# Classification of SD-OCT Volumes with LBP: Application to DME Detection

Guillaume Lemaître*[1,2], Mojdeh Rastgoo[1,2], Joan Massich[2], Fabrice Mériaudeau[2], and Désiré Sidibé[2]

[1] ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain,
[2] LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, 12 rue de la Fonderie, 71200 Le Creusot, France

**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. Our method is based on Local Binary Patterns (LBP) features to describe the texture of Optical Coherence Tomography (OCT) images and we compare different LBP features extraction approaches to compute a single signature for the whole OCT volume. Experimental results with two datasets of respectively 32 and 30 OCT volumes show that regardless of using low or high level representations, features derived from LBP texture have highly discriminative power.
Moreover, the experiments show that the proposed method achieves better classification performances than other recent published works.

**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 [19]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [25]. 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 (CAD) systems devoted to ophthalmology, have been developed focusing on the automatic analysis of fundus

* Corresponding authors: `g.lemaitre58@gmail.com`, `mojdeh.rastgoo@gmail.com`, `mailsik@gmail.com` - Source available in GitHub [13]

images [1, 22]. 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. However, 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 [11]. 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. 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/seconds with an axial resolution ranging from $3.5\,\mu m$ to $6\,\mu m$ [7].

I don't think thats the way to introduce it

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 [8, 12], ~~and~~ sik (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 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 [20]. 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 pre-processing steps for OCT data, and different classifiers.

In the following of this paper, fist in Sect. 2 a summary of the related studies is presented.

This paper is organized as follows, Section 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 directions are drawn in Sect. 5.

## 2 Background

review, add previous version of the paper

~~Quellec *et al.* proposed a method for the identification of fluid-filled regions in SD-OCT images of the macula based on texture features extracted in the pre-segmented retinal layers [17].~~ sik (Isn't this out of focus?)

This section reviews up to our knowledge the works straightly addressing the problem of classifying OCTvolume classification as normal or abnormal. A summary can be found in Table 1.

Srinivasan *et al.* [21] proposed a classification method to distinguish DME, 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) 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.* also proposed a method for OCT images classification using the BoW models [24]. The method starts with the detection and selection of keypoints in each individual B-scan by keeping the most salient points corresponding to the top 3% of the vertical gradient values. Then, a texton of size $9 \times 9$ pixels is extracted around each keypoint, and Principal Component Analysis (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 using $k$-means clustering, and the obtained codebook from the training is used to represent each OCT volume as a feature vector occurrence histogram. Finally, this histogram is 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 Curve (AUC) of 0.984 with a dataset of 384 OCT volumes.

Liu *et al.* proposes a methodology for detecting macular pathology in OCT images using LBP and gradient information as attributes [14]. The method starts by aligning and flattening the images, then a 3-level multi-scale spatial pyramid is created and edge and LBP histograms are extracted in each block at every level of the pyramid. All obtained histograms are concatenated into a global descriptor whose dimensions are reduced using PCA. Finally a SVM is used as classifier. The method achieved good results in detection OCT scan containing different pathology such as DME or AMD, with an AUC of 0.93 using a dataset of 326 OCT scans.

Lemaître *et al.* propose

summary of previous best result

## 3 Materials and Methods

In my opinion this section should contain a description of the things we use in a generic manner. Describe the data, all the blocks in the figure but not their relations

The proposed method, as well as, its experimental set-up for OCT volume classification are outlined in Fig. 1. The methodology is formulated as a standard classification procedure, through 5 steps which are explained in the following. First, the OCT volumes are pre-processed as presented in details in Sect. 3.1.

**Table 1.** Other methodologies overview

| | **Srinivasan *et al.* [21]** | **Venhuizen *et al.* [24]** | **Liu *et al.* [14]** | **Lemaître *et al.* [?]** |
|---|---|---|---|---|
| task | DMEvs AMDvs normal | ??? | ??? | DMEvs normal |
| data size | 45 | 384 | 326 | ?? |
| pre-processing | filtering + flatten | ?? | ?? | filtering |
| features | HOG | BoWof texton keypoints | ?? | ?? |
| classifier | SVM | random forest | ?? | ?? |
| rate | 100% 100% 86.7% | AUC98.4% | ?? | ?? |

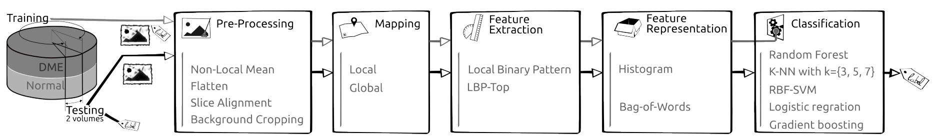



**Fig. 1.** Machine learning classification basic scheme

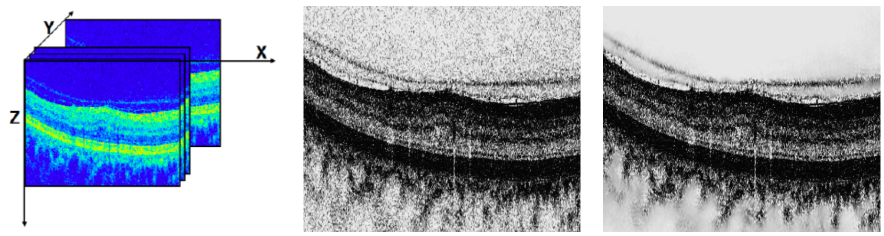
The mapping stage is used to determine a discrete set of elements (or structures) which is used for representing the OCT volume. Thereafter, two mapping strategies are defined: (i) *global* and (ii) *local* mapping. In the global mapping approach, a single structure is computed for the image/volume while in the local mapping, a set of structures is defined by sliding a window through the image/volume. Then, a descriptor is computed for each structure. The feature extraction and representation are presented in depth in Sect. 3.2 and Sect. 3.3, respectively. Finally the classification step is presented in Sect. 3.4.

### 3.1 Image pre-processing

describe the new preprocessing strategies. They should all be explained as independent stuff, and in experimentation we'll describe which are we using for each experiment scenario.

**Non-Local Means (NL-means)** OCT images are known to be affected by a speckle noise [18]. Subsequently, NL-means [5] filtering has been successfully used in Ultra-Sound (US) images to filter similar noise [9] and is used in our framework to denoise each B-scan (i.e. each $x-z$ slice) of the OCT volumes (see in Fig. 2(a)). NL-means filtering offers the advantage to use all the possible self-predictions that the image can provide rather than local or frequency filters such as Gaussian, anisotropic or Wiener filters [5]. An example of filtering using NL-means filter on OCT image is depicted in Fig 2(b) and Fig. 2(c).

**Flatten** Texture descriptors characterize spatial arrangement of intensities. Therefore, to ensure a consistent characterization of the tissue disposition regardless of the location within the retina, the natural curvature of the retina needs to be taken into account. This can be done in different manners: using a




(a) (b) (c)

**Fig. 2.** OCT: (a) Organization of the OCT data - (b) Original image - (c) NL-means filtering.

descriptor that allow for more characterization, using a rotation invariant descriptor, or by unfolding the curvature of the retina.

This process of unfolding the curvature of the retina is known as image flattening. When flattening, an estimation of the Retinal Pigment Epithelium (RPE)layer is used to modify the volume by imposing that the RPEshould be flat.

**Slice alignment**

**Background cropping**

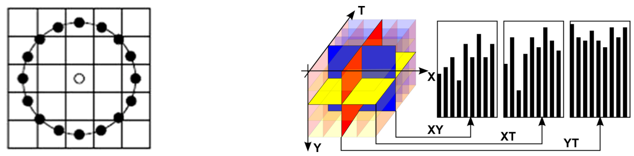
### 3.2 Features extraction

In this research we chose to extract simple and efficient LBP texture features with regards to each OCT slice and volumes. LBP is a texture descriptor based on the signs of the differences of a central pixel with respect to its neighboring pixels [16]. These differences are encoded in terms of binary patterns as in Eq. (1):

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

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 $R$. Figure 3(a) illustrates the meaning of $P$ and $R$.

Ojala *et al.* further extend the original LBP formulation to achieve rotation invariance at the expense of limiting the texture description to the notion of circular "uniformity" [16]. Volume encoding is later proposed by Zhao *et al.* by computing LBP descriptors in each orthogonal planes, so called LBP from Three Orthogonal Planes (LBP-TOP) [26].




(a) (b)

**Fig. 3.** The different LBP descriptors: (a) LBP with $(R = 2, P = 16)$ - (b) LBP-TOP [26].

### 3.3 Feature representation

Each OCT volume can be described by its texture and we employed two strategies.

**Low-level representation** The texture descriptor of an OCT volume is defined as the concatenation of the LBP histograms. Therefore, for the LBP-TOP, the feature descriptor is computed through the concatenation of the LBP histograms of the three orthogonal planes and for the LBP, the descriptor is defined either through concatenation of the LBP histograms per each B-scan (*gocal*-mapping), or per each patch (P) (*local*-mapping).

*Global* mapping considers to extract the features from the 2D B-scans for LBP and 3D volume for LBP-TOP. Therefore for a volume with $d$ B-scans, the *global*-LBP feature leads to a final descriptor of size $d \times LBP_{hist}$ and *global*-LBP-TOP feature returns the final descriptor of size $3 \times LBP_{hist}$. Here $LBP_{hist}$ refers to the size of uniform and rotation invariant LBP histogram which it's number of bins depends on the number of sampling points in a given neighborhood.
In *local* mapping, the features are extracted from a set of 2D patches for LBP and a set of sub-volumes for LBP-TOP. Considering a $(m \times m)$ patch (P) for 2D LBP and $(m \times m \times m)$ sub-volume for LBP-TOP. Using these elements, the *local*-LBP approach provides a final descriptor of size $(d \times N \times LBP_{hist})$ while *local*-LBP-TOP results in a final descriptor of size $\frac{d}{m} \times N \times 3 \times LBP_{hist}$. Here $N$ is the total number of elements in each B-scan, and 3D volume, respectively.

**High-level representation** According to the chosen mapping strategy, the low-level representation can lead to a high dimensional feature space. High-level representation simplifies this high dimensional feature space into a more discriminant lower space. BoW approach is used for this purpose [20]. This model represents the features by creating a visual dictionary or "codebook", 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.

### 3.4 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.

**$k$-Nearest Neighbor (NN)** is a method in which the labelling decision depends of its is one the simplest supervised machine learning classification methods. In this method a new unlabeled vector is assigned to the most represented class from $k$ nearest-neighbors in the features space. To avoid a tie case, the parameter $k$ is set to an odd number.

**Logistic Regression (LR)** [10] is another supervised learning which can provide associated probability of each prediction. As the name suggests, this classifier uses a logistic function, to estimate the probabilities. By defining the posterior probability for one of the classes, using logistic function (see Eq. 2), the probability of other class is defined as $p(c_2|x_i) = 1 - p(c_1|x_i)$.

$$p(c_1|x_i) = \frac{1}{1 + \exp(-w^T x_i)} \tag{2}$$

Here $w$ is a vector of regression parameters, which allows to obtain a linear combination of the input feature vector $x_i$. Using this model the unlabeled sample is assigned to the class which maximizes the posterior probability.

$$C(x_i) = \arg\max_k p(C = k|x_i) \tag{3}$$

In this method, finding an optimal set of parameters for $w$ is essential. The vectors of parameters $w$ can be inferred by finding the maximum likelihood estimates via optimization methods such as quasi-Newton method [6].

**Random Forest (RF)** is an ensemble of decision trees and was introduced by [4]. The ensemble uses each tree to predict an output and finalizes the ultimate prediction by aggregating the outputs of all tress. This classifier learns the data by training multiple decision trees on bootstrap samples of the original data. Each bootstrap of $D$ dimension is used for training one decision tree and at each node, the best split among randomly ($d << D$) selected subset of descriptors is chosen. Each tree is grown to its maximum length without any pruning. In the prediction stage a sample is voted by each tree and it is labeled by considering the majority of the votes.

**Gradient Boosting (GB)** is a generalization form of AdaBoost (AdB), which is able to use real-value weak learners and minimizes different loss functions [27]. Gradient Boosting (GB) builds the ensemble in a greedy manner. It iteratively selects the best pair of real-valued weak learners and adjust their weights so that they minimize a given differentiable loss function. Common choice for the weak learner is decision stumps or regression trees while the loss function is generally an exponential loss or a logarithmic loss [3]. This minimization is carried out via gradient descent or quadratic approximation.

**Support Vector Machines (SVM)** [23] is a sparse kernel method which aims to separate two classes by finding the best hyperplane which maximizes the margin between the two classes. Maximizing the margin is equivalent to minimizing the norm of the normal vector of the hyperplane:

$$\min_{\mathbf{w},\omega_0} \frac{1}{2}\|\mathbf{w}\|^2 \qquad \text{s.t.} \quad y_i(\mathbf{w}^T\mathbf{x_i} + \omega_0) \geq 1, i = 1:N \tag{4}$$

This constraint intends to force all the point to be in the correct side of the decision boundary (hyperplane) with a minimum distance of 1. This assumption is only valid if the data is linearly separable. Thus for general cases a slack variable $\xi_i \geq 0$ is introduced, which is $\xi_i = 0$, if the points are on/or inside the correct margin boundary, is $0 < \xi_i \leq 1$ if the points are inside the margin but in the correct side of the decision boundary and otherwise if they lie in the wrong side of decision boundary is $\xi_i > 1$. This assumption introduces the *soft margin constraints*. Therefore the optimization problem of SVM classifier is presented by:

$$\begin{aligned} &\min_{\mathbf{w},\omega_0,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{s.t.} \quad \xi_i \geq 0, \quad y_i(x_i^T\mathbf{w} + \omega_0) \geq 1 - \xi_i, i = 1:N \end{aligned} \tag{5}$$

In Eq. 5, the $\sum_i \xi_i$ term, describes the upper bound on the number of misclassified points and $C$ is the regularization parameter that controls the tolerance of the classifiers on the number of errors [15].

## 4 Experiments and Validation

### 4.1 Datasets

In this work, we use two different dataset. Although our framework is implemented based on our dataset (SERI), we use duke dataset as well for further validation of proposed framework.

**SERI** - datasets were acquired by Singapore Eye Research Institute (SERI), 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-sane with dimension of $512 \times 1024$ pixels. All SD-OCT images are read and assessed by trained graders and identifies as normal or DME cases based on evaluation of retinal thickening, hard exudates, intraretinal cystoid space formation and subretinal fluid.

**Duke** - datasets published by Srinivasan et al. [21] were acquired in Institutional Review Board-approved protocols using Spectralis SD-OCT (Heidelberg Engineering Inc., Heidelberg, Germany) imaging at Duke University, Harvard University and the University of Michigan. This datasets consist of 45 OCT volumes (15 AMD, 15 DME and 15 normal). In this study we only consider a subset of the original data containing 15 DME and 15 normal OCT volumes.

### 4.2 Experiments & Results

To evaluate the effects and influence of different stages of our framework, we perform various tests. Majority of the experiments are performed on SERI dataset and for the sake of comparison, the optimal configurations are evaluated on Duke dataset. 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, two different mapping strategies, *local* and *global*, are used.

In this research we consider a $(7 \times 7)$ patch (P) for 2D LBP and $(7 \times 7 \times 7)$ sub-volume for LBP-TOP.

The *global* and *local* extracted features are then presented in low or high representation. As previously mentioned, BoW approach is used for high-level representation. In this regard, to find the optimal number of "visual-words" ($k$), we perform a preliminary experiment.

Using BoW approach to find the optimal number of "visual-words" ($k$), we perform a preliminary classification where different number of words randomly are selected. In this setup, the words are randomly selected using $k$-means++ algorithm [2]. This test is performed for all the high-level feature sets considering different pre-processing steps. The obtained results of this test is represented in experiment #1. Beside this experiment, four other experiments are performed which are explained in the following. For the sake of comparison all the experiment are performed using all the aforementioned classifiers. However only the relevant part and results related to our experiment is represented within the paper, while the rest are mentioned in the appendix.

The pipeline in each experiment is evaluated using Leave-One-Patient Out Cross-Validation (LOPO-CV) strategy. In this validation, at each round a pair DME-normal volume is selected for testing while the rest of the 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. However, and despite this limitation, LOPO-CV has been employed due to the small size of the dataset.

The obtained results of all the experiments except experiment #1 is represented in terms of SE and SP. These statistics are driven from the confusion matrix (see Fig. 4).

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

**Fig. 4.** Confusion matrix with truly and falsely positive detected samples (TP, FP) in the first row, from left to right and the falsely and truly negative detected samples (FN, TN) in the second row, from left to right.

The former evaluate the performance of the classifier with respect to the positive class, while the later evaluate it's performance with respect to negative class. These measurements are formulated as:

$$SE = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{6}$$

In experiment #1 we use Accuracy (ACC) and F1-score (F1) instead. Accuracy is used to have a overall sense of classifier performance, and F1 is used to see the trade off between SE and precision. Equation. 7 shows the formulation of these measurements.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad F1 = \frac{2TP}{2TP + FP + FN} \tag{7}$$

The performed experiments are listed in the following:

**Experiment #1** is carried on SERI dataset. This experiment is conducted to find the optimal number of words for BoW high-level feature representation. *Global* and *local*-LBP and *local*-LBP-TOP feature descriptors are re-represented using BoW approach. In this experiment, the BoW algorithm is performed using various number of words in the range of $\{10, 20, 30, \cdots, 100, 200, \cdots, 500, 1000\}$. The words are randomly selected using $k$-means++ algorithm and the features are mapped to their nearest word to create the final clusters. In this experiment in order to asses the effect of number of words, simple linear classifier such as Logistic Regression (LR) is used. The number of words associated with highest ACC and F1 score, is selected as the optimum number of words. Table 2 shows the obtained results of this experiment. The optimum number of words and the achieved ACC and F1 of LR classifier for different configurations are listed in this table.
Figure. **??** shows the obtained graphs for some of the configurations.

**Experiment #2** is performed for high-level features on SERI dataset as well. In this experiment using the optimum number of words which are obtained

**Table 2.** The obtained results of experiment #1. Optimum number of words for each configuration as a result of LR Classification, for high-level representation of *global* and *local*-LBP, and *local*-LBP-TOP features with different pre-processing. The pre-processing includes: NF, F, F+A, and F+A+C.

| Features | Pre-processing | $8^{riu2}$ | | | $16^{riu2}$ | | | $24^{riu2}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ACC% | F1% | W# | ACC% | F1% | W# | ACC% | F1% | W# |
| *global*-LBP | | | | | | | | | | |
| | NF | 81.2 | 78.5 | 500 | 62.5 | 58.06 | 80 | 62.5 | 62.5 | 80 |
| | F | 71.9 | 71 | 400 | 68.7 | 66.7 | 300 | 68.7 | 66.7 | 300 |
| | FA | 71.9 | 71 | 500 | 71.9 | 71 | 200 | 75 | 68.7 | 500 |
| | F+A+C | 75 | 73.3 | 500 | 78.1 | 75.8 | 500 | 68.7 | 68.7 | 90 |
| *local*-LBP | | | | | | | | | | |
| | NF | 75 | 75 | 70 | 65.6 | 64.5 | 90 | 62.5 | 60 | 30 |
| | F | 75 | 73.3 | 30 | 71.8 | 61 | 70 | 62.5 | 62.5 | 100 |
| | FA | 75 | 69 | 40 | 71.9 | 71 | 200 | 68.7 | 66.7 | 10 |
| | F+A+C | 68.7 | 68.7 | 300 | 65.6 | 64.5 | 100 | 65.6 | 64.5 | 100 |
| *local*-LBP-TOP | | | | | | | | | | |
| | NF | 68.7 | 68.7 | 400 | 75 | 75 | 500 | 71.9 | 71 | 60 |
| | F | 68.7 | 68.7 | 300 | 68.7 | 66.7 | 50 | 75 | 76.5 | 80 |
| | F+A | 75 | 73.3 | 100 | 75 | 73.3 | 90 | 75 | 69 | 70 |
| | F+A+C | 71.9 | 69 | 400 | 75 | 73.3 | 100 | 75 | 73.3 | 60 |

from the previous experiment, the low-level feature sets with regards to different pre-processing configurations are re-represented using BoW and $k$-means clustering approach and are classified using $k$-Nearest Neighbor (NN) classifier. $k$-NN classifier was chosen since it provides a sense of Table **??** shows the obtained results from this experiment.

**Experiment #3** is dedicated to compare the performance of different classifiers. This experiment is also performed on SERI dataset and for high-level features. The best configurations of feature set and pre-processing based on the previous experiment is classified with different classifiers such as RF, GB, SVM, LR and $k$-NN. The obtained results of this experiment is listed in Table. **??**.

**Experiment #4** is conducted for low-level features. The *global* LBP and LBP-TOP features are classified using the same classifiers as previous experiments. The obtained results from this experiment is listed in Table. **??**.

**Experiment #5** is performed to evaluate the performance of our proposed framework considering different dataset. In this regards, the best configuration in terms of low-level or high-level representations, number of words and classifiers obtained from the previous experiments are performed on Duke dataset. A comparison of the obtained results from SERI and Duke datasets are presented in Table. **??**.

## 5 Conclusions

The work presented here addresses the automatic classification of SD-OCT data to identify subjects with DME versus normal. Based on the reported results, the low level volume 3D features and high level 2D features using patches achieve the most desirable results in the experimental setup presented here. The comparison against different datasets and methodologies, highlights that: regardless of using low or high level representations, volume signatures derived from LBP texture show high discriminative power for distinguishing DME vs normal volumes.

## References


1. Abramoff, M.D., Garvin, M.K., Sonka, M.: Retinal image analysis: a review. IEEE Review Biomed. Eng. 3, 169–208 (2010)
2. Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. pp. 1027–1035. Society for Industrial and Applied Mathematics (2007)
3. Becker, C., Rigamonti, R., Lepetit, V., Fua, P.: Supervised feature learning for curvilinear structure segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 526–533. Springer (2013)
4. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
5. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 60–65. IEEE (2005)
6. Byrd, R.H., Nocedal, J., Schnabel, R.B.: Representations of quasi-newton matrices and their use in limited memory methods. Mathematical Programming 63(1-3), 129–156 (1994)
7. Chen, T.C., Cense, B., Pierce, M.C., Nassif, N., Park, B.H., Yun, S.H., White, B.R., Bouma, B.E., Tearney, G.J., de Boer, J.F.: Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophtalmic imaging. Arch. Ophtalmol. 123(12), 1715–1720 (2005)
8. Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S.: Automatic segmentation of seven retinal layers in sd-oct images congruent with expert manual segmentation. Optic Express 18(18), 19413–19428 (2010)
9. Coupe, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. IEEE TIP pp. 2221–2229 (Oct 2009)
10. Cox, D.R.: The regression analysis of binary sequences. Journal of the Royal Statistical Society. Series B (Methodological) pp. 215–242 (1958)
11. Early Treatment Diabetic Retinopathy Study Group: Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1. Arch. Ophtalmol. 103(12), 1796–1806 (1985)
12. Kafieh, R., Rabbani, H., Abramoff, M.D., Sonka, M.: Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map. Medical Image Analysis 17, 907–928 (2013)
13. Lemaître, G., Rastgoo, M., Massich, J.: retinopathy: Miccai-omia-2015 (Jul 2015), http://dx.doi.org/10.5281/zenodo.22195
14. Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J.S., M., R.J.: 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, 748–759 (2011)

15. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT press (2012)
16. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(7), 971–987 (2002)
17. Quellec, G., Lee, K., Dolejsi, M., Garvin, M.K., Abramoff, M.D., Sonka, M.: Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula. IEEE Trans. on Medical Imaging 29, 1321–1330 (2010)
18. Schmitt, J.M., Xiang, S., Yung, K.M.: Speckle in optical coherence tomography. Journal of biomedical optics 4(1), 95–105 (1999)
19. Sharma, S., Oliver-Hernandez, A., Liu, W., Walt, J.: The impact of diabetic retinopathy on health-related quality of life. Curr. Op. Ophtal. 16, 155–159 (2005)
20. Sivic, J., Zisserman, A.: Video google: a text retrieva approach to object matching in videos. In: IEEE ICCV. pp. 1470–1477 (2003)
21. Srinivasan, P.P., Kim, L.A., Metttu, P.S., Cousins, S.W., Comer, G.M., Izatt, J.A., Farsiu, S.: Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. Biomedical Optical Express 5(10), 3568–3577 (2014)
22. Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J., al Diri, B., Cheung, C., Wong, D., Abramoff, M., Lim, G., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H.F., Meriaudeau, F., Quellec, G., MacGillivray, T., Dhillon, B.: Validation retinal fundus image analysis algorithms: issues and proposal. Investigative Ophthalmology & Visual Science 54(5), 3546–3569 (2013)
23. Vapnik, V., Lerner, A.: Generalized portrait method for pattern recognition. Automation and Remote Control 24(6), 774–780 (1963)
24. Venhuizen, F.G., van Ginneken, B., Bloemen, B., van Grisven, M.J.P.P., Philipsen, R., C., H., Theelen, T., Sanchez, C.I.: Automated age-related macular degeneration classification in oct using unsupervised feature learning. In: SPIE Medical Imaging. vol. 9414, p. 94141l (2015)
25. Wild, S., Roglic, G., Green, A., Sicree, R., King, H.: Global prevalence of diabetes estimates for the year 2000 and projections for 2030. Diabetes Care 27(5), 1047–1053 (2004)
26. Zhao, G., Ahonen, T., Matas, J., Pietikäinen, M.: Rotation-invariant image and video description with local binary pattern features. Image Processing, IEEE Transactions on 21(4), 1465–1477 (2012)
27. Zheng, Z., Zha, H., Zhang, T., Chapelle, O., Chen, K., Sun, G.: A general boosting method and its application to learning ranking functions for web search. In: Advances in neural information processing systems. pp. 1697–1704 (2007)