







Automatic identification and characterization of the epiretinal membrane in OCT images

SERGIO BAAMONDE,^{1,2}  JOAQUIM DE MOURA,^{1,2,*}  JORGE NOVO,^{1,2}  PABLO CHARLÓN,³ AND MARCOS ORTEGA^{1,2} 

¹Department of Computer Science, University of A Coruña, 15071 A Coruña, Spain

²CITIC-Research Center of Information and Communication Technologies, University of A Coruña, 15071 A Coruña, Spain

³Instituto Oftalmológico Victoria de Rojas, A Coruña, Spain

*joaquim.demoura@udc.es

Abstract: Optical coherence tomography (OCT) is a medical image modality that is used to capture, non-invasively, high-resolution cross-sectional images of the retinal tissue. These images constitute a suitable scenario for the diagnosis of relevant eye diseases like the vitreomacular traction or the diabetic retinopathy. The identification of the epiretinal membrane (ERM) is a relevant issue as its presence constitutes a symptom of diseases like the macular edema, deteriorating the vision quality of the patients. This work presents an automatic methodology for the identification of the ERM presence in OCT scans. Initially, a complete and heterogeneous set of features was defined to capture the properties of the ERM in the OCT scans. Selected features went through a feature selection process to further improve the method efficiency. Additionally, representative classifiers were trained and tested to measure the suitability of the proposed approach. The method was tested with a dataset of 285 OCT scans labeled by a specialist. In particular, 3,600 samples were equally extracted from the dataset, representing zones with and without ERM presence. Different experiments were conducted to reach the most suitable approach. Finally, selected classifiers were trained and compared using different metrics, providing in the best configuration an accuracy of 89.35%.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

The epiretinal membrane (ERM), also known as macular pucker, is a pathological condition that makes the vitreous shrink, being detached in advanced stages from the retinal layers. This fibrocellular proliferation produces the vision loss, being its presence mainly associated to diseases like the macular edema or metamorphopsia [1]. It is usually asymptomatic and does not present a clear and unique identifiable cause. The most common etiology of the ERM is idiopathic, without ocular abnormality. However, its peeling is clinically recommended to improve the visual quality of the patient, while being a very cost-effective procedure [2]. Surgical interventions can also lead to the ERM development, with further complications. Additionally, other ocular diseases (vitreous hemorrhage, diabetic retinopathy, retinal angioma, among others) that can induce the ERM appearance can also cause the proliferation of secondary ERMs.

Among the most significant diseases that are normally related with the ERM presence, we can find the diabetic macular edema (DME), one of the main causes of blindness in developed countries. Studies have found that the ERM appears in a 27% to 34% of the eye fundus of patients with DME [3,4] and that the ERM peeling together with vitrectomy surgery improves the visual acuity and the degree of macular edema [5].

Advances in the diagnosis and understanding of the DME disease have been made in the recent decades [6], including the proposal of new therapies besides macular laser therapy. Also, retinal image analysis using Optical Coherence Tomography (OCT) scans became popular for the pathological DME identification [7] and characterization using both traditional classification

techniques [8] or more novel ones, such as deep learning strategies [9,10]. Research in these scenarios contributes to the improvement of diagnosis, prognosis and monitoring tasks of the DME disease.

The OCT image properties were clinically used to categorize the DME pathological condition. As reference, Otani *et al.* [11] proposed a defined set of patterns in the OCT scans of patients with DME: sponge-like retinal swelling, cystoid macular edema and serous retinal detachment. Panozzo *et al.* [12] refined this classification by considering five parameters: retinal thickness, volume, morphology, diffusion and epiretinal traction. While considering the epiretinal traction, they proposed four grades of severity depending of the characteristics of the ERM: hyper-reflectivity, continuity, points of adhesion and retinal distortion. In this context, this work faces the automatic identification of the ERM presence using the analysis of OCT scans.

The OCT image modality increased its popularity over the recent years as it offers cross-sectional visualizations of the retinal tissues in a non-invasive way [13,14]. OCT scans are frequently used in the detection and analysis of retinal diseases that produce structural alterations of the retina [15] as happens, for example, with macular edemas or cysts [16]. OCT is widely used to detect the presence of the ERM, since its appearance is indicated by a wrinkling on the retinal surface, or as a bright layer superposed with the inner limiting membrane (ILM), the topmost layer of the retina, illustrated in Fig. 1. Given the ERM presence could cause ocular complications, an early identification of the ERM could help to prevent further damage. However, the current few existing works of this issue are manual and clinical proposals that are mainly focused on the use of OCT scans as a way of visualizing and analyzing the status of the retina before and after ERM surgery [17–20], being only used as a supplementary information technique. In the work of Michalewski *et al.* [21], the identification of the ERM presence is manually done by the specialist. Other studies are focused on the establishment of a precise set of features that allow the unequivocal recognition of the ERM. Wilkins *et al.* [22] developed a method based on the detection of the ERM using manual markers in the OCT scans that are placed by the specialist, allowing the computer software to measure the macular thickness around the selected area. To date, we are not aware of other traditional or novel computational proposals without the need of the user interaction to perform the ERM detection, as is the case of the proposed method of this work. Additionally, this proposal also considers the identification and characterization

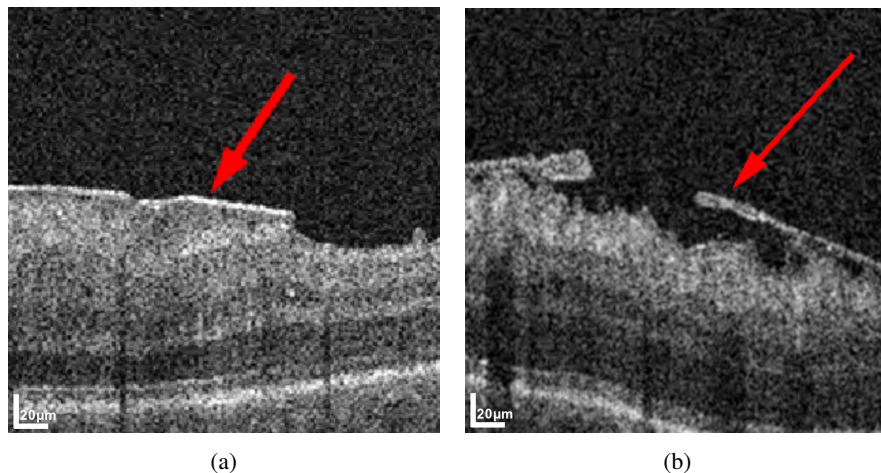


Fig. 1. Different epiretinal membrane appearances in the OCT scans. The arrow indicates the specific localization of the ERM, a reflective layer on the top of the ILM. (a) ERM attached to the retinal surface. (b) Detached ERM, situated on the background.

of the ERM presence in the different degrees of pathological attachment that may be present, hardening significantly the issue.

Given that, this work presents an automatic methodology for the detection of the ERM using OCT images, autonomously and without the need of placing manual markers for the initialization of the method, reducing the amount of manual labor required for a successful identification. The methodology identifies the presence of the ILM layer to posteriorly analyze the neighboring region and determine the presence or non-presence of the ERM. For this identification, intensity, texture and domain-related clinical features were exploited with representative classifiers that discriminate the ERM presence. Additionally, feature selectors were applied to the defined feature set to identify those with the highest discriminating power and optimize the efficiency of the proposal. The robustness and suitability of the proposed feature set and methodology was demonstrated by the similar accuracy results obtained across different classifier types and configurations. Partial preliminary results were presented in [23,24], demonstrating the suitability of the designed strategy.

This paper is organized in the following parts: Section 2 explains the main characteristics of the proposed methodology. Section 3 showcases the practical results, validation and discussion of the experiments. Finally, Section 4 presents the conclusions of this proposal and suggests possible future work on the matter.

2. Methodology

The proposed system receives OCT images as input. These OCT scans provide meaningful information to make an assumption about the presence or absence of ERM in patient's retina. Generally, we illustrate the proposed methodology in Fig. 2.

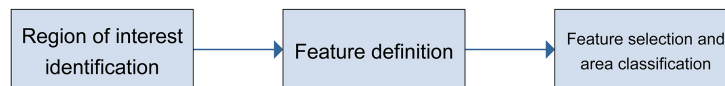


Fig. 2. Main structure of the proposed methodology.

Firstly, the ILM, the layer where the ERM may appear, is localized using an active contour model [25]. The resulting connected points that identify the ILM layer are used as centers of a series of windows, from which a set of features is extracted, representing the information of each window. This information is, then, exploited to discriminate the ERM presence. Afterwards, a process of feature selection is applied to the entire feature set to identify the optimal feature subset with the highest discriminative power to improve the accuracy and efficiency of the identification process. Finally, the feature subsets are used as input of representative classifiers that identify the presence or absence of the ERM at each point.

Additionally, we designed the classification stage with two different approaches:

- A two-class classification process (presence and non-presence of ERM).
- A three-class classification process (presence of ERM on top of the retina, presence of ERM detached from the retina, non-presence of ERM).

We tested in the experiments the feasibility and suitability of each approach.

2.1. Localization of the ILM layer and definition of the region of interest

The first task to achieve is the localization of the ILM layer given that over it we determine the pathological presence of the ERM. We consider this task as the extraction of the first found layer, independently of the tissue. This layer is typically the ILM layer, but in some particular cases of

OCT scans with necrosis or hyperplasia tissues the ILM could be located under these tissues. Despite that, the first found layer is the one used as reference to analyze and isolate the ERM presence over it. Therefore, for simplification, given the common scenario, we refer to it as the ILM layer segmentation. Hence, in order to localize precisely this layer, we use in this work a method based on the use of an active contour model to identify the ILM retinal layer boundaries. This approach already proved its suitability in segmentation issues using OCT scans [26,27].

A preprocessing step was performed to unequivocally define the region of interest (ROI) where the identification process is done. This step is defined by the automatic trimming of the black borders around the OCT surface and the resizing of the resulting images to unify the image dimensions across the entire used dataset. Afterwards, we used an active contour model that is guided by internal (E_{int}) and external (E_{ext}) forces (Eq. 1), converging the model around the features (edges) on our target images.

$$E_{total} = \int E_{int}(v(s)) + E_{ext}(v(s))ds \quad (1)$$

The internal energy represents the shape of the model whereas the external energy governs the attraction of the model to the image characteristics. The external energy is used predominantly in the first steps of the process to further attract the model to the ILM, while the internal energy is used in the last steps to accurately segment the layer (Fig. 3).

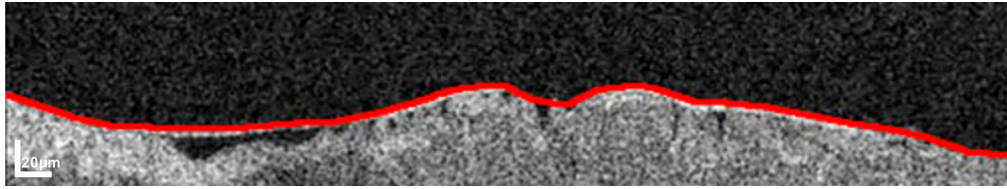


Fig. 3. ILM segmentation by the active contour model. The red points represent the ILM layer identification.

Based on the clinical knowledge, our hypothesis is that the ERM presence on an OCT scan can be determined by local features based on the image properties around the vertical vicinity of the ILM layer. For instance, regions with a higher contrast of intensities are associated with the ERM presence, whereas the absence of the ERM is correlated to lower intensity profiles around the ILM area. In order to exploit this hypothesis, we compute local features in areas surrounding the extracted points of the obtained ILM. More precisely, we segmented the adjacent area in 5 different square shaped sub-windows (Fig. 4) to define a set of features that are adapted to our problem. The size of these regions determines the amount of information that is analyzed. The calculation of the optimal region size is posteriorly determined in the experiments that were conducted in this work.

2.2. Feature definition

Using the defined vertical windows over each analyzed point, we implemented a complete set of features that include intensity and texture analysis in combination with domain-related properties in order to capture the characteristics of the ERM presence and maximize its separability with respect to the normal appearance of the ILM region.

The defined feature set includes a range of 432 to 482 features. The amount of features is correlated to the number of bins of the intensity histogram of each window, which is explained below. The used features belong to the following main groups:

Intensity global features. We obtained 13 global features containing general information about the intensity of the analyzed region: *maximum, minimum, mean, median, standard*

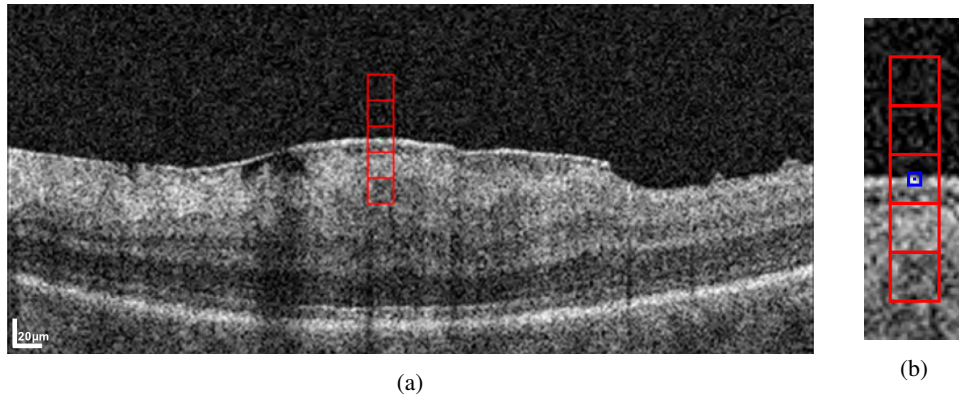


Fig. 4. Definition of the region of interest around an analyzed point. (a): Definition of the vertical window around the central point, with a size of 17×85 pixels. (b): Zoom on the vertical window region and the corresponding 17×17 square sub-windows. The small blue square represents the center of the entire window in the identified ILM position.

deviation, variance, first quartile, third quartile, skewness and maximum likelihood estimate (considering a normal distribution). These properties try to capture general intensity deviations with the presence of the ERM.

Window features. As indicated, each vertical rectangular window is divided in five square-shaped sub-windows. From each sub-window, we calculated its histogram, dividing the data between a number of bins determined by the experiments. The number of features in this section varies between $5 \times \min(N_{\text{bins}})$ and $5 \times \max(N_{\text{bins}})$, where the minimum and the maximum number of bins are studied in the experiments. These features provide useful information about the amount of low and high luminosity pixels on each sub-window and, consequently, if the ERM is present or absent because of its hyper-reflectivity.

The central window contains data about the area that is close to the ERM, while the rest of the windows provide information about the surrounding areas (above and under the ILM) that are also affected by the ERM presence. Only 5 windows have been considered since, as shown in Fig. 4, the inclusion of more regions above or under the identified ILM do not contribute with any further meaningful information as they are too far from any hypothetical ERM presence. These domain-related features were specifically designed to characterize the most complex cases, as regions with strong light reflection.

Principal Component Analysis (PCA) features. PCA is used to obtain information about the existence of a spatial correlation between the different samples. We obtained the eigenvalues of the window since they provide information about intensity changes in different directions [28]. The 4 highest and the 2 lowest eigenvalues were selected. Additionally, some ratios between them are also calculated and included in the feature set.

Gray-level Intensity Histogram (GLIH). Besides the histogram from each sub-window, we also obtained the full histogram from the complete retinal window. Afterwards, we extracted *skewness, kurtosis, energy* and *entropy* metrics, with the same purpose of capturing global deviations on the intensity profiles when the ERM is present.

Histogram of Oriented Gradients (HOG). The histogram of oriented gradients [29] is useful in our approximation by capturing existing gradients and their orientations where the ERM is present. 9 HOG windows were used and 9 bins, adding 81 features to the analysis.

Gray-level Co-occurrence Matrix (GLCM). To determine if a spatial relation exists between pixels over the vertical window and capture texture patterns [30], we use Gray-level Co-occurrence

Matrix features. More precisely, we included a total of 16 features by considering a distance of 2 pixels and 4 directions [31].

Gabor features. Other powerful texture-based features that were frequently used in medical imaging solutions [32,33]. We included 160 Gabor features to capture pathological and normal texture patterns over the analyzed window.

Local Binary Patterns (LBP). The analysis with Local Binary Patterns contributes to the pattern detection in the analyzed window. They are often used in contribution with HOG descriptors, improving considerably the accuracy in many cases [34]. 64 features were calculated and included in the feature set.

Laws features. Laws masks [35] can contribute to the detection of line patterns in the image by calculating texture energy transforms. We obtained and included a total of 28 Laws features in the analysis.

2.3. Feature selection

Given the large amount of features that were analyzed, we complemented their definition with a feature selection process. This way, we analyzed and selected the most relevant features, to filter the most useful information for the discrimination in the classification process. Filtering the original set of features results in the optimization of the process, since unnecessary data is removed before introducing it to the classifiers.

Feature selection has been performed by the use of Correlation Feature Selection (CFS) algorithm [36], which evaluates the worth of a subset of attributes by taking into account the predictive ability of each independent feature, but also the redundancy between features. The CFS algorithm operates under the hypothesis that good feature subsets should contain features that have a high correlation with the classification, but low correlation between them, calculated by the evaluation of a metric score.

In this proposal, we used a best first search strategy to obtain the optimal feature subset. This strategy allows backtracking when performing a forward search through the feature set when considering local changes (i.e. addition or deletion of a feature) to the current feature subset and how the merit score changes for each different feature subset.

In contrast with the typical feature selection methods (wrappers and filters), CFS algorithm can automatically select the optimal number of features with a low execution time. Considering these factors, we opted to use the CFS algorithm in order to select features that are highly correlated with the aimed classes, but also features with low inter-correlation.

2.4. Classification

We tested representative models with our dataset to measure the accuracy of the results by introducing different classifier types that have been proved useful in image analysis processes:

Multilayer Perceptron [37]. It works by fitting the weight of its layers to the response from the following layers by using backpropagation procedures.

Naive Bayes [38]. It is a probabilistic classifier that assumes independence between features, working well with datasets with and without noise.

K-Nearest Neighbors [39]. It is among the simplest machine learning algorithms by assigning to each sample the class voted by the majority of its neighbors.

Random Forest [40]. It creates decision trees and averages multiple deep decision trees to reduce the variance of the model.

The most accurate trained models are finally exploited to produce intuitive pathological color maps. In particular, the method uses the ERM identifications to represent in clear color codes the presence or absence of the ERM over the ILM, using as reference the OCT images. Figure 5 portrays representative output ERM depictions for a particular OCT image. With this, the

specialists are able to easily observe a graphical representation of the pathological ERM presence and adequately diagnose the revised patients.

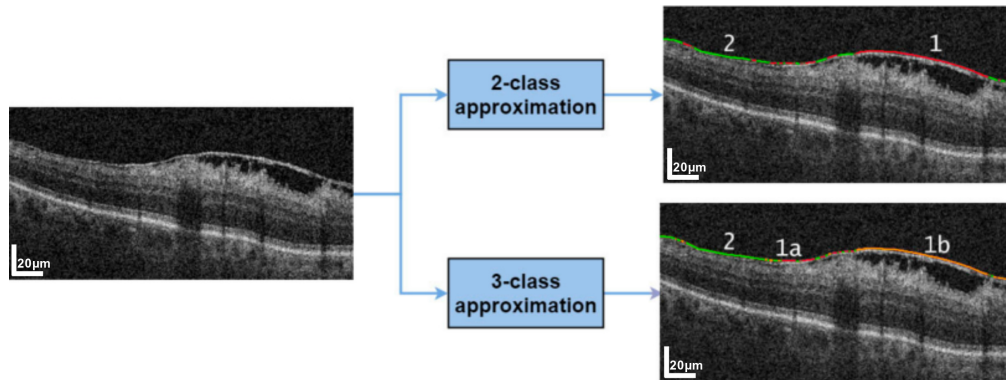


Fig. 5. 2-class and 3-class approximations. (1) Red symbolizes the ERM presence. (1a) Red symbolizes the ERM on top of the ILM. (1b) Yellow symbolizes the ERM detached from the ILM. (2) Green symbolizes the ERM absence.

2.5. 2-class and 3-class approximations

As indicated, to achieve the goal of this work of identifying the ERM presence in the OCT scans, we used 2 classes to classify each point of the dataset: *membrane* and *non-membrane* classes (ERM presence and ERM absence, respectively). This approximation serves the purpose of identifying the ERM and discarding points that do not contain the ERM.

Simultaneously, we also proposed another approximation by further splitting the *membrane* identification class in 2 more precise subclasses, illustrated in Fig. 5: *membrane* and *floating membrane*. The *floating membrane* class represents the ERM points that are detached from the ILM layer in the scans and, consequently, possess slightly different features compared to points situated on the top of the ILM.

3. Results and discussion

The proposed method was validated using a set of 285 OCT images that cover 238×244 mm and with a resolution of 96 pixels per inch that were obtained from the tomograph CIRRUSTMHD-OCT Zeiss using Spectral Domain Technology. The original pixel dimensions of the scans are 902×924 pixels before any preprocessing. After removal of superfluous information from the scans and re-scaling, the resulting images are 501×480 pixel sized. We obtained scans from healthy and pathological cases centered on the macula covering a 6mm of resolution, while also considering information from both eyes. All the procedures that were performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The methodology was implemented using MATLAB for the preprocessing and feature extraction steps and the Weka 3 [41] workbench for the feature selection and model training tasks. Access to the implementation would be granted to the clinical research and practitioner community by contact with varpa@udc.es.

Regarding the parameter set, we used the following configurations for the selected representative models: for the Multilayer Perceptron, we used a learning rate of 0.3, a momentum of 0.2 and a number of hidden layers equal to $(n_{attributes} + n_{classes})/2$; for the Naive Bayes classifier, we assumed a Gaussian (normal) distribution for the values associated with each class; in regards to

the K-Nearest Neighbors classifier, we used the euclidean distance metric. Additionally, after performing multiple tests with different values for the number of neighbors, we determined a value of 15 for this parameter. Finally, for the Random Forest algorithm, we obtain 100 trees with unlimited tree depth.

Each OCT image has been manually labeled by an expert clinician, marking the areas where the ERM is present. Afterwards, the process considers the remaining areas without ERM presence. Based on the resulting ground truth, we obtained a set of samples for each class that were randomly obtained and where half of the samples represent points with ERM presence and the other half points without ERM presence. In the case of the 3-class approximation, the set of points with ERM presence is also split in half between *membrane* and *floating membrane* classes. The criteria used to determine the *floating membrane* class instances is the presence of hyporeflective spaces under visually identified ERM, in order to be able to discern if the ERM is in contact with the retinal surface.

To validate the results, in each classification task we performed a 10-fold cross-validation and averaged the results over 10 iterations in order to avoid overfitting to the training set. Furthermore, in order to verify the adequacy of the results we crafted multiple and independent test sets comprised of the same number of samples as the training set. The number of selected samples varies depending on the classification task at hand, while the precise number of samples used in the final classification task is determined empirically in the first experiment by testing multiple datasets with different number of samples.

3.1. Feature selection

As said, we analyzed the defined set of features with the CFS feature selector. With regards to the selection process, we tested the accuracy of the approach when varying the number of selected features. We found the optimal feature number to be 36 features, as shown in Fig. 6. The most selected features (Fig. 7) belong to window features, being coherent with the premise that intensity information above, below and over the ILM layer provides useful information about the ERM presence or absence. Most of the selected window features belong to the central sub-window, since this is the window where the ERM is specifically placed, being the information in this region very significant to the discrimination power of the problem.

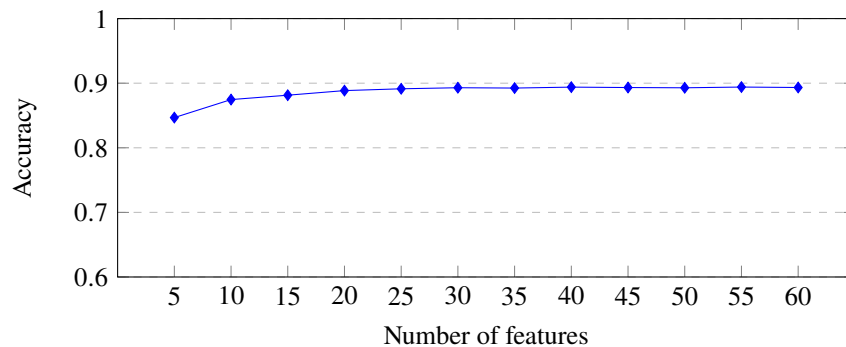


Fig. 6. Evolution of the accuracy when varying the number of features that were progressively selected with the CFS algorithm.

HOG features were almost equally represented as window features, meaning they also provide important information about the changes of intensity in the defined region and their orientations. More precisely, the contrast between the retina and the background of the OCT scans is different when the ERM is present because of the possibility of behaving as an hyper-reflecting layer over the ILM layer. The rest of the categories play secondary roles given that the number of

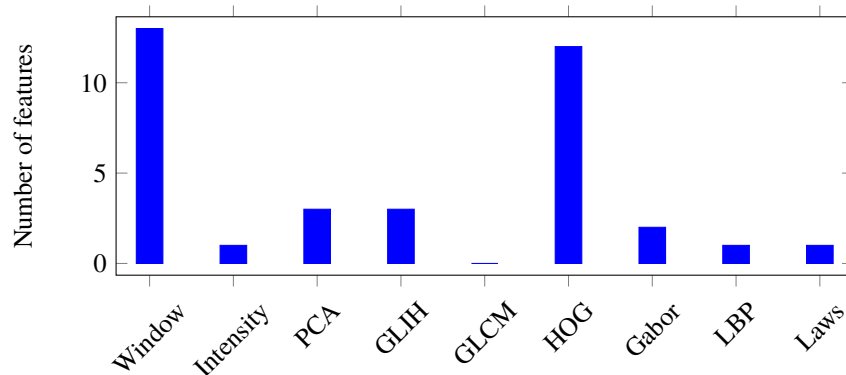


Fig. 7. The 36 selected features, grouped by categories.

contributed features is consistently lower than the aforementioned features, but also helping in the discrimination process.

3.2. The proposed experiments

To showcase the validity of the constructed approach, we prepared a set of experiments to empirically determine the best initial configuration for the whole process. These experiments were performed independently from each other, to avoid spurious correlations that could appear. To this end, we used the accuracy metric (Eq. 2) as our gold standard, using the ERM and non-ERM instances as positive and negative classes, respectively.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The first experiment aimed to calculate the most optimal number of samples that will serve as input to the classifiers. By testing multiple values for this number, we obtained different values of the average accuracy and the training computational time. We selected the most accurate configuration within the limits of a reasonable execution time. This experiment also showcased the differences in accuracy between the different tested approximations. In the end, we validated the obtained results for each value with a test set comprised of the same number of samples, selecting the most accurate configuration to perform further refinement.

The second experiment is part of the main experimentation work: a extensive set of classifiers with different configurations were tested with all the constructed datasets to check the accuracy and other relevant metrics. A small extension of the experiment has been done for the best classifiers to refine their performance and accuracy.

3.2.1. First experiment: quantity of samples

We firstly analyzed the amount of samples that are suitable for the learning stage of the methodology. In particular, we randomly selected progressive larger sample sets with ERM and non-ERM presence. For this step, we used a Random Forest classifier based on its previous satisfactory results with values of $W_{size} = 15$ and $N_{bins} = 13$ as a configuration that offered a correct performance. Figure 8 presents the accuracy results of the use of the different constructed datasets. As we can see, the accuracy is stabilized around 3,600 samples, without obtaining significant improvements with larger sample datasets.

Additionally, given we calculated the mean accuracy over a given number of repetitions, we also analyzed the variability of these performance rates that can be seen in Fig. 9. There, we can observe that higher number of samples decrease the spread of accuracy value, proving that lower

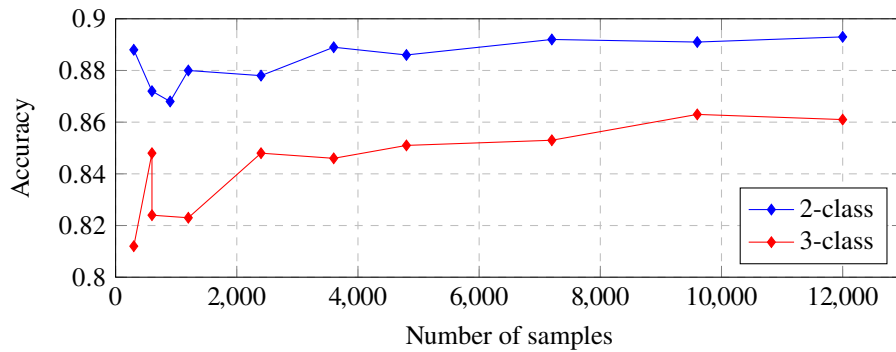


Fig. 8. Evolution of the accuracy when increasing the number of used samples.

number of samples may provide higher single overall accuracies but the added variability can contaminate the results.

Given all the factors that are considered (computational cost, accuracy and stability of the performance), we determined a number of samples of 3,600 as suitable for this approach, where 1,800 samples represent ERM instances (both with and without retinal detachment) and 1,800 samples represent non-ERM instances, both being randomly selected from the available dataset. Likewise, we also used the same number of samples for the test sets used in the next classification steps. Additionally, as Fig. 9 illustrated, the variability of the accuracy of the model in the 3,600 samples configuration is suitable for the proposed approach, guaranteeing that any subset of 3,600 samples is representative of the complete dataset.

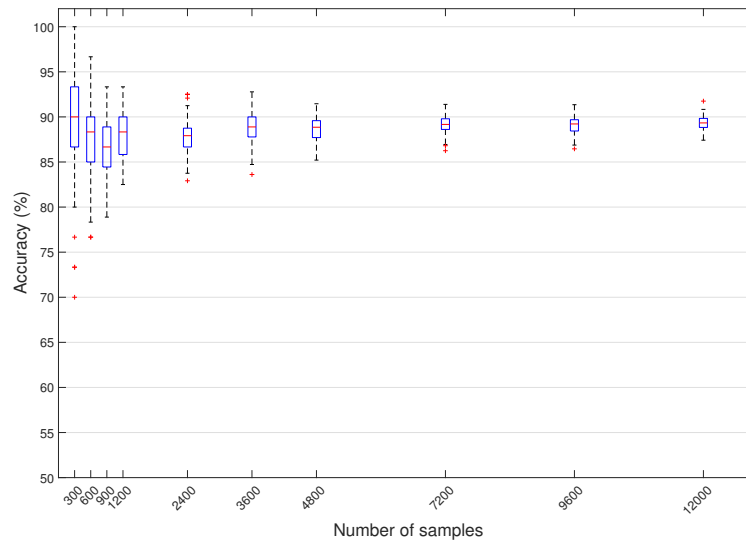
Additionally, we performed this analysis with both 2-class and 3-class classification approaches. Once again, observing the results of Figs. 8 and 9, in terms of accuracy, we can see that the 3-class approach offered acceptable results with a best mean value of 84.56%, but the 2-class approach reached better values, providing a best mean value of 88.93%.

In general, from the obtained results we can assume that the analysis of the problem using 2 classes provides better accuracy instead of the 3-class approximation. The binary nature of the 2-class approximation implies that the classes are more easily separated, obtaining a higher overall accuracy, as well as being more approximated to the clinical evaluation of the specialists. For each pair of experiments with the same number of samples, the 2-class approximation achieves a higher mean accuracy than the 3-class approximation. In Fig. 10, we can see a representative example of both classification approaches. For a question of simplicity, further experiments were conducted with the 2-class approximation.

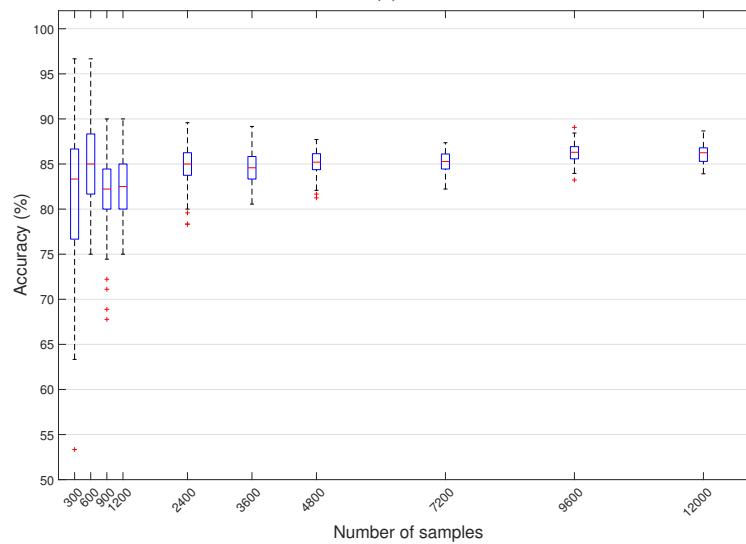
3.2.2. Second experiment: classifier accuracy

Having defined the optimal previous configuration, we proceeded to analyze the methodology using the specified 4 representative classifiers. This set of classifiers were frequently used in other imaging approaches, being their optimal configurations empirically established. We used Random Forest, Naive Bayes, Multilayer Perceptron and k-Nearest Neighbor (kNN) classifiers. For the kNN classifier, a wide range of configurations were tested, being the optimal values of k located between 11 and 19, approximately, for what we used an average value of $k = 15$.

Table 1 showcases the result of the classification process. The results highlighting the best configuration for each type of classifier are represented in bold. As previously explained, the sample subset used in the classification task is representative of the total dataset. In order to avoid overfitting to the classification set, each experiment was done using 10-fold cross validation and averaging the results using 10 different iterations, for a total of 100 classification tasks. In this case, Random Forest and kNN classifiers achieved higher overall accuracy values than



(a)



(b)

Fig. 9. Boxplot graph showing different values of accuracy in different configurations. (a) Accuracy for the 2-class approximation. (b) Accuracy for the 3-class approximation.

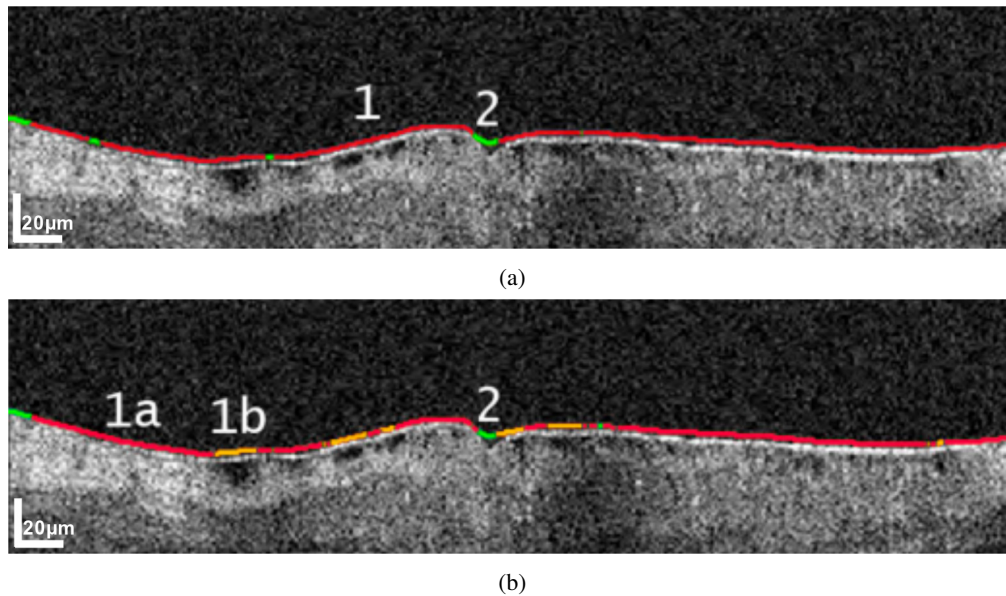


Fig. 10. ERM detection with different approaches. (1) Red, the ERM is present. (1a) Red, the ERM is present and situated on top of the ILM. (1b) Yellow, the ERM is present but detached from the ILM. (2) Green, the ERM is absent.

Naive Bayes and Multilayer Perceptron classifiers. Table 1 also illustrates the mean standard deviation (SD) across the entire cross-validation process for each configuration. The low SD scores obtained in every configuration are associated with low overfitting from the training set, whereas the similarity of the SD scores for all the different configurations illustrate the stability of the results for all classifier types and configurations. We expanded the results by testing more configurations with Random Forest and kNN methods, that are depicted in Table 2.

Table 1. 2-class classification accuracy results.

N_{bins}	Random Forest			Naive Bayes		
	W_{size}			W_{size}		
	13	15	17	13	15	17
13	87.62%±0.15	88.94%±0.19	88.43%±0.20	84.92%±0.18	83.93%±0.21	84.46%±0.15
15	87.81%±0.21	88.70%±0.28	89.35% ±0.22	84.09%±0.19	84.34%±0.23	84.57%±0.21
17	87.84%±0.21	87.56%±0.21	88.43%±0.22	85.53% ±0.12	83.61%±0.13	84.34%±0.21
19	87.99%±0.14	88.06%±0.30	89.12%±0.18	84.22%±0.17	83.49%±0.13	83.72%±0.15
	Multilayer Perceptron			kNN		
13	84.63%±0.44	85.84%±0.43	86.60%±0.41	86.97%±0.15	88.01%±0.18	88.04%±0.24
15	85.15%±0.61	86.57%±0.40	86.75%±0.33	86.87%±0.23	88.35%±0.18	88.97% ±0.19
17	85.32%±0.70	84.30%±0.51	86.16%±0.45	87.99%±0.20	86.88%±0.25	87.80%±0.23
19	84.85%±0.25	85.84%±0.36	86.92% ±0.33	86.53%±0.14	87.17%±0.24	88.87%±0.24

3.3. Discussion

From the results shown in Table 2, we can conclude that the most accurate configuration is defined by a $W_{\text{size}} = 17$ and $N_{\text{bins}} = 15$ with the Random Forest algorithm. Regarding kNN,

Table 2. 2-class classification expanded results using Random Forest and kNN classifiers.

N_{bins}	Random Forest					kNN				
	W_{size}					W_{size}				
	11	13	15	17	19	11	13	15	17	19
11	88.28%	88.44%	88.45%	88.81%	88.66%	87.01%	87.80%	88.27%	87.64%	88.39%
13	87.69%	87.62%	88.94%	88.43%	88.61%	86.61%	86.97%	88.01%	88.04%	88.17%
15	88.33%	87.81%	88.70%	89.35%	88.89%	87.36%	86.87%	88.35%	88.97%	88.57%
17	88.24%	87.84%	87.56%	88.43%	87.94%	86.94%	87.99%	86.68%	87.80%	87.54%
19	87.79%	87.99%	88.06%	89.12%	88.89%	86.30%	86.53%	87.17%	88.77%	88.42%
21	88.09%	88.53%	88.38%	88.90%	88.58%	87.59%	87.60%	87.98%	88.63%	87.84%

the results are close to the ones provided by the Random Forest classifier, also obtaining the most accuracy with the mentioned parameters. Extreme values in both parameters decrease the accuracy of the classifier, since small values imply that not enough relevant information is being inputted to the classifier, while large values add noise that interferes with the classification.

As said, using the trained models, we applied them to entire OCT images, representing the results in intuitive colors that indicate the presence and non presence of the ERM, overlapping the scans for a better inspection by the user. This way, the clinicians can inspect the result in an intuitive and easy way, facilitating the ERM identification and, therefore, the diagnostic process. Figure 11 presents some representative examples of the results of the proposed method. The results are very precise in areas where the ERM is highly reflective, while in other specific areas the detection fails sometimes because the luminosity is not enough to provide a perfect classification (Fig. 12). Moreover, the reflective properties of the ILM can produce a bright region when the OCT laser falls perpendicularly upon the ILM, affecting the ERM detection process by introducing erroneous instances of ILM being classified as ERM (Fig. 13). Despite that, many of this extremely complex cases were correctly handled thanks to the defined feature set of the method.

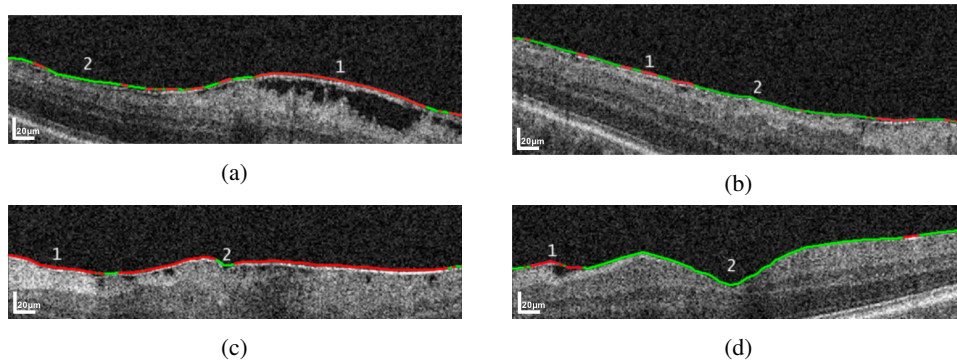


Fig. 11. Results of the method over representative OCT images. (1) Red symbolizes the ERM presence. (2) Green symbolizes the ERM absence.

With respect to the region of the eye fundus that is analyzed in the process of ERM identification, we want to emphasize that this retinal region of the tested OCT images is centered on the macula covering a 6mm of resolution, configuration frequently used by the clinical specialists in the analysis, diagnosis and treatment of this relevant eye disease.

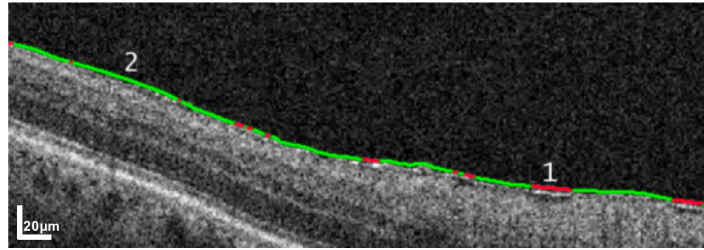


Fig. 12. Erroneous classification of the ERM case in an OCT image. Multiple ERM points possess low luminosity and they are not able to be identified by the method.

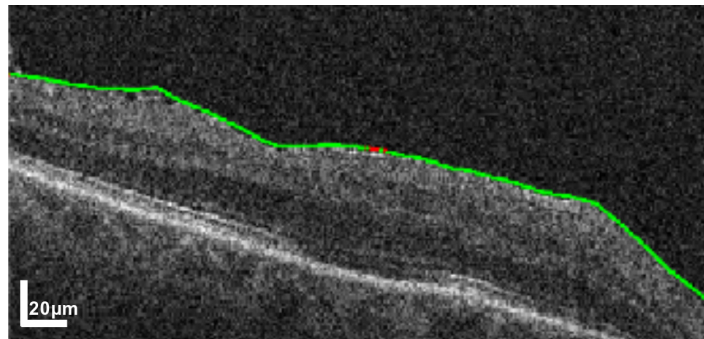


Fig. 13. ILM instance being erroneously classified as a ERM point caused by the high reflectivity properties of the ILM.

4. Conclusions

In ophthalmology, the pathological identification of the ERM in OCT images is a highly important issue in the analysis of relevant diseases such as diabetic macular edema or metamorphopsia, given that its presence on locations of the eye fundus provokes damages in the vision. For that reasons, its adequate identification helps in the prevention of further complications that could occur if the ERM is not early identified and treated. The proposed segmentation techniques for the independent analysis of the ERM and the retina could be used to correlate the anatomical and functional alterations in patients with ERM in order to better understand this pathological condition, its evolution and effects on the vision. Moreover, the possibility of being implemented and used in intrasurgical OCT could help, in a non-invasive way, the surgeon in the decision making process.

In this work, we presented an automatic methodology for the identification and characterization of the ERM presence in OCT images. We aimed to obtain an automatic method without any human interaction, in contrast with the few previous works of the field that were based on manual initializations of markers on the surface of the scan. In particular, the methodology is refined and improved by using learning strategies that extract and exploit useful information with a complete set of intensity, texture and domain-related characteristics to discriminate the ERM presence from the normal retinal tissue.

We used a feature set ranging from 432 to 482 features. These features were refined using a feature selection method to reduce the data dimensionality and avoid unnecessary and useless information by the system. Finally, representative classifiers were validated using an iterative cross-validation method and tested with selected representative datasets in order to measure the performance of the constructed system. Two different experiments were done to solve the class

split problematic and to determine empirically the most optimal number of samples to be used. The experiments demonstrated the suitability of the constructed methodology, providing a best accuracy of 89.35% in the validation set with the Random Forest classifier. The trained models, finally, are applied to entire OCT images, presenting graphically the results in intuitive colors for a better inspection of the clinical users and facilitate, therefore, the diagnostic process.

Future works plan the introduction of parallel computing techniques which contribute to the reduction of the processing time of the method. Likewise, GPU computing algorithms can ease the burden of the processor when doing image analysis and reduce the overall execution time. Furthermore, improvements in the accuracy of the method could be achieved by increasing the quality of the images and by introducing new classifier types which are more adequate to this problematic. To this end, we will validate the suitability of novel deep learning procedures for this proposal, since they have shown promising results for other problematics in similar research fields. In addition, the availability and analysis of complete OCT cubes is interesting with the potential of extending the used feature set to a 3D domain, which may facilitate the most complex identifications.

Funding

Ministerio de Economía, Industria y Competitividad, Gobierno de España (DPI2015-69948-R); Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia (ED431C 2016-047, ED431G/01); Instituto de Salud Carlos III (ISCIII) (DTS18/00136).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. C. A. Medina, J. H. Townsend, and A. D. Singh, eds. *Manual of Retinal Diseases* (Cham, 2016).
2. O. P. Gupta, G. C. Brown, and M. M. Brown, "A Value-Based Medicine Cost-Utility Analysis of Idiopathic Epiretinal Membrane Surgery," *Am. J. Ophthalmol.* **145**(5), 923–928.e1 (2008).
3. T. Yamamoto, N. Akabane, and S. Takeuchi, "Vitreotomy for diabetic macular edema: The role of posterior vitreous detachment and epimacular membrane," *Am. J. Ophthalmol.* **132**(3), 369–377 (2001).
4. N. G. Ghazi, J. B. Ciralsky, S. M. Shah, P. A. Campochiaro, and J. A. Haller, "Optical Coherence Tomography Findings in Persistent Diabetic Macular Edema: The Vitreomacular Interface," *Am. J. Ophthalmol.* **144**(5), 747–754.e2 (2007).
5. A. Gandorfer, M. Rohleder, S. Grosselfinger, C. Haritoglou, M. Ulbig, and A. Kampik, "Epiretinal pathology of diffuse diabetic macular edema associated with vitreomacular traction," *Am. J. Ophthalmol.* **139**(4), 638–652 (2005).
6. B. Bahrami, M. Zhu, T. Hong, and A. Chang, "Diabetic macular oedema: Pathophysiology, management challenges and treatment resistance," *Diabetologia* **59**(8), 1594–1608 (2016).
7. D. Sidibé, S. Sankar, G. Lemaître, M. Rastgoo, J. Massich, C. Y. Cheung, G. S. W. Tan, D. Milea, E. Lamoureux, T. Y. Wong, and F. Mériaudeau, "An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images," *Comput. Methods Programs Biomed.* **139**, 109–117 (2017).
8. G. Samagaio, A. Estévez, J. de Moura, J. Novo, M. Ortega, and M. Fernández, "Automatic Identification of Macular Edema in Optical Coherence Tomography Images," in *13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, (VISAPP, 2018), pp. 533–540.
9. A. Vahadane, A. Joshi, K. Madan, and T. R. Dastidar, "Detection of diabetic macular edema in optical coherence tomography scans using patch based deep learning," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (IEEE, 2018), pp. 1427–1430.
10. K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," *Biomed. Opt. Express* **9**(12), 6205–6221 (2018).
11. T. Otani, S. Kishi, and Y. Maruyama, "Patterns of diabetic macular edema with optical coherence tomography," *Am. J. Ophthalmol.* **127**(6), 688–693 (1999).
12. G. Panozzo, B. Parolini, E. Gusson, A. Mercanti, S. Pinackatt, G. Bertoldo, and S. Pignatto, "Diabetic macular edema: An OCT-based classification," *Semin. Ophthalmol.* **19**(1-2), 13–20 (2004).
13. J. G. Fujimoto, C. Pitris, S. A. Boppart, and M. E. Brezinski, "Optical Coherence Tomography: An Emerging Technology for Biomedical Imaging and Optical Biopsy," *Neoplasia (N. Y., NY, U. S.)* **2**(1-2), 9–25 (2000).

14. F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning," *Graefes' Arch. Clin. Exp. Ophthalmol.* **257**(3), 495–505 (2019).
15. Y. Hiram, A. Tsujikawa, M. Sasahara, N. Gotoh, H. Tamura, A. Otani, M. Mandai, and N. Yoshimura, "Alterations of retinal pigment epithelium in central serous chorioretinopathy," *Clin. & Exp. Ophthalmol.* **35**(3), 225–230 (2007).
16. P. L. Vidal, J. de Moura, J. Novo, M. G. Penedo, and M. Ortega, "Intraretinal fluid identification via enhanced maps using optical coherence tomography images," *Biomed. Opt. Express* **9**(10), 4730–4754 (2018).
17. M. H. Suh, J. M. Seo, K. H. Park, and H. G. Yu, "Associations Between Macular Findings by Optical Coherence Tomography and Visual Outcomes After Epiretinal Membrane Removal," *Am. J. Ophthalmol.* **147**(3), 473–480.e3 (2009).
18. M. Inoue, S. Morita, Y. Watanabe, T. Kaneko, S. Yamane, S. Kobayashi, A. Arakawa, and K. Kadonosono, "Inner Segment/Outer Segment Junction Assessed by Spectral-Domain Optical Coherence Tomography in Patients with Idiopathic Epiretinal Membrane," *Am. J. Ophthalmol.* **150**(6), 834–839 (2010).
19. P. Massin, C. Allouch, B. Haouchine, F. Metge, M. Paques, L. Tanguy, A. Erginay, and A. Gaudric, "Optical coherence tomography of idiopathic macular epiretinal membranes before and after surgery," *Am. J. Ophthalmol.* **130**(6), 732–739 (2000).
20. C. A. Puliafito, M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, J. A. Izatt, E. A. Swanson, and J. G. Fujimoto, "Imaging of Macular Diseases with Optical Coherence Tomography," *Ophthalmology* **102**(2), 217–229 (1995).
21. J. Michalewski, Z. Michalewska, S. Cisiecki, and J. Nawrocki, "Morphologically functional correlations of macular pathology connected with epiretinal membrane formation in spectral optical coherence tomography (SOCT)," *Graefes' Arch. Clin. Exp. Ophthalmol.* **245**(11), 1623–1631 (2007).
22. J. R. Wilkins, C. A. Puliafito, M. R. Hee, J. S. Duker, E. Reichel, J. G. Coker, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto, "Characterization of Epiretinal Membranes Using Optical Coherence Tomography," *Ophthalmology* **103**(12), 2142–2151 (1996).
23. S. Baamonde, J. de Moura, J. Novo, J. Rouco, and M. Ortega, "Feature Definition and Selection for Epiretinal Membrane Characterization in Optical Coherence Tomography Images," in *Image Analysis and Processing - ICIAP 2017*, (2017), pp. 456–466.
24. S. Baamonde, J. de Moura, J. Novo, and M. Ortega, "Automatic Detection of Epiretinal Membrane in OCT Images by Means of Local Luminosity Patterns," in *Advances in Computational Intelligence*, (2017), pp. 222–235.
25. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.* **1**(4), 321–331 (1988).
26. A. González-López, J. de Moura, J. Novo, M. Ortega, and M. Penedo, "Robust segmentation of retinal layers in optical coherence tomography images based on a multistage active contour model," *Heliyon* **5**(2), e01271 (2019).
27. K. Gawlik, F. Hausser, F. Paul, A. U. Brandt, and E. M. Kadas, "Active contour method for ILM segmentation in ONH volume scans in retinal OCT," *Biomed. Opt. Express* **9**(12), 6497–6518 (2018).
28. D. Nandi, A. S. Ashour, S. Samanta, S. Chakraborty, M. A. Salem, and N. Dey, "Principal component analysis in medical image processing: A study," *Int. J. Image Min.* **1**(1), 65–86 (2015).
29. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1 (2005), pp. 886–893.
30. B. Ramamurthy and K. R. Chandran, "Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms," *J. Comput. Sci.* **8**(7), 1070–1076 (2012).
31. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man. Cybern. SMC-3(6), 610–621 (1973).*
32. T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Inf. Retr.* **11**(2), 77–107 (2008).
33. M. Park, J. S. Jin, and L. S. Wilson, "Fast content-based image retrieval using quasi-Gabor filter and reduction of image feature dimension," in *Proceedings Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, (2002), pp. 178–182.
34. X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th International Conference on Computer Vision*, (2009), pp. 32–39.
35. K. I. Laws, "Textured image segmentation," Tech. Rep., DTIC Document (1980).
36. M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. thesis, The University of Waikato (1999).
37. M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmos. Environ.* **32**(14–15), 2627–2636 (1998).
38. I. Rish, "An empirical study of the naive bayes classifier," Tech. rep. (2001).
39. N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *Am. Stat.* **46**(3), 175–185 (1992).
40. T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1 (1995), pp. 278–282.
41. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD explorations newsletter* **11**(1), 10–18 (2009).