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 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 $\$500\,\mathrm{M}$ [1]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over $300\,\mathrm{M}$ 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]. For many years, fundus images have been the modality of choice to reveal most of eye pathologies [4, 5]. However, Optical Coherence Tomography (OCT) has recently shown to provide useful additional information about cross-sectional retinal morphology [6] (see Fig. []), reflected by the growing interest in developing methodologies for this modality. In this sense, great efforts have 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 Spectral Domain OCT (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 [11], the fact that OCT suffers from high levels of noise and the overall image quality is inconsistent [12]. (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. II. 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 targeted pathology. The methods are categorized in terms of their learning strategy, namely supervised or semi-supervised learning.

A. Supervised methods

Supervised learning is based on a fully annotated and labeled training set. In this approach, the labeled training data are used to train the classifier function later used for prediction. Figure 1 illustrates a prevalent framework for supervised learning. Each SD-OCT volume undergoes: (i) *pre-processing* to reduce noise and other acquisition deficiencies which alter the images; (ii) *feature detection* to quantify visual cues like appearance, texture, shape, etc.; (iii) *mapping* in which a sample is either considered as whole (i.e., global) or partitioned into a set of sub-elements (i.e., local dense/sparse

patches, pyramid, etc.); (iv) *feature representation* to associate a descriptor (e.g., concatenation, statistics, histogram, Principal Component Analysis (PCA), Bag-of-Words (BoW), etc.) for each element from the *mapping-stage*. This descriptor packages the visual cues related to the sample; (v) *classification* to determine the associated class of each sample.

Venhuizen *et al.* propose a classification method to distinguish between Age-related Macular Degeneration (AMD) and normal SD-OCT volumes using BoW models [9]. A set of keypoints are detected and selected at each individual B-scan, by keeping the salient points included in the top 3% of the vertical gradient values. Around each of these keypoints, a $9\,\mathrm{px} \times 9\,\mathrm{px}$ texton is extracted, generating a feature vector of 81 dimensions, later reduce to 9 using PCA. All extracted feature vectors are used to create a codebook using *k*-means clustering. Then, each OCT volume descriptor is represented as a histogram that captures the codebook occurrences and are classified by a Random Forest (RF) composed 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.* propose a classification method to distinguish DME, AMD, and normal SD-OCT volumes [14]. Each OCT slice is pre-processed using Block Matching 3D filtering (BM3D) to reduce the speckle noise and is flattened to reduce the inter-patient retinal curvature variations. A multi-resolution pyramid is generated for each pre-processed slice and a Histogram of Oriented Gradients (HOG) feature is computed for each layer. These features are classified using a linear Support Vector Machines (SVM). Note that each individual B-scan is classified into one of the three categories, namely DME, AMD, and normal, and a volume is label to a given class by taking the majority vote of all B-scans. This method is also tested using a publicly available dataset, composed of 45 patients equally subdivided into the three targeted classes. Correct classification rates of 100%, 100% and 86.67% are obtained for normal, DME, and AMD patients, respectively.

Extending the previous work, Alsaih *et al.* aggregate Local Binary Patterns (LBP) to HOG in order to add texture information and reduce the number of dimension using PCA [15].

Lemaître *et al.* propose a method based on LBP features to describe the texture of OCT images and dictionary learning using the BoW models [16]. 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 a discrete set of structures: (i) local corresponding to patches, or (ii) global corresponding to volume slices or the whole volume. According to the chosen mapping, LBP or LBP from Three Orthogonal Planes (LBP-TOP) texture features are extracted and represent each volume through histogram, PCA, or BoW representation. The final feature descriptors are classified using RF classifier. This methodology is 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 at classifying B-scan rather than volume. The classification goal is to distinguish between macular pathology and normal OCT B-scan images using LBP and gradient information as attributes [10]. Each OCT slice are flattened before to create a 3-level mutlti-scale spatial pyramid. From each layer of this pyramid, edges are extracted and LBP descriptors are computed for the flattened slice and the edge map. 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. A detection rate with an AUC of 0.93 is achieved, using a dataset of 326 OCT scans with various pathologies.

Albarrak *et al.* propose another classification framework to differentiate AMD and normal volumes [17]. Each OCT slice undergoes two pre-processing routines: (i) a joint denoising and cropping step using the split Bregman isotropic total variation algorithm and (ii) a flattening step by fitting a second-order polynomial using a least-square approach. Then, LBP-TOP and HOG combined with LBP-TOP features are extracted from individual sub-volumes from each original cropped volume. These features are concatenated into a single feature vector per OCT volume and its dimension is reduced using PCA. Finally, a Bayesian network classifier is used to classify the volumes. The classification performance of the framework in terms of SE and SP achieve 92.4% and 90.5%, respectively, outperforming the method of Liu *et al.* [10], using a dataset composed of 140 OCT volumes.

Anantrasirichai *et al.* propose to detect glaucoma in OCT images based on a variety of texture descriptor [18]. The texture information is described through LBP, Gray-level co-occurrence matrix (GLCM), wavelet, granulometry, run length measures, and intensity level distributions in combination with retinal layer thickness estimation, without any pre-processing. Each feature vector is projected using PCA before to be classified using an SVM with both linear and RBF kernel. The best performance is obtained by combining the layer thickness information with the texture descriptors to get an Accuracy (ACC) of 81.95%.

B. Semi-supervised methods

An example of semi-supervised approach for SD-OCT classification is recently proposed by Sankar *et al.* [19]. The proposed method is based on appearance modeling of normal OCT images using Gaussian Mixture Model (GMM) and anomaly detection. The abnormal B-scans are detected as outliers to the fitted GMM and volume classification is performed based on the number of detected outliers in the volume.

This approach differs from supervised approaches since the B-scan detection method does not require a labeled training set of B-scans. This method starts by pre-processing the B-scans using resizing, flattening and denoising (NLM filter). The features are extracted by taking the intensity information of each B-scan and applying PCA to reduce their dimension.

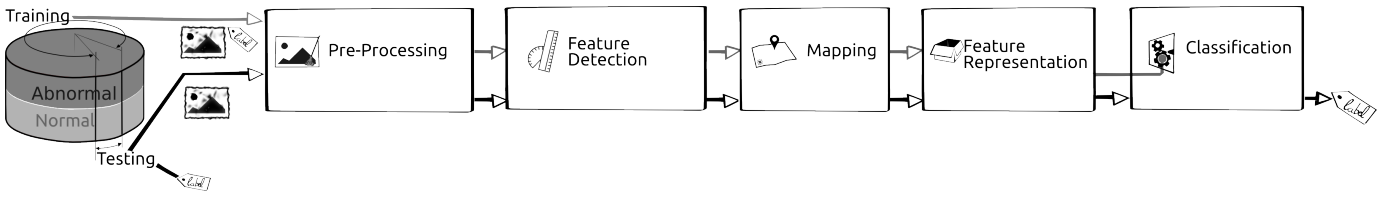


Fig. 1. Common framework

The feature space is then modeled using GMM. In the testing stage, for the new B-scan, the features are extracted in a similar way and they are classified as normal or abnormal based on their Mahalanobis distance to the GMM. Finally the volume classification is performed considering the umber 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.

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

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

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.

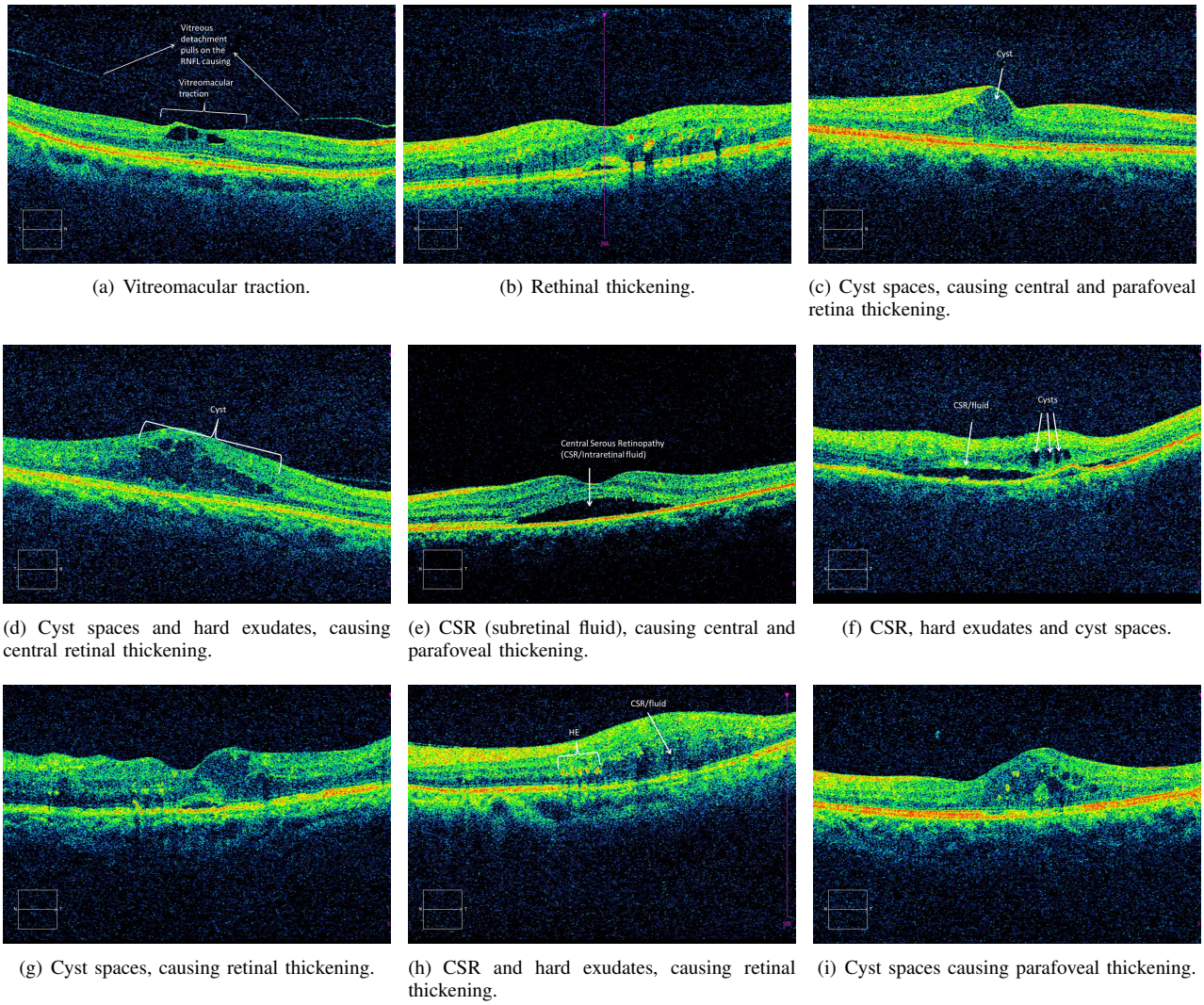
$$SE = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{1}$$

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

[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.



(a) Vitreomacular traction.

(b) Rethinal thickening.

(c) Cyst spaces, causing central and parafoveal retina thickening.

(d) Cyst spaces and hard exudates, causing central retinal thickening.

(e) CSR (subretinal fluid), causing central and parafoveal thickening.

(f) CSR, hard exudates and cyst spaces.

(g) Cyst spaces, causing retinal thickening.

(h) CSR and hard exudates, causing retinal thickening.

(i) Cyst spaces causing parafoveal thickening.

Fig. 2. Examples of DME cases in SERI dataset.

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	68.8	61.5	75.0	87.5	93.8
SP	93.8	58.8	87.5	75.0	80.0

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.

[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: **********