) is concerned with the discovery of frequently occurring sub-graphs in a given collection of graphs  $D$ . A subgraph  $g$  is interesting if its support (occurrence count),  $sup(g)$ , in  $D$  is greater than a predefined support threshold. Given a graph dataset,  $D$ , the support of a sub-graph  $g$  in dataset  $D$  is formalised as:

$$sup_D(g) = \frac{freq_D(g)}{|D|} \quad (15)$$

$$freq_D(g) = |\{G_i \in D | g \subseteq G_i\}| \quad (16)$$

A sub-graph  $g$  is frequent if and only if  $sup(g) \geq the support threshold$ . The FSM problem is directed at finding all frequent sub-graphs in  $D$ . There are many different FSM algorithms that have been reported on in the literature. With respect to the proposed tree based approach described in this section the popular gSpan [41] FSM algorithm was used as the foundation for the proposed WFSM algorithm. The idea behind the proposed WFSM algorithm was the observation that some objects in an image can generally be assumed to be more important than others. With respect to the work presented in this paper it was conjectured that, nodes that are some “distance” away from their parent are more informative than those that are not. In the context of the work described here, such distance is measured by considering the difference of average colour intensity between a parent and its child nodes, normalised to the average colour intensity of the parent. The intuition here was that normal retinal pixels have similar colour intensity, while a substantial difference in intensity may indicate the presence of drusen. Thus, the quality of the information in the un-weighted tree representation can be improved by assigning weights to nodes and edges according to this distance measure.

The specific graph weighting scheme adopted with regard to the WFSM algorithm advocated in this chapter is based on the work described in [42]. In the context of the proposed approach, two weights were assigned to each sub-graph  $g$ :

- 1) Node weights: Since abnormalities in retinal images commonly appear to be brighter than normal retina, higher value weights are assigned to such nodes (as these nodes are deemed to be more important).
- 2) Edge weights: The edge weight (that defines the relationship between a child and its parent) is defined by the distance measure described above.

Given a graph dataset,  $D = G_1, G_2, \dots, G_z$ , each node of a graph contains an average intensity value for a region within the image  $I$  it represents. A scheme to compute graph weights similar to that described in [42] was adopted. The node (and edge) weights for  $g$  were calculated by dividing the sum of the average node (and edge) weights in the

graphs that contained  $g$  with the sum of the average node (and edge) weights of all the graphs in  $D$ .

It is suggested that the utilisation of node and edge weights together can reduce the computational cost of FSM, as less frequent sub-graphs will be identified. To extract frequent sub-trees (image features) that are useful for classification, a WFSM algorithm, an extension of the well-known gSpan algorithm, was defined. The WFSM algorithm operated in a similar manner to that described in [42], but took both node and edge weightings into consideration (rather than node or edge weightings). A sub-graph  $g$  is weighted frequent, with respect to  $D$ , if it satisfies the following two conditions:

$$(C1) N_D(g) \times \text{sup}(g) \geq \sigma, \quad (C2) E_D(g) \geq \lambda \quad (17)$$

where  $N_D(g)$  is the node weighting for  $g$  in  $D$ ,  $E_D(g)$  is the edge weighting for  $g$  in  $D$ ,  $\sigma$  denotes a predefined weighted minimum support threshold and  $\lambda$  denotes a weighted minimum edge threshold.

The output of the application of the weighted frequent subgraph mining algorithm was then a set of Weighted Frequent Sub-Trees (WFSTs). In order to allow the application of existing classification algorithms to the identified WFSTs, feature vectors were built from them. The identified set of WFSTs was first used to define a feature space. Each image was then represented by a single feature vector comprised of some subset of the WFSTs in the feature space. In this manner the input set can be translated into a two dimensional binary-valued table of size  $z \times h$ ; of which the number of rows,  $z$ , represented the number of images and  $h$  the number of identified WFSTs. An additional class label column was added.

### C. Feature Selection and Classification

The number of features discovered by the WFST mining algorithm, as described above, is determined by both  $\sigma$  and  $\lambda$  values. Previous work conducted by the author, and presented in [32], [33], [39], demonstrated that relatively low  $\sigma$  and  $\lambda$  values were required in order to generate a sufficient number of WFSTs. Setting low threshold values however results in large numbers of WFSTs, of which many were found to be redundant and/or ineffective in terms of the desired classification task. Thus, a feature selection process was applied to the discovered features. The input to the feature ranking algorithm (similar to the approach described in Section 5.2) was the set of identified WFSTs, and the output was a ranked list of WFSTs sorted in descending order according to their weights. The feature selection process was then concluded by selecting the top  $k$  WFSTs, consequently the size of the feature space was significantly reduced.

The final stage of the proposed tree based retinal image classification process was the classification stage. As described above, each image was represented by a feature vector of WFSTs. Any appropriate classification technique

could then be applied. In the context of the work described here a SVM technique was used.

Some preliminary evaluation (see [32], [39]), which has been conducted using smaller data set and applied to binary classification problems, indicated that the best results were produced using a maximum decomposition of 7 ( $D_{max} = 7$ ),  $\sigma = 10\%$  and  $\lambda \leq 40\%$ . Overall, the application of feature selection produced a better performance than when feature selection was not used, however  $k$  was best set at between 1000 and 4000.

## VII. EVALUATION

This section presents an overview of the evaluation conducted with respect to the three approaches considered above. The section is divided into two subsections. The evaluation in terms of AMD classification is reported in Sub-section 7.1, where five metrics were used to compare the operation of the proposed approaches: (i) sensitivity, (ii) specificity, (iii) accuracy, (iv) Area Under the receiver operating Characteristic Curve (AUC) and (v) the False Negative Rate (FNR). Note that the evaluation of the proposed approaches was conducted using Ten-fold Cross Validation (TCV). The TCV was repeated five times and the training and test images for each TCV were randomised. Average results are thus presented in Sub-section 7.1. Sub-section 7.2 then presents a discussion of the statistical significance analysis conducted (ANOVA and Tukey testing).

### A. Evaluation in Terms of AMD Classification

Table 1 presents the results using the three proposed techniques in the context of a binary classification problem (AMD vs. non-AMD). Note that the results were generated using the best parameter settings as identified from previous experimentation (and as noted above). Table II presents the results obtained in the context of a multiple class classification setting (AMD, DR and "normal"). The right most column shows the FNR produced by the proposed approaches. The best results are indicated in bold font. From Table 1 it can be seen that the Tabular and Tree approaches produced high classification performances of greater than 85% accuracy and greater than 90% AUC. The best recorded accuracy and AUC of 99.9% and the lowest FNR value of 1.0% were obtained using the Tree approach. These are excellent results. The best sensitivity and specificity were also produced by the Tree based approach. The Time Series approach produced the worst results. From Table 2 it can be seen that, as might be expected, an overall lower performance was recorded compared to the binary setting with the exception of the sensitivity to identify AMD (Sens. AMD) and FNR with respect to the Tree based representation. Overall the Tree approach outperformed the Time Series and Tabular approaches with respect to all the evaluation metrics used. These results indicate that using the proposed tree representation, coupled with a weighted frequent sub-graph mining algorithm is the most appropriate

TABLE I. CLASSIFICATION PERFORMANCE AMD V. NON-AMD

Ap.	Sens	Spec	Acc	AUC	FNR
Time Series	74.5	60.4	69.3	73.2	25.5
Tabular	92.3	78.7	87.3	93.2	7.7
Tree	<b>99.0</b>	<b>100.0</b>	<b>99.9</b>	<b>99.9</b>	<b>1.0</b>

TABLE II. CLASSIFICATION PERFORMANCE MULTICLASS SETTING

Ap.	Sens-AMD	Sens-other	Spec	Acc	AUC	FNR
Time Series	57.0	48.0	50.5	52.4	70.7	43.0
Tabular	77.3	75.3	49.2	69.7	84.7	22.7
Tree	<b>99.5</b>	<b>97.1</b>	<b>96.8</b>	<b>98.1</b>	<b>98.5</b>	<b>0.5</b>

with respect to the classification of the retinal images for the purposes of the evaluation. The Tree approach also produced the most reliable results, with a high sensitivity value that would avoid AMD patients being mistakenly screened as being healthy.

As already noted in Section 2 there is very little comparable reported work on the classification (screening) of retina images for AMD. The authors have only been able to identify four instances of comparable work, namely: (i) Brandon and Hoover [1], (ii) Chaum et al. [11], (iii) Agurto et al. [10] and (iv) Cheng et al. [13]. Direct comparison with this reported work is not possible because the data sets used in each case are not in the public domain, except Brandon and Hoover that used the STARE data set. However, with respect to this reported work, it can be observed that:

- 1) The evaluation presented in Brandon and Hoover was applied not only to AMD screening (AMD vs non-AMD), but also to grade the detected AMD. The reported overall accuracy obtained was 90% on 97 images. The AUC metric was not used.
- 2) The work of Chaum et al. was applied in a multiclass setting. The overall reported classification accuracy was 91.3% on 395 images. In their evaluation, 48 images (12.2% of the total images used) were classified as “unknown” and excluded from the accuracy calculation. If this number was included as miss-classifications, the accuracy will be lower.
- 3) Agurto et al. reported a best recorded AUC value of 84% to identify AMD images against non-AMD images from normal eyes and eyes with DR. They also presented the results of applying their approach to AMD images that featured only drusen, as a result of which the recorded AUC value decreased to 77%. No classification accuracy was reported.
- 4) The results reported in Cheng et al. were generated from the classification of AMD images against non-AMD; 350 images were used. Only sensitivity and specificity were recorded, where the best of each were 86.3% and 91.9% respectively.

- 5) Mookiah et al. [12] reported an average accuracy of 95.07% and 95% for ARIA and STARE datasets respectively.

Thus, from the above, it is suggested that the proposed approaches presented in this paper, in particular the Tree approach, produced a comparable performance to those associated with the existing work reported in the literature.

### B. Statistical Comparison

The comparison with respect to AMD screening reported in the foregoing section shows that the best classification performance was produced by the third approach, the Tree based representation. In this section, the results of an Analysis of Variance (ANOVA) test [43] are presented which was used to demonstrate that this result is indeed significant. The ANOVA test operates in terms of the mean of the “accuracies” produced using  $k$  different classifiers. The means of the accuracies of the compared classifiers are said to be different if the between classifiers variability is significantly larger than the within classifiers variability; if this is the case, the null hypothesis can be rejected [43], [44]. This is indicated by the resulting  $p$  value; in the context of the work presented in this paper the  $p$  value corresponds to the probability that all classifiers produced the same mean. From the literature, classifiers are deemed to be significantly different if  $p \leq 0.05$  [45]. The generated accuracy for each run of the cross validation was taken as a sample for the statistical testing. Thus, the number of samples used,  $n$ , was  $10 \times 5 = 50$  for each classifier. Experiments were conducted in terms of both binary classification and multi-class classification. Although there is insufficient space in this paper to present full details of the ANOVA testing conducted, in both cases the differences in accuracy between the proposed approaches, according to the ANOVA test, were highly significant such that the null hypothesis was rejected at  $p < 0.005$  for both the binary and the multi-class contexts.

In order to identify the difference in the operation of the classifiers the Tukey post hoc test was applied [46]. A Tukey test performs multiple pairwise classifier comparisons by calculating the differences between the means of the compared classifiers. The best performing classifier is then identified if the computed differences are large enough. In this context the Critical Difference (CD) value with respect to the binary classification scenario was calculated at  $CD_B = 3.1$ , and for the multi-class scenario at  $CD_M = 2.7568$ . Summaries of the results produced using the Tukey test are presented in Tables 3 and 4. In the Tables  $\bar{x}_A - \bar{x}_B$  indicates the difference between the mean accuracy using approach  $A$  and approach  $B$ . From the tables, the differences between the approaches were all greater than the computed  $CD_B$  and  $CD_M$  values. Thus it can be concluded that the difference between the approaches, in all cases, is statistically significant. In the case of binary classification (Table 3) the Tabular approach performed better than the Time Series approach, while the Tree

TABLE III. TUKEY TEST AMD VS. NON-AMD

Comparison A vs. B	$\bar{X}_A - \bar{X}_B$
Tree vs. Tabular	99.9 – 87.3 = 12.6
Tree vs. Time Series	99.9 – 669.3 = 30.6
Tabular vs. Time Series	87.3 – 69.3 = 18

TABLE IV. TUKEY TEST MULTICLASS SETTING

Comparison A vs. B	$\bar{X}_A - \bar{X}_B$
Tree vs. Tabular	94.8 – 69.5 = 25.4
Tree vs. Time Series	94.8 – 52.4 = 42.5
Tabular vs. Time Series	69.5 – 52.4 = 17.1

approach produced the greatest difference. In the case of multi-class classification (Table 4) the Tree based approach produced the greatest differences and hence the best performance. Thus, in conclusion, the conducted comparisons clearly demonstrated that the proposed Tree based approach outperformed the other two approaches.

VIII. CONCLUSION

The work described in this paper compared the operation of three alternative representations to support retina image classification with respect to screening processes. The first used a time series representation which was coupled with a CBR approach to classification. The second used a tabular representation containing purely statistical information to which standard classification techniques could be applied. The third used a hierarchical decomposition mechanism to construct a tree representation (one per image) which was coupled with a weighted sub-graph mining technique to generate a feature vector representation to which standard classification techniques could again be applied. Excellent results were produced. Sensitivity, specificity and accuracy rates of 99% and over were recorded. Further evaluation indicated that the results were statistically significant. Overall the tree representation was found to be the most effective. The excellent results obtained indicated that classification systems are ideally suited to large scale AMD screening processes. More specifically that the tree based approach produced the most effective results in comparison with the time series and tabular approaches also considered.

ACKNOWLEDGMENT

The work presented in this paper has partly been supported financially by the Ministry of Higher Education Malaysia (grant number RAG009-TK-2012) and Artificial Intelligence Research Unit, UMS.

REFERENCES

[1] L. Brandon and A. Hoover, "Drusen detection in a retinal image using multi-level analysis," in Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2003, pp. 618–625.

[2] D. Freund, N. Bressler, and P. Burlina, "Automated detection of drusen in the macula," in Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 61–64.

[3] C. Kose, U. Sevik, and O. Gencalioglu, "Automatic segmentation of age-related macular degeneration in retinal fundus images," Computers in Biology and Medicine, vol. 38, pp. 611–619, 2008.

[4] K. Rapantzikos, M. Zervakis, and K. Balas, "Detection and segmentation of drusen deposits on human retina: potential in the diagnosis of age-related macular degeneration," Medical Image Analysis, vol. 7, pp. 95–108, 2003.

[5] Z. Sbeh, L. Cohen, G. Mimoun, and G. Coscas, "A new approach of geodesic reconstruction for drusen segmentation in eye fundus images," IEEE Transactions on Medical Imaging, vol. 20, no. 12, pp. 1321–1333, 2001.

[6] D. Minassian and A. Reidy, "Future sight loss UK (2): an epidemiological and economic model for sight loss in the decade 2010-2020," Royal National Institute of Blind People, Tech. Rep., 2009.

[7] A. Loewenstein, "The significance of early detection of age-related macular degeneration," The Journal of Retinal and Vitreous Diseases, vol. 27, no. 7, pp. 873–878, 2007.

[8] A. Elsayed, F. Coenen, M. Garcia-Finana, and V. Sluming, "Classification of MRI brain scan data using shape criteria," in Annals of the British Machine Vision Association, 2011, pp. 1–14.

[9] C. K'ose, U. S. evik, and O. Genc, alio'glu, "A statistical segmentation method for measuring age-related macular degeneration in retinal fundus images," Journal of Medical Systems, vol. 34, no. 1, pp. 1–13, 2008.

[10] C. Agurto, E. Barriga, V. Murray, S. Nemeth, R. Crammer, W. Bauman, G. Zamora, M. Pattichis, and P. Soliz, "Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images," Investigative Ophthalmology & Visual Science, vol. 52, pp. 5862–5871, 2011.

[11] E. Chaum, T. Karnowski, V. Govindasamy, M. Abdelrahman, and K. Tobin, "Automated diagnosis of retinopathy by content-based image retrieval," Retina, vol. 28, no. 10, pp. 1463–1477, 2008.

[12] M. Mookiah, U. Acharya, J. Koh, V. Chandran, C. Chua, J. Tan, C. Lim, E. Ng, K. Noronha, L. Tong, and A. Laude, "Automated diagnosis of age-related macular degeneration using greyscale features from digital fundus images," Computers in Biology and Medicine, vol. 53, no. 2014, pp. 55–64, 2014.

[13] J. Cheng, D. Wong, X. Cheng, J. Liu, N. Tan, M. Bhargava, C. Cheung, and T. Wong, "Early age-related macular degeneration detection by focal biologically inspired feature," in Proceedings of 19th IEEE International Conference on Image Processing, 2012, pp. 2805–2808.

[14] R. Gonzalez and R. Woods, Digital image processing, R. Gonzalez and R. Woods, Eds. Pearson Prentice Hall, 2008.

[15] A. Osareh, "Automated identification of diabetic retinal exudates and the optic disc," Ph.D. dissertation, University of Bristol, UK, 2004.

[16] M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," Medical Image Analysis, vol. 9, pp. 179–190, 2005.

[17] S. Pizer, E. Amburn, J. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. Romeny, J. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Processing, vol. 39, no. 3, pp. 355–368, 1987.

[18] K. Zuiderveld, Contrast limited adaptive histogram equalization, ser. Academic Press Graphics Gems Series. Academic Press Professional, Inc., 1994, pp. 474–485.

[19] J. Soares, J. Leandro, R. C. Jr., H. Jelinek, and M. Cree, "Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification," IEEE Transactions on Medical Imaging, vol. 25, no. 9, pp. 1214–1222, 2006.

[20] M. Sofka and C. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," IEEE Transactions on Medical Imaging, vol. 25, no. 12, pp. 1531–1546, 2006.

[21] J. Kolodner, Case-based reasoning. Morgan Kaufmann, 1993.

- [22] M. Hijazi, F. Coenen, and Y. Zheng, "Retinal image classification for the screening of age-related macular degeneration," in The 30th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, 2010, pp. 325–338.
- [23] M. Hijazi, F. Coenen, and Y. Zheng, "A histogram approach for the screening of age-related macular degeneration," in Medical Image Understanding and Analysis 2009. BMVA, 2009, pp. 154–158.
- [24] A. Mahfouz and A. Fahmy, "Ultrafast localisation of the optic disc using dimensionality reduction of the search space," in proceedings of Medical Image Computing and Computer-Assisted Intervention, 2009, pp. 985–992.
- [25] W. Hsu, S. Chua, and H. Pung, "An integrated color-spatial approach to content-based image retrieval," in Proceedings of the Third International Conference on Multimedia, 1995, pp. 305–313.
- [26] B. Ooi, K. Tan, T. Chua, and W. Hsu, "Fast image retrieval using color-spatial information," The International Journal of Very Large Data Bases, vol. 7, no. 7, pp. 115–128, 1998.
- [27] D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in AAAI Workshop on Knowledge Discovery in Databases, 1994, pp. 229–248.
- [28] C. Myers and L. Rabiner, "A comparative study of several dynamic time warping algorithms for connected word recognition," The Bell System Technical Journal, vol. 60, no. 7, pp. 1389–1409, 1981.
- [29] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. Ratanamahatana, "Fast time series classification using numerosity reduction," in Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 133–1040.
- [30] L. Fritsche, A. Schlaefer, K. Budde, K. Schroeter, and H. Neumayer, "Recognition of critical situations from time series of laboratory results by case-based reasoning," Journal of the American Medical Informatics Association, vol. 9, pp. 520–528, 2002.
- [31] T. Rath and R. Manmatha, "Word image matching using dynamic time warping," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, pp. 521–527.
- [32] M. Hijazi, F. Coenen, and Y. Zheng, "Data mining techniques for the screening of age-related macular degeneration," Journal of Knowledge Based Systems, vol. 29, pp. 83–92, 2012.
- [33] Y. Zheng, M. Hijazi, and F. Coenen, "Automated "disease/ no disease" grading of age-related macular degeneration by an image mining approach," Investigative Ophthalmology and Visual Science, vol. 53, no. 13, pp. 8310–8318, 2012.
- [34] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Transactions on Medical Imaging, vol. 8, no. 3, pp. 263–269, 1989.
- [35] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, pp. 610–621, 1973.
- [36] Y. Chang and C. Lin, "Feature ranking using linear SVM," in WCCI2008, 2008, pp. 53–64.
- [37] C. Jiang, F. Coenen, and M. Zito, "Frequent sub-graph mining on edge weighted graphs," in Proceedings of 12th Int. Conf. on Data Warehousing and Knowledge Discovery, 2010, pp. 77–88.
- [38] C. Jiang, F. Coenen, R. Sanderson, and M. Zito, "Text classification using graph mining-based feature extraction," in Knowledge-Based Systems, vol. 23, 2010, pp. 302–308.
- [39] M. Hijazi, C. Jiang, F. Coenen, and Y. Zheng, "Image classification for age-related macular degeneration screening using hierarchical image decompositions and graph mining," in Proceedings of ECML PKDD 2011 Machine Learning and Knowledge Discovery in Databases, D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, Eds., vol. 2. Springer, 2011, pp. 65–80.
- [40] F. Golchin and K. Paliwal, "Quadtree-based classification in subband image coding," Digital Signal Processing, vol. 13, pp. 656–668, 2003.
- [41] X. Yan and J. Han, "gSpan: graph-based substructure pattern mining," in IEEE Conference on Data Mining, 2002, pp. 721–724.
- [42] C. Jiang and F. Coenen, "Graph-based image classification by weighting scheme," in AI2008, 2008, pp. 63–76.
- [43] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," Journal of Machine Learning Research, vol. 7, pp. 1–30, 2006.
- [44] D. Sheskin, Handbook of parametric and nonparametric statistical procedures, 3rd ed. Chapman & Hall/ CRC, 2004.
- [45] S. Szalberg, "On comparing classifiers: pitfalls to avoid and a recommended approach," Data Mining and Knowledge Discovery, vol. 1, pp. 317–328, 1997.
- [46] J. Tukey, "Comparing individual means in the analysis of variance," Biometrics, vol. 5, pp. 99–114, 1949.