Classifying DME vs Normal SD-OCT volumes: A review

Joan Massich*, Mojdeh Rastgoo*†, Guillaume Lemaître*†, Carol Y. Cheung*,
Tien Y. Wong†, Désiré Sidibé†, Fabrice Mériaudeau*§
*LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté,
12 rue de la Fonderie, 71200 Le Creusot, France
†ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain
‡Singapore Eye Research Institute, Singapore National Eye Center, Singapore
§Centre for Intelligent Signal and Imaging Research (CISIR), Electrical & Electronic Engineering Department,
Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia
¶Corresponding author: joan.massich@u-bourgogne.fr

***Abstract*—This article reviews the current state of automatically classify Spectral Domain OCT (SD-OCT) data to identify Diabetic Macular Edema (DME) versus normal subjects. Addressing this classification problem has valuable interest since early detection and treatment of DME play a major role to prevent eye adverse effects such as blindness.**

Despite previous works addressing this problem, this article points out the lack of publicly available data and benchmarking suited for this particular task of identify DME *vs. normal* SD-OCT volumes. The main contribution of this article is to cover these deficiencies by providing a common benchmark, dataset, and a collection of our own implementation of the most relevant methodologies found in the literature.

***Index Terms*—**

I. Introduction

Eye diseases such as Diabetic Retinopathy (DR) and 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 [1]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [2]. Given this scenario, early detection and treatment of DR and DME play a major role to prevent adverse effects such as blindness. 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 [3]. Fundus images which have proven to be very useful in revealing most of the eye pathologies [4, 5] are not as good as Optical Coherence Tomography (OCT) images which provide information about cross-sectional retinal morphology [6] (see Sect. **??** for some image examples). Therefore the growing interest in developing methodologies for OCT data. In this sense, great effort has been placed in retinal layers segmentation, which is a necessary step for retinal thickness measurements [7, 8].

However, few studies have addressed the specific problem of DME automatic detection in OCT, revealing large ground to be covered in terms of: (i) manipulating OCT volumes, (ii) finding radiomix pathology signs, or (iii) appropriated classification strategies.

Advances in any of those regards is of great interest since (i) manual evaluation of SD-OCT volumetric scans is expensive and time consuming [9]. (ii) SD-OCT acquisition has some deficiencies due to eye movement during the scan [10], the reflectivity nature of the retina [**?**], the fact that OCT suffers from high levels of noise and the overall image quality is inconsistent [**?**]. (iii) Coexistence of multiple pathologies [10], easy to miss pathology signs [9], or large variability within the pathology treats which difficult to obtain proper radiomix to facilitate the task of classification.

The rest of this article is structured as follows: Section II offers a general idea of the literature state-of-the-art in SD-OCT volume classification. Section III reviews some publicly available datasets and states the need for another one that suits the classification task here described. Section IV proposes an experimental benchmark to compare different methodologies presented in Sect. **??**. Section V reports and discusses the obtained results, while Sect. VI wraps up our thoughts regarding this work and its possible direction.

II. Background

This section reviews works straightly addressing the problem of classifying OCT volumes as normal or abnormal, regardless of the target pathology. The methods are categorized in terms of its learning strategy, namely: supervised or semi-supervised.

A. Supervised methods

Supervised classification is based on full annotated and labeled training set. In such methods the labeled training data is used to train the classifier function, which is latter used for prediction. Figure 1 describes a prevalent structure for supervised classification. The volumes undergo: (i) *Pre-processing* to reduce the natural noise of the images and correct acquisition deficiencies; (ii) *Feature detection* to quantify visual cues like appearance, texture, shape, etc. (iii) *Mapping* to determine the discrete set of elements (structures) to represent the sample to be classified (i.e.B-scan/volume); (iv) *Feature representation* to associate a descriptor for each element from the *mapping-stage*. This descriptor packages the visual cues associated to the sample. (v) *Classification*.

Venhuizen *et al.* propose a classification method to distinguish between Age-related Macular Degeneration (AMD) and normal SD-OCT volumes using Bag-of-Words (BoW) models [9]. The method detects and selects a set of keypoints at each individual B-scan. Essentially, keeping the salient points comprised at the top 3% of the vertical gradient values. Then, a texton of size 9×9 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. 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 is tested using a publicly available dataset of 384 OCT volumes [13], achieving an Area Under the Curve (AUC) of 0.984.

Srinivasan *et al.* [14] propose a classification method to distinguish DME, AMD and normal SD-OCT volumes. The OCT images are pre-processed by first enhancing sparsity in a transform-domain (BM3D [**?**]), to reduce their speckle noise, and then by flattening the retinal curvature to reduce the inter-patient variations. Histogram of Oriented Gradients (HOG) features are then extracted from multi-resolution pyramid of each pre-processed slice of a volume. These features are classified using a linear Support Vector Machines (SVM). Note that the method classifies each individual B-scan into one of three categories, i.e.DME, AMD, and normal, and then classifies a volume based on the number of B-scans in each category. This method is also tested using a publicly available dataset, composed of 45 patients equally subdivided into the three target classes. The method achieves a correct classification rate of 100%, 100% and 86.67% for normal, DME and AMD patients, respectively.

Alsaih *et al.* [15] extended Srinivasan *et al.* [14] by (i) incorporating Local Binary Patterns (LBP) to the feature detection stage; and (ii) adding PCA to the feature representation step, as proposed by Venhuizen *et al.* [9];

Lemaitre *et al.* [16] propose a method based on LBP features to describe the texture of OCT images and dictionary learning using the BoW models [**?**]. Note that using BoW and dictionary learning contrary to [14] the classification is performed per volume, rather than B-scan. In this method the OCT images are first pre-processed using Non-Local Means (NLM) filtering, to reduce the speckle noise. Then the volumes are mapped into discrete set of structures namely: local, when these structures correspond to patches; or global, when they correspond to volume slices or the whole volume. According to different mapping, LBP or LBP from Three Orthogonal Planes (LBP-TOP) texture features are extracted and represented (per volume) using histogram, PCA or BoW. The final feature descriptors per volumes are classified using RF classifier. This methodology was tested against Venhuizen *et al.* [9] using public and non-public datasets showing an improvement within the results by achieving a Sensitivity (SE) of 87.5% and a Specificity (SP) of 75%.

Liu *et al.* propose a methodology aiming for B-scan classification, rather than volume classification. The classification goal is to distinguish between macular pathology and normal OCT B-scan images using LBP and gradient information as attributes [10]. 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 block of every level of the pyramid. All the obtained histograms are concatenated 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 of OCT scan containing different pathology such as DME or AMD, with an AUC of 0.93 using a dataset of 326 OCT scans.

Albarrak *et al.* [17] propose another volumetric classification framework for differentiating AMD and normal volumes. The author propose to flatten the volume of interest (VOI) from each OCT volume as a pre-processing step and extract LBP-TOP and HOG+LBP-TOP features from individual sub-volumes within each VOI. The extracted features were concatenated into a single feature vector per OCT volume and presented in lower dimensions using PCA. Finally a Bayesian network classifier was used for classifying the volumes. Testing their proposed method and comparing with [10] using 140 OCT volumes, they achieved the highest SE and SP of 92.4% and 90.5%, respectively.

Anantrasirichai *et al.* [18] propose to detect glaucoma in OCT images based on a variety of texture measures. The images are described in terms of LBP, Gray-level co-occurrence matrix (GLCM), wavelet, granulometry, run length measures, and intensity level distributions in combination with retinal layer thickness without any pre-processing. Using PCA and linear and kernel-SVM classifier, the authors compared the performance of individual features and their combinations. Testing with rather a small dataset of 24 OCT voluems, their proposed method achieved an Accuracy (ACC) of 81.95% while using layer thickness information.

B. Semi-supervised methods

Sankar *et al.* [19] propose to use a semi-supervised strategy to classify DME *vs. normal* OCT volumes. The main differences between this method and the afore stated methodologies, with the fully supervised methodologies pipeline, are (i) only *normal* volumes are needed to train the system; and (ii) the training volumes are not used for building the classification function. The proposed method is based on appearance modeling of normal OCT images using Gaussian Mixture Model (GMM) and consider a DME any volume containing two consecutive B-scans not following such model.

During the training stage, volumes only undertake the first two steps of the schema presented in Fig. 1. The volumes are pre-processed using: NLM denoising, flattening and resizing of the B-scans to ensure homogeneous dimension across all volumes. Finally, an appearance model to be used for feature

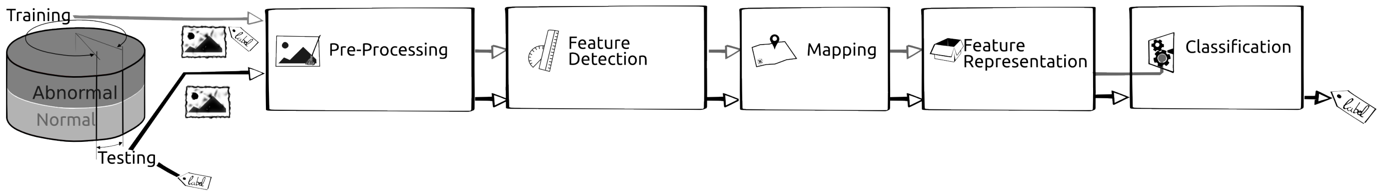


Fig. 1. Common framework

detection at testing stage is created as follows: (i) intensity values of each B-scans for all *normal* volumes within the training set are projected into a lower space using PCA. (ii) GMM to describe the resulting space is created.

During the testing stage, the testing volume is pre-processed in the same manner than during training. The detected feature for each B-scan corresponds to whether the B-scan follows or not the GMM learned during the training stage. To do so, the B-scan is projected to the same lower space and then compared to the learned GMM using the Mahalanobis distance. The mapping is global and the feature representation corresponds to the concatenation of the outlier detection used as feature detection. The final decision is hard-coded and based on the number of outliers (abnormal) B-scans per volume.

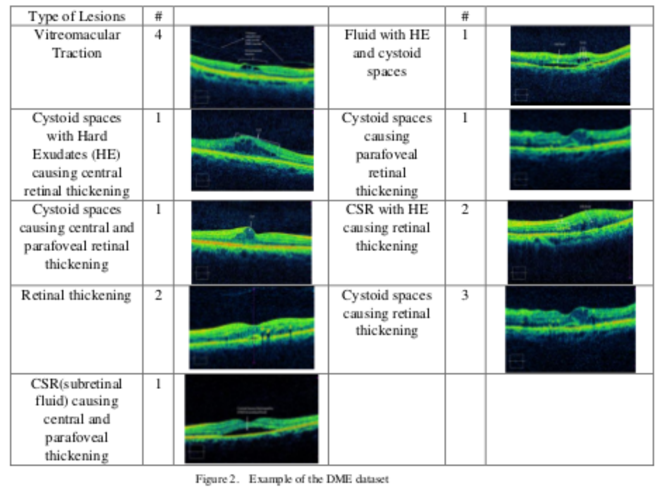
III. Data

In order to compare different methodologies, the first requirement is access to a common pull of images. Despite the fact that lack of public data is a common claim within the medical image community [20], the community developing methodologies for SD-OCT imagery has public data available [13, 14], mainly gathered at *Duke University*. However this data has deficiencies that makes it unsuitable for our problem. Venhuizen *et. al.* test using a large public dataset of 384 OCT annotated volumes of AMD *vs. normal* cases. Despite the interest of testing against a large dataset, our goal remains not to detect AMD but to study the detection of DME. Srinivasan *et al.* [14] also test using a public dataset from *Duke Univeristy*, this time containing AMD, DME and *normal* volumes. However, the volumes of this dataset have been manipulated using pre-processing, realignment, cropping, etc. and there original data is not available making the dataset unsuited for our purposes.

Therefore, we use the Singapore Eye Research Institute (SERI) dataset [21] to conduct this study (see fig. 2). This dataset was acquired by the SERI, using CIRRUS TM (Carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT device. The dataset consists of 32 OCT volumes (16 DME and 16 normal cases). Each volume contains 128 B-scan with resolution of 512 $\times$ 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 cystoid space formation and subretinal fluid.

IV. Experimental Setup

The experimental set-up is summarized in table II. Where the most relevant works in Sect. II are formulated as the as the 5-steps standard classification procedure described in Fig. 1.



Type of Lesions	#			#	
Vitreomacular Traction	4		Fluid with HE and cystoid spaces	1	
Cystoid spaces with Hard Exudates (HE) causing central retinal thickening	1		Cystoid spaces causing parafoveal retinal thickening	1	
Cystoid spaces causing central and parafoveal retinal thickening	1		CSR with HE causing retinal thickening	2	
Retinal thickening	2		Cystoid spaces causing retinal thickening	3	
CSR(subretinal fluid) causing central and parafoveal thickening	1				

Figure 2. Example of the DME dataset

Fig. 2. SERI dataset description.

TABLE II
Correspondence between the most relevant methodologies reviewed in Sect. II and the proposed experimental framework.

Ref	Pre-processing	Features	Mapping	Representation	Classification
Venhuizen *et al.* [9, 24]		Texton	Local	BoW, PCA	RF
Srinivasan *et al.* [14, 25]	De-noise Flatten Cropped	HOG	Global		linear-SVM
Lemaitre *et al.* [16, 26]	De-noised	LBP LBP-TOP	Local Global	PCA, BoW, Histogram	RF
Alsaih *et al.* [15]	*********	LBP LBP-TOP	***** ******	PCA, BoW, *********	RF
Sankar *et al.* [19, 27]	De-noised Flatten Cropped	Pixel -intensities	Global	PCA	Mahalanobis -distance to GMM

A. Implementation details

For reproductivity purposes, the experimentation described in this work can be found in [22], where the image processing and Machine Learning (ML) rapid pipeline prototyping library *Protoclass* [23] has been used to implement the methodologies in Tab. II in accordance to proposed experimentation framework. Each methodology implementation can be seen as a plug-in to experiment in [22], while references to stand-alone implementation of these methodologies can be found in Tab. II. All the repositories are publicly available and provided with

TABLE I
SUMMARY OF THE CLASSIFICATION PERFORMANCE IN TERMS OF SE AND SP IN (%).

	Srinivasan *et al.* [14]	Venhuizen *et al.* [9]	Alsaih *et al.* [15]	Lemaitre *et al.* [16]	Sankar *et al.* [19]
SE	61.5	68.8	75.0	61.3	93.8
SP	58.8	93.8	87.5	83.8	80.0

tests to ensure that our implementation agrees with the results reported by the original works. [1]

B. Evaluation

All the experiments are evaluated in terms of SE and SP (see Eq. 1) using the Leave-One-Patient Out Cross-Validation (LOPO-CV) strategy, in line with [16]. 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.

$$SE = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{1}$$

The use of LOPO-CV implies that at each round, a pair DME-normal volume is selected for testing while the remaining volumes are used for training. Subsequently, no SE or SP variance can be reported. However, LOPO-CV strategy has been adopted despite this limitation due to the reduced size of the dataset.

V. RESULTS AND DISCUSSION

Table I shows the results obtained in terms of SE and SP, where Sankar *et. al.* achieve the best performance, revealing that DME lesions can be modeled accurately based only on a set of normal SD-OCT volumes.

VI. CONCLUSION AND FURTHER WORK

The work here presented states the relevance of developing methodologies to automatically differentiate DME *vs. normal* SD-OCT scans. This article offers an overview of the state-of-the-art methodologies and provides a public benchmarking to facilitate further studies. In this regard, there are two crucial aspects to improve the work here presented: (i) enlarge the dataset. (ii) reach out for the original authors of those methods in Sect. II that could not be included in Sect. IV because our implementation could not be tested against the original data or we did not achieve the original results.

REFERENCES

[1] S. Sharma, A. Oliver-Hernandez, W. Liu, and J. Walt, "The impact of diabetic retinopathy on health-related quality of life," *Current Opinion in Ophtalmology*, vol. 16, pp. 155–159, 2005.

[2] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.

[1]Note that methodologies where this quality control could not had been enforced have been discarded for experimentation and only reviewed based on the results reported by the original work and compiled in Sect. II.

[3] Early Treatment Diabetic Retinopathy Study Group, "Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1," *JAMA Ophthalmology*, vol. 103, no. 12, pp. 1796–1806, 1985.

[4] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Computers in Biology and Medicine*, vol. 43, no. 12, pp. 2136–2155, 2013.

[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, T. MacGillivray, and B. Dhillon, "Validation retinal fundus image analysis algorithms: issues and proposal," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 5, pp. 3546–3569, 2013.

[6] Y. T. Wang, M. Tadarati, Y. Wolfson, S. B. Bressler, and N. M. Bressler, "Comparison of Prevalence of Diabetic Macular Edema Based on Monocular Fundus Photography vs Optical Coherence Tomography," *JAMA Ophthalmology*, pp. 1–7, Dec 2015.

[7] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in sd-oct images congruent with expert manual segmentation," *Optic Express*, vol. 18, no. 18, pp. 19 413–19 428, 2010.

[8] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map," *Medical Image Analysis*, vol. 17, pp. 907–928, 2013.

[9] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. P. P. van Grisven, R. Philipsen, C. Hoyng, T. Theelen, and C. I. Sanchez, "Automated age-related macular degeneration classification in OCT using unsupervised feature learning," in *SPIE Medical Imaging*, vol. 9414, 2015, p. 94141l.

[10] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and 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*, vol. 15, pp. 748–759, 2011.

[11] J. S. Schuman, C. A. Puliafito, J. G. Fujimoto, and J. S. Duker, *Optical coherence tomography of ocular diseases*. SLACK incorporated, 2004.

[12] P. Barnum, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman, "Local quality assessment for optical coherence tomography," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008, pp. 392–395.

[13] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, C. A. Toth, A.-R. E. D. S. . A. S. D. O. C. T. S. Group *et al.*, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, 2014.

[14] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical Optical Express*, vol. 5, no. 10, pp. 3568–3577, 2014.

[15] K. Alsaih, G. Lemaitre, J. Massich, M. Rastgoo, D. Sidibe, and F. Meriaudeau, "alsaih-2016-aug: Icpr 2016," Apr. 2016. [Online]. Available: http://dx.doi.org/10.5281/zenodo.49499

[16] G. Lemaitre, M. Rastgoo, J. Massich, S. Sankar, F. Meriaudeau, and D. Sidibe, "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.

[17] A. Albarrak, F. Coenen, and Y. Zheng, "Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction," in *Proceedings of 2013 International Conference on Medical Image, Understanding and Analysis*, 2013, pp. 59–64.

[18] N. Anantrasirichai, A. Achim, J. E. Morgan, I. Erchova, and L. Nicholson, "Svm-based texture classification in optical coherence tomography," in *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2013, pp. 1332–1335.

[19] S. Sankar, D. Sidibé, Y. Cheung, T. Wong, E. Lamoureux, D. Milea, and F. Meriaudeau, "Classification of sd-oct volumes for dme detection: an anomaly detection approach," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 97 852O–97 852O.

[20] M. L. Giger, H.-P. Chan, and J. Boone, "Anniversary paper: History and status of CAD and quantitative image analysis: the role of medical physics and AAPM," *Medical physics*, vol. 35, no. 12, p. 5799, 2008.

[21] G. Lemaitre, J. Massich, M. Rastgoo, S. Sankar, D. Sidibe, . ******, . ******, . ******, and F. Meriaudeau, "srinivasan-2014-oct: Icpr 2016," Apr. 2016. [Online]. Available: **

[22] J. Massich, G. Lemaitre, and M. Rastgoo, "******************************," Apr. 2016. [Online]. Available: **

[23] G. Lemaitre, "Protoclass, a rapid prototyping tool for image processing and machine learning," Apr. 2016. [Online]. Available: **

[24] G. and Lemaitre, J. Massich, and M. Rastgoo, "******************************," Feb. 2016. [Online]. Available: **

[25] G. Lemaitre, J. Massich, and M. Rastgoo, "srinivasan-2014-oct: Icpr 2016," Apr. 2016. [Online]. Available: http://dx.doi.org/10.5281/zenodo.49669

[26] ——, "******************************," Apr. 2016. [Online]. Available: **

[27] ——, "******************************," Feb. 2016. [Online]. Available: **