Age-related Macular Degeneration Screening using Data Mining Approaches

Mohd Hanafi Ahmad Hijazi∗, Frans Coenen† and Yalin Zheng‡

∗Applied Computing Group, School of Engineering and Information Technology, Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia
Email: hanafi@ums.edu.my

†Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK
Email: coenen@liverpool.ac.uk

‡Department of Eye and Vision Science, Institute of Ageing and Chronic Disease, University of Liverpool, UCD Building, Liverpool L69 3GA, UK
Email: yalin.zheng@liverpool.ac.uk

Abstract—This paper investigates the use of tabular approach to classifying retinal images according to whether they feature Age-related Macular Degeneration (AMD), a retina condition that causes blindness in old age. The novelty of the proposed approach is that it is not founded on feature segmentation, instead entire image encodings are used. Features in the form of statistical parameters extracted directly and indirectly from the images are considered. For the evaluation two publically available, retinal fundus image data sets were used. The evaluation was conducted in the context of AMD screening. Excellent results were produced: Sensitivity and AUC of 90% and over were recorded for binary-class classification problem.

Keywords—Age-related Macular Degeneration; Data Mining; Decision Support Techniques; Classification; Retinal Image;

I. INTRODUCTION

Age-related Macular Degeneration (AMD), is a condition where the delicate cells of the macula become damaged (and stop functioning properly) in the later stages of life, it is the leading cause of adult blindness in the UK, typically affecting people who are aged 50 years and over. In the UK, it is estimated that in 2020 this age group will comprise a population of 25 million people and more than 7% of these are expected to be affected [1]. AMD is currently incurable and causes permanent total blindness. However, there are new treatments that may stem the onset of advanced AMD if detected at a sufficiently early stage [2]. The diagnosis of AMD is typically undertaken through the careful inspection of retinal images by trained clinicians. Figure 1 shows some example images. Figure 1(a) presents a normal retina image, Figure 1(b) a retina that displays signs of early stage AMD and (c) a retina that features neovascular AMD.

This paper proposes a mechanism for AMD screening using image mining approaches. The idea is to identify the best mechanism to screen AMD that avoid or minimize the use of segmentation techniques, the technology on which retina image analysis normally relies [3]–[7], but instead use alternative “whole image” representation techniques that do not rely on segmentation. Earlier attempts to use similar approaches have been reported in [8] where two techniques were evaluated. The first technique represents images in terms of point series and uses a Case Based Reasoning (CBR) approach to classify retina images. The second techniques uses a hierarchical decomposition mechanism to generate a set of trees, one per image in the training set, which are then processed using a frequent subgraph mining technique to produce a feature space (to which established classification techniques can again be applied). Both performed well for binary classification problem, in particular the tree based approach. However, the computational cost to identify features is high.

The main contributions of this paper are:
• The mechanism for generating feature vectors using statistical parameters extracted directly and indirectly from images.
• The empirical comparison results using image mining approaches conducted on binary and multi-class classification settings for AMD screening.

The rest of the paper is organised as follows. Section II presents the background to the work described in this paper.
Data preparation is presented in Section III. The proposed image classification approach is described in Section IV. Section V presents a comparison between the proposed approach and other approaches found in the literature. Some conclusions are provided in Section VI.

II. PREVIOUS WORK

The diagnosis of AMD is typically undertaken by manual inspection of retinal images by trained clinicians. In most cases, an early indicator of AMD is the presence of drusen, yellowish-white subretinal deposits, on the macula as shown in Figure 1(b). The presence of large and numerous drusen indicate an early sign of AMD. Drusen can be categorised as either hard and soft drusen. Hard drusen have well-defined borders, while soft drusen tend to blend into the retinal background.

The earliest work reported in the literature concerning the automated or semi-automated diagnosis of AMD is that of [7] who used mathematical morphology to detect drusen. Other work on the identification of drusen in retina images has focuses on segmentation coupled with image enhancement approaches [5], [6], [9]. The work described in [6] adopted a multilevel histogram equalisation technique to enhance the image contrast followed by drusen segmentation using both global and local thresholds. A different concept, founded on the use of histograms for AMD screening is proposed in this paper. In [5], [9] a two phased approach was proposed involving inverse drusen segmentation within the macular area. In [10] a signal based approach called AM-FM was proposed to generate multi-scale features to represent drusen signatures. Images were partitioned into sub-regions and features were extracted from each sub-region. A wavelet analysis technique to extract drusen patterns, and a multilevel classification for drusen categorisation were described in [3], where a set of rules was used to identify potential drusen pixels. In [11], a content-based image retrieval technique was employed to obtain a probability value for the presence of a particular pathology. Segmentation of objects was first conducted after which features extraction was applied to the identified objects.

Of the reported work found in the literature that the authors are aware of, only four reports [3], [10]–[12] extend drusen detection and segmentation to distinguish retinal images according to whether they exhibit AMD or not. However, this work (unlike the work described in this paper) first required the identification (segmentation) of AMD pathologies (drusen) using image processing and content based image retrieval techniques. This feature identification is often founded on some form of segmentation, a subject of much continuing investigation and research.

Earlier attempts to limit the use of image segmentation for AMD screening using point and tree based approaches have been reported in [8]. The tree based approach has been found to performed well on binary-class setting. The only drawback is the computational cost is high (combinatorial complexity) to extract the features. The point based approach, which consumed less computational cost, on the other hand produced lower classification performance. The approach proposed in this paper is thus designed to produce high classification performance with lower computational cost.

III. DATA PREPARATION

Two publicly available retinal image datasets were used to evaluate the proposed approach: (i) the ARIA and (ii) the STructured Analysis of the Retina (STARE) datasets. Both data sets featured “normal” retinae, retinae that showed signs of AMD and retinae that feature Diabetic Retinopathy (DR). DR is another retina condition that leads to blindness, which is also typically identified through screening. Thus the data sets could be used for binary classification purposes (AMD vs. non-AMD) or multi-class classification purposes. ARIA has a total of 220 manually labeled images. Of these, 101 featured AMD, 59 featured DR and 60 were “normal”. A total of 174 STARE images were acquired for the work described, of these, 64 featured AMD, 72 DR and 38 “normal”. The datasets were combined to produce a single large dataset comprising 394 images, of which 165 featured AMD, 131 DR and 98 neither AMD nor DR.

The retinal images were pre-processed to enhanced the images and remove blood vessels. Blood vessels are a common structure found in retinal images and thus are deemed to be insignificant with respect to distinguishing between AMD and non-AMD images. Blood vessel removal was thus deemed desirable. Another common retinal structure is the Optic Disc (OD). However, it is difficult to achieve high OD localisation accuracy in the case of retina images that feature severely damaged retinae or images of low appearance quality. Thus the routine localisation and removal of the OD was omitted from the image pre-processing task. However, as will become apparent later in this paper, one of the proposed approaches does adopt OD removal under certain conditions. Complete details concerning the nature of the applied image pre-processing can be found in the authors’ earlier work [13].

IV. CLASSIFIER GENERATION

The proposed approach is founded on a tabular representation that utilises the basic 2-D array image format. The proposed tabular representation utilised both colour and spatial information to identify image features (defining them in terms of statistical parameters) which can be extracted either directly or indirectly from the representation. Two parameter extraction strategies were considered: (i) global extraction where the entire image is taken into consideration, and (ii) local extraction by partitioning the image down to some pre-prescribed level of decomposition ($D_{max}$) and extracting parameters on a region by region basis. We refer
to the first strategy as S1 and the second as S2. In both cases a feature selection process was applied where the top \( K \) features were chosen. The rest of this sub-section is arranged as follows. Sub-section IV-1 considers the adopted features and IV-2 the feature selection process.

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

- **Features generated directly from the pixel colour information contained in the image (six features).** The six colour features extracted were the average values for each of the RGB colour channels (red, green and blue) and the HSI components (hue, saturation and intensity). These values were computed directly from a 2-D array colour representation of each image.

- **Features generated from a “colour histogram” representing the colour information contained in the image (two features).** The two histogram based features were: (i) histogram spread and (ii) histogram skewness, \( h_{\text{spread}} \) and \( h_{\text{skew}} \), were computed as follows:

\[
h_{\text{spread}} = \frac{1}{|h|} \sum_{i=1}^{\max} (\hat{h}_i - \bar{h})^2
\]

\[
h_{\text{skew}} = \frac{1}{|h|} \sum_{i=1}^{\max} (\hat{h}_i - \bar{h})^3
\]

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

- **Features generated from the co-occurrence matrices representing the image (three features).** A co-occurrence matrix is a matrix that represents image texture information in the form of the number of occurrences of immediately adjacent intensity values that appear in a given direction \( P \) [15], [16]. Four possible different directions can be used to define \( P: 0^\circ, 45^\circ, 90^\circ \) or \( 135^\circ \). With respect to the approach described in this section, four co-occurrence matrices (one for each \( P \) direction) were generated for each image. Three textural features were then extracted from each matrix: (i) correlation, (ii) energy and (iii) entropy.

- **Features generated using a wavelet transform (four features).** A single level 2-D Discrete Wavelet Transform (DWT) was employed to generate the four wavelet based features used. The features were extracted by computing the average of four types of DWT coefficient as described in [15]. The most common Haar wavelet filter was employed in this paper. The resulting features were kept in a tabular form where each column represents a feature, and each row an image.

The first strategy (S1) was to extract these features with respect to the entire image. The second strategy (S2) was to first partition each image into \( R \) sub-regions using a quad-tree image decomposition technique, the features of interest were then generated from each sub-region. Note that in the context of the work presented in this paper, the decomposition of an image was conducted until some predefined maximum depth, \( D_{\text{max}} \), was reached. Since quad-trees are more suited to square images, the image size was first expanded so that both the height and width of the images were identical. In the context of the work described in this paper, the dimensions of each retinal image were fixed to \( 768 \times 768 \) pixels. This was achieved by expanding the images with zero valued pixels. The extracted feature vectors were then arranged according to the order of the sub-regions that they represent in an ascending manner, such that the features of the first sub-region formed the first 15 features, the second sub-region formed the next 15 features, while the \( R^{th} \) sub-region formed the last 15 features. The time complexity to generate features is thus linear, depending on the value of \( R \). Figure 2 shows the sub-region ordering of an image using a quad-tree of depth, \( D_{\text{max}} = 2 \). The value of \( R \) is thus determined by the value of \( D_{\text{max}} \) such that \( R \approx 4^{D_{\text{max}}} \).

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

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

2) Feature Selection: The next step was to reduce the number of extracted features; the aim being to prune the feature space so as to increase the classification efficiency (through removal of redundant or insignificant features) while at the same time maximising the classification accuracy. The adopted feature selection process comprised a feature ranking strategy based on the discriminatory power of each feature and selection of the best top \( K \) performing features. The feature ranking mechanism employed used Support Vector Machine (SVM) weights to rank features [17]. The main advantage of this approach was its implementational simplicity and effectiveness in identifying relevant
features.

V. RESULT AND DISCUSSION

This section presents an overview of the evaluation conducted with respect to the proposed approach considered above, Tabular, compared with the other approaches reported in [8], Point Series and Tree. Five metrics were used to compare the operation of the proposed approaches: (i) sensitivity (Sens), (ii) specificity (Spec), (iii) accuracy (Acc), (iv) Area Under the receiver operating Characteristic Curve (AUC) and (v) the False Negative Rate (FNR). Note that the evaluation of the proposed approaches was conducted using Ten-fold Cross Validation (TCV). The TCV was repeated five times and the training and test images for each TCV were randomised. Average results are thus presented in this section.

Table I presents the results obtained using the three techniques in the context of a binary classification problem (AMD vs. non-AMD). Note that the results were generated using the best parameter settings as identified from previous experiments conducted by the authors. Table II presents the results obtained in the context of a multiple class classification setting (AMD, DR and “normal”). The right most column shows the FNR produced by the proposed approaches. Best results are indicated in bold font. From Table I it can be seen that the Tabular and Tree approaches produced high classification performances of greater than 85% accuracy and greater than 90% AUC. The Tabular hold the advantage where the computational complexity to identify features is lower (linear complexity) than the Tree based (combinatorial explosion complexity). The best sensitivity and specificity were produced using the Tree based approach. The Point Series approach produced the worst results.

From Table II it can be seen that, as might be expected, an overall lower performance was recorded compared to the binary setting with the exception of the sensitivity to identify AMD (Sens-AMD) and FNR with respect to the Tree based representation. The proposed Tabular approach was not performed well in particular with respect to the classification specificity. Overall the Tree based approach outperformed the Point Series and Tabular approaches with respect to all the evaluation metrics used. These results indicated that using the tree representation, coupled with a weighted frequent sub-graph mining algorithm, is the most appropriate with respect to the classification of the retinal images for the purposes of AMD screening. The image decomposition technique used to generate the trees could contribute to such performances, where image details central to the retinal images were extracted more effectively than the quadtree image decomposition used in this paper. The Tree based approach also produced the most reliable results, with a high sensitivity value that would avoid AMD patients being mistakenly screened as being healthy. The only drawback of the Tree based approach is its computational cost to extract features is combinatorial explosion. Once features are identified, the cost is however similar to the other approaches.

Based on the presented results, the Tabular approach could be used to screen AMD from non-AMD images. Once the AMD images have been screened, the Tree based approach could be used to grade the severity of the AMD.

As already noted in Section 2 there is very little comparable reported work on the classification (screening) of retina images for AMD. The authors have only been able to identify four instances of comparable work, namely: (i) Brandon and Hoover [3], (ii) Chaum et al. [11], (iii) Agurto et al. [10] and (iv) Cheng et al. [12]. Direct comparison with this reported work is not possible because the data sets used in each case are not in the public domain, except in the case of Brandon and Hoover who used the STARE data set. However, with respect to the reported work, it can be observed that the evaluation presented in Brandon and Hoover (which was applied not only to AMD screening, but also to grade the detected AMD) achieved an overall accuracy of 90% on 97 images. The work of Chaum et al. reported an overall classification accuracy of 91.3% on 395 images. Agurto et al. reported a best recorded AUC value of 84% to identify AMD images against non-AMD images from normal eyes and eyes with DR. The results reported in Cheng et al., where only sensitivity and specificity were recorded, produced 86.3% and 91.9% respectively on 350 images.

VI. CONCLUSION

The work described in this paper proposed an alternative representation to support retina image classification with respect to AMD screening processes. The approach used a tabular representation containing purely statistical information to which standard classification techniques were applied. Comparison with two other similar approaches, the point series and tree, was also conducted. The proposed
tabular approach performance was comparable to the tree based approach on binary-class setting (sensitivity and AUC of greater than 90% each), with lower computational cost (linear complexity), and thus deemed appropriate to screen the AMD from non-AMD images. Once the AMD images are identified, the tree based approach could be used to grade the severity of AMD. Comparison with conventional approaches to AMD screening also indicated that the proposed image classification systems are ideally suited to large scale AMD screening. Future directions include provide reasoning to clinicians on how a decision was reached, and to investigate the implication of image registration to the classification performance.

ACKNOWLEDGMENT

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

REFERENCES


