

Classification of SD-OCT Volumes using Local Binary Patterns: Experimental Validation for DME Detection

Guillaume Lemaître^{1,a,*}, Mojdeh Rastgoo^{1,a,*}, Joan Massich^{1,*}, Carol Y. Cheung^c, Tien Y. Wong^c, Ecosse Lamoureux^c, Dan Milea^c, Fabrice Mériaudeau¹, Désiré Sidibé¹

^a*ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain*

^b*LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, 12 rue de la Fonderie, 71200 Le Creusot, France*

^c*Singapore Eye Research Institute, Singapore National Eye Center, Singapore*

Abstract

This paper addresses the problem of automatic classification of Spectral Domain OCT (SD-OCT) data for automatic identification of patients with Diabetic Macular Edema (DME) versus normal subjects. Optical Coherence Tomography (OCT) has been a valuable diagnostic tool for DME, which is among the most common causes of irreversible vision loss in individuals with diabetes. Here, a classification framework with five distinctive steps is proposed and we present an extensive study of each step. Our method considers combination of various pre-processings in conjunction with Local Binary Patterns (LBP) features and different mapping strategies. Using linear and non-linear classifiers, we tested the developed framework on a balanced cohort of 32 patients.

Experimental results show that the proposed method outperforms the previous studies by achieving a Sensitivity (SE) and Specificity (SP) of 81.2% and 93.7%, respectively. Our study concludes that the 3D features and high-level representation of 2D features using patches achieve the best results. However, the effects of pre-processing is inconsistent with respect to different classifiers

☆Document source available in GitHub [1]

*Corresponding author

Email addresses: g.lemaître58@gmail.com (Guillaume Lemaître),
mojdeh.rastgoo@gmail.com (Mojdeh Rastgoo), joan.massich@u-bourgogne.fr
(Joan Massich)

and feature configurations.

Keywords: Diabetic Macular Edema, Optical Coherence Tomography, DME, OCT, LBP

1. Introduction

Eye diseases such as Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) are the most common causes of irreversible vision loss in individuals with diabetes. Just in United States alone, health care and associated costs related to eye diseases are estimated at almost \$500 M [2]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [3]. Early detection and treatment of DR and DME play a major role to prevent adverse effects such as blindness. Indeed, the detection and diagnosis of retinal diseases are based on the detection of vascular abnormalities or lesions in the retina.

In past decades, Computer Aided Diagnosis systems devoted to ophthalmology, have been developed focusing on the automatic analysis of fundus images [4, 5]. However, the use of fundus photography is limited to the detection of signs which are correlated with retinal thickening such as hard and soft exudates, hemorrhages or micro-aneurysms except if multiple fundus images are used [6]. However, only qualitative information can be inferred. Moreover, DME is characterized as an increase in retinal thickness within 1 disk diameter of the fovea center with or without hard exudates and sometimes associated with cysts [7]. Therefore, fundus photography cannot always identify the clinical signs of DME; for example cysts, which are not visible in the retinal surface. In addition, it does not provide any quantitative measurements of retina thickness or information about cross-sectional retinal morphology.

Recently, Optical Coherence Tomography (OCT) has been widely used as a valuable diagnosis tool for DME detection. OCT is based on optical reflectivity and produces cross-sectional and three-dimensional images of the central retina, thus allowing quantitative retinal thickness and structure measurements.



Figure 1: Example of SD-OCT images for normal (a) and DME patients (b)-(c) with cyst and exudate, respectively.

The new generation of OCT imaging, namely Spectral Domain OCT (SD-OCT) offers higher resolution and faster image acquisition over conventional time domain OCT. SD-OCT can produce 27,000 to 40,000 A-scans/second with an axial resolution ranging from $3.5\mu\text{m}$ to $6\mu\text{m}$ [8]. Figure 1 shows one normal B-scan and two abnormal B-scans.

Many of the previous works on OCT image analysis have focused on the problem of retinal layers segmentation, which is a necessary step for retinal thickness measurements [9, 10]. However, few have addressed the specific problem of DME and its associated features detection from OCT images.

In this research we focus on the latter problem and propose an automatic framework for identification of DME patients versus normal subjects using OCT volumes. The proposed method, which is an extension of our previous work [11], is based on Local Binary Patterns (LBP) features to describe the texture of OCT images and dictionary learning using the Bag-of-Words (BoW) models [12]. We propose to extract 2D and 3D LBP features from OCT images and volumes, respectively. The LBP descriptors are further extracted from the entire sample or local patches within individual samples. In this research beside

the comparison of 2D and 3D features, we also compare the effects of common
45 pre-processing steps for OCT data, study the optimal configuration regarding
the BoW approach in conjunction with different base classifiers.

This paper is organized as follows: Sect. 2 presents a summary of the related
studies; the proposed framework is explained in Sect. 3, while the experiments
and results are discussed in Sect. 4. Finally, the conclusion and avenue for future
50 directions are drawn in Sect. 6.

2. Related Work

This section reviews the works straightly addressing the problem of classify-
ing OCT volumes as normal or abnormal. A summary can be found in Table 1.

Srinivasan *et al.* [13] proposed a classification method to distinguish DME,
55 Age-related Macular Degeneration (AMD) and normal SD-OCT volumes. The
OCT images are pre-processed by reducing the speckle noise by enhancing the
sparsity in a transform-domain and flattening the retinal curvature to reduce
the inter-patient variations. Then, Histogram of Oriented Gradients (HOG) are
extracted for each slice of a volume and a linear Support Vector Machines (SVM)
60 is used for classification. On a dataset of 45 patients equally subdivided into the
three aforementioned classes, this method leads to a correct classification rate
of 100%, 100% and 86.67% for normal, DME and AMD patients, respectively.

Venhuizen *et al.* proposed a method for OCT images classification using
the BoW models [14]. The method starts with the detection and selection of
65 keypoints in each individual B-scan, by keeping the most salient points corre-
sponding to the top 3% of the vertical gradient values. Then, a texton of size
 9×9 pixels is extracted around each keypoint, and Principal Component Anal-
ysis (PCA) is applied to reduce the dimension of every texton to get a feature
vector of size 9. All extracted feature vectors are used to create a codebook
70 using k -means clustering. Then, each OCT volume is represented in terms of
this codebook and is characterized as a histogram that captures the codebook
occurrences. These histograms are used as feature vector to train a Random
Forest (RF) with a maximum of 100 trees. The method was used to classify

OCT volumes between AMD and normal cases and achieved an Area Under the
75 Curve (AUC) of 0.984 with a dataset of 384 OCT volumes.

Liu *et al.* proposed a methodology for detecting macular pathology in OCT
images using LBP and gradient information as attributes [15]. The method
starts by aligning and flattening the images and creating a 3-level multi-scale
spatial pyramid. The edge and LBP histograms are then extracted from each
80 block of every level of the pyramid. All the obtained histograms are concate-
nated into a global descriptor whose dimensions are reduced using PCA. Finally
a SVM with an Radial Basis Function (RBF) kernel is used as classifier. The
method achieved good results in detection OCT scan containing different pathol-
ogy such as DME or AMD, with an AUC of 0.93 using a dataset of 326 OCT
85 scans.

As stated in the previous section, our current research is an extension of our
previous work [1] with further contributions and evaluations at every stages of
our classification framework.

Ref	Diseases			Data size	Pre-processing				Features	Representation	Classifier	Evaluation		
	AMD	DME	Normal		De-noise	Flatten	Aligning	Cropping				Sensitivity (SE)	Specificity (SP)	AUC
[13]	✓	✓	✓	45	✓	✓		✓	HOG		linear-SVM	86.7%,100%,100%		
[14]	✓		✓	384					Texton	BoW, PCA	RF			0.984
[15]	✓	✓	✓	326		✓	✓		Edge, LBP	PCA	SVM-RBF			0.93
[11]		✓	✓	62	✓				LBP-LBP-TOP	PCA, BoW, histogram	RF	87.5%	75%	

9

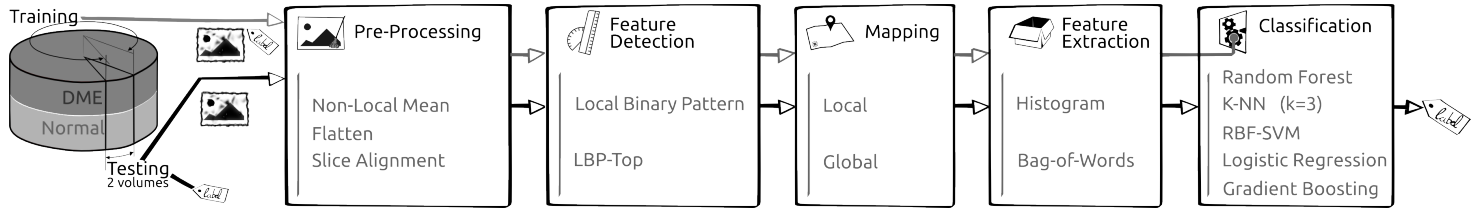


Figure 2: Our proposed classification pipeline.

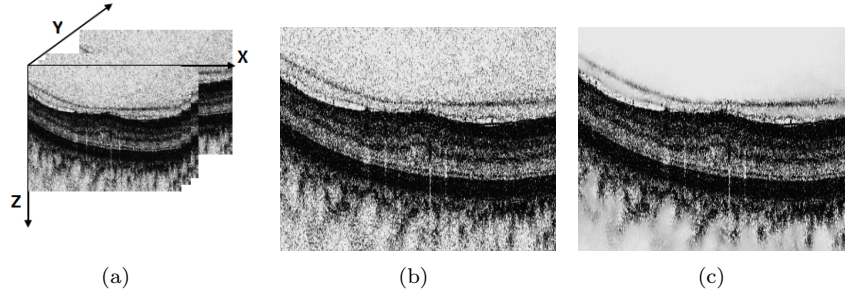


Figure 3: OCT: (a) Organization of the OCT data - (b) Original image - (c) NL-means filtering. Note that the images have been negated for visualization purposes.

3. Materials and Methods

90 The proposed method, as well as, its experimental set-up for OCT volume classification are outlined in Fig. 2. The methodology is formulated as a standard classification procedure which consists of five steps. First, the OCT volumes are pre-processed as presented in details in Sect. 3.1. Then, LBP and LBP-TOP features are detected, mapped and extracted as discussed in depth in Sect. 3.2, 95 Sect. 3.3, and Sect. 3.4, respectively. Finally, the classification step is presented in Sect. 3.5.

3.1. Image pre-processing

This section describes the set of pre-processing techniques which aim at enhancing the OCT volume. The influence of these pre-processing methods and 100 their possible combinations are extensively studied in Sect. 4.3-??.

3.1.1. Non-Local Means (NL-means)

OCT images suffer from speckle noise, like other image modalities such as Ultra-Sound (US) [16]. The OCT volumes are enhanced by denoising each B-scan (i.e. each $x - z$ slice) using the NL-means [17], as shown in Fig. 3. 105 NL-means has been successfully applied to US images to reduce speckle noise and outperforms other common denoising methods [18]. NL-means filtering preserves fine structures as well as flat zones, by using all the possible self-predictions that the image can provide rather than local or frequency filters such as Gaussian, anisotropic, or Wiener filters [17].



Figure 4: Flattening procedure: (a) original image, (b) thresholding, (c) median filtering, (d) curve fitting, (e) warping, (f) flatten image.

3.1.2. Flattening

Textural descriptors characterize spatial arrangement of intensities. However, the OCT scans suffer from large type of variations: inclination angles, positioning, and natural curvature of the retina [15]. Therefore, these variations have to be taken into account to ensure a consistent characterization of the tissue disposition, regardless of the location in the retina. This invariance can be achieved from different manners: (i) using a rotation invariant descriptor (cf. Sect. 3.2), or (ii) by unfolding the curvature of the retina. This latter correction is known as image flattening which theoretically consists of two distinct steps: (i) estimate and fit the curvature of the Retinal Pigment Epithelium (RPE) and (ii) warp the OCT volume such that the RPE becomes flat.

Our correction is similar to the one of Liu *et al.* [15]: each B-scan is thresholded using Otsu's method followed by a median filtering to detect the different retina layers (see Fig 4(c) and Fig 4(b)). Then, a morphological closing and opening is applied to fill the holes and the resulting area is fitted using a second-order polynomial (see Fig. 4(d)). Finally, the scan is warped such that the curve becomes a line as presented in Fig. 4(e) and Fig. 4(f).

Table 2: Number of patterns ($LBP_{\#pat}$) for different sampling points and radius ($\{P, R\}$) of the LBP descriptor.

	Sampling point for a radius ($\{P, R\}$)		
	$\{8, 1\}$	$\{16, 2\}$	$\{24, 3\}$
$LBP_{\#pat}$	10	18	26

3.1.3. Slice alignment

The flattening correction does not enforce an alignment through the OCT volume. Thus, in addition to the flattening correction, the warped curves of each B-scan are positioned at the same altitude in the z axis.

3.2. Feature detection

In this research, we choose to detect simple and efficient LBP texture features with regards to each OCT slice and volume. LBP is a texture descriptor based on the signs of the differences of a central pixel with respect to its neighboring pixels [19]. These differences are encoded in terms of binary patterns as in Eq. (1):

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where g_c , g_p are the intensities of the central pixel and a given neighbor pixel, respectively. P is the number of sampling points in the circle of radius R . Ojala *et al.* further extended the original LBP formulation to achieve rotation invariance at the expense of limiting the texture description to the notion of circular “uniformity” [19]. Volume encoding is later proposed by Zhao *et al.* by computing LBP descriptors in three orthogonal planes, so called LBP-TOP [20].

In this research, we consider rotation invariant and uniform LBP and LBP-TOP features with various sampling points (i.e., $\{8, 16, 24\}$) with respect to different radius, (i.e., $\{1, 2, 3\}$). The number of patterns ($LBP_{\#pat}$) in regards with each configuration is reported in Table 2.



Figure 5: *Global* (a)-(b) and *local* (c)-(d) mapping for LBP and LBP-TOP features (2D B-scan and 3D volume, respectively).

3.3. Mapping

The mapping stage is used to partition the previously computed feature images to later extract the final descriptor as presented in the next section. For this work, two mapping strategies are defined: (i) *global* and (ii) *local* mapping.

Global mapping considers to extract the final descriptors from the 2D feature image for LBP and 3D volume for LBP-TOP. Therefore, for a volume with d slices, the *global*-LBP mapping will lead to the extraction of d elements. While the *global*-LBP-TOP represents the whole volume as a single element. The *global* mapping for 2D images and 3D volume is shown in Fig. 5(a) and 5(b).

Local mapping considers to extract the final descriptors from a set of $(m \times m)$ 2D patches for LBP and a set of $(m \times m \times m)$ sub-volumes for LBP-TOP. Given N and N' the total number of 2D patches and 3D sub-volumes respectively, the *local*-LBP approach provides $N \times d$ elements, while *local*-LBP-TOP provides N' elements. This mapping is illustrated in Fig. 5(c) and 5(d).

3.4. Feature representation

Two strategies are used to describe each OCT volume texture.

Low-level representation The texture descriptor of an OCT volume is defined as the concatenation of the LBP histograms with the *global*-mapping.

The LBP histograms are extracted from the previously detected LBP images (see Sect. 3.2). Therefore, the LBP-TOP final descriptor is computed through the concatenation of the LBP histograms of the three orthogonal planes with the final size of $3 \times LBP_{\#pat}$. Similarly, the LBP descriptor is defined through concatenation of the LBP histograms per each slice with the final size of $d \times LBP_{\#pat}$.

High-level representation The concatenation of histograms employed in the low-level representation in conjunction with either *global*- or *local*-mapping can lead to a high dimensional feature space. For instance, *local*-mapping results to a size of $N \times d \times LBP_{\#path}$ for the final LBP descriptor and $N' \times LBP_{\#path}$ for the final LBP-TOP descriptor. High-level representation simplifies this high dimensional feature space into a more discriminant lower space. BoW approach is used for this purpose [12]. This model represents the features by creating a codebook or visual dictionary, from the set of low-level features. The set of low-level features are clustered using k -means to create the codebook with k clusters or visual words. After creating the codebook, each of the training example is represented as a histogram of size k . The histogram is obtained by calculating the frequency of occurrences of each of the k words in the extracted features from the training example.

3.5. Classification

Classification corresponds to the mapping of a set of inputs \mathbf{x} into a set of categorical outputs \mathbf{y} using a linear or non-linear function $f(\cdot)$. In supervised learning methods, this function is defined by providing a training set of N samples \mathbf{x}_{tr} with their associated labels \mathbf{y}_{tr} . In the remainder of this section, we briefly summarize the supervised classification methods used in the experiments. Details regarding the parameters used in our experiments are provided in Sect. 4.

k -Nearest Neighbor (NN) is a non-parametric classification method in which an unlabeled feature vector x is assigned to the majority class of its k

nearest-neighbors from the training set. To avoid a tie case, the parameter k is set to an odd number.

Logistic Regression (LR) is a linear classifier which uses the logistic function to estimate the probability of x to belong to a particular class c_i [21]. Thus, the posterior probability is expressed as:

$$p(c_i|x) = \frac{1}{1 + \exp(-w^T x)} \quad (2)$$

where w is a vector of the regression parameters to obtain a linear combination of the input feature vector x . The vector w can be inferred by finding the maximum likelihood estimates via optimization methods such as quasi-Newton method [22]. Once the vector w is found, an unlabeled feature vector is assigned to the class which maximizes the posterior probability.

Random Forest (RF) is an ensemble of decision trees [23] which generalizes the classification process by applying two types of randomization: at the tree level, each tree is fed by a bootstrap made of S' samples which are built from the original data of size S such that $S = S'$, and at the node level, a subset of feature dimensions m is randomly selected from the original dimension M such that $m \ll M$. The trees in RF are grown to their maximum length without any pruning. In the testing stage, each tree in the ensemble casts a unit vote in the final prediction and the final prediction is based on combination of all the votes.

Gradient Boosting (GB) is a reformulation of AdaBoost [24] in which the problem of finding an ensemble of real-valued weak learners is tackled as a numerical optimization [25]. A strong learner is built by iteratively finding the best pair of real-valued weak learner function and its corresponding weight which minimizes a given differentiable loss function. Common choice for weak learners is decision stumps or regression trees while the loss function is generally an exponential or logarithmic loss [26], minimized via gradient descent or quadratic approximation.

Support Vector Machines (SVM) is a sparse kernel classification method

which aims at finding the best linear hyperplane which separates two classes by maximizing the margin between them [27]. SVM becomes a non-linear classifier by using the kernel trick [28] which consists in replacing each inner product by a non-linear kernel function such as RBF or polynomial kernels.

4. Experiments

An experimental suit composed of three experiments is designed to test the influence of the different blocks composing our framework in comparison to our previous work [11] (baseline). Table 3 summarizes the different aspects of each experiment as well as the baseline [11].

Lemaitre *et al.* [11] proposed to pre-process the volumes using NL-means denoising without any flattening or aligning. Then using local and global mapping LBP and LBP-TOP features were extracted. BoW approach with fixed codebook size of 32 words were then used as a high representation of the feature. Finally the RF classifier was used to classify the the volumes. This framework was tested using two datasets of Duke and SERI and was compared to method proposed by Venhuizen *et al.* [14]. The obtained results highlighted the discriminative power of LBP descriptors and showed that the proposed framework outperforms the others. However while comparing the obtained results using the two dataset of SERI and Duke, substantial difference were observed. This was attributed to the fact that the volumes in Duke dataset were provided with embedded pre-processing steps.

Considering the conclusion drawn from our baseline, three experiments are designed such as:

- Experiment #1 is intended to find the optimal number of words (unlike [11]) and its effect with respect to different configurations (pre-processing which consists of denoising, flattening, and aligning along different mapping).

- Using the optimal number of words for high level representation of the features, experiment #2 investigates the performance of different configurations and classifiers.
- Finally experiment #3 is defined to investigate the benefits of different configurations and performance of different classifiers on low level represented features.

The detail explanation of each experiment is presented in the following sub-sections while the reminder of this section details the common configuration parameters across the three experiments.

All the experiments are performed using our own dataset SERI. Acquisition details regarding SERI is reported in Sect. 4.1. For all the experiments, LBP and LBP-TOP features are extracted for different sampling points of 8, 16, and 24 for radius of 1, 2, and 3, respectively. As previously mentioned, the *local*- and *global*-mapping strategies are used. The partitioning for *local*-mapping is set to (7×7) patch for 2D LBP and $(7 \times 7 \times 7)$ sub-volume for LBP-TOP.

The three experiments are evaluated using Leave-One-Patient Out Cross-Validation (LOPO-CV) strategy. At each round, a pair DME-normal volume is selected for testing while the remaining volumes are used for training. The use of this method implies that no variance in terms of Sensitivity (SE) and Specificity (SP) can be reported. Despite this limitation, LOPO-CV has been employed due to the small size of the dataset. All the experiments are evaluated in terms of SE and SP, which are statistics driven from the confusion matrix (see Fig. 6) as stated in Eq. 3. The SE evaluates the performance of the classifier with respect to the positive class, while the SP evaluates its performance with respect to negative class.

$$SE = \frac{TP}{TP + FN} \quad SP = \frac{TN}{TN + FP} \quad (3)$$

Only in experiment #1, beside SE and SP additional metrics such as Accuracy (ACC) and F1-score (F1) are applied. ACC is used to have an overall sense of the classifier performance and F1 is used to see the trade off between

		Actual	
		A+	A-
Predicted	P+	True Positive (TP)	False Positive (FP)
	P-	False Negative (FN)	True Negative (TN)

Figure 6: Confusion matrix with true and false positive detected samples (TP, FP) in the first row, from left to right and the false and true negative detected samples (FN, TN) in the second row, from left to right.

SE and precision.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad F1 = \frac{2TP}{2TP + FP + FN} \quad (4)$$

4.1. SERI-Dataset

This data was acquired by the Singapore Eye Research Institute (SERI),
 270 using CIRRUS TM (Carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT device.
 The datasets consist of 32 OCT volumes (16 DME and 16 normal cases). Each
 volume contains 128 B-scan with resolution of 512×1024 pixels. All SD-OCT
 images are read and assessed by trained graders and identified as normal or
 DME cases based on evaluation of retinal thickening, hard exudates, intraretinal
 275 cystoid space formation and subretinal fluid.

Table 3: The outline and summary of the performed experiments. \sim indicate that common configuration applies.

	Dataset	Pre-processing	Features	Mapping	Representation	Classification	Evaluation
Common:	SERI	NL-means	LBP,LBP-TOP $P = \{8, 16, 24\}$ $R = \{1, 2, 3\}$				LOPO-CV SE, SP
Baseline [11]: Goal: Evaluation of features, mapping and representation	+ Duke	\sim	\sim	<i>global</i> <i>local</i>	BoW Histogram	RF	+ [14]
Experiment#1: Goal: Finding the optimum number of words	\sim	+ F + F+A	\sim	<i>global</i> <i>local</i>	BoW $k \in K$	LR	+ACC, F1
Experiment#2: Goal: Evaluation of different pre-processing for high-level features	\sim	+F +F+A	\sim	<i>global</i> <i>local</i>	BoW optimal k	3-NN RF SVM GB	\sim
Experiment#3: Goal: Evaluation of different pre-processing for low-level features	\sim	+F +F+A	\sim	<i>global</i>	Histogram	3-NN RF SVM GB	\sim

4.2. Experiment #1

In order to determine the optimal size of the codebook when using BoW, this experiment evaluates several codebook sizes on SERI dataset.

Several pre-processing strategies are evaluated: (i) NL-means, (ii) a combination of NL-means and flattening, and (iii) a combination of NL-means, flattening and aligning. LBP and LBP-TOP descriptors are detected using the default configuration. Volumes are represented using BoW, where the codebook size ranging for $k \in \{10, 20, 30, \dots, 100, 200, \dots, 500, 1000\}$. Finally, the volumes are classified using LR. The choice of this linear classifier avoids that the results get boosted by the classifier. In this manner any improvement would be linked to the pre-processing and the size of the codebook.

The usual build of the codebook consists of clustering the samples in the feature space using k -means (see Sect.3.4). However, this operation is rather computationally expensive and convergence of the k -means algorithm for all codebook sizes is not granted. Nonetheless, Nowak *et al.* [29] pointed out that randomly generated codebooks can be used at the expenses of accuracy. Thus, the codebook are randomly generated since the final aim is to asses the influence of codebook size and not the performance of the framework. For this experiment, the codebook building is carried out using random initialization k -means++ algorithm [30], which is usually used as a k -means initialization algorithm.

Figure 7 shows the ACC and F1 score graphs obtained for a single case¹ in [1], while the optimal number of words for all the configurations are reported in a compact manner in Appendix - Table ??.

4.3. Experiment #2

After studying the impact of the codebook size in Sect.4.3, this experiment explores the improvement associated with: (i) different pre-processings, and (ii) using larger range of classifiers (i.e. linear and non-linear) for high represented features.

¹<http://tinyurl.com/jeczfh6>

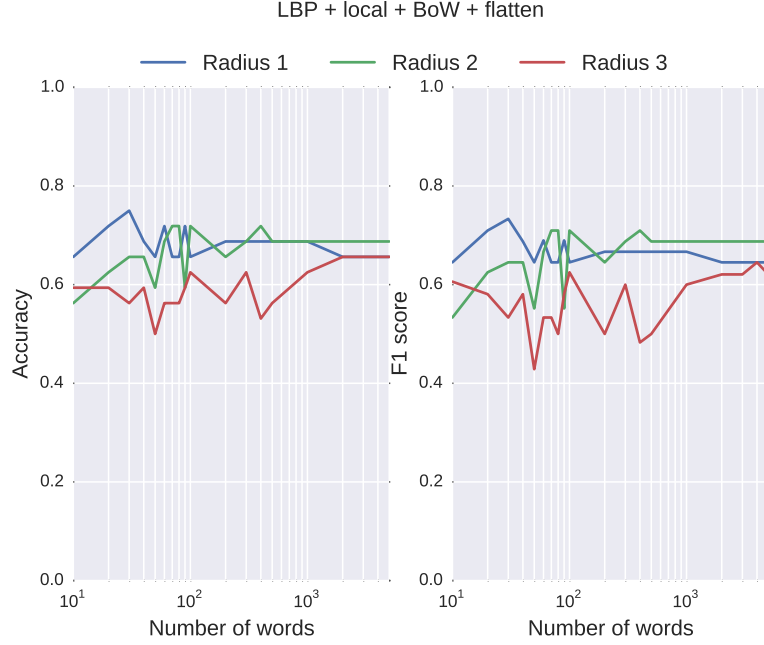


Figure 7: The performance of LR with NL-means+F pre-processing for different P and R .

Similar pre-processing strategies to the previous experiment are evaluated
 305 (NLM, NLM+F, and NLM+F+A). In this experiment the codebooks for the
 BoW representation of LBP and LBP-TOP features are computed using regular
 k -means algorithm which is initialized using k -means++, where k is chosen
 according to the findings of *Experiment #1*. Finally, the volumes are classified
 using k -NN, RF, GB, and SVM. Regarding the classification strategies, k -NN
 310 classifier is trained by considering the 3 nearest neighbor. The RF and GB
 classifier are trained using 100 un-pruned trees, while SVM classifier is trained
 with RBF kernel and its parameters C , and γ are optimized through grid-
 search. Complete list of the obtained results from this experiment are shown
 in Appendix - Table ??, while the most relevant configurations are shaded and
 315 the highest results are highlighted in **bold**.

4.4. Experiment #3

This experiment replicates the *Experiment #2* for the case of low-level represented features from the volumes.

The same pre-processing strategies (NLM, NLM+F, and NLM+F+A) are investigated. LBP and LBP-TOP descriptors are detected using the default configuration. Volumes are represented using low-level feature representation of the *global* mapping. Finally, the volumes are classified using k -NN, RF, GB, and SVM, similarly to *Experiment #3*.

5. Results

Table 4 represent a summary and comparison of the obtained results from experiment #2 & #3 with respect to the baseline. The complete list of obtained result regarding each experiment can be found in Appendix (see Table ??, ??, and ??). This table illustrates the highest 25 performances from highest to lowest considering their achieved SE and SP. The related experiment, choice of pre-processing, type and configuration of the features, choice of mapping and representation, choice of classifier, and finally if required size of codebook is illustrated in this table.

Table 4: Summary of all the results. The best results for each experiment are denoted in bold.

Experiment	Evaluation		Pre-processing	Feat. Detection			Mapping	Feat. Representation	Classifier	#words
	SE	SP		Type	{8, 1}	{16, 2}				
#2	81.2	93.7	NLM+F	LBP	✓		local	High	SVM	30
#2	75.0	93.7	NLM+F+A	LBP	✓		local	High	SVM	40
#2	75.0	93.7	NLM	LBP	✓		local	High	SVM	70
#2	75.0	100	NLM	LBP-TOP		✓	local	High	SVM	500
#2	81.2	87.5	NLM	LBP-TOP	✓		local	High	SVM	400
#2	81.2	87.5	NLM+F+A	LBP-TOP		✓	local	High	RF	90
#2	81.2	81.2	NLM	LBP	✓		local	High	RF	70
#3	81.2	81.2	NLM	LBP-TOP			global	Low	RF	
#2	81.2	81.2	NLM+F	LBP-TOP	✓		local	High	SVM	300
#3	81.2	81.2	NLM+F+A	LBP-TOP			global	Low	GB	
#3	81.2	81.2	NLM+F	LBP-TOP			global	Low	RF	
#2	75.0	87.5	NLM	LBP	✓		local	High	k-NN	70
Lemaitre <i>et al.</i> [11]	75.0	87.5	NLM	LBP	✓		local	High	RF	32
Lemaitre <i>et al.</i> [11]	75.0	87.5	NLM	LBP-TOP		✓	global	Low	RF	
#2	68.7	93.7	NLM	LBP	✓		global	High	RF	500
#3	75	81.2	NLM+F+A	LBP-TOP			global	Low	RF	
#2	68.7	81.2	NLM	LBP-TOP		✓	local	High	RF	500
#3	62.5	93.7	NLM	LBP-TOP		✓	global	Low	SVM	
#3	68.7	87.5	NLM	LBP-TOP		✓	global	Low	RF	
#3	68.7	81.2	NLM	LBP-TOP			global	Low	RF	
#3	75.0	75.0	NLM	LBP-TOP			global	Low	RF	
#3	68.7	75.0	NLM+F	LBP-TOP	✓		global	Low	SVM	
#3	56.2	75.0	NLM	LBP			global	Low	RF	
#3	56.2	75.0	NLM+F	LBP		✓	global	Low	k-NN	
#3	56.2	75.0	NLM+F+A	LBP		✓	global	Low	k-NN	
Venhuizen <i>et al.</i> [14]	61.5	58.8								

6. Conclusions

The work presented here addresses automatic classification of SD-OCT volumes as normal or DME. In this regard, an extensive study is carried out covering (i) the effects of different pre-processing steps, (ii) the influence of different mapping and feature extraction strategies, (iii) the impact of the codebook size in BoW, and (iv) the comparison of different classification strategies.

Table 4 summarizes the most relevant aspects of all the experimentations, showing a clear improvement with respect to previous studies [11, 14]. Based on the reported results, the 3D features and high-level representation of 2D features using patches achieve the best results.

Conflict of interest statement

The authors declare no conflict of interest.

References

- [1] G. Lemaître, M. Rastgoo, J. Massich, retinopathy: Jo-omia-2015 (Nov. 2015). doi:10.5281/zenodo.34277.
URL <http://dx.doi.org/10.5281/zenodo.34277>
- [2] S. Sharma, A. Oliver-Hernandez, W. Liu, J. Walt, The impact of diabetic retinopathy on health-related quality of life, Curr. Op. Ophtal. 16 (2005) 155–159.
- [3] S. Wild, G. Roglic, A. Green, R. Sicree, H. King, Global prevalence of diabetes estimates for the year 2000 and projections for 2030, Diabetes Care 27 (5) (2004) 1047–1053.
- [4] M. D. Abramoff, M. K. Garvin, M. Sonka, Retinal image analysis: a review, IEEE Review Biomed. Eng. 3 (2010) 169–208.
- [5] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. Hubschman, B. al Diri, C. Cheung, D. Wong, M. Abramoff, G. Lim, D. Kumar, P. Burlina, N. M. Bressler, H. F. Jelinek, F. Meriaudeau, G. Quellec,

- 360 T. MacGillivray, B. Dhillon, Validation retinal fundus image analysis algorithms: issues and proposal, *Investigative Ophthalmology & Visual Science* 54 (5) (2013) 3546–3569.
- [6] L. Giancardo, F. Meriaudeau, T. P. Karnowski, K. W. Tobin Jr, E. Grisan, P. Favaro, A. Ruggeri, E. Chaum, Textureless macula swelling detection
365 with multiple retinal fundus images, *Biomedical Engineering, IEEE Transactions on* 58 (3) (2011) 795–799.
- [7] Early Treatment Diabetic Retinopathy Study Group, Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1, *Arch. Ophtalmol.* 103 (12) (1985) 1796–1806.
- 370 [8] T. C. Chen, B. Cense, M. C. Pierce, N. Nassif, B. H. Park, S. H. Yun, B. R. White, B. E. Bouma, G. J. Tearney, J. F. de Boer, Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophtalmic imaging, *Arch. Ophtalmol.* 123 (12) (2005) 1715–1720.
- [9] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, S. Farsiu, Auto-
375 matic segmentation of seven retinal layers in sd-oct images congruent with expert manual segmentation, *Optic Express* 18 (18) (2010) 19413–19428.
- [10] R. Kafieh, H. Rabbani, M. D. Abramoff, M. Sonka, Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map, *Medical Image Analysis* 17 (2013) 907–928.
- 380 [11] G. Lemaître, M. Rastgoo, J. Massich, S. Sankar, F. Mériaudeau, D. Sidibé, Classification of SD-OCT volumes with LBP: Application to dme detection, in: *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Ophthalmic Medical Image Analysis Workshop (OMIA)*, 2015.
- 385 [12] J. Sivic, A. Zisserman, Video google: a text retrieval approach to object matching in videos, in: *IEEE ICCV*, 2003, pp. 1470–1477.

- [13] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, S. Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, *Biomedical Optical Express* 5 (10) (2014) 3568–3577.
- [14] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. P. P. van Grisen, R. Philipsen, C. Hoyng, T. Theelen, C. I. Sanchez, Automated age-related macular degeneration classification in OCT using unsupervised feature learning, in: *SPIE Medical Imaging*, Vol. 9414, 2015, p. 941411.
- [15] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, R. 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 (2011) 748–759.
- [16] J. M. Schmitt, S. Xiang, K. M. Yung, Speckle in optical coherence tomography, *Journal of biomedical optics* 4 (1) (1999) 95–105.
- [17] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, IEEE, 2005, pp. 60–65.
- [18] P. Coupe, P. Hellier, C. Kervrann, C. Barillot, Nonlocal means-based speckle filtering for ultrasound images, *IEEE TIP* (2009) 2221–2229.
- [19] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (7) (2002) 971–987.
- [20] G. Zhao, T. Ahonen, J. Matas, M. Pietikäinen, Rotation-invariant image and video description with local binary pattern features, *Image Processing, IEEE Transactions on* 21 (4) (2012) 1465–1477.
- [21] D. R. Cox, The regression analysis of binary sequences, *Journal of the Royal Statistical Society. Series B (Methodological)* (1958) 215–242.

- 415 [22] R. H. Byrd, J. Nocedal, R. B. Schnabel, Representations of quasi-newton matrices and their use in limited memory methods, *Mathematical Programming* 63 (1-3) (1994) 129–156.
- [23] L. Breiman, Random forests, *Machine learning* 45 (1) (2001) 5–32.
- [24] J. H. Friedman, Stochastic gradient boosting, *Computational Statistics & Data Analysis* 38 (4) (2002) 367–378.
- 420 [25] G. Lemaitre, J. Massich, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, D. Sidibe, F. Meriaudeau, A boosting approach for prostate cancer detection using multi-parametric mri, in: *International Conference on Quality Control and Artificial Vision (QCAV2015)*, SPIE, 2015.
- 425 [26] C. Becker, R. Rigamonti, V. Lepetit, P. Fua, Supervised feature learning for curvilinear structure segmentation, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, Springer, 2013, pp. 526–533.
- [27] V. Vapnik, A. Lerner, Generalized portrait method for pattern recognition, *Automation and Remote Control* 24 (6) (1963) 774–780.
- 430 [28] A. Aizerman, E. M. Braverman, L. I. Rozoner, Theoretical foundations of the potential function method in pattern recognition learning, *Automation and Remote Control* 25 (1964) 821–837.
- [29] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: *Computer Vision–ECCV 2006*, Springer, 2006, pp. 490–503.
- 435 [30] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
- 440