

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*

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.

Key words: Image classification, spatial-histograms, image decomposition,

*Corresponding author

Email addresses: m.ahmad-hijazi@liverpool.ac.uk (Mohd Hanafi Ahmad Hijazi), coenen@liverpool.ac.uk (Frans Coenen), yalin.zheng@liverpool.ac.uk (Yalin Zheng)

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 mechanisms 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 $3 \times$

3 grid describing 9 regions, while the second applied an angular partitioning describing 8 regions. Both approaches produced better results than using colour histograms in isolation. The second proposed AMD screening technique 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 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 subdivision 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 Figure 1(a), the macula is a small area approximately 5 to 5.5mm in diameter centered on the fovea. The fovea is a concave central retinal depression approximately 1.5mm 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 ($63\mu\text{m}$ in diameter) or a few medium (63 to $124\mu\text{m}$) sized drusen or retinal pigmentary abnormalities. The presence of at least one large ($124\mu\text{m}$) 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 center of the macula. *Choroidal neovascularisation* characterizes 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. Figure 1(a) shows an example of normal retinal image with the macula circled. A retina image that features drusen is given in Figure 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, 10, 11, 12, 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 segmen-

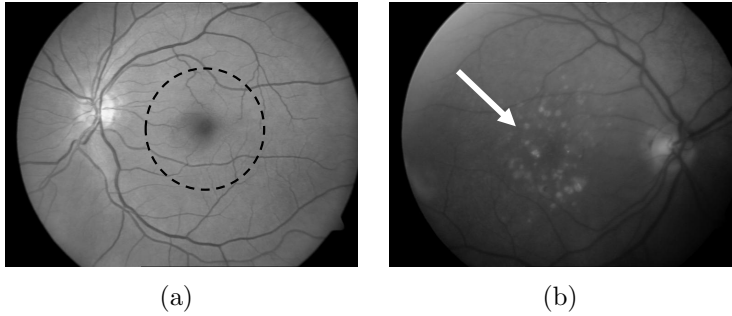


Figure 1: Illustration of fundus images in grayscale: (a) Normal and (b) AMD.

tation coupled with image enhancement approaches [11, 12, 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. However, all the previous work is focused on the detection of drusen using segmentation, a challenging task given the inconsistent visual appearance of drusen and other lesions. The clarity, colour, luminosity and texture of images are affected by several factors during the image acquisition process, such as involuntary eye movement and the media opacity of the subject.

The distinction between the work described here and previous approaches is that we make no attempt to locate and isolate (segment) drusen within retinal images. Instead we propose two techniques whereby retina images can

be classified that obviate the need for the specific identification of drusen. The first extends the use of individual colour channel histograms [15] to a spatial-histogram based approach. Spatial-histograms extend the concept of simple colour histograms by including spatial pixel information [6, 16, 17] and have been shown to perform well in region-based tracking [6], object detection [17] and image retrieval [16]. We described the generated histograms using X-Y curves. The proposed classification is then conducted using a CBR approach. CBR is a well established AI technique with an associated, well established, body of literature. Recommended reference works include [18] and [19]. A specific application of CBR to signals and images have been comprehensively described in [20]. CBR has also been widely used in medical applications [21], as well as other domains. Recent work on the utilisation of CBR for classification in the medical domain can be found in [22]. One particular issue for CBR is the similarity measures used to identify the “most similar cases” [20]. In the context of the work described here Dynamic Time Warping (DTW) [23, 24] is used. DTW is a time series analysis approach to measuring the distance between two different curves (i.e. histograms).

The second approach is founded on the concept of hierarchical decomposition applied to retina images so that each image is described using a tree structure. Hierarchical data structures have been widely applied in various domains, such as image segmentation [25], image coding [26] and image classification [27]. One main advantage of this type of data structure is its ability to represent different parts of images using different levels of granularity. Detailed parts of the image are “tessellated” down to a high level of granularity (depth) while less detailed parts are tessellated to a fairly coarse level of granularity. This gives rise to an efficient representation with associated improved execution times [28]. The most commonly used image decomposition techniques result in a *quadtree* representation. A given “space” is initially split into four quadrants. Each quadrant is then decomposed further into sub-quadrants, and so on, until a certain level of granularity is reached, or homogeneous sub-quadrants are arrived at.

The term “graph mining” refers to the collection of data mining techniques that are used to find interesting patterns in tree/graph data. From the literature two broad types of graph mining can be identified: transaction and single graph mining [7]. Transaction graph mining is directed at the discovery of frequently occurring sub-graphs that exist across a collection of graphs. The process is similar to the processes found in Association Rule Mining (ARM). Thus we can apply Frequent Sub-Graph (FSG) mining

techniques to our tree represented images to discover frequently occurring sub-trees which can then be used as features with which our retina images can be encoded. The application of FSG algorithms entails a significant computational overhead. To reduce this overhead Weighted FSG (WFSG) can be applied, the objective being to focus on the identification of those frequent sub-graphs that are likely to be the most significant according to some weighting scheme.

4. Image Pre-processing

This section describes the nature of the image pre-processing that we applied in order to represent the retinal images in a meaningful form ready for image mining. The preprocessing comprised:

1. **Image Enhancement:** Normalisation and enhancement of the image contrast.
2. **Object Segmentation:** Identification of blood vessels.
3. **Noise Reduction:** Removal of the identified blood vessel pixels from the retina images.

Image enhancement and the segmentation of blood vessels (for noise removal) are described in more detail in the following two sub-sections respectively.

4.1. Image Enhancement

The quality of the retinal images is often severely affected by factors such as colour variance and non-uniform illumination [29, 30], which are difficult to control. In the context of the proposed AMD classification this will lead to the introduction of inaccuracies, and hamper the associated identification and localisation of retinal common structures such as retinal blood vessels.

The adopted image enhancement comprised colour and illumination normalisation and contrast enhancement. Colour normalisation was applied first, followed by illumination normalisation and then contrast enhancement to increase the “visibility” of the main retinal anatomy (i.e. blood vessels). Due to the colour variation between different retinal images, colour normalisation must be performed prior to image enhancement. To normalise the colours featured in retinal images a *histogram specification* approach was applied [31]. First, a reference image that represents the best colour distribution and contrast is selected by a trained clinician. Then, the Red-Green-Blue

(RGB) colour histograms of the reference image are generated. Finally, the RGB histograms of other images are extracted and each of these histograms is tuned to match the reference image histograms. Figure 2(b) shows the normalised colour fundus retinal image (the colour image has been converted to grayscale only for illustrative purposes) of the original image (Figure 2(a)).

Once the colour is normalised, illumination normalisation is applied so as to reduce the luminosity variations. An approach, to estimate the luminosity and contrast variability of the retinal image based on the image background colour, proposed by Foracchia et al. [29] was adopted. This approach estimates the original image pixels, $\bar{I}(x, y)$, as follows:

$$\bar{I}(x, y) = \frac{I(x, y) - \bar{L}(x, y)}{\bar{C}(x, y)}, \quad (1)$$

where I is the observed image, and \bar{L} and \bar{C} are the estimations of luminosity and contrast, calculated in the neighbourhood N of each pixel. One drawback of this approach is that drusen that are larger than the window size N , used for the estimation, are smoothed in the normalisation process. However, the authors found that this disadvantage could be limited by setting the \bar{C} value to 1, thereby excluding the contrast estimation. Contrast normalisation was then conducted using Contrast Limited Adaptive Histogram Equalisation (CLAHE) as described in [32]. Figures 2 (c) and (d) show the illumination normalised and contrast enhanced retinal image given in Figure 2(a).

4.2. Object Segmentation

The presence of retinal anatomies, such as blood vessels and the optic disc, sometimes hampers the detection of drusen. The authors' own experiments have indicated that the removal of blood vessel pixels from retina images can improve classification accuracy [15]. This has also been observed more generally by other researchers in the field ([12, 33, 13]). We refer to the process of removing specific retinal anatomy as "noise removal".

To segment the retinal blood vessels 2-D Gabor wavelet filters [34] were applied. A pixel is classified as *vessel* or *non-vessel* by means of a Bayesian classifier with a class-conditional probability density function, generated using the Gaussian mixture model. As a result a "retinal vessels" binary representation is generated for each image which is then applied as a "mask" to the enhanced retinal images and consequently the blood vessels pixel values replaced with a "null" value. The outcomes of these processes is visualised in Figures 2 (e) and (f).

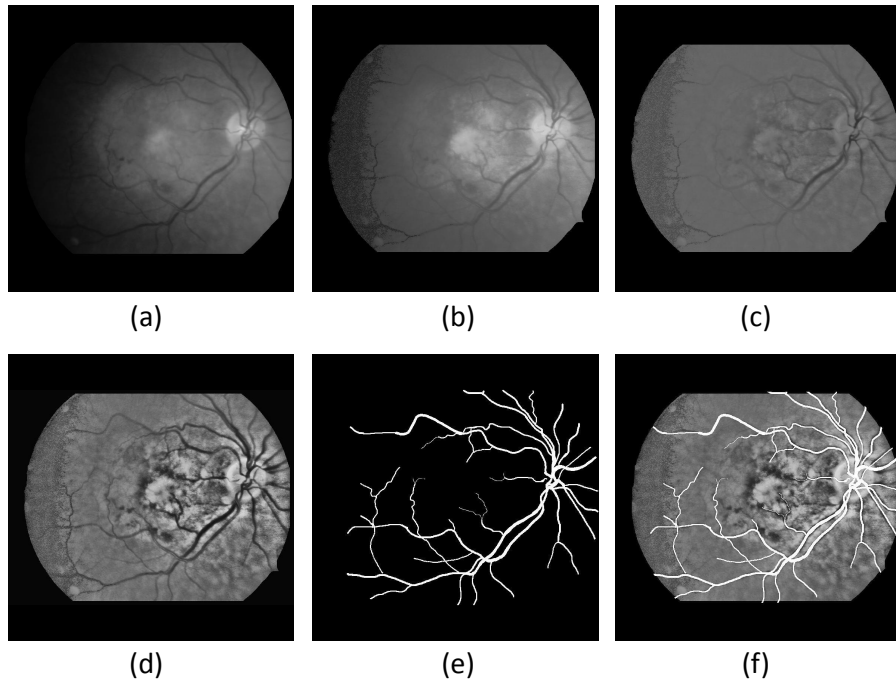


Figure 2: (a) Original image, (b) colour normalised, (c) illumination normalised, (d) contrast enhanced, (e) retinal blood vessels segmented, and (f) noise (blood vessels) removed.

The optic disc was however left untouched as experiments conducted by the authors, reported in [15], indicated that removal of the optic disc only resulted in increased accuracy with respect to a minority of retina images and decreased accuracy with respect to the majority.

5. AMD Screening Using Spatial Histograms

An overview of the proposed retinal image AMD classification mechanism founded on the use of spatial histograms is presented in this section (the hierarchical decomposition based technique is described in the following section). The hierarchical histogram based approach can be viewed as comprising two stages: Case Base (CB) generation and image classification.

CB generation comprises three sub-stages: (i) image preprocessing, (ii) histogram generation and (iii) feature selection. The image preprocessing that was undertaken has already been discussed in Section 4. The next step is to generate the spatial-histograms. In order to make the representation more tractable, colour quantisation was applied to the preprocessed images to reduce the overall dimensionality (number of colours). To generate the

histograms the quantised colour retinal images were first partitioned into N *regions* and spatial-histograms extracted for each region. The idea here was that the presence of drusen is often regionalised and consequently we may be more interested in some regions than others. Section 5.1 gives more detail of the technique used to generate the spatial-histograms. During feature selection the spatial-histograms (regions) that feature the best discriminatory power (in the context of AMD classification) are identified. The regions are ranked according to their discriminatory power and the top T selected. This process also ensured that the size (number of pixels) of each region/histogram does not bias the resulting classification. The feature selection was conducted using a *class separability* measure which was applied to the collection of histograms representing each retina image and the most appropriate histograms selected. The selected spatial-histograms were then combined and stored in the form of time series curves (one per image). The feature selection process is discussed in further detail in Section 5.2. The image classification task is detailed in Section 5.3.

5.1. Spatial Histogram Generation

Colour histograms have been widely used as a simple way of representing images for object identification and retrieval [35, 36]. The main advantage is their robustness against object changes in terms of shape and position within images. The main disadvantage is the loss of spatial information between pixels and colours, thus images with similar histograms may have very different appearances [16, 17]. In some images, the colour distribution of pixels at different sections of an image may be an essential feature that should be included in the image representation. In the context of AMD classification there are a significant number of cases where the AMD images have almost similar colour histograms to the normal ones. The fact that drusen pixel colours are very similar to the colours of pixels adjacent to the retinal blood vessels boundaries (as well as the optic disc), may thus lead to classification errors. A spatial-histogram [4, 5] representation is therefore suggested.

The spatial information contained within an image can be captured by preserving the objects texture and shape using templates [17]; as well as by partitioning the image into regions, based on the chosen colour values, and recording the regions location for each of the chosen colours [4]. The utilisation of texture and shape to extract spatial information is hampered by the nature of the AMD featured images where no common textures and

shapes exist, other than the main retinal structures. Therefore, a method to generate colour distribution for each region [16] has been applied in the work described here as it is conjectured that similar regions, in two different classes of retinal image, will have different colour distributions. The generation of spatial-histograms consisted of several steps. First, the number of colours was reduced to make the computational cost more feasible. The minimum variance quantisation technique [37], with dithering [38] (implementation using Matlab¹ function *rgb2ind*), was used to reduce the image colours to C colours. A careful selection of C value is essential as it will affect the quality of the generated histograms (as demonstrated in Section 7). The colour quantisation was applied on the global colour space, instead of the local colour space, in order to standardise the colour mapping. Thus, all images referenced a similar colour map.

Once the colour quantisation was complete each image was partitioned into N similar sized regions, $R = \{r_1, r_2, \dots, r_N\}$, and a spatial-histogram generated for each. Two image partitioning approach were implemented. A grid based approach that partitioned an image into a (3×3) grid to give 9 equal sized regions; and an angular partitioning, with 8 radii values (45° , 90° , 135° , 180° , 225° , 270° , 315° and 360°) to partition each image into 8 regions. Figure 3 gives an example of each image partitioning approach. The black coloured pixels surrounding the retina, as well as the retinal blood vessels were omitted from the generation of the spatial-histograms. The set of spatial histograms for a given image m is defined as:

$$h_m = \{sh_1^m, sh_2^m, \dots, sh_N^m\} \quad (2)$$

where sh_n^m is the spatial-histogram generated for region n , ($1 \leq n \leq N$) in image m with C bins. The histogram value for colour c in histogram sh_n^m is then given by:

$$sh_n^m(c) = \alpha \quad (3)$$

where α is the c -th bin count in region t of image m ($0 \leq c < C$). The size of each image spatial-histograms, h_m , for an image m is equivalent to $C \times N$; the number of colours, C , multiplied by the number of regions, N . The complete set of histograms representing an image set is then defined as $H = \{h_1, h_2, \dots, h_M\}$, where M is the number of images.

¹<http://www.mathworks.com>

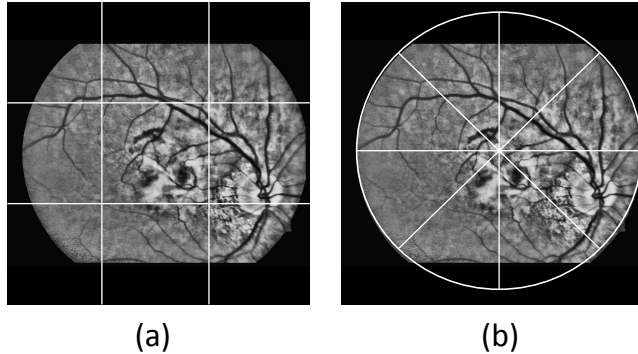


Figure 3: Retinal image partitioning using: (a) 3 x 3 image tessellation and (b) angular partitioning.

5.2. Feature Selection

Feature selection is a process to reduce the number of features contained in a feature space by removing irrelevant or redundant features [39, 40, 41]. By selecting only those features that have a strong discriminatory power between classes, the computational cost of classification can be considerably reduced while at the same time maximising classification accuracy [39]. Common feature selection techniques [40, 41] include the χ^2 measure, mutual information, Odds Ratio and Principal Component Analysis.

With respect to the AMD screening process described here a class separability method [39] that estimates the effectiveness of a features ability to distinguish between classes, using the Kullback-Leibler (KL) distance measure, was adopted. This was a two stage process. First an average *signature*, γ_n , histogram was generated for each region with respect to each class as follows:

$$\gamma_n^a = \frac{1}{p} \sum_{j=1}^p sh_n^j \quad (4)$$

where n is the region identifier, a is a class label and p is the number of training set images labelled as class a . The *class separability*, $dist_n$, is then calculated by:

$$dist_n = \sum_{a=1}^d \sum_{b=1}^d \delta_n(a, b) \quad (5)$$

where d is the number of classes and $\delta_n(a, b)$ is the KL distance, between histograms of γ_n , corresponding to classes a and b , described as:

$$\delta_n(a, b) = \sum_{i=1}^C p_n(\gamma_n^a(i)) \log \left(\frac{p_n(\gamma_n^a(i))}{p_n(\gamma_n^b(i))} \right) \quad (6)$$

where C is the number of bins (colours) in the histogram, and $p_n(\gamma_n^a(i))$ is the probability that the n -th feature takes a value in the i -th bin of the signature spatial-histogram γ_n given a class a . The probability, p_n was calculated by dividing each bin count of γ_n by the total number of elements in γ_n .

The features are then sorted in descending order of $dist_n$; the top T features with the highest $dist_n$ provided the best separation between classes and were therefore selected. However, the selection of the value for T is domain dependent and might not work well if only one single T value is used (as demonstrated in Section 7). The other regions were omitted from further processing. Thus, the size of h_m has been reduced to only $C \times T$. These histograms then make up the CB for the CBR process.

5.3. Retinal Image Classification using CBR and DTW

Given a new set of potential AMD images these may be classified using the CB developed as described in the foregoing subsections. As noted previously, the histograms in the CB may be viewed as time series. Similarity checking may therefore be conducted using time series analysis techniques. For the AMD screening a Dynamic Time Warping (DTW) technique [23, 24] was adopted. DTW is a time series analysis technique that measures the distance between two time series through the generation of a *warping path* between these sequences. Given two time series, $T = \{t_1, t_2, \dots, t_m\}$ and $\bar{T} = \{\bar{t}_1, \bar{t}_2, \dots, \bar{t}_n\}$, a matrix of size $m \times n$ will be formed. The distance between t_i and \bar{t}_j , $d(t_i, \bar{t}_j)$, where $0 \leq i < m$ and $0 \leq j < n$ for all i and j is computed using the Euclidean distance similarity measure (other similarity measure methods can also be applied). The minimal warping path is computed by summing up the minimal d for each matrix grid point thus giving a distance between T and \bar{T} . More details of the DTW approach with respect to retinal image classification can be found in [42].

6. AMD Screening Using Hierarchical Decomposition

As in the case of the spatial histogram based technique, the hierarchical decomposition based technique commences with retinal image cleaning (as

described in Section 4). The process then proceeds with the decomposition of the image, this is described in detail in Section 6.1 below; the result is a collection of tree represented images (one per image). Next WFSG is applied to the tree represented data (this is described in Section 6.2). The identified frequent sub-trees are then used to define the elements of a feature space that is used to encode the individual input images in the form of feature vectors itemising the frequent sub-graphs that occur in each image. The adopted feature selection process is described in Section 6.3. Once the feature vector representation has been generated we can apply established classification techniques (see Section 6.4).

6.1. Image Decomposition

As noted above hierarchical image decomposition is a well established technique [26, 28, 25]. The distinguishing and novel feature of the proposed approach is that the partitioning is conducted in an interleaving angular and circular manner. During angular partitioning the decomposition is defined by two radii describing a minor arc on the circumference of the image “disc”. Circular decomposition is defined by a pair of arcs radiating out from the center of the retina disc. Individual regions identified during the decomposition are thus delimited by a pair of radii and a pair of arcs. Figure 4(a) shows an example of a partitioning that might be applied to an image; Figure 4(b) presents the associated Tree storage structure. The proposed image decomposition proceeds as follows. First we identify the retina region of interest (ROI) by using the image background to exclude areas outside of the field of view of the fundus (the retina disc) and the blood vessel pixels within the fundus. The actual decomposition then 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 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 the circular decomposition. The decomposition continuous in this manner.

The RGB (red, green and blue) colour model is used to extract the pixel intensity values, which means each pixel will have three intensity values associated with it. Thus initially three trees are generated which are then merged. As already noted the hierarchical image partitioning commences with circular partitioning. On the following iteration angular partitioning is applied.

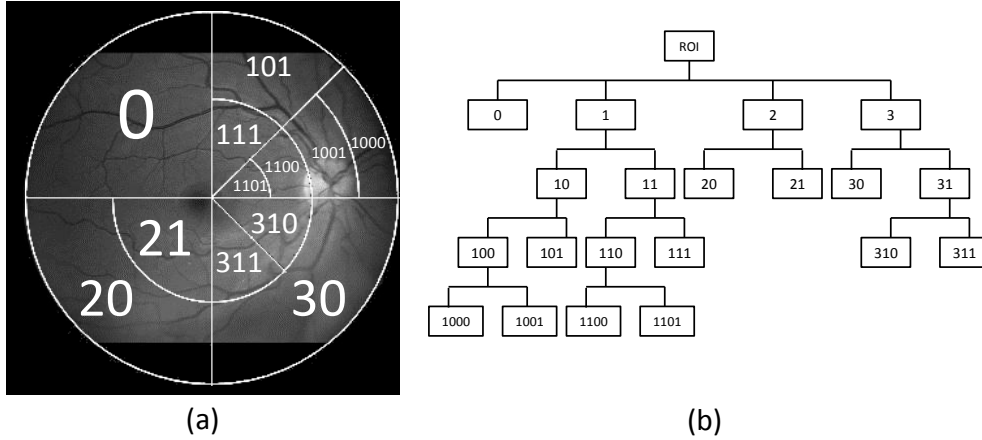


Figure 4: An example of (a) circular and angular image decomposition, (b) The associate tree data structure.

Both circular and angular partitioning will then be called alternately until uniform regions are arrived at or a maximum “depth” of decomposition is reached. Throughout the process the tree data structures are continuously updated. At the end of the decomposition process the three trees are merged to form a single tree.

6.2. Weighted Frequent Sub-Graph Mining

On completion of image decomposition the input image set is represented as a collection of trees. Each tree is defined as follows: $T = (V, E, L_V, L_E, u)$; where V and E are sets of vertices and edges respectively, L_V and L_E are sets of labels for vertex and edge respectively, while u defines a label mapping function.

In [7] it was suggested that for many applications, such as image mining, some vertices have more significance associated with them than others. In the case of the hierarchical decomposition described in this paper, vertices that feature a significant difference in colour intensity, in comparison to their parent, were deemed to be more significant (than the parent). The underpinning philosophy is that normal retinal background pixels have a similar colour intensity, while a significant difference in intensity is likely to indicate the presence of drusen. A weighting scheme was therefore applied to the tree representation so as to enhance the quality of the information contained within it. Thus, in the tree representation, the strength of each vertex $v \in V$

was weighted by the average colour intensity value of the region represented by the vertex, and the strength of each edge $e \in E$, e_w was weighted by:

$$e_w = \sqrt{(I_{par} - I_v)^2} \quad (7)$$

where I_v is the average colour intensity value for vertex v and I_{par} is the average colour intensity value of its parent.

By adding vertex and edge weights into the tree representation, the weighted tree representation was able to capture more image information than comparable unweighted approaches. A Weighted Frequent Sub-Graph (WFSG) mining algorithm, an extension of the well-known gSpan algorithm [43], was then applied to the tree data so as to identify frequently occurring trees within the dataset. The algorithm operated in a similar manner to that described in [7], but utilised both vertex and edge weightings. In the context of WFSG mining, a sub-graph pattern g is considered to be “interesting”, if it satisfies the following two conditions:

$$(\mathbf{C1})V_{wr} \times sup(g) \geq \sigma, \quad (\mathbf{C2})E_{wr} \geq \lambda \quad (8)$$

Where: V_{wr} denotes the vertex weighting, $sup(g)$ denotes the support (i.e. frequency) of g , and σ denotes a *minimum support* threshold, E_{wr} denotes the edge weighting, and λ denotes a *minimum weight* threshold. Both the V_{wr} and E_{wr} are computed using a similar scheme to that described in [7].

The number of patterns discovered by the WFSG mining algorithm is thus determined by both the σ and λ values. According to initial experiments conducted by the authors, relatively low σ and λ threshold values are required, in order to extract a sufficient number of patterns (frequent sub-graphs). However, setting low threshold values still tends to results in a substantial number of patterns. Therefore a feature selection was applied as described in the following subsection.

6.3. Feature Selection

To reduce the number of identified frequent sub-graphs to a manageable number a feature selection strategy was applied so as to identify those sub-graphs that displayed a strong discriminatory power (i.e. sub-graphs that were likely to produce good classification results). A feature ranking mechanism was therefore applied that used linear Support Vector Machine (SVM) weights to rank features as proposed in [44]. The main advantage of this approach is its implementation simplicity and effectiveness in determining

relevant features. The identified sub-graphs were ranked by first calculating their weights using the L2-regularized L2-loss SVM model, and then sorting them in descending order according to their absolute value [45, 46]. The selection of only the top K sub-graphs for classification then concluded the feature selection process.

6.4. Retinal Image Classification

The final stage of the proposed hierarchical decomposition based AMD classification process is the classification stage. The identified top K frequent sub-graphs were used to define a feature space. Each image was then defined, in terms of this feature space, using a feature vector representation. Any appropriate classification technique could then be applied. In the reported experiments (Section 7) a Support Vector Machine (SVM) [47] was adopted. This was chosen because it is frequently acknowledged to be one of the most effective classification method in machine learning. The SVM used in the evaluation was built using LibSVM [48] and a radial basis function kernel.

7. Experimental Setup and Evaluation

To evaluate the proposed AMD screening techniques a collection of 161 retinal images, acquired as part of the ARIA² project, were used. The collection was manually pre-labelled, and included 101 AMD images and 60 non-AMD images. The experiments described in this section were designed to evaluate the performance of the proposed approaches. Three metrics were used for the evaluation: *Specificity*, *Sensitivity* and *Accuracy*. All experiments were conducted using Tenfold Cross Validation (TCV) whereby the dataset was randomly divided into equal sized “tenths”; and on each TCV iteration, one tenth was used as the test set while the remainder was used as the training set. The objectives of the experiments may be summarised as follows:

1. Spatial Histogram Based Technique.

- a) **Number of Bins Parameter (C):** To determine the minimum number of bins for the histograms, with respect to colour quantisation, such that classification accuracy would not be adversely affected.

²http://www.eyecharity.com/aria_online

Table 1: Classification results for a range of C values (number of colour quantisation output bins)

C	Specificity (%)	Sensitivity (%)	Accuracy (%)
32	58	70	65
64	63	73	67
128	54	79	68
256	60	73	68

- b) **Number of Regions Parameter (T):** To determine the most appropriate setting for the T parameter, the threshold that determines the number of regions to be included in the final representation during feature selection.
2. **Hierarchical Decomposition Based Technique.**
- a) **Depth of Decomposition Parameter (D_{max}):** To analyse the effect that the depth of decomposition had on classification accuracy.
 - b) **Number of Selected Frequent Sub-Graph Parameter (K):** To analyse the performance of the hierarchical decomposition technique with respect to the size of the feature space (the number of frequent subgraphs selected to be included in the feature space).
3. **Overall Comparison:** Overall comparison of the operation of the two proposed techniques.

The experiments are described in more detail in the following five subsections.

7.1. Number of Bins Parameter (C)

The first set of experiments considered the number of output bins required for colour image quantisation with respect to the spatial histogram based technique. The aim was to determine the minimum number of required bins while at the same time maximising classification accuracy. Experiments using C values of 32, 64, 128 and 256 were conducted (with $N = 1$). Table 1 shows the classification results obtained. The results clearly indicate that the higher the C value, up to $C = 128$, the better the classification accuracy. This was expected, as low numbers of colour bins will tend to group different coloured pixels in to the same bin, and consequently reduce the discriminative power of the colour representation.

7.2. Number of Regions Parameter (T)

The results presented in the foregoing subsection were generated by setting the number of regions parameter T to one. The experiment described in this sub-section considered the effect on classification accuracy according to the number of regions considered (the T parameter). Using the proposed grid and angular partitioning 9 and 8 regions were generated respectively. Spatial-histograms were then generated as described in Section 5. For the experiments described here C parameter values of 32, 64 and 128 were used; $C = 256$ was omitted from further analysis as it did not give any significant improved performance over $C = 128$, and also because it would introduce a significant computational overhead. The retinal image classification was performed using the top- T regions that had the highest discriminatory capability.

Table 2: Classification results with $C = 32$, a range of T values and grid (*grid*) and angular partitioned (*ang*)

T	SH-dimension	Specificity (%)		Sensitivity (%)		Accuracy (%)	
		<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>
1	32	61	59	78	67	71	65
2	64	63	65	68	71	66	68
3	96	65	61	70	77	67	69
4	128	68	63	75	77	72	70
5	160	63	66	77	80	70	74
6	192	67	65	73	77	71	71
7	224	61	65	72	74	68	70
8	256	61	63	73	71	69	68
9	288	65	-	78	-	73	-

Tables 2, 3 and 4 compare the operation of the spatial histogram based approach using various combinations of T and C parameter values. In the tables the *SH-dimension* column indicates the total number of bins (dimensions) used for the spatial-histogram representation (calculated by multiplying the C parameter by the T parameter). In Table 2 (32 bins) the best results were obtained using $T = 5$ for angular partitioning, and $T = 9$ for grid partitioning, with an overall accuracy of 74%. Different results are shown in Table 3 (64 bins) with the best overall accuracy of 72% when $T = 2, 3$ and 7 for angular partitioning, while grid partitioning produced the best result with $T = 6$. Table 4 shows that the best performance was obtained with $T = 5$

Table 3: Classification results with $C = 64$, a range of T values and grid (*grid*) and angular partitioned (*ang*)

T	SH-dimension	Specificity (%)		Sensitivity (%)		Accuracy (%)	
		<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>
1	64	56	63	71	72	64	69
2	128	61	70	74	75	70	72
3	192	54	65	74	77	65	72
4	256	63	60	74	77	69	69
5	320	65	63	74	75	70	70
6	384	63	60	77	76	72	69
7	448	63	67	73	77	69	72
8	512	63	69	74	75	70	71
9	576	61	-	77	-	70	-

for grid partitioning with an overall accuracy of 74%. The best specificity of 70% was recorded with $T = 2$ and 64 colour bins (angular partitioning), and the best sensitivity of 86% with $T = 5$ and 128 colour bins. The results reported in Table 4 contains only six T values (1 to 6) as the machine memory required for the classification process increases quadratically with the size of the colour bins. Thus the process was stopped at $T = 6$.

Table 4: Classification results with $C = 128$, a range of T values and grid (*grid*) and angular partitioned (*ang*)

T	SH-dimension	Specificity (%)		Sensitivity (%)		Accuracy (%)	
		<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>	<i>grid</i>	<i>ang</i>
1	128	52	48	78	77	65	66
2	256	44	56	81	76	68	69
3	384	52	48	84	79	70	68
4	512	58	46	84	78	72	67
5	640	56	44	86	80	74	65
6	768	50	46	81	84	69	68

7.3. Depth of Decomposition Parameter (D_{max})

Table 5 shows the performances of the proposed hierarchical decomposition approach when using three different levels of decomposition (values for D_{max}). Feature selection was not applied in these experiments. F denotes the size of the feature space in terms of the number of identified frequent

sub-trees, while *Sens*, *Spec* and *Acc* refer to sensitivity, specificity and accuracy. Each σ value was tested against a range of λ values (20, 40, 60 and 80), however in the table (because of space limitations) only the best performing λ value associated with each σ value is recorded. Inspection of Tables 5 indicates that the best accuracy was achieved using $D_{max} = 5$ (75%). The best sensitivity and specificity that were obtained were 100% (for all shown D_{max} values) and 63% ($D_{max}=5$) respectively. The best sensitivity and specificity occurred using different σ and λ .

Table 5: TCV classification results obtained using different values for D_{max}

σ (%)	$D_{max} = 5$					$D_{max} = 6$					$D_{max} = 7$				
	λ	F	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	λ	F	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	λ	F	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>
10	20	1177	82	63	75	20	3762	86	55	74	40	36540	100	15	66
20	20	656	86	33	66	20	3089	94	28	70	20	25234	89	50	74
30	20	403	84	52	72	60	1336	92	23	66	20	13540	83	47	70
40	80	126	99	5	64	20	1144	94	17	65	20	8241	92	23	66
50	80	126	99	5	64	80	511	92	13	63	20	5341	95	20	67
60	20	140	100	3	64	80	511	92	13	63	80	3235	99	7	65
70	20	105	100	3	64	20	410	98	8	65	20	2472	100	3	64
80	20	80	99	7	65	20	280	100	0	63	20	1610	94	20	66
90	20	56	98	7	64	20	188	100	0	63	20	949	100	7	65

7.4. Number of Selected Frequent Sub-Graph Parameter (K)

Tables 6 shows the performances of the proposed hierarchical decomposition approach with respect to different values of K . Recall that the size of the feature space was determined by selecting only the top K features defined as a percentage (P) of $|F|$ where F is the set of features. Experiments using a variety of P values were conducted, however, only the results using $D_{max} = 7$ and P values of 0.05, 0.2 and 0.6 are presented in the table because these produced the best classification performances. Inspection of the tables indicates how the performance changes as the size of the feature space is reduced.

Inspection of Table 6 demonstrates that the best results were obtained using lower numbers of features, where $P = 0.05$ (equating to $K = 1262$). A 100% best accuracy was obtained. The highest sensitivity and specificity was also 100%. All of the best results were generated using various values of σ and λ . The results shows that accuracy increased as the σ value decreased for all K values. The results produced show that, as might be expected, the larger the feature space the better the classification performance. It should be noted that the results reported in Section 7.3, where feature selection

Table 6: TCV classification results using $D_{max} = 7$ and a range of K values

σ (%)	$P_{0.05}$					$P_{0.2}$					$P_{0.6}$				
	λ	K	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	λ	K	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	λ	K	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>
10	80	162	100	0	63	20	13101	100	80	93	20	39303	97	65	85
20	20	1262	100	100	100	60	2085	100	100	100	20	15141	98	82	92
30	20	677	99	97	98	60	2085	100	100	100	20	8125	98	78	91
40	20	412	100	2	63	20	1678	99	72	89	20	4945	96	42	76
50	80	162	100	0	63	20	1068	100	17	69	20	3205	96	27	70
60	80	162	100	0	63	80	647	100	0	63	20	2175	99	7	65
70	20	124	100	0	63	20	495	100	0	63	20	1484	100	2	63
80	20	81	100	5	65	20	322	98	25	71	20	967	97	20	68
90	20	48	100	0	63	20	190	100	3	64	20	570	100	7	65

Table 7: Comparison of the proposed Spatial Histogram and Hierarchical Decomposition based AMD screening approaches

Approach	Features	Sensitivity	Specificity	Accuracy
Spatial Histogram	640	86	56	74
Hierarchical decomposition	1262	100	100	100

was not applied are not as good as those reported in Table 6. This clearly demonstrates that feature selection improves the classification performance.

7.5. Overall Comparison

Table 7 compare the best classification results obtained using both the spatial histogram based approach and the hierarchical decomposition approach. The hierarchical decomposition based approach produced much better results than the spatial histogram based approach with a recorded accuracy of 100% with respect to the ARIA dataset used in the evaluation. Indicating that the hierarchical decomposition based approach clearly outperforms the spatial histogram based approach. Best results using the histogram approach were produced using $T = 5$ and $C = 128$. Best results using the decomposition approach were produced using $D_{max} = 7$, $P = 0.05$ and $K = 1262$ (based on the lowest value of K).

8. Conclusion

Two approaches to retinal image classification for AMD screening have been described. In the first approach the images were represented in the form of spatial-histograms that stored the colour information of the images while at the same time maintaining the spatial information of each colour

value. A feature selection strategy, to identify regions in an image that have strong discriminative power to separate classes, was applied to remove irrelevant features, as well as reducing the overall computational cost. The second approach was founded on a novel hierarchical circular and angular image decomposition technique. The decomposition resulted in a tree data structure to which a WFSG mining technique was applied so as to identify frequent occurring sub-trees. The generated frequent sub-trees were then used to recast the input data (the training set) into a feature vector representation. A classifier could then be built using this feature vector representation as the input data. For evaluation purpose the proposed approaches were applied to publicly available retinal fundus images. A 100% accuracy, sensitivity and specificity was produced using the hierarchical decomposition technique.

References

- [1] R. D. Jager, W. F. Mieler, J. W. Mieler, Age-related macular degeneration, *The New England Journal of Medicine* 358 (24) (2008) 2606–2617.
- [2] U. M. Fayyad, P. Smyth, N. Weir, S. Djorgovski, Automated analysis and exploration of image databases: Results, progress, and challenges, *Journal of Intelligent Information Systems* 4 (1995) 7–25.
- [3] W. Hsu, M. L. Lee, J. Zhang, Image mining: Trends and developments, *Intelligent Information Systems* 19 (1) (2002) 7–23.
- [4] W. Hsu, S. T. Chua, H. 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.
- [5] B. C. Ooi, K.-L. Tan, T. S. Chua, W. Hsu, Fast image retrieval using color-spatial information, *The International Journal of Very Large Data Bases* 7 (7) (1998) 115–128.
- [6] S. T. Birchfield, S. Rangarajan, Spatial histograms for region-based tracking, *ETRI Journal* 29 (5) (2007) 697–699.
- [7] C. Jiang, F. Coenen, Graph-based image classification by weighting scheme, in: *AI2008*, 2008, pp. 63–76.
- [8] P. T. V. M. de Jong, Age-related macular degeneration, *The New England Journal of Medicine* 355 (14) (2006) 1474–1485.

- [9] L. Brandon, A. Hoover, Drusen detection in a retinal image using multi-level analysis, in: *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, Springer-Verlag, 2003, pp. 618–625.
- [10] D. E. Freund, N. Bressler, 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.
- [11] C. Köse, U. Şevik, O. Gençalioğlu, Automatic segmentation of age-related macular degeneration in retinal fundus images, *Computers in Biology and Medicine* 38 (2008) 611–619.
- [12] C. Köse, U. Şevik, O. Gençalioğlu, A statistical segmentation method for measuring age-related macular degeneration in retinal fundus images, *Journal of Medical Systems* 34 (1) (2008) 1–13.
- [13] K. Rapantzikos, M. Zervakis, K. Balas, Detection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration, *Medical Image Analysis* 7 (2003) 95–108.
- [14] Z. B. Sbeh, L. D. Cohen, G. Mimoun, G. Coscas, A new approach of geodesic reconstruction for drusen segmentation in eye fundus images, *IEEE Transactions on Medical Imaging* 20 (12) (2001) 1321–1333.
- [15] M. H. A. Hijazi, F. Coenen, Y. Zheng, Retinal image classification using a histogram based approach, in: *Proc. International Joint Conference on Neural Networks*, IEEE, 2010, pp. 3501–3507.
- [16] H.-C. Wu, C.-C. Chang, An image retrieval method based on color-complexity and spatial-histogram features, *Fundamenta Informaticae* 76 (2007) 481–493.
- [17] H. Zhang, W. Gao, X. Chen, D. Zhao, Object detection using spatial histograms features, *Image and Vision Computing* 24 (2006) 327–341.
- [18] D. B. Leake, *Case-based reasoning: Experiences, lessons and future directions*, AAAI Press/MIT Press, 1996.
- [19] J. Kolodner, *Case-based reasoning*, Morgan Kaufmann, 1993.

- [20] P. Perner, Introduction to case-based reasoning for signals and images, *Studies in Computational Intelligence* 73 (2008) 1–24.
- [21] A. Holt, I. Bichindaritz, R. Schmidt, P. Perner, Medical applications in case-based reasoning, *The Knowledge Engineering Review* 20 (2005) 289–292.
- [22] A. Elsayed, F. Coenen, C. Jiang, M. Garcia-Finana, V. Sluming, Region of interest based image classification using time series analysis, in: *IEEE International Joint Conference on Neural Networks*, 2010, pp. 3465–3470.
- [23] D. J. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series, in: *AAAI Workshop on Knowledge Discovery in Databases*, 1994, pp. 229–248.
- [24] C. S. Myers, L. R. Rabiner, A comparative study of several dynamic time-warping algorithms for connected word recognition, *The Bell System Technical Journal* 60 (7) (1981) 1389–1409.
- [25] M. Spann, R. Wilson, A quad-tree approach to image segmentation which combines statistical and spatial information, *Pattern Recognition* 18 (1985) 257–269.
- [26] F. Golchin, K. K. Paliwal, Quadtree-based classification in subband image coding, *Digital Signal Processing* 13 (2003) 656–668.
- [27] A. Elsayed, F. Coenen, C. Jiang, M. Garcia-Finana, V. Sluming, Corpus callosum MR image classification, *Knowledge Based Systems* 23 (4) (2010) 330–336.
- [28] H. Samet, The quadtree and related hierarchical data structures, *ACM Computing Surveys* 16 (2) (1984) 187–260.
- [29] M. Foracchia, E. Grisan, A. Ruggeri, Luminosity and contrast normalization in retinal images, *Medical Image Analysis* 9 (2005) 179–190.
- [30] A. Osareh, Automated identification of diabetic retinal exudates and the optic disc, Ph.D. thesis, University of Bristol, UK (2004).
- [31] R. C. Gonzalez, R. E. Woods, *Digital image processing*, Pearson Prentice Hall, 2008.

- [32] K. Zuiderveld, Contrast limited adaptive histogram equalization, Academic Press Graphics Gems Series, Academic Press Professional, Inc., 1994, pp. 474–485.
- [33] N. Patton, T. M. Aslam, T. MacGillivray, Retinal image analysis: Concepts, applications and potential, *Progress in Retinal and Eye Research* 25 (2006) 99–127.
- [34] J. V. B. Soares, J. J. G. Leandro, R. M. C. Jr., H. F. Jelinek, M. J. Cree, Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification, *IEEE Transactions on Medical Imaging* 25 (9) (2006) 1214–1222.
- [35] R. Brunelli, O. Mich, Histograms analysis for image retrieval, *Pattern Recognition Letters* 34 (2001) 1625–1637.
- [36] M. J. Swain, D. H. Ballard, Color indexing, *International Journal of Computer Vision* 7 (1) (1991) 11–31.
- [37] X. Wu, *Graphic Gems II*, Elsevier Science and Technology, 1991, Ch. Efficient statistical computations for optimal color quantization, pp. 126–133.
- [38] R. W. Floyd, L. Steinberg, An adaptive algorithm for spatial greyscale, *Society for Information Display* 17 (2) (1976) 75–77.
- [39] E. Cantu-Paz, Feature subset selection, class separability, and genetic algorithms, in: *Proceedings of Genetic and Evolutionary Computation Conference, 2004*, pp. 959–970.
- [40] E. Cantu-Paz, S. Newsam, C. Kamath, Feature selection in scientific applications, in: *Proceedings of 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004*, pp. 788–793.
- [41] G. Forman, An extensive empirical study of feature selection metrics for text classification, *Journal of Medical Learning Research* 3 (2003) 1289–1305.
- [42] M. H. A. Hijazi, F. Coenen, Y. Zheng, A histogram based approach for the screening of age-related macular degeneration, in: *Medical Image Understanding and Analysis 2009, BMVA, 2009*, pp. 154–158.

- [43] X. Yan, J. Han, gSpan: Graph-based substructure pattern mining, in: IEEE Conference on Data Mining, 2002, pp. 721–724.
- [44] Y.-W. Chang, C.-J. Lin, Feature ranking using linear SVM, in: WCCI2008, 2008, pp. 53–64.
- [45] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, Liblinear: A library for large linear classification, *Journal of Machine Learning Research* 9 (2008) 1871–1874.
- [46] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, S. Sundararajan, A dual coordinate descent method for large-scale linear SVM, in: *International Conference on Machine Learning*, 2008, pp. 408–415.
- [47] C. Cortes, V. Vapnik, Support -vector network, *Machine Learning* 20 (1995) 273–297.
- [48] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (2001).