HAL open science

Classification of endomicroscopic images of the lung based on random subwindows and extra-trees

Chesner Désir, Caroline Petitjean, Laurent Heutte, Mathieu Salaün, Luc Thiberville

To cite this version:

Chesner Désir, Caroline Petitjean, Laurent Heutte, Mathieu Salaün, Luc Thiberville. Classification of endomicroscopic images of the lung based on random subwindows and extra-trees. IEEE Transactions on Biomedical Engineering, 2012, Volume 59, (Issue 9), pp. 2677- 2683. 10.1109/TBME.2012.2204747 . hal-00796593

HAL Id: hal-00796593

https://hal.science/hal-00796593v1

Submitted on 5 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification of endomicroscopic images of the lung based on random subwindows and extra-trees

Chesner Désir, Caroline Petitjean, Laurent Heutte, Mathieu Salaün, Luc Thiberville

***Abstract*—Recently, the in-vivo imaging of pulmonary alveoli was made possible thanks to confocal microscopy. For these images, we wish to aid the clinician by developing a computer-aided diagnosis system, able to discriminate between healthy and pathological subjects. The lack of expertise currently available on these images has first led us to choose a generic approach, based on pixel-value description of randomly extracted subwindows and decision tree ensemble for classification (extra-trees). In order to deal with the great complexity of our images, we adapt this method by introducing a texture-based description of the subwindows, based on Local Binary Patterns (LBP). We show through our experimental protocol that this adaptation is a promising way to classify FCFM images. In addition, we introduce a rejection mechanism on the classifier output to prevent non detection errors.**

***Index Terms*—Medical imaging, confocal fluorescence microscopy, image classification, extra-trees, reject rule.**

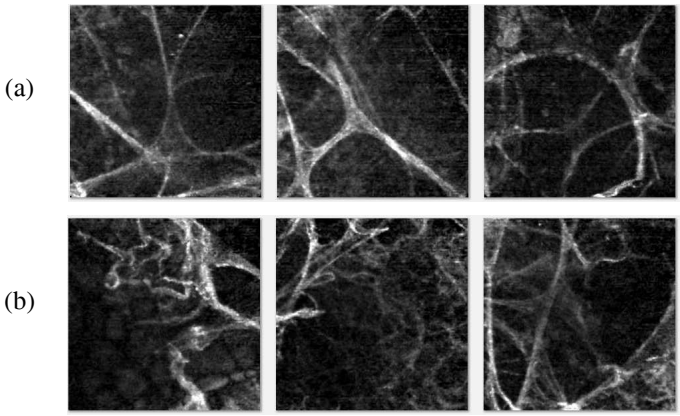

Fig. 1. FCFM images of non-smoking (a) healthy and (b) pathological subjects, showing the alveolar structure. Note how the meshing is tangled and altered for pathological cases (b).

I. Introduction

Whereas the conduction airways can be explored during bronchoscopy, the alveolar region is currently investigated only in vitro, using invasive techniques such as open lung biopsies. A new endoscopic technique, called Fibered Confocal Fluorescence Microscopy[1] (FCFM), has recently been developed that enables the real-time visualisation of the alveoli in vivo [1]. The technique is based on the principle of fluorescence confocal microscopy, where the microscope objective is replaced by a fiberoptic miniprobe, made of thousands of fiber cores. The miniprobe can be introduced into the 2 mm working channel of a flexible bronchoscope to produce in vivo endomicroscopic imaging of the human respiratory tract. This promising technique, whose applications for lung exploration are currently under investigation, could replace lung biopsy in the future and might prove to be helpful in a large variety of diseases, including interstitial lung diseases [2].

The system produces 420x420 pixel images (Fig. 1 and 2), with a circular 600 μm diameter field of view, an approximate spatial resolution of 1μm per pixel and a depth of focus of the probe of 0-50 μm. In smokers, FCFM images are quite different notably because of the presence of macrophages (cells which digest cellular debris inside the alveoli), which are made visible because of tobacco-tar induced fluorescence (they appear as white spots 2). When images are acquired on healthy subjects, the alveolar structure appears as a network of (almost) continuous lines, but it can be altered by distal lung pathologies. As shown in Fig. 1 and 2, FCFM images exhibit some complexity; there is a large intraclass and relatively small interclass variability in the images. The possible confusion originates mostly from the elastin fiber framework which is not always clearly visible and from the noise present in the images.

Interpreting the lung FCFM images represents several levels of difficulties for the clinician, the first being to be able to differentiate normal images from real pathological images. In order to solve this problem, we have recorded in-vivo FCFM images from non-smoking and smoking healthy volunteers, that are considered by definition as “normal”. We also recorded images from patients suffering from several diagnosed diffuse interstitial - or parenchymal - lung diseases, obtained from a lung segment that appeared abnormal on the chest CT Scan. Alveolar images from this latter group have been selected by the clinicians for being visually different from the normal images and considered abnormal. Because normality criteria based on the confocal image by itself are currently not known, we chose an objective definition for normality as “any alveolar confocal image taken from an healthy subject”. Our aim is to provide the clinician with a computer aided-diagnosis (CAD) tool, so as to help him to analyze these new images. For instance, such a strategy may help to the automated recognition of the area to biopsy during endoscopy.

We describe in this paper such a CAD tool, able to classify FCFM images as healthy or pathological. A state-of-the-art efficient classification technique called extra-trees [3], which

C. Désir, C. Petitjean, L. Heutte, M. Salaün and L. Thiberville are with the LITIS EA 4108, Université de Rouen, 76801 Saint-Etienne-du-Rouvray, France. M. Salaün and L. Thiberville are also with the Clinique Pneumologique, Rouen University Hospital, 76031 Rouen, France.
e-mail: caroline.petitjean@univ-rouen.fr

[1]FCFM is also refered to as probe-based Confocal Laser Endomicroscopy (pCLE), especially in gastrointestinal imaging.

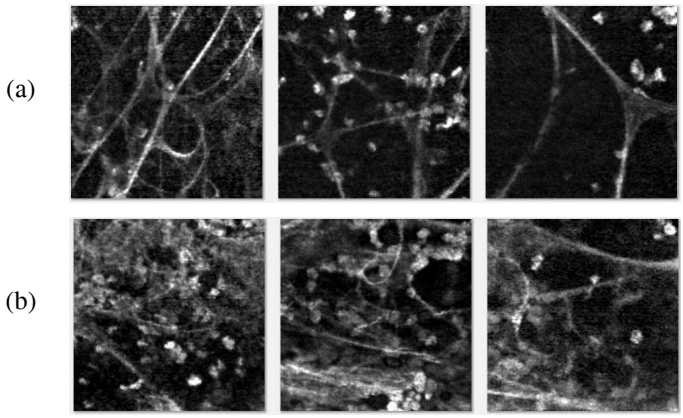

Fig. 2. FCFM images of smoking (a) healthy and (b) pathological subjects, showing the alveolar structure. White spots are macrophages, visible here due to tobacco-tar induced fluorescence. Note how the meshing is tangled and altered for pathological cases (b).

has proved its efficiency on various image classification tasks [4], is implemented. This classification method is combined with a feature extraction technique based on the random extraction of subwindows from the image . The extraction of subwindows artificially increases the training sample size and allows to take into account more local characteristics. In order to capture the complex nature of FCFM images, we introduce high-level descriptors to describe the subwindows. The aim is to benefit from the association of the informative nature of texture description and the high precision and computational efficiency of the extra-trees.

A brief visual analysis of FCFM images reveals that in pathological cases, the disorganization of the alveolar meshing is characterized by the numerous irregularities and the tangle of the fibered structures, in all possible directions. The distributions of local texture patterns in healthy and pathological images are thus expected to be different and could be adequately described by rotation and scale invariant texture descriptors, such as the Local Binary Patterns (LBP) operator. Their ability to characterize the spatial configuration of local image texture and to recognize certain local binary texture patterns will help to apprehend the complex, textured nature of FCFM images. This new approach to characterize the images, that combines texture-based description with the random extraction of subwindows, is the first part of our contribution. It is shown to have encouraging results on FCFM images. The second point of our contribution concerns the introduction of a rejection mechanism on the classifier output that allows to improve the practical use of the system. We indeed show how such a mechanism can help reducing the non detection rate of the system, thus increasing its reliability.

The remainder of the paper is organized as follows: we review in Section II related works in the field of image classification and introduce our approach. The main steps of the FCFM image classification system are described in Section III. We present in Section IV the experimental protocol and the results. Finally we conclude in Section V by drawing some future works on this challenging task of classification.

II. Related works in image classification

A classification system usually includes a feature extraction step and a classification step.

A. Feature extraction

The aim of the feature extraction step is to feed the classifier with relevant features that must be chosen to maximize interclass variance and minimize intraclass variance. The choice of a suitable feature space is often problem-dependent and features are therefore usually adapted to the image content. For instance, blood cell images can be described by frequency analysis, using Fourier-Mellin transform [5]; protein subcellular images can be described using Haralick statistical parameters on cooccurrence matrices, Zernike moment features and features derived from morphological image processing [6]. An alternative approach in feature extraction consists in using unspecific features such as raw pixel values or generic texture descriptors, such as the LBP operator [7] for example, which is particularly suitable for image description and has been widely used for different image classification tasks [8]. In order to locally characterize the image, subwindows can also be extracted and described with a feature vector. The sampling of subwindows may be either deterministic or random, this latter solution having proved its efficiency for image characterization in various image classification tasks, when a lot of subwindows are sampled [9]. Random selection also allows to bypass costly steps such as keypoint detection or exhaustive window sliding on all pixels. Such a sampling method is very efficient computationally and remains generic since it does not require to design a dedicated keypoint detection for a specific application.

B. Classification

In the past decades, extensive work has been done in the machine learning and pattern recognition communities to design efficient classifiers. Recently, ensemble methods such as bagging [10] or boosting [11] have been of main interest especially in medical imaging [12]. Among these methods, ensembles of decision trees, such as random forests [13], have proved their efficiency over single classifiers on various classification tasks [14].

Random Forests (RF) are a family of classifier ensemble methods that use randomization to produce a diverse pool of individual classifiers. It can be defined as a generic principle of classifier combination that uses L tree-structured base classifiers $\{h(x, \Theta_i), i = 1, ...L\}$ where h is a tree-based classifier, Θ_i a family of i.i.d random vectors, and x an input feature vector. The particularity of this kind of ensemble is that each decision tree is built from a random vector of parameters. RF have been shown to be particularly competitive with SVM and boosting, one of the most efficient learning principles [13], [11]. Other randomization mechanisms may be exploited in RF: bagging (bootstrap-aggregating) for training each individual tree in the ensemble on a different replica of the training set (a bootstrap), obtained by subsampling the training set at random with replacement; and random feature

selection, that selects a random subspace at each node of the tree. The Extremely Randomized Decision Trees, or extra-trees [3], are based on another randomization process. At each node of an extra-tree, splitting rules are randomly drawn and the rule that performs best according to a score computation is then chosen to be associated to that node. This allows to drastically reduce the complexity of the induction process, increase the speed of training, and weaken the correlation between the induced decision trees. The whole available learning set is provided to each individual decision tree, i.e. there is no bagging.

C. Proposed approach

We have chosen the extra-trees due to their interesting statistical and computational properties. Considering the bias-variance tradeoff, extra-trees are expected to better reduce the variance than classical randomized methods such as the Forest-RI approach [13] as the randomization process is stronger and to minimize the bias as it uses the whole available learning set for the induction of each tree, contrary to classical approaches such as bagging [13]. The computational load of the algorithm is also reduced due to the simplicity of the node splitting procedure as it does not search for a local optimal cut-point.

The extra-trees approach, in combination with an original feature extraction technique, has proved its efficiency on various image classification tasks and compares well with state of the art classification systems [4]. In these works, the authors propose to extract a huge number of subwindows from the images of the learning set at random locations and with random sizes, and to label each of them with the class label of the original image. This new learning set, i.e. the set of subwindows, is used to train the extra-trees. The classification of a test image follows the same principle: subwindows are extracted from the image and a majority vote on the decisions assigned to the subwindows is used to decide for the class of the image. We propose to use this generic framework to classify FCFM images with fast extra-trees. The approach in [4] used raw pixel description of the subwindows. In our study, we first test this low-level based description for FCFM images but we suggest that a higher level descriptor, namely LBP, can be efficiently used to describe subwindows without loss of genericity. Additionally, we show that the classification of FCFM images can be made more reliable by taking benefit from the posterior probability assigned to an image. Indeed, by introducing a threshold on posterior probabilities, we give our system the ability to detect (i.e. reject) images that are likely to be false negative and thus to classify FCFM images with no error.

III. Our classification approach

Our classification system is composed of three main steps (Fig. 3):

1) random extraction of N_w subwindows of different sizes and locations from the image;
2) characterization of each subwindow by means of a feature vector that is classified using extra-trees;
3) combination of probabilities output by extra-trees on each extracted subwindow through a majority vote to either assign a class to the input image, or to reject it.

A. Feature extraction in FCFM images

In the original work presented in [4], the authors have proposed a low-level description of the subwindows, claiming that it makes their approach more generic. Indeed, the subwindows extracted from the images are chosen at random locations with random sizes; then they are resized with bilinear interpolation to the empirical resolution 16x16 pixels, in order to obtain a fixed size feature vector. Downsampling to 16x16 implies a loss of details, which is thought to be compensated for with a large number ($N_w = 100$) of subwindows extracted in the image. This genericity has been of particular interest for our application since no a priori knowledge was available on FCFM images. However, we suggest to go further in the description of the subwindows by investigating the use of a classical texture descriptor, namely the LBP operator, in order to feed the system with more relevant features. The LBP [7] consists of computing the distribution of binary patterns. These patterns are obtained by thresholding the neighbouring pixels compared to the central pixel, for each pixel of the subwindow, and coding them with powers of 2. The circular neighbourhood can be characterized with a number P of pixels and a radius R. Each pixel c is assigned the following number:

$$b(c) = \sum_{p=0}^{P-1} s(g_p - g_c).2^p$$

where g_c is the gray value of the center pixel, g_p defines the gray value of neighbor p and $s(x)$ is the thresholding function defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The final characterization is the distribution of b values over the whole subwindow. Combining the responses of LBP operators computed with different values for P and R can also be of interest. This multiresolution approach has been chosen for our study, using standard values ($P = 8$, $R = 1$) and ($P = 16$, $R = 2$). Since the LBP operator yields $P + 2$ values [7], each subwindow is characterized by a 28-dimensional feature vector.

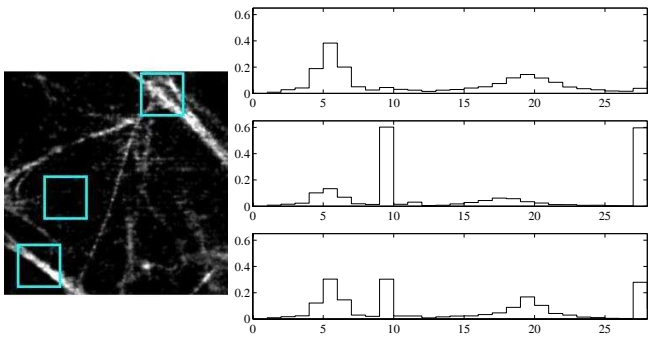

Fig. 4. Subwindows extracted from an FCFM image and their associated distributions of LBP patterns.

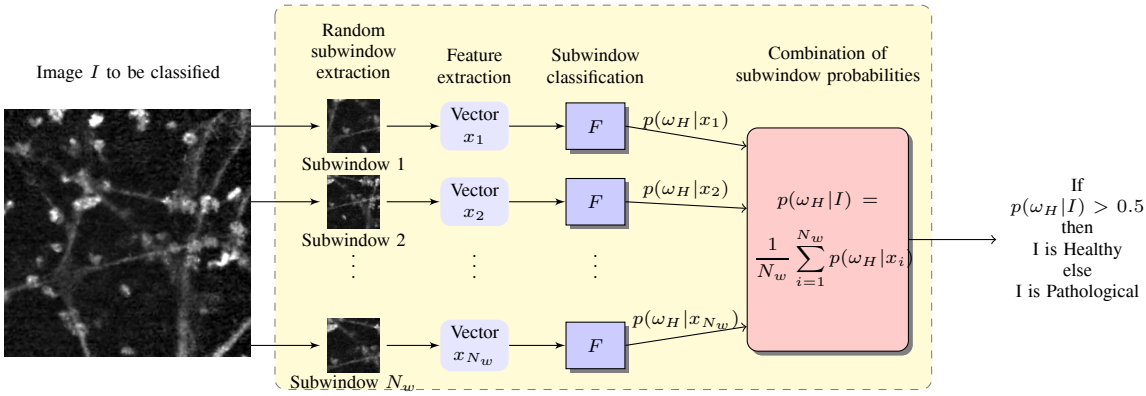

Fig. 3. Our classification approach based on random subwindows extraction and extra-trees. N_w is the number of extracted subwindows from image I, F the forest of L extra-trees and ω_H the Healthy class.

B. Classification based on extra-trees

In the classification step we use an ensemble of Extremely Randomized Trees [3] induced on the generated learning set of subwindows. Each tree is induced by performing at each node the following steps: (i) randomly pick K different features from the whole feature vector, (ii) compute one random split per feature and (iii) select the one that maximizes a score measure based on a normalized version of the Shannon information gain [15]. K acts as a parameter that enables to control the degree of randomization for inducting the tree: choosing a high value for K offers the opportunity to filter irrelevant variables and thus makes the tree classification process more reliable while introducing more correlation between trees in the forest; on the contrary, a low value for K will provide more randomized trees and thus decrease the correlation within the forest. Yet the classification process will take benefit from the ensemble of trees.

Performance of extra-trees also depend on the adjustement of two parameters, L, the number of trees in the ensemble, and n_{min}, the minimum sample size for splitting a node. n_{min} controls the strength of averaging output noise; L, the number of trees in the ensemble, determines the strength of the variance reduction of the forest. As usual, these parameters can be adapted to the problem specificities either manually or in an automatic way (e.g. by cross-validation).

However, it has been shown in [3] that the default values of the two main parameters, i.e. K and n_{min}, are near-optimal for a broad range of typical classification problems. The choice of these default values is discussed in section IV-B.

C. Combination of output probabilities and proposed rejection mechanism

The probabilities output by extra-trees on each extracted subwindow are combined through a majority vote to assign, to the input image to classify, a class, "healthy" (denoted ω_H) or "pathological" (denoted ω_P): let x_i be the feature vector describing a subwindow, l_i^H the number of trees assigning the class ω_H to x_i; the probability of the subwindow to be assigned the "healthy" class is:

$$p(\omega_H|x_i) = \frac{l_i^H}{L}$$

where L is the number of trees in the forest. The posterior probability of ω_H assigned to the image I is then obtained by averaging over all N_w subwindows:

$$p(\omega_H|I) = \frac{1}{N_w}\sum_{i=1}^{N_w} p(\omega_H|x_i)$$

where N_w is the number of subwindows extracted from the image. The same reasoning applies to the computation of probability of the pathological class.

The fraction of trees that voted for a particular class can be used to handle the reliability of the extra-trees decision. Indeed, reliability of automatic pattern recognition systems is a critical issue in many practical applications. A rejection threshold is commonly used to reach a trade-off between error and reject probabilities. From the practical point of view, it offers the possibility of refusing to assign the examined pattern to any class, possibly prompting for a further investigation by another system or by a human supervisor [16]. In our medical context, providing our FCFM image classification system with a reject option would enable to filter the doubtful (low confidence) images that could further be handled either by the clinician who will make his/her own diagnosis or by an additional, more accurate and usually more costly, classification system. The combination of probabilities output by the extra-trees offers a means to introduce a rejection mechanism.

Indeed, the probabilities of each class (healthy or pathological) enable to measure the level of confidence associated to the classification. For instance, a test image that has been classified by the extra-trees with a probability of 0.9 means that a great number of subwindows have been assigned a high probability of being labeled healthy or pathological: such a high level of agreement between subwindows, knowing their random and independent process of extraction, suggests that the class has been assigned to the image with a high level of confidence. On the contrary, a probability near 0.5 should be considered as a low level of confidence decision and suggests that the decision associated to the image should be rejected in order to avoid a non detection or false alarm error. Consequently, thresholding the probability assigned to an image should enable to minimize the non detection error rate of the system.

Note that the proposed rejection mechanism could also be controlled another way. For example, instead of determining the posterior probability of the image by averaging the probabilities assigned to the extracted subwindows, one could filter first subwindows whose predictions are "uncertain", i.e. remove from the final vote subwindows whose probability is under a given threshold, in order to keep only strong decisions, which boils down to voting on highly confident subwindows [17].

IV. Experiments

A. Image databases and subwindow extraction

In-vivo FCFM images from non-smoking and smoking healthy volunteers, that are considered by definition as "normal", have been recorded. We also recorded images from patients suffering from several diagnosed diffuse intersitial - or parenchymal - lung diseases, obtained from a lung segment that appeared abnormal on the chest CT Scan. Due to the great difference between images acquired on smoker and non-smoker subjects, our experiments have been conducted separately on those two groups. The non-smoker (NS) image database and the smoker (S) image database are composed of respectively 133 and 93 images, assigned to two different classes: "healthy" (31 for NS and 60 for S) and "pathological" (102 for NS and 33 for S). In order to assess the generalization error of the system, a 10-fold cross validation process is used. The principle is to partition the complete image dataset into 10 complementary subsets. The training is performed on 9 subsets and testing is done on the remaining subset to assess classification performance. To reduce variability, 10 rounds of cross-validation are performed using the 10 different partitions, and the classification results are averaged over the rounds. In order to deal with the imbalanced datasets, an identical number of subwindows (namely 80000, i.e. 40000 per class) is extracted from the images present in the 9 folds, during the learning phase [4]. The 80000 subwindow extraction process is thus renewed for each different training partition. During the test phase, $N_w = 100$ subwindows are randomly extracted from each test image of the remaining set.

B. Extra-trees parameterization

As mentioned previously, performance of extra-trees depend on the adjustement of three parameters, namely K, n_{min} and L. The tree depth is implicitly fixed by n_{min}, which is the minimum number of samples required for splitting a node: the smaller the value of n_{min}, the deeper the tree. In our experiments, we have chosen $n_{min} = 2$ (as for normal tree classifier) [3]. Regarding parameter L (number of trees in the forest), it is well-known that for randomized ensembles the classification error is a monotically decreasing function of L: the higher the value of L, the better the accuracy. Nevertheless, choosing an appropriate value for L is not so critical: given the computational efficiency of the tree growing algorithm, a high value for L may be chosen. In this study, an ensemble of 30 trees has been found to be large enough and that convergence is reached well before $L = 30$ (typically between 15 and 20 trees could be enough). Parameter K, the number of variables randomly selected at each node during the tree growing, determines the strength of the variable selection process and is usually set to $\sqrt{M}$ where M is the dimension of the feature vector [13]. This default setting is the best compromise for a wide range of classification problems [18], [19], [3]. Let us recall that $M = 256$ for the raw pixel feature space since each subwindow has been resized to 16x16 pixels whereas $M = 28$ for the LBP operator. Note that the resizing operation of the subwindow is performed only for the raw pixel space.

TABLE I
Classification rates (mean ± stdev) for non-smoker and smoker groups

	Non-smoker	Smoker
Random subw. + raw pixel + ET [4]	$80.0 \pm 9.2\%$	$73.3 \pm 13.3\%$
Random subw. + LBP + ET (our method)	$92.3 \pm 8.8\%$	$96.6 \pm 5.4\%$

C. Experimental results

Raw pixel vs. LBP characterization: Results obtained on the FCFM image database for the previously defined feature spaces (raw pixel values and LBP) are compared. Mean classification rates and standard deviations obtained over the 10 folds are provided in Table I. For the FCFM image database, we can observe in Table I that the classification rates obtained with raw pixel values are rather low, underlining the difficulty for the classifier to cope with such a low level description of FCFM images. Though it has been shown that this classification approach is efficient [4], describing the subwindows with raw pixel values does not seem to be adequate for our application. The image content requires a higher level of description as confirmed by the higher classification rates obtained with LBP descriptors. LBP descriptors perform better on smoker images than on non-smoker images, whereas raw pixels perform worse, as can be noticed also on ROC curves[2] (Fig. 5). This phenomenon could be related to the presence of several additional corners in smoker images, because of the macrophages: they are clearly visible on healthy images and somehow hidden when the alveolar structure is disorganized because of a pathology.

Random-sized subwindows vs. constrained and small-sized subwindows: One advantage of extracting random subwindows is that no additional hyperparameter value (such as subwindow size) has to be set. The other advantage is that random subwindows are able to catch both large and small patterns in the image and that their combination brings additional information to discriminate between classes. To gain more insight on this issue, we have investigated the effect of constraining vs. not constraining the size of subwindows. Yet it is well-known that small local patches can give good results for texture classification tasks [22]. We have thus compared the extraction of constrained and small-sized subwindows vs.

[2]A single ROC curve can be computed with a 10-fold protocol by merging all test folds into one large single test set, and sorting the instances by their assigned score [20]. A ROC curve can then be plotted using a generation algorithm such as [21].

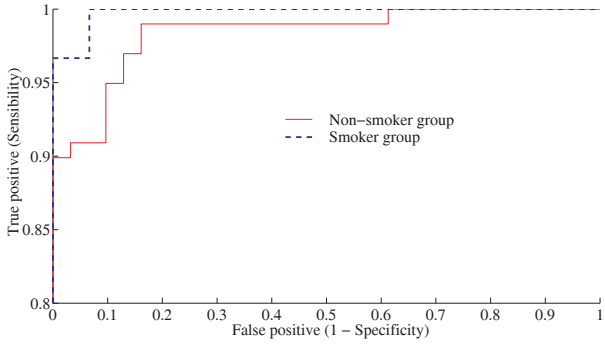

Fig. 5. ROC curves with random subwindows + LBP + ET

TABLE II
COMPARISON IN CLASSIFICATION RATES (MEAN ± STDEV) BETWEEN SMALL SUBWINDOWS AND RANDOM SUBWINDOWS EXTRACTION.

	Non-smoker	Smoker
4x4 subwindows	76.1±12.4%	90.0±9.7%
8x8 subwindows	78.4±13.4%	92.2±9.1%
16x16 subwindows	82.3±10.2%	94.4±5.8%
Random subwindows	92.3 ± 8.8%	96.6 ± 5.4%

unconstrained, random-sized subwindows, using the same protocol as described in Section IV-A, with an LBP description of each subwindow. Table II shows that the extraction of random subwindows provides the best results, compared to small-size subwindows. This can be explained by the fact that the fibered structure of FCFM images is randomly tangled and differs from textured images, that exhibit periodical patterns. The periodicity property is important to explain why small patches contain sufficient information for successful discrimination [22]. Unconstrained random-sized subwindows thus seem to be the best solution for FCFM images.

Reduction of non detection with a rejection mechanism: For assessing the ability of our method to minimize the non detection rate, we show in Fig. 6 the distribution of the confidence values associated to the decision of the forest, i.e. when the class with highest posterior probability is retained (no rejection rule is applied). These distributions are drawn for the two groups, smoker and non smoker, on the test images over the 10 folds.

As shown in Fig. 6, the posterior probabilities assigned by the classifier to the correctly classified (either pathological or healthy) test images vary from 0.51 up to 0.98 for the non smoker group and from 0.53 up to 0.99 for the smoker group (red circles). More interestingly, without rejection rule, the posterior probabilities assigned by the classifier to the incorrectly classified test images (either pathological or healthy) range from 0.53 up to 0.89 for the non smoker group and from 0.52 up to 0.61 for the smoker group (blue crosses). This means that if a reject threshold on posterior probabilities (e.g. 0.89 for the non smoker group, 0.61 for the smoker group) is properly adjusted, accepting only decisions above this threshold will make the decision process more reliable, as errors will be eliminated. False alarm and non detection rates will then be equal to zero. The rejection threshold will thus be useful to filter doubtful images that could further be handled either by the clinician who will therefore focus only on these images, or by an additional but more costly classification system. This additional handling of doubtful images would allow to make a decision on, in particular, rejected pathological images that would not have been detected otherwise. An alternative to an automatic rejection process of low confidence images could also consist in providing the clinician with the system decision, i.e. the probability associated to the healthy or pathological class, along with the rejection threshold, so as to warn him of a possible error risk.

From Table III, one can see that the application of such a rejection rule with our approach (RS+ET) on the non smoker group turns into rejecting 36.1% of the test images (i.e. 48/133) leading to a correct classification rate of 63.9% (i.e. 85/133) with no error. Of course, this null error rate can only be achieved at the expense of rejecting some test images that would have been correctly classified otherwise (i.e. 38/48) but this is the price worth paying for eliminating non detection errors. Similarly, for the smoker group, applying such a rejection rule on the posterior probabilities enables to classify with no error 94.6% of the test images (i.e. 88/93) while rejecting 5.4% (i.e. 5/93). A null detection rate for the smoker group can thus be achieved at the expense of rejecting 2 correctly classified test images. All in all, introducing a rejection mechanism enables to correctly classify 76.6% of the test images of the two groups with no error (neither false alarms nor non detections) whereas accepting, without reject, all decisions of the automatic system would lead to a 5.9% error rate.

Random subwindows vs. global image descriptor: In order to assess the interest of introducing a random subwindow description of the image, we compared our random subwindow based approach (RS+ET) to the image-based approach (I+ET) that relies on an LBP-based description of the image, without random subwindow extraction. Results with the I+ET approach have been obtained using a leave-one-out protocol and are reported in Table III. As classification rates of RS+ET and I+ET are close, it is not possible to conclude on the superiority of the RS+ET approach. Nonetheless, the RS-based approach is more efficient for rejection: less images have to be rejected to reach a null error rate, compared to the I+ET approach. This advantage of the RS-based approach might originate from the redundancy (by extracting numerous overlapping subwindows) and diversity (by means of random locations and sizes) induced by the subwindow extraction process. These properties lead to better estimated posterior probabilities, that enable to better control the rejection of doubtful images.

It is worth noticing that the generalization of the rejection threshold on other datasets of FCFM images holds of course if and only if enough training data are used to estimate an accurate threshold and if the data, on which the rejection threshold will be applied in practical use, share the same statistical properties as the training data. At present, we do not have at our disposal enough data to make the evaluation of the rejection threshold more robust but we have shown that our method offers a means to go one step further in making the system more reliable by minimizing non detection errors. Gathering more labeled images is one of our future works.

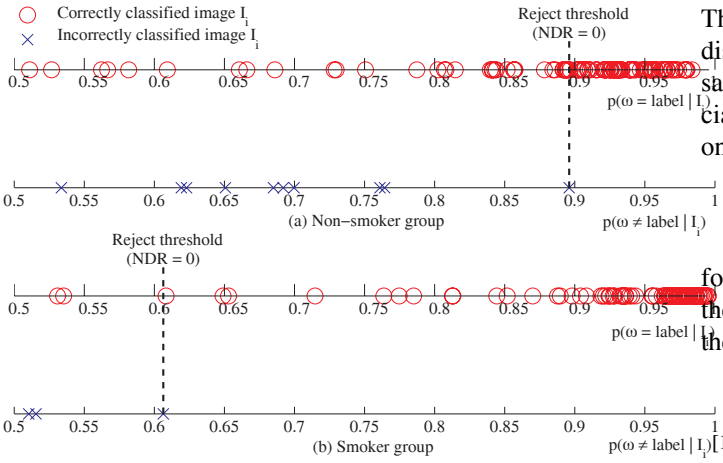

Fig. 6. Distribution of the probabilities associated to correctly and incorrectly classified images. "label" denotes the ground truth associated to each image.

TABLE III
MEAN CLASSIFICATION RATES OF THE RANDOM SUBWINDOW BASED APPROACH (RS+ET) VS. THE IMAGE-BASED APPROACH (I+ET)

			Correct	Error	Reject
NS	RS+ET	without reject	92.3%	7.7%	0%
		with reject	63.9%	0%	36.1%
	I+ET	without reject	92.1%	7.9%	0%
		with reject	0%	0%	100%
S	RS+ET	without reject	96.6%	3.4%	0%
		with reject	94.6%	0%	5.4%
	I+ET	without reject	95.7%	4.3%	0%
		with reject	88.8%	0%	11.2%

V. CONCLUSION

In this paper, we have presented a classification system, based on the analysis of FCFM images representing the alveolar structure in vivo. This system allows to distinguish healthy cases from pathological ones for computer-aided diagnosis. The proposed classification approach is based on a local characterization of the images and an extra-tree classification system. One major advantage of the approach is that no critical parameter has to be tuned as default settings for extra-trees and unconstrained sizes for extracted subwindows are used, thus maximising the genericity and autonomy of the method. The methodological contribution relies on (i) the extraction of higher level descriptors on random subwindows to describe FCFM images and (ii) a rejection mechanism based on the thresholding of the probabilities output by the classifier, in order to equate non detection to zero. We showed that raw-pixel values description does not allow to account for the complexity of the images, and that the introduction of a more relevant texture descriptor, namely the LBP, allowed to achieve higher classification rates. A study on the level of confidence associated to the decision of the forest has also been conducted that showed that our system is rather accurate for classifying FCFM images and highly reliable as it is possible to handle a reject rule based on probability thresholding to reach a null non-detection rate.

Future works include the validation of these results on larger databases of FCFM images, in order to make the rejection process more robust for practical computer-aided diagnosis. The development of a multi-class classification system able to discriminate different kinds of pathologies (fibrosis, sclerosis, sarcoidosis, abestosis) would also be of interest for the clinician. In terms of applications, we plan to evaluate our method on bigger and different databases of images.

ACKNOWLEDGEMENTS

The authors would like to thank the *Ligue Contre le Cancer* for supporting Chesner Désir. They also would like to thank the anonymous reviewers whose comments helped to improve the manuscript.

REFERENCES

[1] L. Thiberville, S. Moreno-Swirc, T. Vercauteren, E. Peltier, C. Cave, and G. Bourg-Heckly, "In vivo imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy," *Am. J. Respir. Crit. Care Med*, vol. 175, pp. 178–187, 2007.
[2] L. Thiberville, M. Salaün, S. Lachkar, S. Dominique, S. Moreno-Swirc, C. Vever-Bizet, and G. Bourg-Heckly, "Human in vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy," *Eur Respir J.*, vol. 33, no. 5, pp. 974–85, 2009.
[3] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 36(1), pp. 3–42, 2006.
[4] R. Marée, P. Geurts, J. Piater, and L. Wehenkel, "Biomedical image classification with random subwindows and decision trees," in *ICCV Workshop on Computer Vision for Biomedical Image Applications*, ser. LNCS, vol. 3765, 2005, pp. 220–229.
[5] J. Dahmen, J. Hektor, R. Perrey, and H. Ney, "Automatic classification of red blood cells using gaussian mixture densities," in *Bildverarbeitung für die Medizin*, 2000, pp. 331–335.
[6] R. Murphy, M. Velliste, and G. Porreca, "Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images," *The Journal of VLSI Signal Processing*, vol. 35, no. 3, pp. 311–321, 2003.
[7] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," *ECCV*, pp. 404–420, 2000.
[8] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, 2009.
[9] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *European Conference on Computer Vision*, ser. LNCS, vol. 3954, 2006, pp. 490–503.
[10] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 26(2), pp. 123–140, 1996.
[11] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," *Int Conf on Machine Learning*, pp. 148–156, 1996.
[12] Z. Zhou, Y. Jiang, Y. Yang, and S. Chen, "Lung cancer cell identification based on artificial neural network ensembles," *Artificial Intelligence in Medicine*, vol. 24, no. 1, pp. 25–36, 2002.
[13] L. Breiman, "Random forests," *Mach Learn*, vol. 45, pp. 5–32, 2001.
[14] T. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, ser. LNCS, vol. 1857, 2000, pp. 1–15.
[15] L. Wehenkel, *Automatic learning techniques in power systems*. Kluwer Academic Publishers, 1998.
[16] P. Foggia, G. Percannella, C. Sansone, and M. Vento, "On rejecting unreliably classified patterns," in *Multiple Classifier Systems*, ser. LNCS, vol. 4472, 2007, pp. 282–291.
[17] R. Marée, P. Geurts, J. Piater, and L. Wehenkel, "A generic approach for image classification based on decision tree ensembles and local subwindows," in *ACCV*, vol. 2, 2004, pp. 860–865.
[18] S. Bernard, L. Heutte, and S. Adam, "Influence of hyperparameters on random forest accuracy," *Multiple Classifier Systems*, pp. 171–180, 2009.
[19] ——, "Forest-RK: A new random forest induction method," in *ICIC, Shanghai, China*, ser. LNCS, vol. 5227, 2008, pp. 430–437.
[20] P. Flach, "Tutorial on the many faces of ROC analysis in machine learning," in *Int Conf on Machine Learning*, 2004.
[21] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
[22] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.