Dictionary Learning-based Volumetric Image Classification for The Diagnosis of Age-related Macular Degeneration

Abdulrahman Albarrak[1], Frans Coenen[1], and Yalin Zheng[2]

[1] Department of Computer Science, University of Liverpool, United Kingdom
[2] Department of Eye and Vision Science, University of Liverpool, United Kingdom
{A.Albarrak,coenen,yzheng}@liverpool.ac.uk

Abstract. A discriminative dictionary-based approach to supporting the classification of 3D Optical Coherence Tomography (OCT) retinal images, so as to determine the presence of Age-related Macular Degeneration (AMD), is described. AMD is one of the leading causes of blindness in people aged over 50 years. The proposed approach is founded on the concept of a uniform 3D image decomposition into a set of sub-volumes where each sub-volume is described in terms of a "spatial gradient" histogram, which in turn is used to define a set of feature vectors (one per sub-volume). Feature selection is conducted using the maximum sum of the squared values of each feature vector for each sub-volume. After that, a "coding-pooling" framework is applied so that each image is represented as a single feature vector. The "coding-pooling" framework generates a representative subset of feature vectors called a dictionary, and then use this dictionary as a guide for the generation of a single feature vectors for each volume. Experiments conducted using the proposed approach, in comparison with range of alternatives, indicated that the approach outperformed other existing methods with an accuracy of 95.2%, sensitivity of 95.7% and specificity of 94.6%.

Keywords: Data mining, Image decomposition, Spatial gradient histograms, diactioanry learning, Medical image processing, Optical Coherence Tomography

1 Introduction

Age-related Macular Degeneration (AMD) is an eye condition that results in vision loss that typically effects people over fifty years of age. AMD can be identified by inspection of retinal imagery. Traditionally this is conducted using 2-dimensional (2D) colour fundus images and a number of techniques for automating this process have been proposed. Of note in this context is the work of Hijazi et al. [1], Zheng et al. [2], Liu et al. [3] and Gossage et al. [4] where data mining techniques (more specifically classification techniques) have been proposed to support 2D retinal image analysis. However, the "traditional" 2D fundus photography for AMD detection has been superseded by three-dimensional

(3D) Optical Coherence Tomography (OCT) imaging techniques. The use of 3D OCT retinal imagery can provide detailed cross-sectional information for better diagnostic purposes. However, ophthalmologists find that they are now overwhelmed by the large amount of 3D image data in their clinical practice. There is thus a requirement for automated (or semi-automated) decision-support systems for the analysis of 3D OCT retinal image data (with respect to both AMD and other eye conditions).

In order to address the above challenge a new volumetric analysis technique is proposed in this paper for the automated diagnosis (classification) of AMD using 3D OCT retinal image data. The main challenges of this work is how best to extract and represent OCT image features so that a minimal amount of discriminative information will be lost, while at the same time the ensuring compatibility with the generation of effective classifiers. A method for representing 3D OCT images for classification purposes is therefore presented. The proposed method features a 3D decomposition into a collection of sub-volumes such that each sub-volume is defined in terms of a "gradient histogram" vector. We then sum the squared values of each histogram vector and sort them in descending order. We then select the top k vectors. These selected vectors are then translated into a feature vector representation using a coding-pooling framework. The coding part of this mechanism comprises the application of "sparse coding" to form a dictionary (a dictionary is a subset of feature vectors that is representative of a complete set of feature vectors). In the pooling part "spatial max" pooling [5] is applied to the selected feature vectors so as to combined them into a single feature vector to represent a given 3D volume.

The main contributions of the work described are: (i) a mechanism for representing 3D OCT retinal data, using image decomposition and a coding-pooling framework, to support volumetric classification; and (ii) a framework to support automated retinal disease detection using 3D OCT image data.

The remainder of this paper is organised as follows. A brief review of previous research is presented in Section 2. An overview of the application domain is described in Section 3. The design of the proposed approach is described in Section 4. Section 5 then assesses the performance of the proposed approach. Finally, this paper is concluded in Section 6 with a summary of the main findings.

2 Related Work

There has been some previous work directed at feature extraction for 3D image classification. The most common approaches rely on the use of statistical feature extraction and representation such as: (i) Local Phase Quantization (LPQ) [6], (ii) Local Binary Patterns (LBP) [7], and (iii) Scale-invariant feature transform (SIFT) [8]. These methods are used to extract low level image features such as: the frequency of intensity values, changes in the image intensity values and the spatial relationships between neighbouring intensities. LPQ uses the local Fourier transform where by a histogram of the quantised Fourier transform at low frequency is computed [6]. LBPs compute the relationship between each

pixel and its immediate neighbours. However, the generation of 3D rotation invariant LBPs are computationally expensive. Zhao and Pietikainen [7] proposed the use of Three Orthogonal Plane LBPs (LBP-TOP). The LBP-TOP representation considers the calculation of LBPs only with respect to neighbouring voxels located in the XY, XZ and YZ planes. SIFT computes the orientation histograms of the image gradient directions. Most of these methods use a global feature vector which is produced by concatenating a range of feature values. This strategy has the disadvantage that some features may be redundant.

In more advanced approaches, feature selection strategies are employed to select only the most discriminative features. For instance, the concept of Dictionary Learning (DL) has drawn the attention of many researchers in computer vision and image classification [9, 5, 10]. To use DL a subset of a given set of feature vectors is selected to form a dictionary (or codebook). The challenge of DL is the identification of a highly discriminative subset of the available vectors. There are two ways to address this challenge. The first is to use a sparse representation. For instance, in the Maximisation of Mutual Information (MMI) method [11] a Gaussian Process (GP) is combined with k-means Singular Value Decomposition (K-SVD) to optimise the dictionary learning. The second is to use a coding-pooling framework. This framework is argued to be one of the most robust ways to represent images for classification in domains such as face recognition [12]. In the coding part of this framework vectors with similar features are clustered (using, for example, the k-means algorithm), an approach called "vector quantisation" [13]. Alternatively, sparse coding may be used to select a subset of vectors, in this case the selected vectors are said to form a "dictionary" or "codebook". Sparse coding tries to find a vector that represents a group of vectors by measuring the "response" of the vector to the group. The basic idea is to apply sparse coding on a random sample of extracted image descriptors such as SIFT [13], in order to identify a highly discriminative set of features. Then multi-scale spatial max pooling is used to form a single feature vector from a set of sparse feature vectors. These feature vectors can be input into a traditional classifier such as a Support Vector Machine (SVM). An example method that uses the coding-pooling framework is Linear Spatial Pyramid Matching Using Sparse Coding (ScSPM) described in [5] for 2D image classification. Alternatively, Locality-constrained Linear Coding (LLC) may be used instead of sparse coding to achieve the same goal [14]. In this paper, a coding-pooling framework is adopted with respect to the work described in this paper (see below).

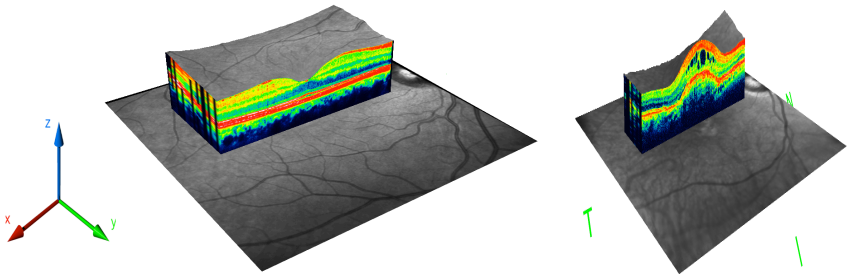
3 Application Domain

The work described in this paper is directed at the detection of retinal disease and in particular the identification of Age-related Macular Degeneration (AMD) in 3D OCT images. Analogous to ultrasound, OCT is a relatively new imaging technique that can capture cross-sectional information of the retina. OCT employs low-coherence light and ultrashort laser pulses to detect the spatial position of tissue and resolve depth information. The employment of light waves

enables acquisition of images (volumes) with very high resolution that can reveal precise details of internal structures. A series of 2D "slices" are acquired to form a 3D cross-sectional volume.

AMD is a condition, typically contracted in old age, which leads to irreversible vision loss at its advanced stages. This loss of vision for AMD patients is due to the damage to the macula, the centre of the retina that facilitates high level visual activities such as reading and recognition of faces [15]. There are some distinct features of AMD that can be readily identified in OCT image data such as: (i) disturbance of the Retinal Pigment Epithelium (RPE) layer underneath the neuro-retina due to the presence of "drusen" (fatty deposits), pigment epithelium detachment or geographic atrophy; (ii) disruption of layered neuro-retinal tissue; (iii) the presence of intra- and sub-retinal fluid and (iv) retinal thickening.

Two example OCT volumes, one normal and one featuring AMD, are presented in Figure 1. Figure 1(a) shows a 3D image of a normal OCT volume and Figure 1(b) shows a 3D OCT volume that features AMD. From these figures it can be seen that there are notable distinctions between the normal and the AMD volumes. The normal volume features smooth and connected layers. However, the AMD volume features disrupted layers and other abnormal patterns, such as thickening of the RPE layer, the presence of intra-retinal fluid, pigment epithelium detachment and some unusual texture patterns.



(a) A normal retina (b) An AMD retina

Fig. 1. The 3D OCT of a "normal" and an AMD retina.

Most of the current macular disease diagnosis tools have been directed at 2D images. Instances can be found in [1], [2], [3] and [4]. In [1] and [2], a graph based image representation method was proposed. The image data is decomposed into a set of quadtrees. These quadtrees are analysed using frequent sub-graph mining. Discriminative frequent sub-graphs are selected and used for generating a feature vector for each image. These feature vectors are then used to generate a classifier. In [3] Liu et al. designed a method for automatic detection of retinal diseases including AMD. A Multi-Scale Spatial Pyramid (MSSP) representation, with different levels, was used. The histograms of the LBPs were applied on each

sub-block of the MSSP at each level. Dimensionality reduction using Principal Component Analysis (PCA) was also utilised. All the generated LBP's were concatenated together so as to build a global feature vector descriptor. The Radial Basis Function (RBF) kernel based SVM classifier was then used for categorising the feature vectors. In [4] a texture based method was employed using a combination of two methods, namely the Spatial Gray-Level Dependence Matrix (SGLDM) and the Discrete Fourier Transform (DFT) for extracting the OCT image features. Then a statistical method was used to extract the features of the SGLDM such as energy, entropy, correlation, local homogeneity, and inertia. A Mahalanobis distance based method was applied to measure the similarities between image features and then a Bayesian classifier was used to differentiate between features. Using SGLDMs different matrices are typically computed according to selected directions, which leads to long feature vectors with some similar values, which may in turn adversely affect classifier performance.

Recently, a 3D method for the classification of OCT was proposed [16] where a 3D decomposition was employed to support image classification which is similar to the proposed approach in this paper. The image is first decomposed into a set of sub-volumes and then a combination of LBP and Histogram of Oriented Gradients (LBP-HOG) used. First the LBP features were computed for every sub-volume. Following this, the gradient of the LBP features were calculated and the complete set of features normalised. These features were then concatenated forming a single feature vector for every image. Principal Component Analysis (PCA) was used to reduce the dimensionality of the feature vector. A classifier was then applied to categorise the feature vectors. In this research, instead of using PCA as in [16], a coding-pooling framework is adopted. In addition, an extension of the graph based methods for 3D classification described in [17, 18] was used.

4 PROPOSED APPROACH

The proposed method relies on the coding-pooling framework described in Section 2. The proposed approach comprises the following stages: (i) preprocessing, (ii) image decomposition and feature extraction, (iii) feature vector coding, (iv) pooling, and (v) classification. Each of these stages is discussed in further detail in the following five sub-sections. Note that the input to the proposed process is a set of 3D OCT images $I = \{i_1, ..., i_n\},\ i_n \in I$, associated with a set of class labels $C = \{c_1, ..., c_n\},\ c_n \in \{-1, 1\}$. The process of feature vector generation is described in Algorithm 1. In the first stage (line 2) the current volume is enhanced forming a Volume of Interest (VOI) and then (stage 2, line 3) this VOI is decomposed to form a set of sub-volumes. Then feature extraction (low level feature extraction) is applied to form a set of Feature Vectors (FV) for each sub-volume in I (lines 4 to 6). The feature vectors are then ranked, using a "max-sum" calculation, and the top k selected (line 8). A dictionary is then built using sparse coding (stage 3, line 9). Following this, pooling is used to form

the final feature representation (stage 4, lines 10 to 12) which is then fed into a classifier generator (stage 5).

Algorithm 1 Pseudocode for the proposed feature vector generation approach

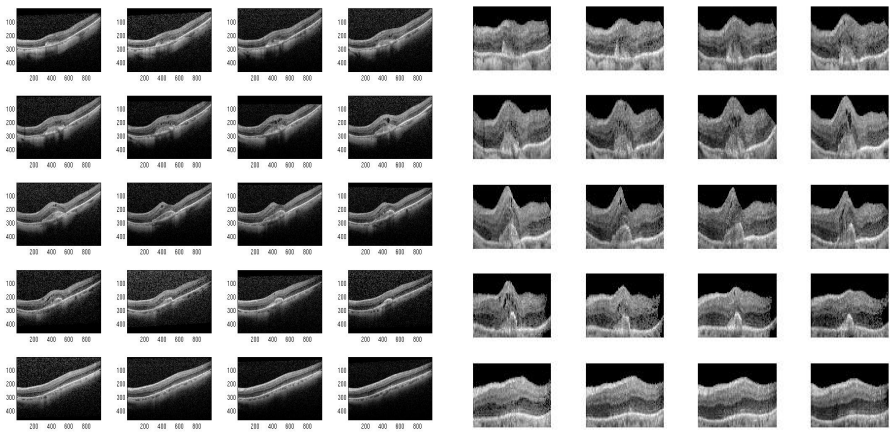
Input: Volumes I
Output: FeatureVector fv_{z_i}
1: **for each** Volume i_n **do**
2: VOI $\leftarrow$ preprocessing(Volume i_n)
3: Sub-Volumes $\leftarrow$ decomposition(VOI)
4: **for each** Sub-Volume j **do**
5: FV $\leftarrow$ feature-extraction(Sub-Volume j)
6: **end for**
7: **end for**
8: $X_n \leftarrow$ MaxSum(FV)
9: (V, U) $\leftarrow$ Coding(X_n) Using algorithm 2
10: **for each** Volume i_n **do**
11: $fv_{z_i} \leftarrow$ pooling ($X_n(i_n)$, V, $U(i_n)$)
12: **end for**

4.1 Preprocessing

The quality of OCT 3D images is usually affected by factors such as unwanted structures and alignments. The preprocessing stage is thus intended to improve the quality of the image and reference the images to a uniform coordinate system. Three steps are applied during the preprocessing stage. First of all, in order to remove unwanted structures from the input retinal images, two steps are used. Firstly a Split Bregman Isotropic Total Variation (ITV) algorithm, developed by [19], is applied to every slice of the 3D volume. Secondly, morphological operators are used: (morphological opening is applied in order to remove small objects not connected to the main retina and morphological closing to fill gaps).

The next step is to flatten the retina object. Flattening is applied using a second order polynomial least-square curve fitting procedure [3] according to the nature of the mean surface of the retina (defined according to the top and bottom retina surfaces). In order to do this we select the slice where the top and bottom surfaces of the volume (retina) are furthest apart and consider these two layers in terms of two vectors made up of voxel values. These two vectors are used to define the "middle" vector which is then used as a reference for flattening the entire retinal volume.

Finally, after flattening, a Volume of Interest (VOI) is defined. The basic idea is to select the top and bottom surface for every slice. These surfaces are then used to define the minimum value for the top and the maximum value for the bottom of the VOI. Figure 2 shows two sets of images describing a 3D OCT retina volume before preprocessing (Sub-figure 2(a)) and the VOI after preprocessing (Sub-figure 2(b)).



(a) Giving 3D image (b) After pre-processing

Fig. 2. Example of preprocessed image.

4.2 Image Decomposition and Feature Extraction

During the image decomposition and feature extraction stage each volume is decomposed into a set of sub-volumes with a "patch size" of $16 \times 16 \times 16$ voxels. For each sub-volume the gradients ∇x, ∇y, ∇z are computed in three dimensions and then normalised to give the sum of the gradients using Equation 1. The "angles" are extracted using Equation 2 [8] to give values of between 0 and 2π, the number of angles was set to 8. For each voxel, an orientation histogram is computed, $hist(angle) = hist(angle) \times magnitude$.

$$magnitude = \sqrt{\nabla x^2 + \nabla y^2 + \nabla z^2} \tag{1}$$

$$angle = atan^{-1}(\frac{\nabla z}{\sqrt{\nabla x^2 + \nabla y^2}}) \tag{2}$$

In order to identify AMD ophthalmologist usually inspect an OCT scan centred on the area around the fovea (the centre of the retina) as this area tends to show more diagnostic information than the remainder of the retina. In order to consider the importance of spatial location, in our approach each sub-volume is given a weight according to its spatial 3D distance from the centre of the retina to the centre of it using Equation 3 where dtc is the sub-volume distance to the centre of the volume and ϵ is a small constant value. A feature vector fv is then formed for each sub-volume: $fv = weights \times hist$. In each case the generated fv is then normalised.

$$weight = \frac{\sqrt{\frac{width}{2} + \frac{height}{2} + \frac{depth}{2}}}{abs(dtc) + \epsilon} \tag{3}$$

4.3 Feature Vector Coding

During stage 3 of the feature vector generation process, sparse coding is applied to the complete set of feature vectors ($FV = \{fv_1, \ldots., fv_m\}$, where m is the total number of feature vectors for all images) describing the entire image set, produced during the previous stage. The objective is to create a dictionary comprised of a representative subset of FV; the feature vectors in the dictionary should be selected so as to maximise the discriminative power of the eventual image classifier. One of the issues with the sparse coding based methods is that they have a high time complexity associated with them, thus in our approach a subset X_n of the FV for a given image n is used. Sparse coding is then applied to X_n. In [5], the authors proposed to generate the set X_n in a random manner. However, it is conjectured in this paper that random sampling may include feature vectors with less discriminative power which could adversely affect the performance of the eventual classifier. In order to address this problem, in the proposed approach, the sum of the squared value of each feature vector in FV is calculated (Equation 4). The set FV is then ranked, in descending order, and the top k selected to form the set X_n.

$$\max_{n} \sum_{i=1}^{m} FV_i \tag{4}$$

Recall that the basic idea behind the concept of sparse coding is to represent a set of feature vectors in terms of a representative sub-set of these vectors. Thus, after the subset of feature vectors X_n has been identified, sparse coding is applied to X_n to form a dictionary V. More specifically the feature vectors in X_n are combined into k representative feature vectors which then defines our dictionary V ($V = \{rfv_1, \ldots, rfv_k\} \subset X_n$). For each $rfv_i \in X_n$ we associate a "feature vector indicator" u_i ($u_i \in U$, where U is the complete set of feature vector indicators) so that each fc_i is linked to the dictionary V (an approach informed by the work of [5]). The problem of optimising the selection of V and U are solved iteratively using a feature-sign search [20]. At the end of this process each feature vector in X_n is associated with its indicator in U and the final dictionary V. Algorithm 2 illustrates how the dictionary V and the complete set of feature vector indicators U are generated. In the algorithm, the dictionary size is set to 1024. The algorithm commences by selecting an initial dictionary randomly $initialV$ from the subset of vectors X_n (line 6). Then the mean values of $initialV$ are subtracted (line 7). In each iteration, in our experiment the number of iterations was set to 5, another random subset of feature vectors $initialB$ is selected (line 12). The gradient is then computed (line 16), which is for measuring the change between the two sets $initialV$ and $initialB$. The gradient is the multiplication of the sparse value of the $initialV$ minus the new set $initialB$. The maximum values of the gradient are selected. The associated indicator U is computed by getting the vectors that minimising the error and maximising the gradient while ensuring the maximum gradient is less than a given threshold (lines 20 to 25). Eventually, dictionary V is generated

by considering the best set of feature vectors from the initial dictionaries $initialV$ and$initialB$ (line 31). Each vector in V is linked to U where the minimum error is reached (lines 26 to 33)).

Algorithm 2 Pseudocode for the dictionary generation approach [5, 20].

Input: X_n
Output: V, U
1: $DictSize = 1024$
2: β = 1e-5
3: $\gamma = 0.15$
4: σ = eye ($DictSize$).
5: $itersNumber = 5$
6: $initialV$ = selected randomaly a subset of size $DictSize$ from X_n
7: $initialV = initialV$ - mean ($initialV$)
8: $initialV = initialV \times diag(1/\sqrt{\sum(initialV.*initialV)})$
9: **for each** i = 1 to $itersNumber$ **do**
10: $A = initialV^T \times initialV + 2 \times \beta \times \sigma$.
11: x = inital a vector with size of A
12: $initialB$ = select randomly from X_n
13: **for each** j = 1 to number vectors in $initialV$ **do**
14: $bb = initialB_j$
15: $b = -initialVT \times bb$
16: $grad = A \times sparse(x) + b$;
17: $ma = \max(abs(grad). * x)$;
18: $x = (\gamma - grad)/A$
19: $EPS = 1e - 9$
20: **while** $ma \leq \gamma \times EPS$ **do**
21: $U_j = \min(0.5 \times x^T \times A \times x + b^T \times x + \gamma \times |x|)$
22: $grad = A \times sparse(x) + b$;
23: $ma = \max(abs(grad). * x)$;
24: $x = (\gamma - grad)/A$
25: **end while**
26: $bu = initialB \times U^T$
27: $uu = U \times U^T$
28: $dual = diag(bu - uu)$
29: $bs = \sum initialB^2$
30: $uinv = inv(uu + diag(dual))$
31: $x = mintrace((bu \times uinv \times bu^T) + bs - \sum dual)$
32: $V = (uu \times diag(x)/bu^T)^T$
33: **end for**
34: **end for**

4.4 Pooling

The next stage of feature vector generation (stage 4) is data pooling where a single feature vector is generated for each 3D OCT retina volume. Data pooling is a statistical method used to concatenate different feature vectors into a single

feature vector. To this end a pooling function is used to combine a collection of different feature vectors into a signal vector. Different factors should be taken into account such as the spatial location and similarities between vectors. Thus, given our subset of feature vectors X_n, and the set of feature vector indicators U linking the elements of X_n with our dictionary V (generated in stage 3 as described in the previous section) the objective is to build a single discriminative feature vector for each volume.

With respect to the work described in this paper spatial max pooling [5] was adopted to encapsulate a set of feature vectors X_i representing a single volume ($X_i \subset X_n$) into a single feature vector guided by the dictionary V. Max pooling recursively computes the histograms of the maximum values for a given set of vectors X_i and their association with elements in V. Equation 5 shows the pooling function used to derived the maximum values with respect to the set U which is then used to generate a new histogram z_r for each region r with M sub-volume. Neighbouring regions are recursively united and then Z_r is applied on each one of them until reaching a final feature vector. On completion of the process a single feature vector fv_{z_i} is obtained that describes each data volume i (retinal volume with respect to the motivation for this paper).

$$z_r = \max\{|u_{1r}|, ...|u_{Mr}|\} \tag{5}$$

4.5 Classification

From the foregoing, a single feature vector (fv_{z_i}) is used to describe each data volume i (retinal volume with respect to the motivation for this paper). For training purposes each feature vector fv_{z_i} was combined with a class label c_i (in our case +1 indicates a retina with AMD, -1 for normal retina, informed by medical retina experts) to create a training set. The resulting representation is compatible with a number of classifier generators. However, linear SVM [21] was used in this paper due to its reasonably good performance in most application. The results of the evaluation are presented in the next section.

5 RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed approach experiments were conducted using 140 3D OCT volumes, 68 "normal" and the remainder AMD. The size of each volume was about $(1024 \times 496\ pixels) \times 19\ slices$ representing a $6 \times 6 \times 2$ mm retinal volume. Ten-fold cross validation was used to evaluate the proposed method. Four metrics were recorded with which to measure the performance of the proposed algorithm: accuracy (Acc) (Equation 6) , sensitivity (Sen) (Equation 7), specificity (Spec) (Equation 8) and the Area Under the receiver operator characteristic Curve (AUC). AUC can reflect the trade-off between sensitivity and specificity.

The experiments were directed at analysing the proposed approach in comparison to approaches using alternative representations taken from the literature,

namely: LBP-TOP [7], LPQ [6], SIFT [8], MMI [11], ScSPM [5] and LBP-HOG [16]. We also compared with MSSP which is a 2D based approach described in [4]. With respect to MSSP, the middle slices from the training and test data were used (because these have a high potential for including part of the fovea, the part of the retina where most indicators of AMD are likely to occur). In addition, the LBP-HOG based methods for 3D image classification [16] was included in the experiments. For each different approach, the SVM parameter was tuned for the best performance.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Sen = \frac{TP}{TP + FN} \tag{7}$$

$$Spec = \frac{TN}{TN + FP}. \tag{8}$$

The results are presented in Table 1. From the table it can be seen that the proposed approach produces better results than the other reported methods. If we compare the accuracy of the proposed method with the other two methods, the proposed method produced a better performance (with a recorded accuracy of 95.2%) while for the LBP-TOP, LPQ, SIFT, MMI, ScSPM, MSSP and LBP-HOG methods accuracy values of 88.6%, 92.9%, 90.4%, 81.3%, 93.8%, 89.8% and 91.4% were recorded respectively. In the context of sensitivity, as shown in Table 1, the proposed approach has a similar result with respect to ScSPM (a recorded sensitivity of 95.7% compared with a recorded sensitivity of 95.3% for ScSPM). MMI performed less well than the proposed approach. In terms of the 3D methods LBP-TOP, LPQ and SIFT, the first two produced a recorded sensitivity of 93.3% compared to 86.0% using SIFT. This sensitivity result for the 2D based method, MSSP, was 87.3%; which was better than the MMI result. In terms of specificity, the proposed approach has a good performance with a recorded specificity of 94.6% with respect to other method: LBP-TOP with 85.0%, LPQ with 91.9%, 3D SIFT with 94.1%, MMI with 79.56%, ScSPM with 92.4% and MSSP with 91.3% and 90.5% for LBP-HOG. The proposed approach also produces better results than the other methods measured by the AUC. The proposed method produced a AUC value of 0.95 while the recorded AUC values with respect to the other methods considered were as follows: LBP-TOP with 0.88, LPQ with 0.92, 3D SIFT with 0.90, MMI with 0.81, ScSPM with 0.93 and MSSP giving 00.89. The AUC value obtained when using LBP-HOG was 0.94.

The results presented in Table 1 also support the idea that by using some guided function such as the sum of the squares with the coding-pooling framework to do image classification, is better than using random selection. In addition it can be noted that discriminative feature extraction methods help to improve the performance of the classifier. Overall LPQ and SIFT, produce a good performance with respect to our dataset. However, the proposed method indicated a better performance.

Table 1. Comparison of the proposed approach with the other methods.

Method	Acc	Sen	Spec	AUC
LBP-TOP	88.6%	**93.3%**	85.0%	0.88
LPQ	**92.9%**	**93.3%**	91.9%	0.92
SIFT	90.4%	86.0%	**94.1%**	0.90
MMI	81.3%	83.3%	79.5%	0.81
ScSPM	**93.8%**	**95.3%**	**92.4%**	**0.93**
MSSP	89.8%	87.3%	91.3%	0.89
LBP-HOG	91.4%	92.4%	90.5%	0.94
Proposed method	**95.2%**	**95.7%**	**94.6%**	**0.95**

6 CONCLUSION

In this paper a new approach to classifying 3D OCT volumes has been proposed with an application in the diagnosis of AMD (AMD vs. non-AMD). More specifically the approach is founded on the use of a coding-pooling framework to identify discriminative features within OCT sub-volumes. The results obtained using the proposed approach demonstrated a better performance in comparison with: LBP-TOP, LPQ, SIFT, sparse based representation methods such as MMI and ScSPM, and the MSSP 2D based method. This research has also shown that using a value, such as the sum of the squares, for the selection of a subset of the available set of feature vectors will improve the performance of the classifier. A further study will be directed at evaluating our approach with respect to different retinal diseases such as diabetic retinopathy. The authors intend to also conduct further research to explore the effect of using different feature selection methods for the coding-pooling framework, and evaluate the proposed approach in larger datasets.

References

1. Hijazi, M.H.A., Coenen, F., Zheng, Y.: Data mining techniques for the screening of age-related macular degeneration. Knowledge-Based Systems **29** (2012) 83–92
2. Zheng, Y., Hijazi, M.H.A., Coenen, F.: Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach. Investigative Ophthalmology & Visual Science **53**(13) (2012) 8310–8318
3. Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J., Rehg, J.M.: Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. Medical Image Analysis **15**(5) (2011) 748–759
4. Gossage, K., Tkaczyk, T., Rodriguez, J., Barton, J.: Texture analysis of optical coherence tomography images: feasibility for tissue classification. Journal of Biomedical Optics **8**(3) (2003) 570–575
5. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2009) 1794–1801

6. Paivarinta, J., Rahtu, E., Heikkilä, J.: Volume local phase quantization for blur-insensitive dynamic texture classification. In: Proceedings of the 17th Scandinavian conference on Image analysis. SCIA'11, Berlin, Heidelberg, Springer-Verlag (2011) 360–369
7. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(6) (2007) 915 –928
8. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition. In: Proceedings of the 15th International Conference on Multimedia. (2007) 357–360
9. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. (2009) 689–696
10. Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher discrimination dictionary learning for sparse representation. In: IEEE International Conference on Computer Vision. (2011) 543–550
11. Qiu, Q., Jiang, Z., Chellappa, R.: Sparse dictionary-based representation and recognition of action attributes. In: IEEE International Conference on Computer Vision. (2011) 707–714
12. Wang, Z., Feng, J., Yan, S., Xi, H.: Linear distance coding for image classification. IEEE Transactions on Image Processing **22**(2) (2013) 537–548
13. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2006) 2169–2178
14. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2010) 3360–3367
15. Jager, R.D., Mieler, W.F., Miller, J.W.: Age-related macular degeneration. New England Journal of Medicine **358**(24) (2008) 2606–2617
16. Albarrak, A., Coenen, F., Zheng, Y.: Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction. In: Proceedings of the 17th Medical Image, Understanding and Analysis Conference. (2013) 59–64
17. Albarrak, A., Coenen, F., Zheng, Y.: Classification of volumetric retinal images using overlapping decomposition and tree analysis. In: IEEE 26th International Symposium on Computer-Based Medical Systems. (2013) 11–16
18. Albarrak, A., Coenen, F., Zheng, Y., Yu, W.: Volumetric image mining based on decomposition and graph analysis: An application to retinal optical coherence tomography. In: IEEE 13th International Symposium on Computational Intelligence and Informatics. (2012) 263–268
19. Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: Segmentation and surface reconstruction. Journal of Scientific Computing **45**(1-3) (2009) 272–293
20. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In Schölkopf, B., Platt, J., Hoffman, T., eds.: Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA (2007) 801–808
21. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machine, 2001. Software available at http://www. csie. ntu. edu. tw/˜cjlin/libsvm