



Feature definition and comprehensive analysis on the robust identification of intraretinal cystoid regions using optical coherence tomography images

Joaquim de Moura^{1,2} · Plácido L. Vidal^{1,2} · Jorge Novo^{1,2} · José Rouco^{1,2} · Manuel G. Penedo^{1,2} · Marcos Ortega^{1,2}

Received: 1 August 2018 / Accepted: 27 August 2021 / Published online: 19 October 2021
© The Author(s) 2021

Abstract

Currently, optical coherence tomography is one of the most used medical imaging modalities, offering cross-sectional representations of the studied tissues. This image modality is specially relevant for the analysis of the retina, since it is the internal part of the human body that allows an almost direct examination without invasive techniques. One of the most representative cases of use of this medical imaging modality is for the identification and characterization of intraretinal fluid accumulations, critical for the diagnosis of one of the main causes of blindness in developed countries: the Diabetic Macular Edema. The study of these fluid accumulations is particularly interesting, both from the point of view of pattern recognition and from the different branches of health sciences. As these fluid accumulations are intermingled with retinal tissues, they present numerous variants according to their severity, and change their appearance depending on the configuration of the device; they are a perfect subject for an in-depth research, as they are considered to be a problem without a strict solution. In this work, we propose a comprehensive and detailed analysis of the patterns that characterize them. We employed a pool of 11 different texture and intensity feature families (giving a total of 510 markers) which we have analyzed using three different feature selection strategies and seven complementary classification algorithms. By doing so, we have been able to narrow down and explain the factors affecting this kind of accumulations and tissue lesions by means of machine learning techniques with a pipeline specially designed for this purpose.

Keywords Optical coherence tomography · Texture analysis · Feature selection · Computer-aided diagnosis · Classification · Feature analysis

✉ Plácido L. Vidal
placido.francisco.lizancos.vidal@udc.es

Joaquim de Moura
joaquim.demoura@udc.es

Jorge Novo
jnovo@udc.es

José Rouco
jose.rouco@udc.es

Manuel G. Penedo
mgpenedo@udc.es

Marcos Ortega
mortega@udc.es

¹ CITIC-Research Center of Information and Communication Technologies, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

² Universidade da Coruña, Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Xubias de Arriba, 84, A Coruña, Spain

1 Introduction

Currently, the study of the retina represents one of the main approaches to the non-invasive study of the human body. The study of the optic disc [32], arterio-venous structures [30] or alterations in the histological composition of the retina [2] are capable of providing useful information not only for diseases affecting the ocular system, but also for pathologies of general impact. At present, of the medical imaging modalities available that focus on the analysis of these structures, we can highlight optical coherence tomography (OCT) [19]. This modality is able to generate, usually by means of interferometry combined with coherent light beams [35], a cross-sectional representation of the main retinal tissues. The relevance of the study of these structures is such that it is possible to obtain information about cardiovascular [20, 45], neurological [12] or even metabolic pathologies; like the nowadays prominent diabetes [44].

Thus, the study presented in this paper addresses the analysis of retinal structures affected by pathological fluid accumulations. Nowadays, pathologies such as age-related macular degeneration (AMD) [21] or diabetic macular edema (DME) [22, 40] represent one of the main causes of blindness in developed countries, both as a consequence of consumer habits and the aging of the human population. These pathologies trigger a degeneration of the delicate retinal tissues that causes the progressive appearance (either by leakage from the damaged arterio-venous structures or by degradation of the natural filtering mechanisms of the retinal tissues) of intraretinal fluid accumulations. These accumulations gradually harm the delicate and complex histological organization of the retina, causing progressive vision loss in the patient. In Fig. 1, we show a representative example of a image where different types of fluid accumulations have completely deformed the retinal architecture. As we can see, the patterns present in a healthy retina and one affected by the aforementioned leakages are completely different. We can even differentiate three types of accumulations in this image. The upper one, with cystic bodies; the intermediate one, with fluid intermixed with retinal tissues with a spongi-form texture; and the lower one, a detachment of the lower retinal layers near the central point of vision (the fovea).

In some cases, these fluid accumulations are very difficult to identify due to the presence of diffuse regions. Moreover, for a correct diagnosis, experts analyze hundreds of images that comprise a complete OCT scan. The number of samples required for analysis, together with the misshapen forms they present (often mistaken for healthy tissues or other pathological bodies) often results in a stronger impact of the subjectivity of the clinical expert. This, consequently, severely hampers clinical follow-up or even early detection: both critical for the correct improvement of the patient and to prevent irreversible damage [3].

It is precisely because of these two factors (the clinical relevance of the fluid patterns as a biomarker for numerous critical pathologies and the level of influence of the subjectivity in the matter) that different automatic methodologies for their analysis have been proposed. These usually follow a basic scheme: first, they remove any unwanted patterns (usually related to noise or artifacts product of the capturing device); second, an initial candidate generation strategy that obtains a set of representative examples (these range from a simple thresholding to even most modern deep learning-based strategies); finally, they perform a false-positive (FP) pruning, optimizing the rough detections into the proper final representation of the fluid accumulations.

To exemplify these cases, we can refer ourselves to two of the first works in the domain, by Wilkings et al. [43] and Roychowdhry et al. [37]. In these works, they detect cystoid fluid accumulations with a thresholding strategy and posterior filtering of FPs based on a set of rules/basic intensity features. The initial artifact removal is based on the reduction of the region of interest to the particular retinal region, completely ignoring the patterns in the choroidal and vitreous humor regions (both the outermost and innermost regions of the human eye: the internal fluid of the eye and the main vascular layer that feeds the retina). Wiclawek [42] and González et al. [16] improved these middle steps by using more advanced initial candidate generation strategies, like using clustering techniques such as the watershed transform to obtain an initial set of candidates.

Regarding noise, an example can be seen in Dehnavi et al. [11], where an initial K-SVD dictionary learning in the Curvelet transform is used to reduce speckle noise. This way, a methodology like the proposed by Wilkings et al. [43] based on empirical thresholdings is more viable, as the initial candidate generation is less prone to FPs (albeit they also preserve the posterior FP filtering step). This strategy was

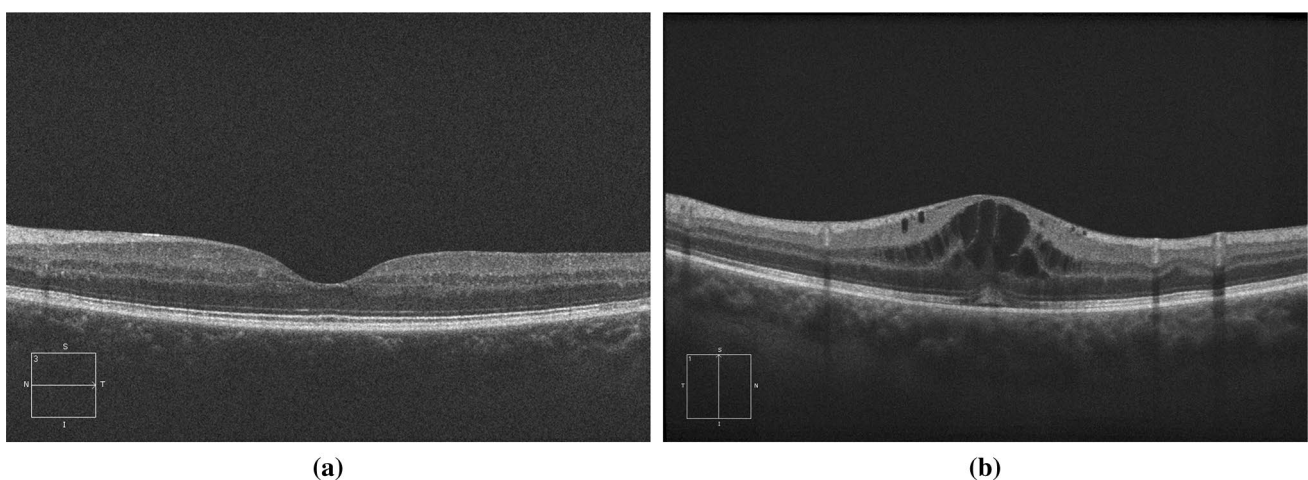


Fig. 1 Macular OCT images. **a** OCT without the presence of cystoid regions. **b** OCT with the presence of cystoid regions

also improved by Wang et al. [41] with fuzzy level sets. The initial candidate selection is obtained from a fuzzy domain to, then, optimized to a precise contour segmentation with level sets. Finally, other works took a more clinically centered approach, focusing on medical knowledge to obtain the candidate selection. For example, Xu et al. [46] used the information of the retinal layers to obtain a defined set of candidates based on a voxel classification, or Lang et al. [26] and Swingle et al. [39], that limited their work to a particular subset of fluid accumulations with their own medical consequences: micro-cystic fluid leakages.

On the one hand, Fig. 2 shows a set of examples with different fluid accumulation patterns. Some of them, like the ones in the first row, present almost clear limits, with high contrast and defined cellular barriers. On the other hand, patterns like the fluid accumulations present in the second row of that same figure are severely intermixed with normal retinal tissues, suffering from contrast issues product of the capturing strategy or their cellular barrier has been broken/still has not been formed, making increasingly difficult the task of diagnosing/correctly segmenting the fluid accumulations (as there may not even be a defined limit to detect). Its in these situations where works of the state of the art, either based on classical learning algorithms or modern deep learning strategies, tend to fail to obtain consistent results with different images (even when belonging to the same patient). In this work, we try to establish a proper baseline for future works regarding what biomarkers and features present in the OCT images are key to the issue at hand. By doing so,

we hope that future pattern recognition strategies can take advantage of our findings to develop more robust computer-aided diagnosis methodologies with less computational resources. Additionally, our findings could be extrapolated to the medical domain, where the patterns found to be relevant for this pathology can be used to better assess and diagnose it as well as others of similar nature. In summary, the main contributions of our work are as follows:

- Methodology to determine the main patterns and components that characterize a pathology in OCT images.
- Analysis methodology easily transferable to other medical imaging modalities and pathologies.
- Implementation of our proposal in the detection of intraretinal fluid, a complex unresolved domain where an early detection is critical.
- Generation of robust and explainable results, thanks to the analysis of a fine-grained and heterogeneous library of 510 feature markers.

This paper is organized in the following sections: Sect. 2, “Methodology,” presents the steps followed to perform the proposed analysis. Section 3, “Resources and configurations,” presents a compendium of all the resources used for the proposed analysis: dataset, classifiers, selectors and feature library (as well as the particular configuration used in each case). Section 4, “Results,” exposes the metrics obtained as a result of following the proposed methodology. These results are further discussed in Sect. 5, “Discussion.”

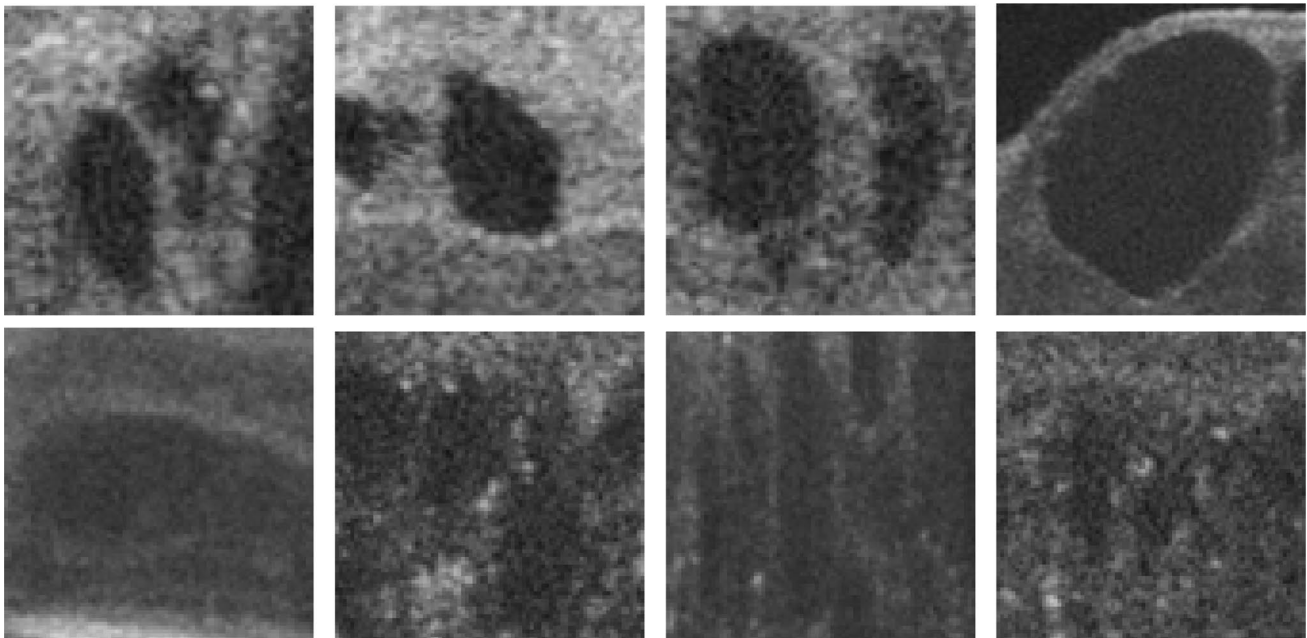


Fig. 2 Examples of cystoid regions with different levels of complexity. 1st row, cystoid regions with clear limits. 2nd row, fluid accumulations mixed with normal retinal tissues

Finally, Sect. 6, “Conclusions,” presents a brief final notes on the presented study.

2 Methodology

In this section, we will proceed to illustrate each of the steps followed to conduct the analysis proposed in this work. These steps are illustrated in Fig. 3. As can be seen in this figure, our analysis is divided into four main stages. The first one consists of a specific sampling to create the dataset to be analyzed. Subsequently, from these samples, we will extract a series of features that will define under different points of view relevant characteristics of the extracted samples.

Once we have collected the features, we will use different feature selection strategies to determine the most relevant ones. Finally, by using different families of classifiers, we will evaluate the contribution of these subgroups of features to the studied problem.

2.1 Dataset creation

First, we will create the dataset. For this, we will need to extract representative samples of the cases to be studied. In this case, the different forms in which the fluid accumulations may appear as well as the different patterns of other tissues present in the retina. Before extracting these

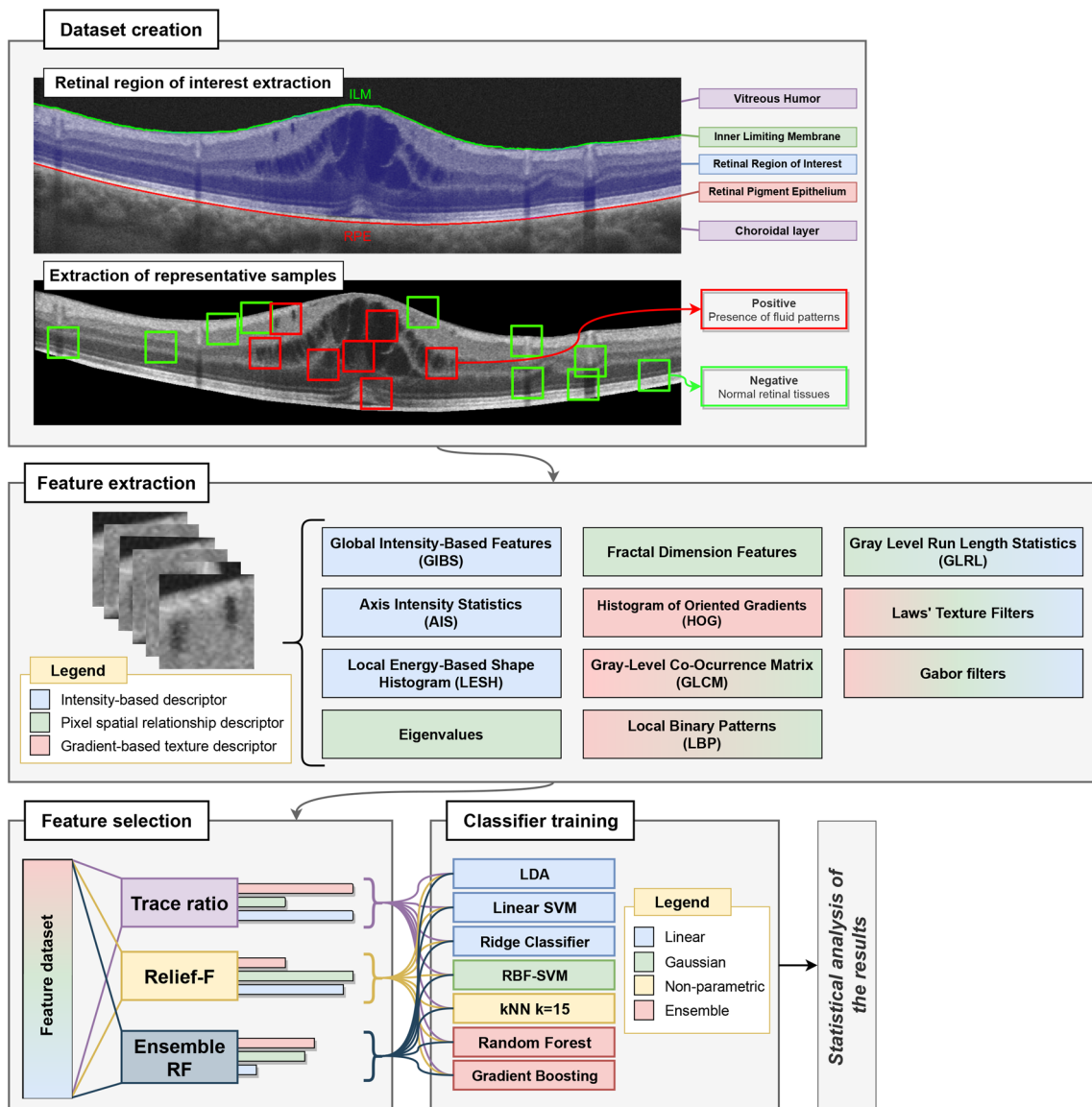


Fig. 3 Main stages of the proposed methodology

samples, we will first remove the non-retinal region to reduce the analysis space.

2.1.1 Retinal region of interest extraction

In this case, this target region of interest is limited by the Inner Limiting Membrane (ILM) and the Retinal Pigment Epithelium (RPE). These retinal layers border the vitreous humor and the choroidal layer, respectively. Eliminating these two regions (the vitreous humor and the choroid) from the analysis is particularly interesting, since the vitreous humor presents fluid patterns similar to those of the studied subject (having similar density) and the choroid presents oval patterns similar to cystic-like patterns also contained within the scope of the study. Thus, by eliminating them, we were able to obtain more significant results in the posterior analysis. These layers have been extracted thanks to the methodology proposed in the work of Chiu et al. [7]. To do so, the vertical gradients of the different retinal layers are used as weights to find the optimal path between both ends of the OCT image. As can be seen in Fig. 3, every layer of the retina (including the limiting ones mentioned above) presents these patterns as a defining property. Thus, using these gradient profiles, we can find the two main layers that delimit the regions of interest in our work.

2.1.2 Extraction of representative samples

Now that we have the regions of interest fully delimited, it is time to extract the samples. Since we intend to examine retinal tissues and fluid accumulations in detail, we used windows of a particular size that exclusively covered each of the two cases mentioned above (and, within each case, the different ways in which the patterns can present themselves). Each of the samples was selected by domain experts in order to cover the different pathological and non-pathological cases needed for the work, thus ensuring the heterogeneity and representativeness of the dataset.

2.2 Feature extraction

To study the patterns present in the retina, we have extracted a comprehensive library of features. These features have been chosen to illustrate three main defining factors of an image: distribution of intensity levels in the sample, spatial relationship of the pixels in the image and organization/orientation of the gradients they exhibit. In the following sections, we will illustrate in more detail why each particular feature class has been chosen.

2.2.1 Intensity-based descriptors

First, and as a basis for the analysis, we focus on features that analyze the non-spatial distribution of the different gray levels. This is because, due to the nature of the fluid accumulations, they appear as regions with high contrasts and low luminance gray levels. On the other hand, for more complex retinal patterns, this family of descriptors can be used as a complement to the other classes considered to significantly improve their ability to characterize retinal fluid regions.

2.2.2 Pixel spatial relationship descriptors

These descriptors based on textural information are useful for determining the distinctive patterns in both considered classes. Fluids present more homogeneous patterns in the inner regions, while more irregular textures as they intermingle with healthy tissues. In addition, each of the retinal layers present unique patterns depending on their biological function. This is also true for other retinal structures, both healthy (such as vessels) and pathological (such as accumulations of dense wastes).

2.3 Gradient-based texture descriptors

Gradient-based descriptors are particularly interesting in this case. Healthy retinal tissues (regions with intact histological organization) tend to exhibit horizontal gradients in a limited range of orientations. On the other hand, fluid accumulations tend to exhibit gradients in almost all directions (being spheroid-like accumulations). In more extreme cases, the very absence of gradients (as in thickened spongiform regions of the retina) can also be a good biomarker. Finally, vessels crossing the retina are easily detectable by this strategy, as they present defined vertical gradients with the shadows they project onto more external layers of the retina.

2.4 Feature selection

After collecting the aforementioned set of features, we will proceed to the selection of the most relevant ones from the total set. As we will demonstrate, having such an exhaustive library, it is common for redundancy to exist in the library. For this reason, by means of feature ranking strategies, we will be able to evaluate which ones provide more information.

In this case, we have evaluated three different methodologies for feature selection. One based on the study of graphs to analyze the inter- and intra-class distance (the Trace ratio estimator), another based on random subsampling with neighbors of the same and different classes to perform a density estimation similar to a k-Nearest Neighbors (ReliefF), and the third based on the training of a pool of decision

trees for the analysis of their relevance (since the features in shallower positions of the tree, in theory, are more representative of each class). Thus, we can see that each selector considers to a different extent the relevance of the sets and the relationship of new features with the previously selected. While the trace ratio is more sensitive to the dataset, finding features less dependent on each other, the Random Forest is less sensitive to the particularities of the dataset, likely generating rankings that consider features that are dependent on the previous ones. In this case, the ReliefF was chosen to act as a middle ground between both extremes.

2.5 Classifier training

Finally, to perform the actual relevance analysis of these features, we will study four families of classifiers. In this way, we intend to analyze different levels of complexity of the studied features, allowing from less to more degrees of freedom and by means of different optimization strategies. In this work, we have considered from Linear and Gaussian models to nonparametric (k Nearest Neighbors) and based on pools of simple classifiers (ensemble). These classifier paradigms were considered as they include models with low fitting capabilities but high generalization (such as linear models) and others with high fitting capacity, but with a high possibility of overfitting to the problem presented (as in the case of ensemble-based classifiers).

The classification was done using a constructed dataset that was randomly divided into a 10-fold cross-validation repeated 10 times for each selector (for a total of 3 selectors) and classifier configuration (for a total of seven classifiers). Additionally, a model was trained for each increasing subset of features to further study the behavior of each selected feature to a maximum of the top 100 selected features (as most selectors failed to properly assign a score to the features from this point onward due to the lack of relevance to the matter).

3 Resources and configurations

In this section, we will detail the resources used and their specifications required to replicate our work. First, in Sect. 3.1 “Dataset,” we explain in detail the characteristics of the employed dataset. Subsequently, in Sect. 3.2 “Feature library,” we explain each of the families of descriptors used in our work and the particular markers. Finally, in Sects. 3.3 and 3.4 “Feature selectors” and “Classifiers,” we present each of the algorithms and classifier configurations implemented in our research to both select the most relevant features and to determine their actual discrimination power.

3.1 Dataset

The dataset used in this work consists of 83 OCT images captured by means of a CIRRUS™ HD-OCT-Carl Zeiss Meditec. In order to better represent the case study, the dataset contains images of both the left and right eye, varying in resolution from 750×500 pixels to 1680×1050 pixels. In addition, they have been captured using different device configurations and used without any kind of preprocessing in order to preserve the original properties and features of the eye fundus tissues. Clinical experts established the ground truth to be used for the extraction of the representative samples from the images, resulting in a total of 806 and 803 windows with and without the presence of cysts, respectively, summing a total of 1609 samples.

From each coordinate in the image deemed as having relevant patterns for the study, different windows sizes were extracted. In this work, the window sizes considered were 11×11 pixels, 15×15 pixels, 21×21 pixels, 31×31 pixels, 41×41 pixels, 51×51 pixels and 61×61 pixels. The non-linear rate of increments in window size has been considered due to the progressively lower information contribution with respect to the one already contained in the smaller sample sizes.

3.2 Feature library

Below, we list and describe each of the feature families used in the development of this manuscript. The index numbers of the descriptor in the final feature vector generated for each sample are also indicated.

3.2.1 Global intensity-based features (GIBS) [1–15]

Basic descriptor of the statistical gray level distribution, ignoring spatial properties: maximum, minimum, mean, median, standard deviation, variance, entropy, 25th & 75th percentile and maximum likelihood estimates for a normal distribution.

3.2.2 Axis intensity statistics (AIS) [16–27]

Minimum, mean and maximum skewness and kurtosis of the sample along both vertical and horizontal axis. Despite technically not considering the spatial distribution of the gray levels, this descriptor compares the tendency of the patterns in each of the axis of the sample toward lighter or darker textures. By doing so, we could be able (in theory) to distinguish between the vertical patterns of a shadow, uniform

shapes of spongiform fluid accumulations and oval/round structures formed by cystoid leakages.

3.2.3 Eigenvalues [28–55]

We calculated the eigenvalues of the analyzed region and selected the 4 highest (λ_{\max_i}) and the 4 lowest (λ_{\min_i}) values. Additionally, the ratios among them and their 25 and 75 percentile were also included as markers to simplify the future abstraction needed to be made by the classifiers. If the sample is not square, it is divided into multiple square windows to extract these eigenvalues.

3.2.4 Local energy-based shape histogram (LESH) [56–183]

Based on the Gabor descriptor [38], we analyze the texture energy using the phase congruency approximation by Kovesi [24, 25] as metric. By finding the maximal phase congruency, this metric becomes quite robust against samples with a high rate of noise [23].

3.2.5 Gray-level co-occurrence matrix (GLCM) [184–199]

This descriptor measures the occurrence of gray levels i and j in pairs of pixels (p_1, p_2) of the analyzed OCT image I , separated by a displacement vector $\delta = (\Delta x, \Delta y)$ into a 2D matrix, given by the following:

$$C_\delta(i, j) = \left| \left\{ p_1, p_2 : I(p_1) = i; I(p_2) = j; p_2 = p_1 \pm \delta \right\} \right| \quad (1)$$

From the resulting matrix, we can extract the contrast, correlation, energy and homogeneity of the textures [18] In this case, the pixel distance was set to two pixels and in four directions: 0° , 45° , 90° and 135° .

3.2.6 Histogram of oriented gradients (HOG) [200–280]

As the name suggests, this descriptor describes the texture present in a particular image or sample by generating a histogram of the detected gradients in the image [9]. Each sample is divided into 9 hog windows, from where the gradient orientations are extracted. A gray level resolution of 9 bins was chosen as a balance between granularity and abstraction of information. This descriptor presents invariance to rotation, translation and scale.

3.2.7 Gabor filters [281–408]

This orientation-sensitive filter works in a similar way to the Fourier transform, but applied to the image domain [14]. This filter is able to decompose the image into its constituents, by generating a library of convolutional masks that specially react to textures in certain orientations and with

certain pixel frequencies. These filters are two-dimensional sinusoids of a given frequency and orientation. Furthermore, a normal distribution of a given mean and standard deviation is used to mask the two-dimensional sinusoid of convolutional filter, to balance the relevance of the central region of the convolutional filter. Thus, following the parametrization defined in [17], we obtain a bank of filters with 8 scales of frequencies and in 8 orientations, invariable to scale or translation. As we obtain the mean and standard deviation of each marker, we obtain a vector of 128 features.

3.2.8 Local binary patterns (LBP)[409–472]

Texture descriptor based on the analysis of a circular neighbor of pixels. Based on the binary comparison between these radial pixels and the center one in illumination, we can obtain a binary descriptor of the underlying texture [33]. In this work, the mean and standard deviation of four filter sizes (4, 8, 12 and 16 pixels) in 8 different equidistant orientations are used. Based on previous experiments, a non-rotation invariant version of the descriptor was employed as the orientations of the textures plays an important role in their class.

3.2.9 Laws texture filters [473–500]

This descriptor consists of a series of 2D convolution filters generated from a set of one dimensional kernels [27]. In this work, we considered the following kernels for window of sizes 3×3 and 5×5 : $L3 = [1 \ 2 \ 1]$, $E3 = [-1 \ 0 \ 1]$, $S3 = [-1 \ 2 \ -1]$, $L5 = [1 \ 4 \ 6 \ 4 \ 1]$, $E5 = [-1 \ -2 \ 0 \ 2 \ 1]$, $S5 = [-1 \ 0 \ 2 \ 0 \ -1]$ and $R5 = [1 \ -4 \ 6 \ -4 \ 1]$. L kernels focus on obtaining the center weighted gray level mean, E kernels respond to edge features, S extract spots and R focus on ripple patterns in the texture. From the result of applying these convolutional kernels, we extract the mean and the standard deviation to generate the final descriptor.

3.2.10 Fractal dimension [501–503]

Descriptor based on the analysis of the texture self-similarity [1]. Using the box-counting algorithm [6] we can study how the patterns are similar to themselves on different scales. We analyze the mean, standard deviation and the lacunarity of the texture (how well the pattern fills the space occupied by the texture).

3.2.11 Gray level run length image statistics (GLRL) [504–510]

This family of features describes the texture in terms of the different homogeneous chains of similar gray-level pixels in a given set of orientations [29]. In our case, the images

were reduced to a set of 51 bins to improve the robustness against intensity variations due to noise or capture device artifacts. The final descriptor consists of the Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-uniformity (GLN), Run Percentage (RP), Run Length Non-uniformity (RLN), Low Gray Level Run Emphasis (LGRE) and High Gray Level Run Emphasis (HGRE).

3.3 Feature selectors

In this work, three feature selector strategies were used: Trace ratio estimator [28, 31], Relief-F [28, 36] and an ensemble-learning model which uses the random forest algorithm. In our case, this ensemble-learning model was based on the randomized trees variation [15, 34] with 1.000 estimators. For the Trace ratio estimation, we used the approach of Nie et al. [31] as it tries to optimize the feature subset score, resulting in more robust feature rankings than other feature selection strategies. Finally, for the Relief-F estimator, we used a number of neighbors of 10, as a higher number of neighbors proved to be detrimental to feature subsets with less feature descriptors.

3.4 Classifiers

To perform the classification stage, seven different classifier configurations were employed based on their classification capabilities and kernel functions: three linear models, a Gaussian, a nonparametric classifier and two ensemble strategies. The three linear models were a Linear Discriminant Analysis or LDA, a Support Vector Machine (SVM) without kernel trick and a Ridge Classifier. The employed LDA uses Bayes' rule to fit a Gaussian density to each class. The main objective of this classifier is to maximize the ratio between the inner scatter and the outer scatter of the classes. On the other hand, the linear SVM [8] establishes a hyperplane between the two most closest samples (support vectors) from both classes, optimized by quadratic programming. Finally, the Ridge-regularized linear model is a linear regression

model, used for classification but with an added constraint to the loss function that prevents overfitting. This is done by imposing a penalty on the size of the linear regression coefficients.

The Gaussian model is based on a SVM by employing the kernel trick. This consists of applying a function to the search space, so the previously non-separable data is now separable in the new synthetic dimension applied to the dataset. In our case, our model optimizes the radial basis function.

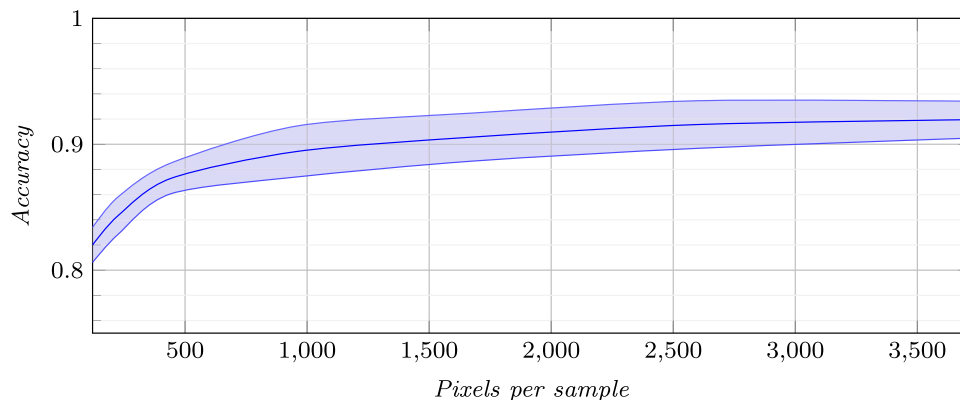
As a nonparametric model, we use a k-Nearest Neighbors (kNN) classifier that approximates the class probability based on the neighbor density that belongs to that same class. We tested different configurations, and for our large library of features a $k = 15$ offered a good balance between generalization and accuracy without overfitting [10]. Additionally, for each query, the points are weighted by the inverse of their distance, making closer points more influential when determining the resulting class.

Finally, as ensemble methods, we used two variations of an implementation of the Random Forest. As its name implies, it creates a forest of classifiers with random subsets of features and picking the best splits [4, 5]. The classifiers are combined by averaging their probabilistic prediction. The Gradient Tree Boosting classifier [13], instead, sequentially builds an additive model. A total of 100 fixed size regression trees are trained and the log-likelihood is optimized in each iteration via gradient descent toward the negative gradient of the loss function.

4 Results

In the first place, in order to begin the analysis, we performed a preliminary initial test with a reduced number of classifiers and features. This experiment was tested with linear, nonparametric and quadratic classifiers and a reduced number of features. The result, presented in Fig. 4, shows how no significant variation is exhibited from a window size

Fig. 4 Mean and standard deviation of the tested models to assess the optimal window size



greater than 2500–3000 pixels. For this reason, larger sizes were not analyzed and we consider 61×61 the optimal size for future experimentation.

First, after feature extraction, we analyzed the correlation between the extracted marker classes. Many of the vectors used consider features similar to each other. For example, LAWS and Gabor include similar convolutional masks in some cases. For this reason, as seen in Fig. 5, the correlation between these classes appears to be strong.

In the same way, we see how Gabor, GLCM and GIBS are very sensitive to intensity changes, so both show high correlation coefficients. However, we can see that in most cases the correlation levels are considerably low (highlighting the case of LESH with correlation levels lower than all the other categories). Another aspect to note is the high degree of autocorrelation of GIBS, LBP and FD. It is normal for a certain degree of correlation to exist within the same family of features when addressing complementary points of view.

However, such high levels of correlation indicate that there is a high level of potential redundancy. This is likely to result in that, within these feature families, few markers will be selected. This is critical especially when a family of features has a low number of markers per se, as is the case with FD.

Nonetheless, correlation does not directly imply relevance for the considered issue. As shown in Fig. 6, while LESH has an extremely low correlation coefficient, the feature selectors have hardly selected a lot of its features as relevant from the total available for this family. This does not necessarily indicate that it is a poor choice as a marker. What these two figures say (Figs. 5 and 6) say is that it is a very heterogeneous marker, and a few of its features are actually relevant to the problem under consideration. On the other hand, the opposite case is the one of FD. This feature has not been chosen at all by any of the three selection strategies. Possibly due to the fact that it is quite correlated with LBP and LAWS.

Fig. 5 Mean correlation by feature classes

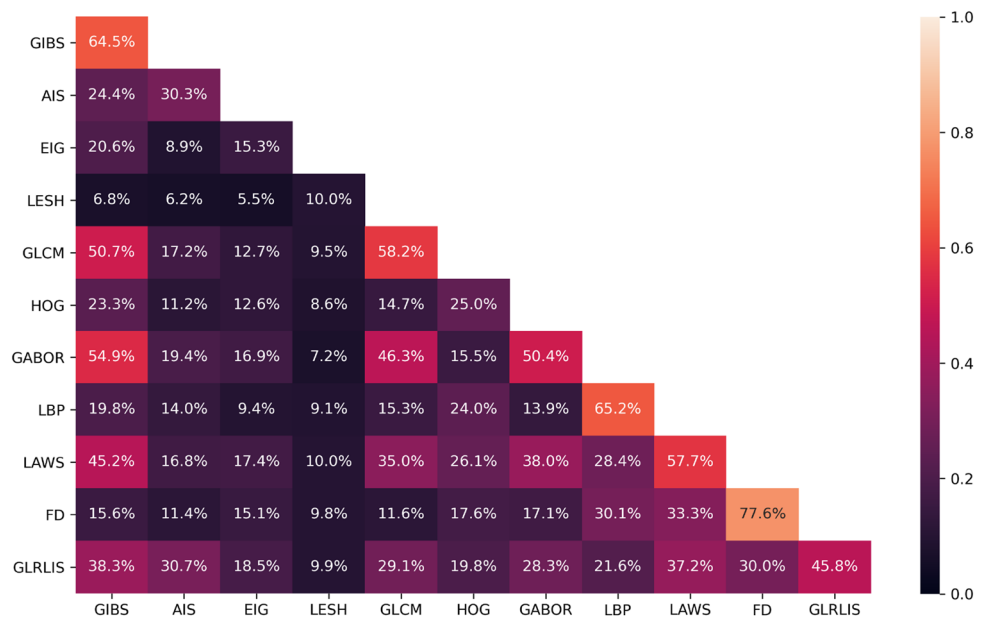
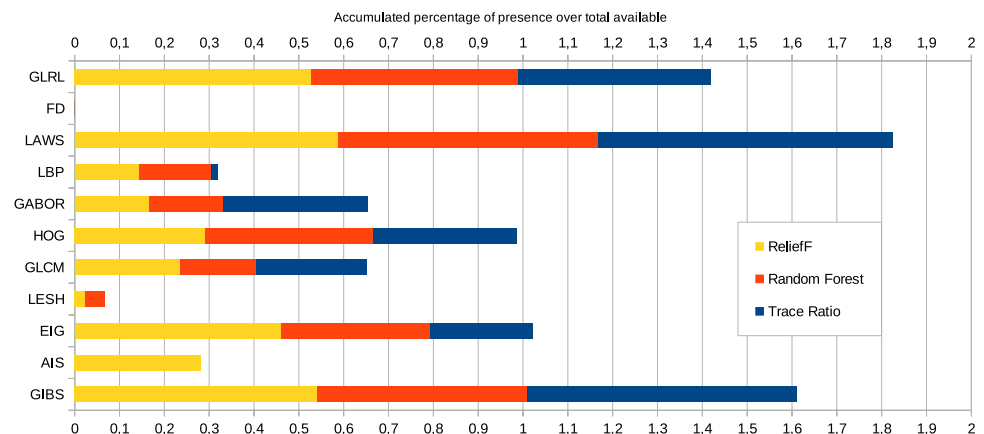


Fig. 6 Percentage of features chosen over total available for each feature family



By examining from another point of view in Fig. 7, we can see how LAWS also suffers in terms of positioning of its markers. In this figure, the score accumulated by the three selectors for each family and LESH (even not taking into account the fact that it has not even been chosen by the Trace ratio) does not usually appear in the top positions. As we mentioned, possibly because it is correlated with better positioned characteristics such as LAWS that return similar and more information with less markers needed (that is, they are more information-efficient). Similarly, we see how the selectors clearly favor GLRL and HOG in the top positions even though HOG was proportionally less chosen. This is a similar case to the one we were previously discussing: a selector from which we do not take many features, but the ones we do take are the ones that are truly relevant to the problem.

As for the training results, we can see in Figs. 8, 9 and 10 the results of the progressive training. For each number of features a model has been trained with the subset of the ranking up to that position. In these figures, we can see three main factors to consider. The first is that classifiers of

any complexity stabilize at 50 features or more. Therefore, from that point onward, we are directly sacrificing generalization capability (by resorting to more complex models) in order to gain a minimum of accuracy. Secondly, we see how complex models such as those based on ensemble classifiers have shown a tendency to obtain better results in the initial training stages, but converging with the simpler linear models toward the end. For this reason, and in order to favor generalization capacity, we consider linear models to be the appropriate ones for the case currently under consideration. Finally, we can see how only the ensemble-based selector has been able to maintain a certain level of stability in early epochs along all classifiers, while with few features the other selectors obtained, in general, worse results.

In Fig. 11, we can see a comparison between the best results of each model and selector together with their standard deviation. We can observe how the linear models have obtained a higher mean accuracy than the rest, while at the same time they have obtained the lowest standard deviations. On the other hand, classifiers such as Random Forest, considered more prone to overfitting, have obtained some of

Fig. 7 Accumulated maximum feature relevance for each feature family

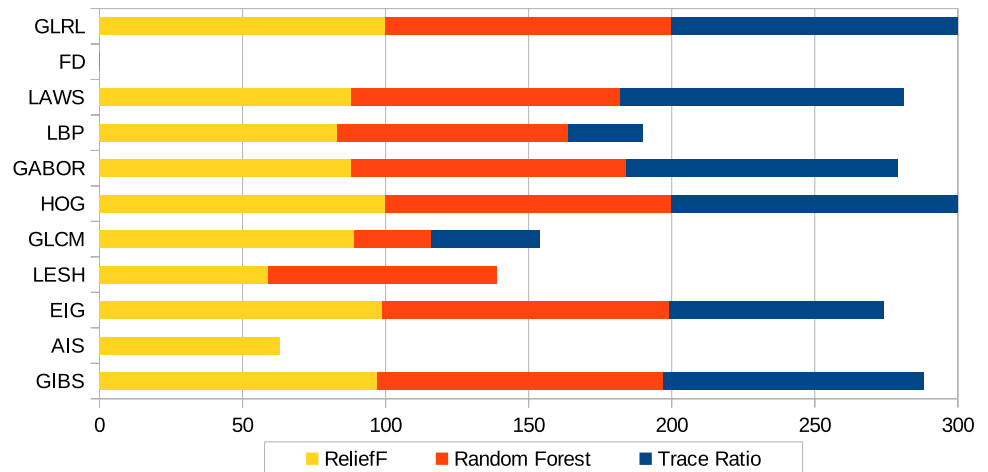


Fig. 8 Evolution of the mean accuracy (and its standard deviation along all repetitions) for each classifier family tested with an increasing subset of features obtained with the trace ratio selector

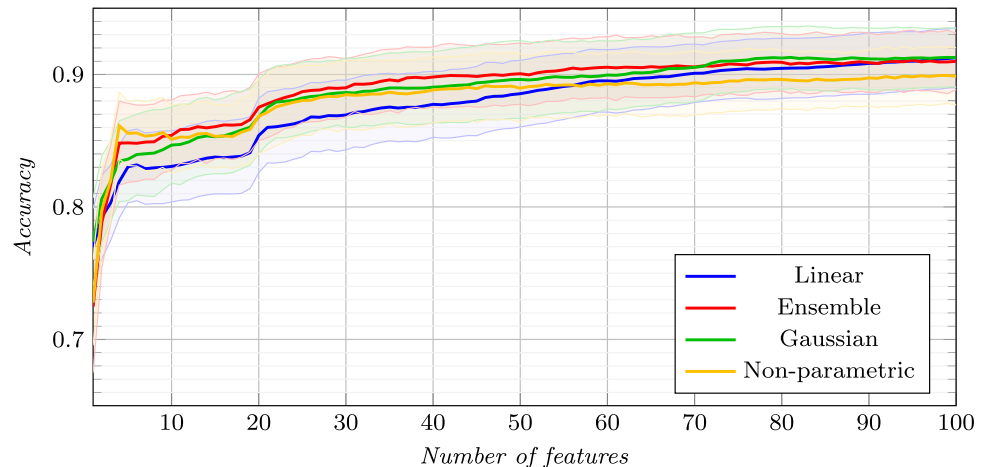


Fig. 9 Evolution of the mean accuracy (and its standard deviation along all repetitions) for each classifier family tested with an increasing subset of features obtained with the random forest selector

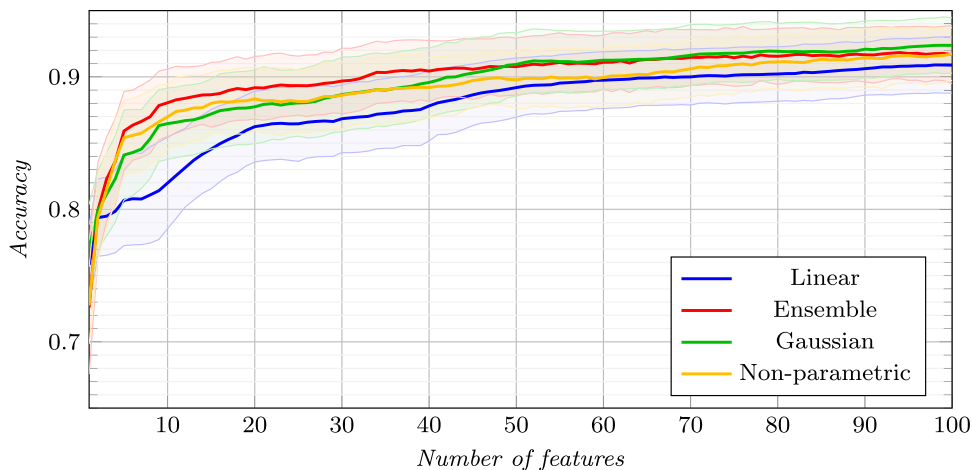


Fig. 10 Evolution of the mean accuracy (and its standard deviation along all repetitions) for each classifier family tested with an increasing subset of features obtained with the ReliefF selector

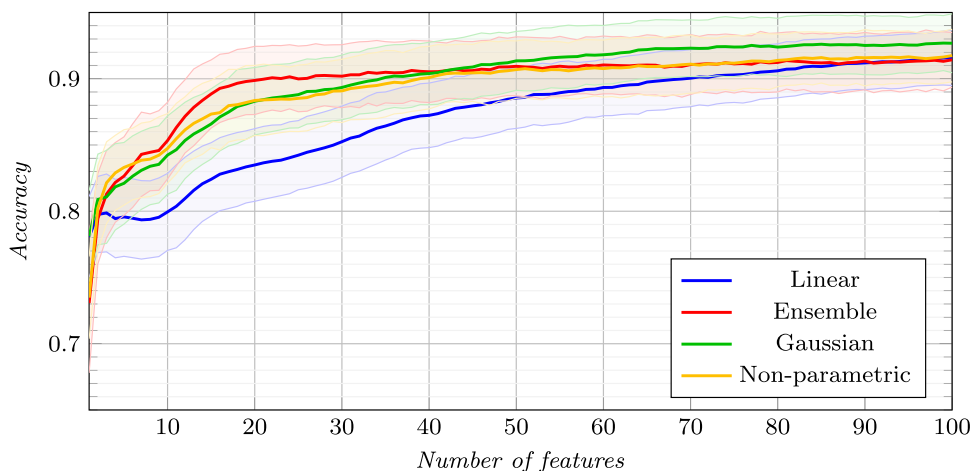
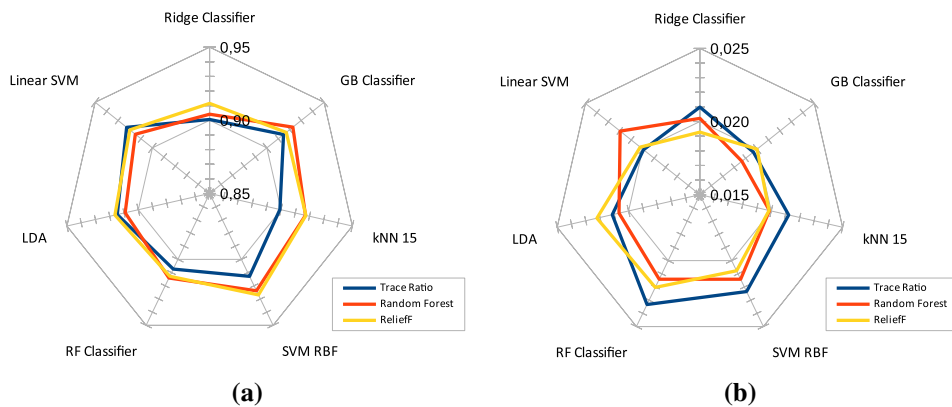


Fig. 11 Best mean accuracy (a) and its standard deviation (b) achieved by each classifier and selector



the worst results in all selectors with a high standard deviation. Ultimately, and taking into account that the models considered for the evaluation in this figure of the best accuracy include all the trained ones (up to 100 features), we can

see how clearly the complex models have ended up being over-trained while the linear models have been able to infer correctly in spite of the higher dimensionality burden. That

is, their own limitations have prevented the dimensionality problem.

5 Discussion

As we could observe, the tests clearly favor feature families based on gradient learning such as HOG, LAWS and GABOR. In addition, we see how GLRL is also in these privileged positions. In Fig. 12, we can see some examples of correctly classified positive and negative samples. By looking at these samples, we can see why the choices were made. As we commented earlier, vertical patterns clearly indicate shadowing, and horizontal patterns tend to belong to healthy tissue samples (being characteristic of the histological organization of the retina). Similarly, cystoid regions ought to be those with patterns that are either undirected (homogeneous tissue) or present in a larger number

of directions (as these tend to have oval or heterogeneous shapes). However, looking at the examples given, we can observe that this is not sufficient for all the cases presented. This is where the other family highlighted in the previous tests comes in: GIBS. The combination of intensity-based samples together with gradient descriptors makes it easy to identify pathological regions. Healthy tissues usually manifest themselves with hyper-reflective patterns, clearer on OCT images (except in rare cases with artifacts). These homogeneous regions in given directions are also the reason for why GLRL was among the top descriptors. Homogeneous patterns in all directions are a probable indicative of a fluid accumulation of considerable size.

In Fig. 13, we can see some samples where the proposed system was not able to correctly predict the sample category. For example, in Fig. 13a, the region is marked as healthy despite having a clear presence of pathological fluid. This is due to the appearance of white dots that are

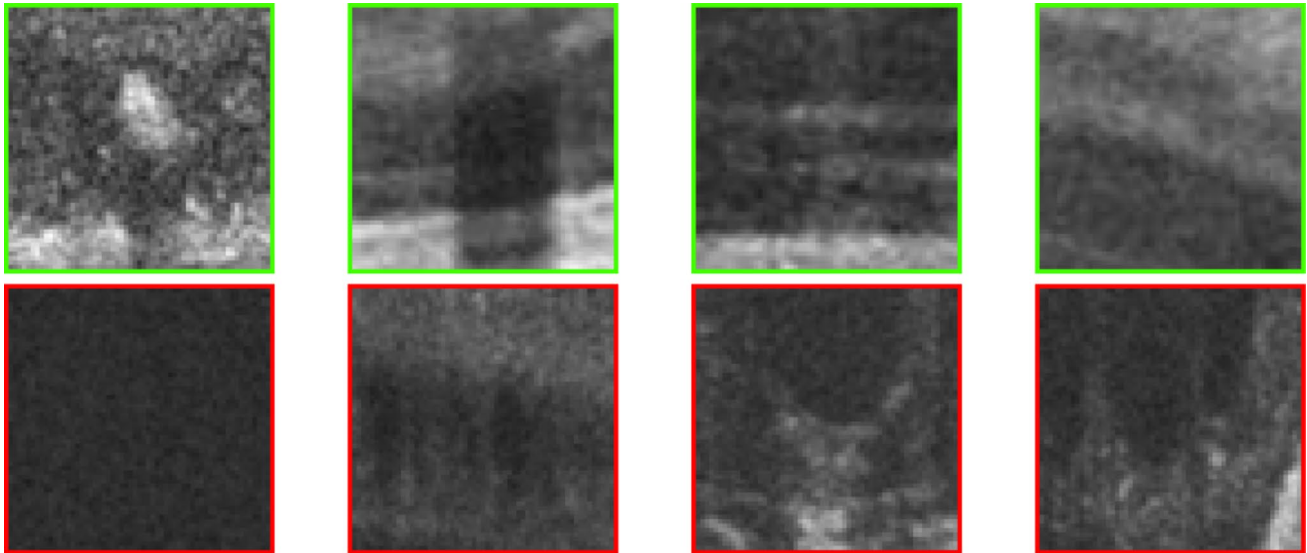


Fig. 12 Examples of regions that were correctly identified. 1st row, non-cystoid regions. 2nd row, cystoid regions

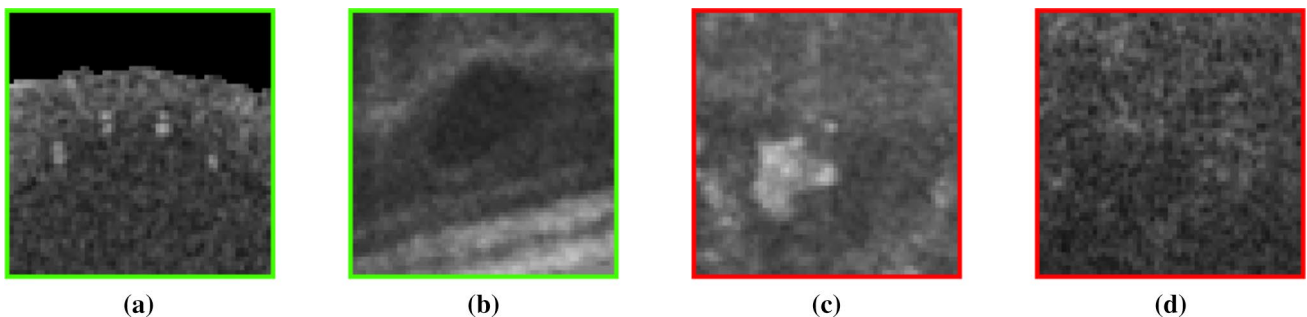


Fig. 13 Examples of regions that were incorrectly identified. **a & b**, Cystoid regions classified as non-cystoid. **c & d**, Non-cystoid regions classified as cystoid

quite characteristic of healthy regions of tissue. However, this mistake is due to the lack of representative samples of this particular case in the training dataset. On the other hand, the results shown in Fig. 13b are the product of the presence of numerous horizontal patterns and a cyst with fairly regular borders. Given that most of the top descriptors examine gradients, an image with low gradients like this one might confuse the classifiers that use almost exclusively this point of view. The next case, Fig. 13c, illustrates a situation in which the sample has just fallen into a particular cyst-like pattern, from the point of view of gradients: an oval-shaped high-contrast artifact. Finally, Fig. 13d presents a texture pattern very similar to that of healthy tissue intermixed with pathological fluid. It is a particular case that even experts in the domain would not be able to differentiate, a borderline case between both categories. Thus, the proposed system acted correctly in benefiting (in case of doubt) the pathological class.

6 Conclusions

One of the main current challenges in medical imaging is the analysis of patterns in structures that, due to their diffuse nature, present a great variability of patterns. This is the case of fluid accumulations in retinal layers. These accumulations, which are the result of diseases considered the main causes of blindness in developed countries, present a wide range of patterns; either due to artifacts produced by the capturing device or complications of the pathology.

These eye diseases require, for proper recovery of the patient, an exhaustive analysis of a large number of OCT images, which is highly reliant on the subjectivity of the clinical expert. Methodologies have been proposed in the state of the art to address this problem, but either they do not offer results that are explainable to experts or the design may result in missing real pathological fluid accumulations.

In this manuscript, we present a complete detailed analysis of what characterizes these pathological intraretinal fluid accumulations. By using a complete library of features, selectors and classifiers, we determine (with very promising metrics in the results) that the minimum amount of information needed for the analysis of these fluid accumulations is a window size of 61×61 pixels. Furthermore, in this context, we considered the optimal number of features for the solution of the problem to be 50, since any addition only serves to decrease significantly the generalization capability at the cost of a negligible gain in classification power. Finally, the most relevant factors for classification of these fluid accumulations were selectors based on gradient orientations (such as Gabor or the Histogram of Oriented Gradients), closely followed by selectors based on purely non-spatial distribution

of gray levels in the samples (such as an analysis of the different quartiles in the image).

With this work, we intend to show not only relevant features for the domain under study, but also a methodological proposal for the analysis in other similar fields, medical imaging modalities and pathologies. Thanks to this analysis we will be able to perform, as future work, more robust systems to the different vicissitudes of capture devices as idiosyncrasies of the studied tissues.

Funding Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project, Ayudas para la formación de profesorado universitario (FPU), grant ref. FPU18/02271; Ministerio de Ciencia e Innovación, Government of Spain through the research project with reference PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva, grant ref. ED431C 2020/24 and through the postdoctoral grant contract ref. ED481B 2021/059; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%). Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding for open access charge: Universidade da Coruña/CISUG.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Al-Kadi O, Watson D (2008) Texture analysis of aggressive and nonaggressive lung tumor CE CT images. *IEEE Trans Biomed Eng* 55(7):1822–1830. <https://doi.org/10.1109/tbme.2008.919735>
2. Baamonde S, de Moura J, Novo J, Ortega M (2017) Automatic detection of epiretinal membrane in OCT images by means of local luminosity patterns. *Adv Comput Intell*. https://doi.org/10.1007/978-3-319-59153-7_20
3. Bogunovic H, Abramoff MD, Zhang L, Sonka M (2014) Prediction of treatment response from retinal OCT in patients with exudative age-related macular degeneration. In: *Proceedings of the Ophthalmic Medical Image Analysis First International Workshop*. University of Iowa. <https://doi.org/10.17077/omia.1018>
4. Breiman L (1998) Arcing classifier (with discussion and a rejoinder by the author). *Ann Stat* 26:3. <https://doi.org/10.1214/aos/1024691079>

5. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/a:1010933404324>
6. Buczkowski S, Kyriacos S, Nekka F, Cartilier L (1998) The modified box-counting method: analysis of some characteristic parameters. *Pattern Recogn* 31(4):411–418. [https://doi.org/10.1016/s0031-3203\(97\)00054-x](https://doi.org/10.1016/s0031-3203(97)00054-x)
7. Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S (2010) Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express* 18(18):19413. <https://doi.org/10.1364/oe.18.019413>
8. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1007/bf00994018>
9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05). IEEE. <https://doi.org/10.1109/cvpr.2005.177>
10. Dasarathy BV (1991) Nearest Neighbor (NN) Norms NN pattern classification techniques. IEEE Computer Society Press
11. Dehnavi A, Esmaeili M, Rabbani H, Hajizadeh F (2016) Three-dimensional segmentation of retinal cysts from spectral-domain optical coherence tomography images by the use of three-dimensional curvelet based k-SVD. *J Med Signals Sensors* 6(3):166. <https://doi.org/10.4103/2228-7477.186878>
12. Elkhatib THM, Hashim NA, Emad EM, Zein H, El-aidy L (2019) Optical coherence tomography and cognitive dysfunction in parkinson disease. *Egypt J Neurol Psychiatry Neurosurg*. <https://doi.org/10.1186/s41983-019-0097-4>
13. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat*. <https://doi.org/10.1214/aos/1013203451>
14. Gabor D (1946) Theory of communication: part 1—the analysis of information. *J Inst Electr Eng Part III Radio Commun Eng* 93(26):429–441. <https://doi.org/10.1049/ji-3-2.1946.0074>
15. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. *Mach Learn* 63(1):3–42. <https://doi.org/10.1007/s10994-006-6226-1>
16. Gonzalez A, Remeseiro B, Ortega M, Penedo MG, Charlon P (2013) Automatic cyst detection in OCT retinal images combining region flooding and texture analysis. In: Proceedings of the 26th IEEE international symposium on computer-based medical systems. IEEE. <https://doi.org/10.1109/cbms.2013.6627825>
17. Haghighat M, Zonouz S, Abdel-Mottaleb M (2015) CloudID: trustworthy cloud-based and cross-enterprise biometric identification. *Expert Syst Appl* 42(21):7905–7916. <https://doi.org/10.1016/j.eswa.2015.06.025>
18. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC* 3(6):610–621. <https://doi.org/10.1109/tsmc.1973.4309314>
19. Huang D, Swanson E, Lin C, Schuman J, Stinson W, Chang W, Hee M, Flotte T, Gregory K, Puliafito C et al (1991) Optical coherence tomography. *Science* 254(5035):1178–1181. <https://doi.org/10.1126/science.1957169>
20. Ikram MK, de Jong FJ, Bos MJ, Vingerling JR, Hofman A, Koudstaal PJ, de Jong PT, Breteler M (2006) Retinal vessel diameters and risk of stroke: the rotterdam study. *Neurology* 66(9):1339–1343. <https://doi.org/10.1212/01.wnl.0000210533.24338.ea>
21. Keane PA, Patel PJ, Liakopoulos S, Heussen FM, Sadda SR, Tufail A (2012) Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol* 57(5):389–414. <https://doi.org/10.1016/j.survophthal.2012.01.006>
22. Kim BY, Smith SD, Kaiser PK (2006) Optical coherence tomographic patterns of diabetic macular edema. *Am J Ophthalmol* 142(3):405–412.e1. <https://doi.org/10.1016/j.ajo.2006.04.023>
23. Kovese P (1999) Image features from phase congruency. *Videre J Comput Vis Res* 1(3):1–26
24. Kovese P (2000) Phase congruency: a low-level image invariant. *Psychol Res* 64(2):136–148. <https://doi.org/10.1007/s004260000024>
25. Kovese P (2003) Phase congruency detects corners and edges. In: The Australian pattern recognition society conference: DICTA, pp. 309–318
26. Lang A, Carass A, Swingle EK, Al-Louzi O, Bhargava P, Saidha S, Ying HS, Calabresi PA, Prince JL (2014) Automatic segmentation of microcystic macular edema in OCT. *Biomed Opt Exp* 6(1):155. <https://doi.org/10.1364/boe.6.000155>
27. Laws KI (1980) Textured image segmentation. University of Southern California Los Angeles Image Processing INST, Tech. rep
28. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H (2018) Feature selection. *ACM Comput Surv* 50(6):1–45. <https://doi.org/10.1145/3136625>
29. Loh HH, Leu JG, Luo R (1988) The analysis of natural textures using run length features. *IEEE Trans Ind Electron* 35(2):323–328. <https://doi.org/10.1109/41.192665>
30. de Moura J, Novo J, Ortega M, Charlón P (2016) 3d retinal vessel tree segmentation and reconstruction with OCT images. In: Lecture Notes in Computer Science, pp. 716–726. Springer International Publishing. https://doi.org/10.1007/978-3-319-41501-7_80
31. Nie F, Xiang S, Jia Y, Zhang C, Yan S (2008) Trace ratio criterion for feature selection. In: D. Fox, C.P. Gomes (eds.), Proceedings of the twenty-third AAAI conference on artificial intelligence, AAAI 2008, Chicago, Illinois, USA, July 13–17, 2008, pp. 671–676. AAAI Press. <http://www.aaai.org/Library/AAAI/2008/aaai08-107.php>
32. Novo J, Penedo MG, Santos J (2008) Optic disc segmentation by means of GA-optimized topological active nets. In: Lecture Notes in Computer Science, pp. 807–816. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-69812-8_80
33. Ojala T, Pietikainen M, Maenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987. <https://doi.org/10.1109/tpami.2002.1017623>
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 2:2825–2830
35. Puzyeyeva O, Lam WC, Flanagan JG, Brent MH, Devenyi RG, Mandelcorn MS, Wong T, Hudson C (2011) High-resolution optical coherence tomography retinal imaging: a case series illustrating potential and limitations. *J Ophthalmol* 2011:1–6. <https://doi.org/10.1155/2011/764183>
36. Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of relief and rrelief. *Mach Learn* 53(1/2):23–69. <https://doi.org/10.1023/a:1025667309714>
37. Roychowdhury S, Koozekanani DD, Radwan S, Parhi KK (2013) Automated localization of cysts in diabetic macular edema using optical coherence tomography images. In: 2013 35th Annual International Conference of the IEEE engineering in medicine and biology society (EMBC). IEEE. <https://doi.org/10.1109/embc.2013.6609778>
38. Sarfraz MS, Hellwich O (2009) On head pose estimation in face recognition. *Commun Comput Inf Sci*. https://doi.org/10.1007/978-3-642-10226-4_13
39. Swingle EK, Lang A, Carass A, Ying HS, Calabresi PA, Prince JL (2014) Microcystic macular edema detection in retina OCT images. In: RC Molthen, JB Weaver (eds.) Medical Imaging 2014: Biomedical applications in molecular, structural, and functional imaging. SPIE. <https://doi.org/10.1117/12.2043910>

40. Trichonas G, Kaiser PK (2014) Optical coherence tomography imaging of macular oedema. *Br J Ophthalmol* 98(Suppl 2):ii24–ii29. <https://doi.org/10.1136/bjophthalmol-2014-305305>
41. Wang J, Zhang M, Pechauer AD, Liu L, Hwang TS, Wilson DJ, Li D, Jia Y (2016) Automated volumetric segmentation of retinal fluid on optical coherence tomography. *Biomed Opt Express* 7(4):1577. <https://doi.org/10.1364/boe.7.001577>
42. Wieclawek W (2015) Automatic cysts detection in optical coherence tomography images. In: 2015 22nd International conference mixed design of integrated circuits & systems (MIXDES). IEEE. <https://doi.org/10.1109/mixdes.2015.7208485>
43. Wilkins GR, Houghton OM, Oldenburg AL (2012) Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans Biomed Eng* 59(4):1109–1114. <https://doi.org/10.1109/tbme.2012.2184759>
44. Wong T, Klein R, Sharrett A, Schmidt M, Pankow J, Couper D, Klein B, Hubbard L, Duncan B, Investigators A (2002) Retinal arteriolar narrowing and risk of diabetes mellitus in middle-aged persons. *J Am Med Assoc* 287:2528–2533
45. Wong TY, Klein R, Sharrett AR, Duncan BB, Couper DJ, Klein BE, Hubbard LD, Nieto FJ (2004) Retinal arteriolar diameter and risk for hypertension. *Ann Intern Med* 140(4):248. <https://doi.org/10.7326/0003-4819-140-4-200402170-00006>
46. Xu X, Lee K, Zhang L, Sonka M, Abramoff MD (2015) Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical OCT data. *IEEE Trans Med Imaging* 34(7):1616–1623. <https://doi.org/10.1109/tmi.2015.2408632>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.