

# Image Classification for Age-related Macular Degeneration Screening Using Hierarchical Image Decompositions and Graph Mining

Mohd Hanafi Ahmad Hijazi<sup>1,3,\*</sup>, Chuntao Jiang<sup>1</sup>, Frans Coenen<sup>1</sup>,  
and Yalin Zheng<sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK

<sup>2</sup> Department of Eye and Vision Science, Institute of Ageing and Chronic Disease,  
University of Liverpool, UCD Building, Liverpool L69 3GA, UK

<sup>3</sup> School of Engineering and Information Technology, Universiti Malaysia Sabah,  
Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia  
{m.ahmad-hijazi, c.jiang, coenen, yalin.zheng}@liverpool.ac.uk

**Abstract.** Age-related Macular Degeneration (AMD) is the most common cause of adult blindness in the developed world. This paper describes a new image mining technique to perform automated detection of AMD from colour fundus photographs. The technique comprises a novel hierarchical image decomposition mechanism founded on a circular and angular partitioning. The resulting decomposition is then stored in a tree structure to which a weighted frequent sub-tree mining algorithm is applied. The identified sub-graphs are then incorporated into a feature vector representation (one vector per image) to which classification techniques can be applied. The results show that the proposed approach performs both efficiently and accurately.

**Keywords:** Hierarchical image decomposition, weighted graph mining, image partitioning, image classification.

## 1 Introduction

Vision loss and blindness may be caused by various factors; Age-related Macular Degeneration (AMD) [17] is the leading cause of adult blindness in the developed world [21]. AMD is currently incurable and causes total blindness. There are new treatments that can stem the onset of AMD if detected at a sufficiently early stage. Drusen, sub-retinal deposits formed by retinal waste, are the first clinical indicator of AMD. The presence of drusen can be detected by inspection of retina images. Substantial work has been directed at applying image processing and content-based image retrieval techniques to support the diagnosis of AMD; however current performance of these techniques is still not sufficient for wide-scale clinical application, largely because of the limitations of the segmentation techniques adopted.

---

\* Corresponding author.

This paper describes an image mining approach to AMD screening where the objective is to classify images as being either AMD or non-AMD. A first attempt at data mining supported AMD screening, that employed a histogram based representation, is described in [14,13]. The objective was to avoid the segmentation difficulties encountered by previous techniques. The technique produced reasonable results. The technique described in this paper is founded on a novel interleaved angular and circular hierarchical decomposition of the image space, the aim being to isolate instances of drusen. The decomposition is stored in a tree data structure to which a weighted frequent sub-tree mining algorithm is applied. The identified frequent sub-trees are then used to define a feature space which can be used to encode an appropriately labelled training set into a set of feature vectors (one per image) to which established classification techniques can be applied. The proposed technique has been evaluated using a sample set of coloured retinal fundus images featuring both AMD images and a control group.

The main contributions of this paper are:

- The proposed circular and angular based hierarchical decomposition.
- The mechanism for generating feature vectors using weighted frequent sub-tree mining.
- The application of the above techniques to AMD screening.

The rest of this paper is organised as follows: some previous works with respect to image decomposition approaches and weighted frequent sub-graph mining is described in Section 2. Section 3 provides a description of the AMD application domain, followed by details of the proposed image classification approach in Section 4. The performance of the proposed approach is extensively evaluated in Section 5 and some conclusions are presented in Section 6.

## 2 Previous Work

Hierarchical data structures have been widely applied in various domains, such as image segmentation [25], image coding [12] and image classification [9]. The main advantage of this type of data structure is that it provides an effective representation of the problem domain that can be readily processed [23]. The most common hierarchical decomposition technique is founded on quadtrees, where the search space is repeatedly quartered until uniform “tiles” are arrived at or a maximum decomposition is reached. In the work described in this paper a new image decomposition technique more suited to retinal images, that uses an alternating angular and circular partitioning, is presented.

Graph mining techniques can be categorised as being either *transaction* based or *single* graph based [18]. Transaction graph mining aims to discover frequently occurring sub-graphs in a given graph data set. Weighted frequent sub-graph mining is founded on the idea that in some cases certain vertices and/or edges in the input graph set can be deemed to be more significant than others. The weighted frequent sub-graph mining has demonstrated its advantages over frequent sub-graph mining in a number of studies [9,18], the main advantage is

that the former spends significantly less run time identifying far fewer patterns (i.e. frequent sub-graphs) than the latter. A variation of a weighted frequent sub-graph mining algorithm [18], founded on the well-known gSpan algorithm [27], is used with respect to the work described in this paper and is described in detail in Subsection 4.2.

Image processing approaches have been widely used in the detection of drusen for AMD diagnosis. The earliest work [24] used a morphological mechanism to localise drusen. Other image processing techniques that have been applied include: (i) histogram-based adaptive local thresholding [22], (ii) region growing [19,20]; (iii) wavelet based feature identification coupled with multilevel classification [3]; (iv) anomaly detection based approaches, that employ Support Vector Data Description (SVDD), to segment anomalous pixels [11]; and (v) signal based approaches, namely amplitude-modulation frequency-modulation (AM-FM), to generate multi-scale features for drusen classification [1,2]. Content-Based Image Retrieval (CBIR) techniques have also been applied. For example, Chaum et al. [6] have applied CBIR to get a probability of the presence of a particular pathology (a confidence threshold is then applied on the generated probabilities to predict the retinal images class).

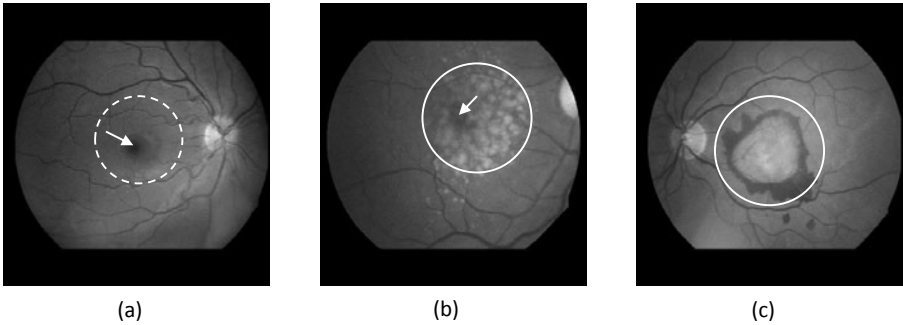
Most of the existing work on AMD diagnosis is founded on drusen detection and segmentation. The authors are aware of only three reports of extending drusen detection to AMD screening [2,3,6]. One issue is the difficulties and challenges in not only identifying the drusen, but also other retinal structures, in particular the optic disc and the macula. The challenge is exacerbated by the natural variation of the appearance of the retina, the image quality and patient factors (e.g. compliance during image acquisition and media clarity of the eye) [6,15]. Alternative techniques, not founded on segmentation, therefore seem desirable. An early attempt in this direction has been proposed in [13] whereby a histogram based representation was used to which Case-Based Reasoning (CBR) was applied to facilitate the classification of “new” cases. Another approach that was built based on work in [13] that uses two “case bases” has been proposed in [15]. Good results were produced, however observations indicated that relying on the retinal image colour distribution alone was not sufficient. Thus, in [14] a spatial histogram technique, that include colour and spatial information, was proposed. The technique gave the best results so far with respect to the test image dataset. The work described in this paper is directed at improving on these results.

### 3 Age-related Macular Degeneration

Some exemplar retinal images are presented in Figure 1. The central region of the human retina is called the *Macula*, which is centered at the *fovea*; this is the place where acute and central vision is made possible. Figure 1(a) shows the fovea (indicated by an arrow) and macula (circled). Damage to the macula causes distortion and loss of the central vision required for (say) reading and writing. There are various reasons why this might happen, one of which is Age-related Macular Degeneration (AMD) where the delicate cells of the macula

becomes damaged and stops functioning at the later stages of life [17]. AMD can be categorised into *early*, *intermediate* and *advanced*. Advanced AMD, which can be further divided into *non-neovascular* and *neovascular*, is when severe vision loss or even total blindness occurs [17]. Although AMD is incurable, the early detection of AMD is desirable as there are new emerging treatments that can be given to patients with AMD to slow down or even halt the progress of the condition.

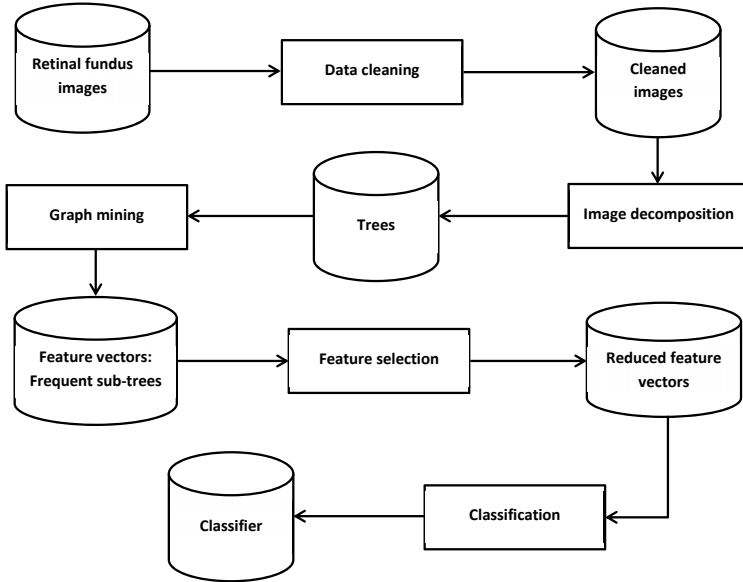
As noted above, early stage AMD can be diagnosed by the existence of *drusen* [8,17], yellowish-white deposits with sizes ranging from less than  $63\ \mu\text{m}$  (small drusen) to greater than  $124\ \mu\text{m}$  (*large* drusen) in diameter [17]. These can be detected through manual inspection, by trained clinicians, of the retinal images collected within screening programmes. This task is however labour intensive given the increasing incidence of AMD. Thus, the automation or semi automation of the process is deemed desirable. Figure 1(b) shows a case of intermediate AMD where the presence of drusen (surrounded by circle) is widespread but the fovea is still visible (dark area pointed by arrow). Figure 1(c) gives an example of advanced neovascular AMD where the fovea is totally obscured resulting in total loss of central vision. Drusen can also be categorised into *hard* or *soft* drusen. Hard drusen is more easy to identify because it has well defined borders, while soft drusen has indistinct edges that blend into the retinal background.



**Fig. 1.** Grayscaled retinal fundus images: (a) normal, (b) intermediate, and (c) advanced neovascular

## 4 AMD Classifier Generation

An overview of the proposed hierarchical decomposition based approach to the generation of AMD classifiers is presented in Figure 2. The approach commences with retinal image cleaning. The quality of the retinal fundus images is often affected by various factors that hinder image classification, such as colour variation and nonuniform illumination. Some image cleaning is therefore required. In the context of the work described in this paper, an approach used in [14] is reused to enhanced the images and to identify blood vessel pixels (which we wish to remove from the image data so as to “clean” the data). The process then proceeds

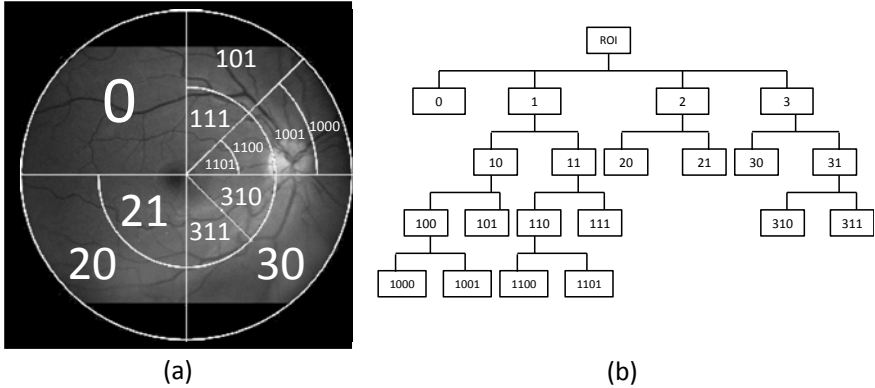


**Fig. 2.** Block diagram of the proposed classifier generation approach

with the decomposition of the image, this is described in detail in Subsection 4.1 below; the result is a collection of tree represented images (one per image). Next the weighted frequent sub-graph (sub-tree) mining approach is applied to the data (the algorithm is detailed in Subsection 4.2). The identified frequent sub-trees then 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 feature selection process is described in Subsection 4.3. Once the feature vector representation has been generated we can apply established classification techniques (see Subsection 4.4).

#### 4.1 Image Decomposition

The proposed image decomposition method is described in this sub-section. As noted above hierarchical image decomposition is a well established technique [12,23,25]. 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 3(a) shows an example of a partitioning that might be applied to an image; Figure 3(b) presents the associated *Tree* storage structure. Note that a numbering convention is used to label individual regions described by nodes in the tree structure.



**Fig. 3.** An example of: (a) circular and angular image decomposition, and (b) the associate tree data structure

Algorithm 1 shows how the interleaved circular and angular partitioning is performed. Given a coloured retinal fundus image,  $I$  with a size of  $X$  pixels. The RGB (red, green and blue) colour model is used to extract the pixels intensity values, which means each pixel will have three intensity values (red, green, blue) associated with it, hence initially three trees are generated which are then merged. The *GetCentroids* method in line 4 uses a retinal image mask,  $M$ , to identify the centroid of the retina disc. The *GetImageBackground* method in line 5 generates a binary format background image,  $imbg$ , to be used to distinguish the background (areas outside of the field of view of the fundus) pixels and the blood vessel pixels,  $V$ , from the retinal pixels.  $imbg$  is defined as:

$$imbg = M \cap RV \quad (1)$$

$$M(x) = \begin{cases} 1, & \text{if } x \text{ is a retina pixel} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$RV(x) = \begin{cases} 0, & \text{if } x \text{ is a blood vessels pixel,} \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

where  $x \in X$ , and  $M$  and  $RV$  are both of size of  $X$  pixels. The image ROI was then identified using the *GetROI* method.

As noted above the proposed hierarchical image partitioning commences with *AngularPartitioning* (line 13 of Algorithm 1). On the next iteration *CircularPartitioning* will be applied. Both *AngularPartitioning* and *CircularPartitioning* will then be called alternately until the  $D_{max}$  tree level is reached or only regions of uniform intensity are left. Throughout the process the tree data structure is continuously appended to. The algorithm ends with the merging of the three trees in  $T$  using the *MergeTrees* method to form a single tree. The merging is done by calculating the Average Intensity Values (*AIV*) for the nodes in  $T$ , defined as:

$$AIV_y = \frac{1}{n} \sum_{k=1}^3 (T_{k_y}) \quad (4)$$

$$T_{k_y} = \frac{1}{z} \sum_{i=1}^z (k_{y_i}) \quad (5)$$

where  $T$  comprises the red, green and blue colour trees,  $y$  is a unique node identifier in  $T_{final}$ ,  $z$  is number of pixels in node  $y$ , and  $n$  is number of occurrences of node  $y$  (whenever  $T_{k_y}$  is not null) in the set of trees  $T$ .

---

**Algorithm 1:** ImageDecomposition
 

---

**Data:** Coloured retinal fundus image  $I$ , retinal image mask  $M$ , retinal blood vessels binary image  $RV$  and  $D_{max}$

**Result:** Image decomposition tree  $T_{final}$

```

1  $c\_count \leftarrow 1$ ;
2  $a\_count \leftarrow 1$ ;
3  $T \leftarrow \{null, null, null\}$ ;
4  $centroid \leftarrow GetCentroid(M)$ ;
5  $imbg \leftarrow GetImageBackground(M, RV)$ ;
6  $roi \leftarrow GetROI(imbg, centroid)$ ;
7 for  $k \leftarrow 0$  to 2 do // Generate trees for each colour channel
8   for  $i \leftarrow 1$  to  $maxDepth$  do // Generate trees for each tree level
9     if  $mod(i/2) = 0$  then
10        $t \leftarrow CircularPartitioning(roi, imbg, c\_count, centroid)$ ;
11        $c\_count \leftarrow c\_count + 1$ ;
12     else
13        $t \leftarrow AngularPartitioning(roi, imbg, a\_count, centroid)$ ;
14        $a\_count \leftarrow a\_count + 1$ ;
15     end
16      $tree \leftarrow UpdateTree(tree, t)$ ;
17   end
18    $T_k \leftarrow tree$ ;
19 end
20  $T_{final} \leftarrow MergeTrees(T_0, T_1, T_2)$ ;

```

---

Algorithm 2 describes the *CircularPartitioning* method. The algorithm returns a set of new nodes  $B$  to be added to the tree structure. The input is the ROI image  $roi$ , the mask  $imbg$ , the level count for the circular partitioning ( $c\_count$ ) and the centroid of the ROI (retina disk). We first identify how many circles,  $m$  are required for the current iteration (line 1). Then we calculate a set of new radii to be included in the partitioning;  $R = \{\rho_0, \rho_1, \dots, \rho_m\}$  describes a sequence of concentric circles. We then (line 3) generate the necessary additional nodes for the tree. Each level  $c\_count$  region is considered in turn and, where appropriate,  $c\_count + 1$  regions are constructed.

---

**Algorithm 2:** CircularPartitioning

---

**Data:** Retinal image *roi*, image background *imbg*, *c\_count* and *centroid*  
**Result:** An array of circular partitioned image regions *B*

```

1  $m \leftarrow 2^{c\_count}$  ; // To calculate number of circles
2  $R \leftarrow GetRadius(imbg, centroid, m)$  ; // To calculate radii values
3  $B \leftarrow SplitImage(roi, imbg, R, centroid)$ ;
4 return  $B$ 
```

---

The *AngularPartitioning* method, as described in Algorithm 3, begins by identifying the number of radii (*m*) that are required (line 1). The radii define the angular partitions which are defined in terms of a set of arcs  $A = \{\alpha_0, \alpha_1, \dots, \alpha_m\}$ . Each arc  $\alpha$  is defined by  $\theta = 2\pi/m$ , used in the *GetTheta* method (line 2). As in the case of the *CircularPartitioning* algorithm, the *SplitImage* method is then called to decompose the image, as indicated, to produce an appropriate set of nodes *B*.

---

**Algorithm 3:** AngularPartitioning

---

**Data:** Retinal image *roi*, image background *imbg*, *a\_count* and *centroid*  
**Result:** An array of angular partitioned image regions *B*

```

1  $m \leftarrow 2 \times 2^{a\_count}$  ; // To calculate number of angular lines
2  $\theta \leftarrow GetTheta(n)$  ; // To calculate the angle between angular lines
3  $B \leftarrow SplitImage(roi, imbg, n, \theta, centroid)$ ;
4 return  $B$ 
```

---

As noted above, an important feature of the hierarchical image decomposition is the selection of a termination criterion,  $\omega$ , which defines the homogeneity of a particular region in an image and is used to determine if further region splitting is required. A common definition for  $\omega$  is in terms of the distance between the highest and lowest intensity values of the pixels in a region *i*. If  $\omega$  is less than a predefined homogeneity threshold,  $\tau$ , no further decomposition of the region *i* will be undertaken.

In this paper, a similar termination criterion to that described in [12] is adopted. The  $\omega$  value is defined according to how well a parent region represents its two child regions' intensity values. If the value (derived from the average intensity values of all pixels in a particular region) of a parent region is similar ( $< \tau$ ) to that of its two child regions, the parent region is deemed to be homogeneous and left unsplit. Otherwise, the parent region will be further partitioned. Thus  $\omega$  can be formalised as:

$$\omega = \frac{1}{s} \sum_{i=1}^s \sqrt{(\mu_p - \mu_i)^2} \quad (6)$$

where *s* is the number of sub-regions,  $\mu_p$  is the average intensity value for the parent region and  $\mu_i$  is the average intensity value for sub-region *i*. Each



identified sub-region is represented as a “node” in a tree data structure where the relationship between sub-regions and their parent form the tree “edges”.

## 4.2 Weighted Frequent Sub-tree (wFST) Mining

After the image decomposition step introduced in the above subsection, images were modelled as a collection of trees. According to [18], each tree is defined as follows:  $T = \{V, E, L_V, L_E, \phi\}$ , where  $L_V$  and  $L_E$  are labels for nodes and edges in  $T$  respectively, and  $\phi$  defines a label mapping function.

In [18] it was suggested that for many applications, such as image mining, some tree nodes have more significance associated with them than other nodes. In the case of the hierarchical decomposition described in this paper, nodes that feature a significant difference in colour intensity when compared to their parent node are deemed to be more significant (than the parent). The underpinning philosophy here 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 node  $v \in V$  was weighted by the average colour intensity value of the region represented by that node  $v$ , and the strength of each edge  $e \in E$ ,  $e_w$ , is weighted by:

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

where  $I_v$  is the average colour intensity value for node  $v$  and  $I_{par}$  is the average colour intensity value for  $v$ 's parent.

By adding node and edge weights into the tree representation, the weighted tree representation was able to capture more image information than the unweighted one. A weighted Frequent Sub-Tree (wFST) mining algorithm, an extension of the well-known gSpan algorithm [27], was then applied to the tree data so as to identify frequently occurring trees within the dataset. The wFST algorithm operated in a similar manner to that described in [18], but utilised both node and edge weightings. In the context of wFST mining, a sub-graph pattern  $g$  is considered to be “interesting”, if it satisfies the following two conditions:

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

Where:  $N_{wr}$  denotes the node weighting,  $\text{sup}(g)$  denotes the support (i.e. frequency) of  $g$ , and  $\sigma$  denotes a *minimum support* threshold,  $E_{wr}$  denotes the edge weighting, and  $\lambda$  denotes a *minimum weight* threshold. Both the  $N_{wr}$  and  $E_{wr}$  are computed using a similar scheme to that described in [18].

The number of patterns discovered by the wFST mining algorithm is thus determined by both the  $\sigma$  and  $\lambda$  values. According to initial experiments conducted by the authors, relatively low  $\sigma$  and  $\lambda$  values are required, in order to extract a sufficient number of image features (frequent sub-trees). However, setting low threshold values results in a substantial number of patterns, of which many are

redundant in terms of the desired classification. Therefore, feature selection was applied to the discovered patterns, this is discussed in the following subsection.

### 4.3 Feature Selection

For the AMD application the number of identified wFSTs was substantial, using low threshold values tens of thousands of wFSTs were identified. To reduce the number of wFSTs to a manageable number a feature selection strategy was applied so as to identify those wFSTs that displayed a strong discriminatory power, which would consequently be able to produce good classification results. A feature ranking mechanism was therefore used, with respect to the AMD application, that used linear Support Vector Machine (SVM) weights to rank features as proposed in [5]. The main advantage of this approach is its implementation simplicity and effectiveness in determining relevant features. The identified wFSTs 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 [10,16]. The selection of only the top  $K$  wFSTs for classification then concluded the feature selection process.

### 4.4 Classification Technique

The final stage of the proposed retinal image classification process was the classification stage. The identified top  $K$  wFSTs 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 5), two different classifier generation techniques were used, Naïve Bayes [26] and Support Vector Machine (SVM) [7]. Naïve Bayes was selected because: (i) it works very well when tested on data with independent attributes [26], and (ii) it does not require user defined parameters. SVM on the other hand was chosen because it is frequently acknowledged to be one of the most effective classification method in machine learning. In this paper, the Naïve Bayes classifier available in Weka [26] was used. The second classifier was built using LibSVM [4] with a radial basis function kernel.

## 5 Evaluation

The proposed AMD screening approach was applied to two retinal fundus images datasets, ARIA<sup>1</sup> and STARE<sup>2</sup>. The ARIA dataset comprises 161 images, of which 101 were AMD and 60 were normal images. The STARE dataset comprised 97 images, of which 59 were AMD and 38 were normal images. The datasets were merged to create an image set comprising 258 images, of which 160 featured AMD and 98 were normal. Sensitivity, specificity and accuracy were used to measure the classification performance.

<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>

In the experiments three values for  $D_{max}$  were used: 5, 6 and 7. The threshold for node splitting,  $\tau$ , was set to 2.5%. To train the LibLINEAR (for feature selection task) and LibSVM classifiers, the default values for the user defined parameters used in Weka were applied, except for the soft margin parameter  $C$  and gamma parameter  $\gamma$  which were determined using the parameter selection tool provided with LibLINEAR and LibSVM [4]. The aim of the experiments was to evaluate: (i) the effect of the value of  $D_{max}$  on the classification results, (ii) how feature selection improved the classification performance, and (iii) how well the proposed approach's performance compared with other AMD classification techniques. Most of the reported experiments were conducted using Ten-fold Cross Validation (TCV).

### 5.1 Performances Using Different Levels of Decomposition

Tables 1 and 2 shows the performances of the proposed approach when using the three different levels of decomposition (values for  $D_{max}$ ) and using Naïve Bayes and LibSVM respectively. Feature selection was not applied in these experiments. Minimum support,  $\sigma$  was used to prune the candidate sub-trees, while the minimum weight,  $\lambda$  was used to further reduce the number of identified frequent sub-trees according to their edge weights.  $F$  denotes the size of the feature space in terms of the number of identified frequent sub-trees, while  $Sens$ ,  $Spec$  and  $Acc$  refers to sensitivity, specificity and accuracy. Each  $\sigma$  value was tested against a range of  $\lambda$  values (20, 40, 60 and 80), however in the table (because of space limitations) only the best performing  $\lambda$  value associated with each  $\sigma$  value is recorded.

**Table 1.** TCV Classification results obtained using different levels of decomposition and Naïve Bayes

$\sigma$ (%)	5					6					7				
	$\lambda$	$F$	$Sens$	$Spec$	$Acc$	$\lambda$	$F$	$Sens$	$Spec$	$Acc$	$\lambda$	$F$	$Sens$	$Spec$	$Acc$
10	40	764	65	60	<b>63</b>	20	7125	64	53	60	80	3433	66	42	57
20	60	291	68	51	62	80	498	70	42	59	80	3433	66	42	57
30	60	291	68	51	62	20	1746	66	49	60	80	3433	66	42	57
40	20	248	66	48	59	80	498	70	42	59	80	3433	66	42	57
50	20	181	69	45	60	80	498	70	42	59	80	3433	66	42	57
60	20	130	67	46	59	20	559	69	45	60	20	3893	66	42	57
70	20	99	69	33	55	20	404	68	34	55	20	2623	69	41	58
80	20	71	80	13	55	20	283	72	27	55	20	1706	70	36	57
90	20	55	86	16	60	20	180	79	22	58	20	955	72	28	55

Inspect of Tables 1 and 2 indicates that the best accuracy was achieved using  $D_{max} = 5$  (63% and 70%). The best sensitivity and specificity were 86% and 60% (Naïve Bayes) and 100% and 43% (LibSVM) respectively. The best sensitivity and specificity for both classifiers occurred using different  $\sigma$  and  $\lambda$  values (but again with  $D_{max} = 5$ ). Experiments were also conducted using an unweighted FST mining algorithm (gSpan), however these indicated that the memory requirements and runtime deemed to be unacceptable.

**Table 2.** TCV Classification results obtained using different levels of decomposition and LibSVM

$\sigma$ (%)	5					6					7				
	$\lambda$	$F$	$Sens$	$Spec$	$Acc$	$\lambda$	$F$	$Sens$	$Spec$	$Acc$	$\lambda$	$F$	$Sens$	$Spec$	$Acc$
10	40	764	86	43	<b>70</b>	20	7125	87	39	69	60	11461	96	15	66
20	20	594	89	37	69	20	3103	91	30	68	60	11461	96	15	66
30	20	365	95	16	65	60	1358	89	33	68	60	11461	96	15	66
40	80	118	100	0	62	20	1135	92	30	68	20	9043	97	8	63
50	80	118	100	0	62	20	779	94	11	62	80	3433	96	8	63
60	20	130	83	30	63	20	559	99	1	62	20	3893	96	11	64
70	20	99	99	4	63	20	404	98	0	60	20	2623	99	4	63
80	20	71	100	0	62	20	283	100	1	62	20	1706	97	8	63
90	20	55	99	3	63	20	180	100	0	62	20	955	96	10	63

## 5.2 Performances of AMD Classification According to the Size of the Identified Feature Space

Tables 3 and 4 shows the performances of the proposed approach with respect to different values of  $K$ , using Naïve Bayes and LibSVM respectively. 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 five different  $P$  values were conducted: 0.05, 0.1, 0.2, 0.4 and 0.6. However, only the results using a  $D_{max} = 7$  and  $P$  values of 0.05, 0.1 and 0.4 are presented in the tables because these produced the best classification performances with respect to both classifiers. Inspection of the tables indicates how the performance changes as the size of the feature space is reduced.

**Table 3.** TCV Classification results using feature selection, a decomposition level ( $D_{max}$ ) of 7, and Naïve Bayes

$\sigma$ (%)	$P_{0.05}$					$P_{0.1}$					$P_{0.4}$				
	$\lambda$	$K$	$Sens$	$Spec$	$Acc$	$\lambda$	$K$	$Sens$	$Spec$	$Acc$	$\lambda$	$K$	$Sens$	$Spec$	$Acc$
10	20	3671	94	96	<b>95</b>	20	7342	92	92	92	40	16278	73	67	71
20	20	1407	91	93	91	20	2814	88	82	85	20	11257	71	65	68
30	20	748	88	82	85	20	1496	86	78	83	20	5983	71	61	67
40	20	452	85	80	83	20	904	84	72	79	20	3618	71	56	65
50	20	291	85	72	80	80	343	86	64	78	20	2330	73	52	65
60	20	195	84	68	78	20	389	86	67	78	20	1558	76	51	66
70	20	131	83	63	75	20	262	83	56	72	20	1050	74	47	64
80	20	85	82	45	68	20	171	81	50	69	20	683	74	45	63
90	20	48	83	37	65	20	96	84	41	67	20	382	78	37	62

Tables 3 and 4 demonstrate that the best results were obtained using lower numbers of features, where  $P = 0.05$  and  $K = 3671$ . The best accuracy for Naïve Bayes was 95% while LibSVM recorded a full 100% accuracy. The highest sensitivity and specificity was 100% (LibSVM). High sensitivity and specificity were also achieved using the Naïve Bayes classifier with a corresponding accuracy of 94% and 96% respectively. All of the best results were generated using  $\sigma = 10\%$

**Table 4.** TCV Classification results using feature selection, a decomposition level ( $D_{max}$ ) of 7, and LibSVM

$\sigma$ (%)	$F_{0.05}$					$F_{0.1}$					$F_{0.4}$				
	$\lambda$	$K$	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	$\lambda$	$K$	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	$\lambda$	$K$	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>
10	20	3671	100	100	<b>100</b>	20	7342	100	100	<b>100</b>	40	16278	99	81	92
20	80	172	100	0	62	80	343	100	0	62	60	4585	99	8	65
30	20	748	99	80	92	20	1496	99	94	97	20	5983	98	70	87
40	80	172	100	0	62	80	172	100	0	62	80	1374	99	5	64
50	20	291	97	54	81	20	583	96	84	91	20	2330	95	56	80
60	80	172	100	0	62	80	172	100	0	62	80	1374	99	5	64
70	20	131	100	0	62	20	262	100	0	62	20	1050	100	2	63
80	20	85	100	0	62	20	171	100	1	62	20	683	99	6	64
90	20	48	100	0	62	20	96	100	1	62	20	382	98	10	65

and  $\lambda = 20\%$  were applied. The accuracy increased as the  $\sigma$  value decreased for all  $K$  values. The results produced show that the larger the feature space the better the classification performance. It should be noted that the results reported in Subsection 5.1, where feature selection was not applied are not as good as those reported here. Feature selection clearly improves the classification performance.

### 5.3 Performance Comparison of AMD Classification Using Various Classification Techniques

Table 5 compares the classification results obtained using the spatial histogram based approach [14] referred to earlier in Section 2 and the AMD screening approach proposed in this paper using the LibSVM classifier. Both approaches were applied to the ARIA dataset. The parameters were set to:  $\sigma = 30\%$ ,  $\lambda = 20\%$  and  $D_{max} = 7$ . These values were chosen as they produced the best classification results through series of experiments conducted. The results clearly indicate the superiority of the proposed approach.

**Table 5.** Comparison of proposed AMD screening approach with alternative histogram based approach

Approach	Features	Sensitivity	Specificity	Accuracy
Histogram based [14]	640	86	56	74
Proposed approach	1354	100	100	100

Table 6 compares the performance of the proposed approach with other reported approaches: Barriga et al. [2], Chaum et al. [6] and Brandon and Hoover [3]. For comparison we also used the leave-one-out testing method used by them. The reported results in [2] were generated using the drusen classification technique described in [1]. For this experiment, the proposed approach utilised the

**Table 6.** Comparison of proposed AMD screening approach with alternative approaches

Approach	Dataset size	Sensitivity	Specificity	Accuracy
Barriga et al. [2]	100	75	50	-
Brandon and Hoover [3]	97	-	-	87
Chaum et al. [6]	395	-	-	88
Proposed approach	258	89	99	93

Naïve Bayes classifier with  $\sigma = 10\%$  and  $\lambda = 20\%$  ( $K = 3671$ ). A decomposition level of  $D_{max} = 7$  was used.

Table 6 includes some missing values because these were not reported in the literature and could not be derived by the authors. The result obtained by Barriga et al. [2] only reported sensitivity and specificity. On the other hand, the work of Brandon and Hoover [3] only reported accuracy<sup>3</sup>, no sensitivity and specificity values were reported. It should also be noted that the approach of Chaum et al. [6] was actually applied in a multi-class setting, of which 12 of the AMD images were classified as “unknown” and excluded from the accuracy calculation (if included this would give an accuracy of 75%). Overall, the results demonstrate that the proposed AMD screening approach outperforms the other approaches by 14% (sensitivity), 49%(specificity) and 5% (accuracy).

## 6 Conclusions

An AMD screening approach founded on a hierarchical circular and angular image decomposition technique has been described. The decomposition results in a tree data structure to which a weighted tree mining technique was applied so as to identify frequent occurring sub-trees. The generated weighted frequent sub-trees were then used to recast the input data (the training set) into a feature vector representation. AMD classifiers were then built using the feature vector representation as the input data. For evaluation purpose the proposed approach was applied to retinal fundus images data from two publically available databases. A 100% accuracy was produced using the LibSVM classifier. A more straightforward and parameter-free classification technique, Naïve Bayes, was also experimented with and also generated good results (95% accuracy). Further experiments demonstrated the superiority of the proposed approach compared to some other reported techniques for AMD detection. Our current work is directed at extending the proposed approach to address multi-class problems. The authors are also interested in grading the severity of AMD, as well as using the approach for the screening of other retinal diseases, such as diabetic retinopathy.

<sup>3</sup> <http://www.parl.clemson.edu/stare/drusen/>

## References

1. Barriga, E.S., Murray, V., Agurto, C., Pattichis, M.S., Russell, S., Abramoff, M.D., Davis, H., Soliz, P.: Multi-scale AM-FM for lesion phenotyping on age-related macular degeneration. In: IEEE International Symposium on Computer-Based Medical Systems, pp. 1–5 (2009)
2. Barriga, E.S., Murray, V., Agurto, C., Pattichis, M.S., Russell, S., Abramoff, M.D., Davis, H., Soliz, P.: Automatic computer-based grading for age-related maculopathy. *Investigative Ophthalmology and Visual Science* 51, E-Abstract 1793 (2010)
3. Brandon, L., Hoover, A.: Drusen detection in a retinal image using multi-level analysis. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 618–625. Springer, Heidelberg (2003)
4. Chang, C-C., Lin, C-J.: LIBSVM: A library for support vector machines (2001) Software, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
5. Chang, Y.-W., Lin, C.-J.: Feature ranking using linear SVM. In: WCCI 2008, pp. 53–64 (2008)
6. Chaum, E., Karnowski, T.P., Priya Govindasamy, V.: Automated diagnosis of retinopathy by content-based image retrieval. *Retina* 28(10), 1463–1477 (2008)
7. Cortes, C., Vapnik, V.: Support -vector network. *Machine Learning* 20, 273–297 (1995)
8. de Jong, P.T.V.M.: Age-related macular degeneration. *The New England Journal of Medicine* 355(14), 1474–1485 (2006)
9. Elsayed, A., Coenen, F., Jiang, C., Garcia-Finana, M., Sluming, V.: Corpus callosum MR image classification. *Knowledge Based Systems* 23(4), 330–336 (2010)
10. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J.: Liblinear: A library for large linear classification. *Journal of Machine Learning Research* 9, 1871–1874 (2008)
11. Freund, D.E., Bressler, N., Burlina, P.: Automated detection of drusen in the macula. In: Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, pp. 61–64 (2009)
12. Golchin, F., Paliwal, K.K.: Quadtree-based classification in subband image coding. *Digital Signal Processing* 13, 656–668 (2003)
13. Hijazi, M.H.A., Coenen, F., Zheng, Y.: Image classification using histograms and time series analysis: A study of age-related macular degeneration screening in retina image data. In: Proceedings of 10th Industrial Conference on Data Mining, pp. 197–209 (2010)
14. Hijazi, M.H.A., Coenen, F., Zheng, Y.: 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, pp. 325–338 (2010)
15. Hijazi, M.H.A., Coenen, F., Zheng, Y.: Retinal image classification using a histogram based approach. In: Proceedings of International Joint Conference on Neural Network 2010 (World Congress on Computational Intelligence 2010), pp. 3501–3507 (2010)
16. Hsieh, C.-J., Chang, K.-W., Lin, C.-J., Sathiy Keerthi, S., Sundararajan, S.: A dual coordinate descent method for large-scale linear SVM. In: International Conference on Machine Learning, pp. 408–415 (2008)
17. Jager, R.D., Mieler, W.F., Mieler, J.W.: Age-related macular degeneration. *The New England Journal of Medicine* 358(24), 2606–2617 (2008)

18. Jiang, C., Coenen, F.: Graph-based image classification by weighting scheme. In: AI 2008, pp. 63–76 (2008)
19. Köse, C., Şevik, U., Gençalioğlu, O.: A statistical segmentation method for measuring age-related macular degeneration in retinal fundus images. *Journal of Medical Systems* 34(1), 1–13 (2008)
20. Köse, C., Şevik, U., Gençalioğlu, O.: Automatic segmentation of age-related macular degeneration in retinal fundus images. *Computers in Biology and Medicine* 38, 611–619 (2008)
21. Minassian, D., Reidy, A.: Future sight loss UK (2): An epidemiological and economic model for sight loss in the decade 2010-2020. Technical report, Royal National Institute of Blind People (2009)
22. Rapantzikos, K., Zervakis, M., Balas, K.: Detection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration. *Medical Image Analysis* 7, 95–108 (2003)
23. Samet, H.: The quadtree and related hierarchical data structures. *ACM Computing Surveys* 16(2), 187–260 (1984)
24. Sbeh, Z.B., Cohen, L.D., Mimoun, G., Coscas, G.: A new approach of geodesic reconstruction for drusen segmentation in eye fundus images. *IEEE Transactions on Medical Imaging* 20(12), 1321–1333 (2001)
25. Spann, M., Wilson, R.: A quad-tree approach to image segmentation which combines statistical and spatial information. *Pattern Recognition* 18, 257–269 (1985)
26. Witten, I., Frank, E.H.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco (2005)
27. Yan, X., Han, J.: gSpan: Graph-based substructure pattern mining. In: *IEEE Conference on Data Mining*, pp. 721–724 (2002)