A probabilistic model of emphysema based on granulometry analysis

J.V. Marcos[a], R. Nava[b], G.Cristobal[a], A. Munoz-Barrutia[c], B. Escalante-Ramírez[d] and C. Ortiz de Solórzano[c]

[a]Instituto de Óptica, Spanish National Research Council (CSIC), Serrano 121, Madrid 28006, Spain

[b]Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, Mexico City, Mexico

[c]Cancer Imaging Laboratory, Center for Applied Medical Research, University of Navarra, Pamplona, Spain

[d]Departamento de Procesamiento de Señales, Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico

ABSTRACT

Emphysema is associated with the destruction of lung parenchyma, resulting in abnormal enlargement of airspaces. Accurate quantification of emphysema is required for a better understanding of the disease as well as for the assessment of drugs and treatments. In the present study, a novel method for emphysema characterization from histological lung images is proposed. Elastase-induced mice were used to simulate the effect of emphysema on the lungs. A database composed of 50 normal and 50 emphysematous lung patches of size 512 x 512 pixels was used in our experiments. The purpose is to automatically identify those patches containing emphysematous tissue. The proposed approach is based on the use of granulometry analysis, which provides the pattern spectrum describing the distribution of airspaces in the lung region under evaluation. The profile of the spectrum was summarized by a set of statistical features. A logistic regression model was then used to estimate the probability for a patch to be emphysematous from this feature set. An accuracy of 87% was achieved by our method in the classification between normal and emphysematous samples. This result shows the utility of our granulometry-based method to quantify the lesions due to emphysema.

Keywords: Emphysema, histology, granulometry, logistic regression

1. INTRODUCTION

Emphysema is defined as the abnormal permanent enlargement of the airspaces distal to the terminal bronchioles, accompanied by destruction of their walls and without obvious fibrosis.[1] Human emphysema was originally described by Ruysch in Amsterdam by the end of the 17th century and in the 19th century by the French physician Laennec, who noted marked variations in the size of the air vesicles.[2] Usually, emphysema manifests as a component of chronic obstructive pulmonary disease (COPD) in smokers. However, emphysematous lung destruction has also been reported in other non-smoking-related disorders such as HIV infection or hypersensitivity pneumonitis.[2] The clinical syndrome of COPD includes airflow obstruction, small airway inflammation and lung parenchyma (alveolar) destruction. In addition, extrapulmonary manifestations such as muscle wasting, osteoporosis and anemia are related to this disease.[2] Mortality and morbidity from COPD is an increasingly serious global health problem. It is worth noting that COPD ranked sixth among the causes of death globally in 1990 and it is expected to be the third most common cause of death in 2020.[3] Therefore, in order to prevent other health complications, accurate characterization of emphysema is required for the development of efficient treatment options.

Corresponding author: J.V. Marcos
E-mail: jvmarcos@gmail.com, Telephone: (34) 91-5616800 (x942319), Fax: (34) 91-5645557

IX International Seminar on Medical Information Processing and Analysis, edited by Jorge Brieva, Boris Escalante-Ramírez, Proc. of SPIE Vol. 8922, 892211 · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2035478

A method for quantifying emphysema is required to evaluate the stage of the disease as well as to assess the benefit derived from experimental treatments. Animal models such as elastase-induced emphysema mice have been previously used for the assessment of therapeutic approaches based on histological *ex vivo* analysis of the lungs. In this context, suitable metrics for the evaluation of the degree of emphysematous lesions from histological samples are required. Several descriptors have been previously defined for this purpose. The mean linear intercept L_m is the simplest technique for emphysema quantification.[4] It is given by the mean length of air segments. To compute L_m, a finite set containing samples of these segments is extracted from the image. For this purpose, a grid is placed on the patch depicting the lung tissue. The intercept between the grid lines and the alveolar tissue walls are detected. An air segment is then identified as the portion of the grid line between two consecutive intercepts. The variable L_m has been the metric of reference for emphysema analysis during the last decades. However, Parameswaran *et al.*[5] showed that it presents two main drawbacks. First, L_m depends on the shape of the airspaces. Hence, even in the case similarly sized airspaces are found in different lung tissue images, the value of L_m may vary from one to another because of their shape. In addition, L_m has shown to be unable to detect emphysematous regions characterized by a heterogeneous distribution of the airspaces. Regions with a single large airspace surrounded by smaller ones will result in a small value of L_m, which could be wrongly interpreted as the absence of emphysema. To overcome these limitations, Parameswaran *et al.*[5] developed a set of descriptors derived from a diameter variable (d), which is obtained by approximating the original airspace by a circle of equal area. The radius of this circle determines the value of the equivalent diameter of the airspace. As a result, the dependence on the shape of the airspace is avoided. The moments of the variable d are proposed as a measure of the degree of emphysema. Recently, its second moment (D_2) has shown to be a useful measure to evaluate the utility of microcomputed tomography (micro-CT) for the quantification of lung damage.[6]

In our study, we propose a novel method for emphysema characterization using granulometry analysis. Granulometry is based on the use of area morphology operators.[7] Unlike standard morphology, area operators do not impose any shape restriction determined by a specific structuring element. Instead, any connected component with an area smaller than that used as threshold will be identified. Therefore, the boundaries of an object in the image are not distorted, resulting in better classification.[8] A pattern spectrum is obtained by computing the difference between successive opening versions of the original image.[9] Subsequently, a set of descriptors or features is extracted from this spectrum in order to perform image classification.[8,10] Previously, image analysis using granulometry has been successfully applied to different scenarios such as industrial production[11] or medical diagnosis.[12]

Granulometry suitably adapts to emphysema characterization since it allows exact quantification of the area of airspaces in the lung image. Hence, accurate statistical modeling of the airspace size can be obtained. In our research, we used a multivariate pattern analysis approach to determine the degree of emphysema in a given tissue sample. It is based on the characterization of the pattern spectrum derived from granulometry analysis. From the segmented version of the lung image, in which alveolar tissue and air are distinguished, granulometry provides the probability density function of the variable s representing the size of the airspace. The first four standard moments of this variable are used as the features. They describe the statistical behavior of the airspace area. Subsequently, the feature vector is fed into a logistic regression model trained with emphysematous and control tissue samples.[13,14] The output of this model can be interpreted as the probability for a lung tissue image to be emphysematous. The purpose of this research work is to provide a novel metric for emphysema assessment. As the equivalent diameter method proposed by Parameswaran *et al.*,[5] our approach does not depend on the shape of the airspaces. In addition, it aims to serve as a preliminary stage to automatically obtain a probability map of emphysema affectation.

The paper is organized as follows. In the second section, a description of the image acquisition process and the dataset used in this study is provided. The third section includes a description of the proposed method for emphysema characterization. The experimental results obtained by this method are provided in the fourth section of the paper. In the fifth section, the properties of the method are discussed, resulting in the main conclusions of the study.

2. IMAGE DATASET

2.1 Animal preparation

All experimental protocols involving animal manipulation were approved by the University of Navarra Experimentation Ethics Committee. To simulate the parenchyma damaged caused by emphysema, treated mice were intratracheally instilled with 6 units per 30 g of porcine pancreatic enastase (PPE, EC134GI, EPC, MI, USA), as described in a previously published protocol.[15] Control animals were instilled with a saline solution.

2.2 Image acquisition

Lung lobe sections were obtained using an automated Axioplan 2ie Zeiss microscope (Carl Zeiss, Jena, Germany). Each slide was initially acquired with a Plan-Neofluar objective (numerical aperture NA = 0.035, magnification 1.25x, pixel resolution 3.546 μm/pixel). The automatic threshold method proposed by Otsu[16] was then applied to detect all tissue areas. The size of the objects was measured and only objects with a reasonable size to represent entire sections of lung lobes were considered for further processing. For each object, a bounding box was created and the coordinates of its four vertices were sent to the microscope. Then, an automatic routine scanned those areas with a Plan-Neofluar objective (NA = 0.3, 10x, 0.725 μm/pixel). Some overlap was allowed between image fields to facilitate the creation of large mosaics. The Stitcher ImageJ plugin[17] was used for it. The resulting mosaics were stored in a server for quantitative analysis. Four lung lobe sections were assessed in the present study. These were identified as emphysematous or normal cases according to the criterion of an expert. A total of 50 normal and 50 emphysematous patches of size 512 x 512 pixels were manually extracted from these sections. Figure 1 shows one of the lung sections and two patches extracted from it.

3. METHODS

The methodology proposed in our study is composed of three stages. In the first one, the original patch containing a region of the lung is segmented in order to differentiate between alveolar tissue and airspaces. The second stage involves the characterization of this patch. For this purpose, feature extraction is performed through granulometry analysis. In the third stage, the feature vector is used to assess the probability that the original lung patch corresponds to an emphysematous region. A description of these stages is provided in the following subsections.

3.1 Image segmentation

Initially, a gray-level version of the original patch is obtained by retaining its green channel as color does not contain useful information for the problem.[4–6] The lung patch is then segmented in order to identify airspaces in the image. Segmentation was performed by following the conventional combination of binarization together with erosion and dilation operations.[18] Original images depict alveolar tissue surrounding airspaces, which are the focus of our analysis. Image thresholding using the Otsu's method[16] was applied to obtain a binary version of the image. Its pixels denote tissue (black) or air (white) elements. To remove artifacts, erosion and dilation operators were applied. As a result, a more accurate definition of the frontiers (tissue) between airspaces is obtained. Figure 2 depicts the results from each step of the segmentation process: original gray-level image, binarization, erosion and dilation.

3.2 Feature extraction

Once segmentation is completed, airspaces are identified in the lung patch. The subsequent stage involves the characterization of these airspaces using granulometry analysis. For accurate emphysema quantification, the purpose is to model the probability density function of the variable s representing the size of an airspace in the image. The pattern spectrum provided by granulometry represents an estimation of this function. Granulometries are obtained by the repeated application of multiscale non-linear filters, which are implemented by means of a morphological operator. In our study, morphological opening at scale s will be used to filter the binary segmented image. This operation is expressed as follows:

$$I_s = X \circ B_s \tag{1}$$

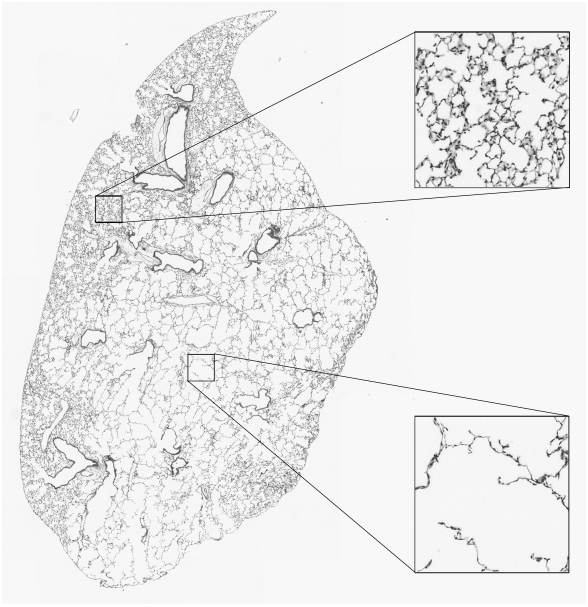

Figure 1. Example of a mosaic depicting a section of the lung and two patches of size 512 x 512 pixels. Top patch: parenchyma tissue without signs of damage; bottom patch: parenchyma tissue severely damaged that reproduces parenchyma destruction occurring in the lungs of patients.

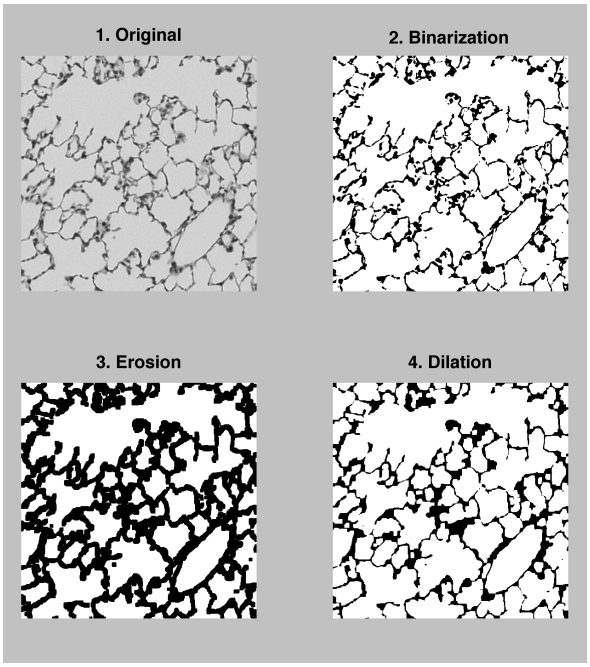

Figure 2. Segmentation of lung patches using binarization, erosion and dilation operators.

where X is the input binary image, B_s denotes the scale operator and I_s is the resulting image. The scale operator B_s will reject connected components with an area smaller than s pixels regardless its shape.[8] In the context of our problem, airspaces (white) of area less than s pixels will be set as tissue samples (black).

As the scale operator B_s is increased, $I_{s+1} = X \circ B_{s+1}$ is a subimage of I_s. The decreasing function $h(s)$ quantifies the total number of pixels remaining after each successive opening. It is worth noting that there exists K $(0 \leq K \leq N^2)$ such that for any $s \geq K$ the function achieves $h(s) = 0$, where the binary image X is supposed to have dimensions N x N pixels. The scale variable s represents the area of the airspaces and can be considered as a random variable. Hence, its cumulative density function $F(s)$ can be obtained as:

$$F(s) = 1 - \frac{h(s)}{h(1)} \tag{2}$$

where $h(1)$ denotes the total area of air in the original binary image. The value of $F(s)$, as expressed by equation (2), denotes the probability of finding an airspace with area equal or less than s in the image. Therefore, the probability density function of this variable is computed by means of the discrete derivative of $F(s)$, which is defined by the following expression:

$$f(s) = F(s) - F(s-1) = -\frac{h(s) - h(s-1)}{h(1)} \quad 1 \leq s \leq N^2 \tag{3}$$

The function $f(s)$ represents the pattern spectrum derived from granulometry analysis of the binary image X.[10] Its value estimates the probability for an airspace of area s to be found in X.

We propose using the first four standard moments to characterize the statistical behavior of s. Therefore, mean (μ_s), variance (σ_s^2), skewness (γ_s) and kurtosis (δ_s) were computed from $f(s)$. They estimate the central tendency, the degree of dispersion, the asymmetry and the peakedness of this function, respectively.[13] An exact definition of these features is given by the following equations:

$$\mu_s = \sum_i s_i f(s_i) \tag{4}$$

$$\sigma_s^2 = \sum_i (s_i - \mu_s)^2 f(s_i) \tag{5}$$

$$\gamma_s = \frac{\sum_i (s_i - \mu_s)^3 f(s_i)}{\sigma_s^3} \tag{6}$$

$$\delta_s = \frac{\sum_i (s_i - \mu_s)^4 f(s_i)}{\sigma_s^4} \tag{7}$$

Therefore, the feature vector $\mathbf{z} = (\mu_s, \sigma_s^2, \gamma_s, \delta_s)$ composed of these four statistical moments was used to describe the information contained in each of the patches.

3.3 Probability estimation and classification

A logistic regression (LR) classifier was used to estimate the probability for a lung patch to be an emphysematous region from its corresponding feature vector $\mathbf{z}$. LR linearly relates $\mathbf{z}$ with a response variable (t) via a link function. The variable t indicates the categorical decision about the input tissue sample. Thus, it represents two possible categories or classes: $t = 1$ (emphysema) or $t = 0$ (normal). The probability density function of t is then modeled by a binomial (Bernoulli) distribution given by:

$$p(t|\pi) = \pi^t (1 - \pi)^{1-t} \tag{8}$$

where π, the expected value of the variable t, provides the probability of being an emphysematous region $(t = 1)$.[13,14]

The aim of LR is to express π as a function of the feature vector $\mathbf{z}$. For this purpose, it is assumed that the value of π depends on the linear combination of the input features, i.e., $\pi = \pi(l)$ with $l = w_0 + \sum_{i=1}^{4} w_i z_i$ and the vector $\mathbf{w} = (w_0, w_1, ..., w_4)$ representing the set of model adaptive parameters or weights. This functional dependence is modelled by LR using the logistic function:[13]

$$\pi(l) = \frac{e^l}{1 + e^l} \tag{9}$$

Classification algorithms based on LR are usually applied to two-class problems as that proposed in the present study. The Bayes' decision rule[14] can be applied since posterior probabilities for both categories are directly obtained. The posterior probability for the emphysema group $p(t = 1|\mathbf{z})$ is given by π. Thus, the posterior probability for the normal tissue category is computed as $p(t = 0|\mathbf{z}) = 1 - p(t = 1|\mathbf{z})$. Therefore, according to the Bayes' rule, an image (identified by the vector $\mathbf{z}$) is assigned to the group for which its corresponding posterior probability is higher. In other words, since π denotes the probability for a lung patch to be emphysematous, it will be considered as emphysema if $p(t = 1|\mathbf{z}) \geq 0.5$ and normal otherwise.

The weight vector $\mathbf{w}$ of the LR model is adjusted (training) from a finite set of sample images from both categories: emphysema and normal. The iterated re-weighted least squares (IRLS) algorithm is used to find the parameters of the LR classifier.[19] It ensures a rapid optimization process according to the maximum likelihood principle.

4. RESULTS

To analyze the utility of the proposed approach in emphysema characterization, the utility of each extracted feature was initially assessed in order to determine its discriminant capability. Subsequently, the most informative features were used to estimate the probability of emphysema using LR analysis.

4.1 Analysis of the extracted features

Initially, we analyzed the information obtained from granulometry analysis of the lung tissue patches. For this purpose, Figure 3 shows an example of four pattern spectra corresponding to four different patches (two of them are emphysematous while the other two are normal tissue samples). The area value is normalized by the size of the window (512 x 512 pixels) used in the analysis. As it can be observed, higher probability is associated with larger airspaces in the case of patches containing emphysematous tissue. In contrast, the probability density function of normal patches tends to be concentrated in small values.

The differences found in the profile of the probability density functions from both tissue types were captured through the first four standard moments. In order to evaluate the discrimination ability of these features, we analyzed the statistical properties of each of them. The non-parametric Kruskal-Wallis test was performed to evaluate the occurrence of significant differences between the distribution of each feature in normal and emphysema groups.[13] Table 1 summarizes the p-value for each of the four features proposed in our study. The results reflect that features (μ_s), (σ_s^2) and (γ_s) provided statistically significant differences (p-value < 0.001) between both groups. However, (δ_s) provided a substantially higher p-value, showing that this feature does not provide useful information to discriminate between emphysematous and normal patches. We then discarded this feature for posterior multivariate analysis using LR.

4.2 Probability estimation and classification results

The three features identified as discriminant were used to estimate the probability of emphysema for a lung patch by means of LR analysis. As described before, the output of the LR model represents the probability that a patch corresponds to an emphysematous area of the lung, which is denoted by $p(t = 1|\mathbf{z})$. Thus, this probability value can be used to classify the patch as emphysematous, $p(t = 1|\mathbf{z}) \geq 0.5$, or normal, $p(t = 1|\mathbf{z}) < 0.5$. Accuracy *Acc* was adopted as the performance measure to evaluate the utility of LR classifiers.[14] It is defined as the probability of correct response and is estimated as the percentage of samples correctly classified:

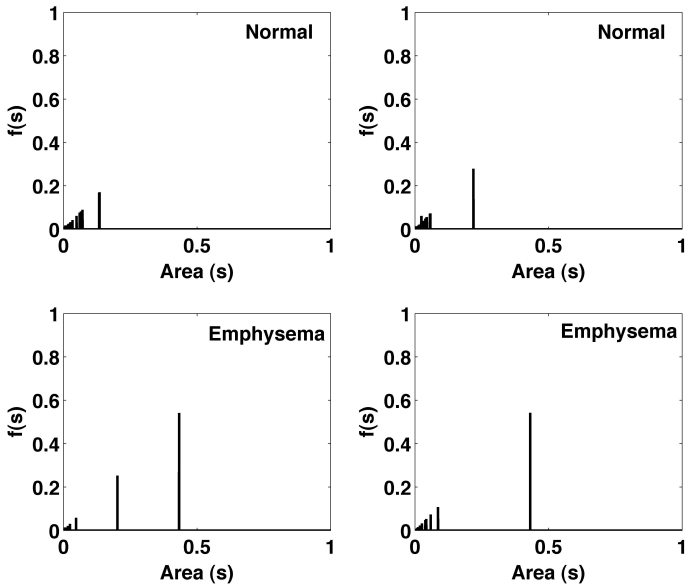

Figure 3. Pattern spectra derived from four patches of the lung parenchyma: two normal and two emphysematous regions.

Table 1. Results from Kruskal-Wallis test for each of the extracted features.

Feature set	***p*-value**
Mean (μ_s)	$4.10e^{-13}(< 0.001)$
Variance (σ_s^2)	$4.45e^{-9}(< 0.001)$
Skewness (γ_s)	$6.47e^{-13}(< 0.001)$
Kurtosis (δ_s)	0.06 (> 0.001)

$$Acc = \frac{T_p + T_n}{T_p + F_n + T_n + F_p} \tag{10}$$

where T_p (true positives) is the number of emphysematous patches correctly classified, T_n (true negatives) is the number of normal patches correctly classified, F_n (false negatives) is the number of misclassified emphysematous patches and F_p (false positives) is the number of misclassified normal patches. In addition, Acc can be expressed in terms of sensitivity (Se) and specificity (Sp). They indicate the number of emphysematous and normal patches correctly classified, respectively. These statistics are respectively estimated as:

$$Se = \frac{T_p}{T_p + F_n} \tag{11}$$

and

$$Sp = \frac{T_n}{T_n + F_p} \tag{12}$$

Additionally, receiver operating characteristic (ROC) analysis was performed.[20] Unlike Acc, Se and Sp, ROC analysis suppresses the requirement for a threshold by appraising the performance of a classifier over its whole range of possible values. A plot of Se versus $1 - Sp$ is made over the evaluated thresholds to obtain the ROC

curve. The area under the ROC curve (AUC) provides a quantitative index for classification performance. This index varies from 0.5 (no apparent accuracy) to 1.0 (perfect accuracy) as the ROC curve moves towards the left and top boundaries of the graph. AUC represents the probability of correct classification for a randomly chosen pairs of samples.

Four different LR-based algorithms were implemented to model the probability associated with emphysema. Three of them had a single input variable by separately using each of the discriminant features. The fourth LR-model corresponds to a multivariate approach based on the combination of the three selected features as inputs. Leave-one-out cross-validation was adopted to estimate the performance of these algorithms. Table 2 shows the classification results achieved in the emphysema detection problem. To compare the diagnostic capability of the four models, Figure 4 depicts the ROC curves provided by them. As can be observed, the multivariate LR model provided the highest performance with a diagnostic accuracy of 87% and an AUC of 0.95. This approach takes into account the information provided by the three selected features to estimate the probability of emphysema. As a result, it substantially improved the individual performance of each feature. It is worth noting that, when features are processed individually, the mean μ_s reached the highest accuracy with a correct classification rate of 80% and an AUC of 0.91. The results indicate that the proposed methodology is capable of accurately identifying emphysema regions.

Table 2. Classification results achieved by the implemented logistic regression models for probability of emphysema. *Se*: sensitivity; *Sp*: specificity; *Acc*: accuracy; AUC: area under the ROC curve.

Feature set	***Se*** **(%)**	***Sp*** **(%)**	***Acc*** **(%)**	***AUC***
Mean (μ_s)	70	90	80	0.91
Variance (σ_s^2)	60	86	73	0.81
Skewness (γ_s)	76	82	79	0.91
All selected (μ_s - σ_s^2 - γ_s)	84	90	87	0.95

5. DISCUSSION AND CONCLUSIONS

A novel method for automatic characterization of emphysema from histological lung analysis has been proposed. The statistical distribution of the airspaces in a lung patch was approximated by means of granulometry analysis. The first four standard moments were proposed to capture the properties of this distribution. From these features, a LR model was implemented to assess the probability for the patch to be emphysematous.

Granulometry has shown to be a useful tool to characterize lung tissue regarding emphysema. In our study, a correct classification rate of 87% and an AUC of 0.95 were achieved on a database composed of 512 x 512 patches from both normal and emphysematous regions of the lung. However, the main attribute of the proposed method is the ability to quantify the probability of a given patch being affected by parenchyma damage. There is an inherent subjectivity on visual evaluation of this disease.[21] Thus, for a given lung patch, the diagnosis may differ between experts. Binary classification between normal and emphysematous categories does not accurately characterize the patch under evaluation. Instead, methods for automatic quantification of emphysema are expected to provide the probability that such a patch is emphysematous. Our approach has been designed following this idea. A LR-model was used to assign the probability of a given patch being emphysematous using a simple feature vector derived from granulometry analysis. Interestingly, the proposed method could be used to build a probability map for the whole slide to identify those regions at higher risk of emphysema.

This probabilistic approach represents a step forward with respect to previous methods for emphysema quantification such as the mean linear intercept (L_m)[4] or the equivalent airspace diameter (d).[5] The latter has been recently pointed out as the most feasible quantitative measure of emphysema.[6] As described by Parameswaran *et al.*,[5] the value of the moments derived from d is not influenced by the distribution of airspace sizes. Hence, it is possible to identify emphysematous lung areas with a heterogeneous distribution of the

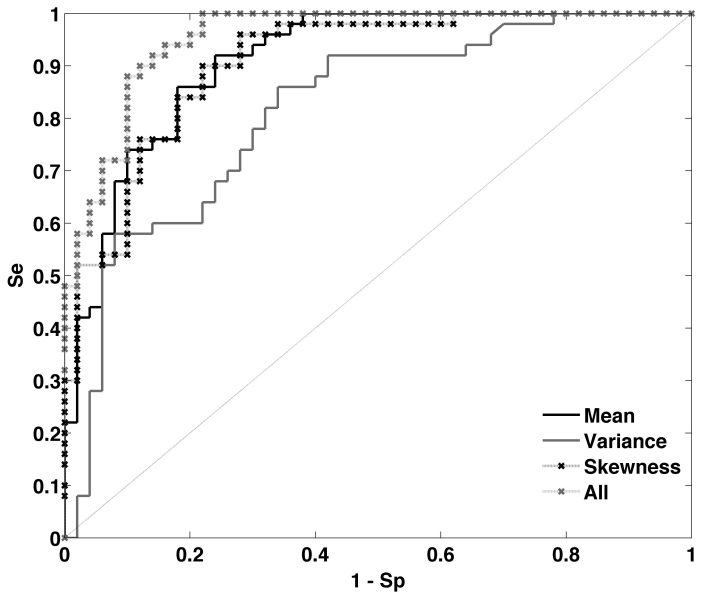

Figure 4. ROC curves provided by each of the logistic regression models implemented in this paper

airspaces (i.e., numerous small airspaces surrounding one large airspace). It is worth noting that this property is also achieved by the pattern spectrum computed by means of granulometry analysis, reflecting the robustness of our method for emphysema characterization. Furthermore, the proposed approach is not influenced by the shape of the airspaces. This behavior is due to the use of morphological area operators, which are not subject to the definition of a structuring element with a specific shape.[8,10] Instead, only the area of connected components is taken into account when morphological opening and closing operations are performed on the original image.

A critical design issue of our method is given by the choice of the window size that defines the lung patch to be analyzed. In our experiments, lung patches of size 512 x 512 pixels were processed. It is worth noting that the selection of the patch size involves a trade-off between the spatial resolution and the quality of the pattern spectrum derived from granulometry analysis. The latter is an approximation to the probability density function of the random variable s that represents the size of the airspaces contained in the image. A more accurate approximation will be obtained when the number of airspaces considered for its computation is higher. To achieve this, larger lung patches must be evaluated. However, increased window sizes result in lower spatial resolution as a larger region of the lung is analyzed. Therefore, further analysis about the influence of the patch size on the proposed characterization method is required.

The main limitation of our research is given by the fact that the presented method does not enable fully automatic analysis of the whole section of the lung tissue. A previous segmentation stage would be needed to achieve this condition. This initial segmentation must distinguish the limits of the lungs from the external zones as well as from the cuts of the lung tissue with blood vessels and bronchioles. Our method for emphysema characterization assumes that lung patches to be processed were previously segmented according to these requirements, i.e., they do not contain any spurious. For this purpose, the dataset evaluated in our experiments was composed of lung patches manually selected from the whole section of lung tissue. Therefore, future research work must address the development of this previous segmentation stage. In addition, the comparison of the proposed approach with the state-of-the-art methods such as the mean linear intercept (L_m) or the equivalent airspace diameter (d) must be considered in further analysis. On the other hand, a study on the number of samples evaluated is required to validate the presented method as a standard metric.

In summary, the proposed method enables emphysema characterization from *ex vivo* histological analysis of the lung. An elastase-induced mouse model of emphysema has been used for the experiments. Our approach is based on granulometry analysis of the lung tissue, which enables exact quantification of the area of airspace enlargements and is not affected by the variability of the airspace size. As an improvement over other previous

metrics, our method estimates the probability for a lung region to be emphysematous. Therefore, it could be used to obtain the probability associated with emphysema for a given region of the lung. The resulting probability map would provide an accurate characterization of its severity, indicating which areas of the lung are the most affected. This result could be of great interest to assess the effect of drugs and treatments for emphysema.

ACKNOWLEDGMENTS

J. V. Marcos is a research fellow at Institute of Optics (CSIC) under the programme Juan de la Cierva (Spanish Ministry of Economy and Competitiveness). This work has been partially sponsored by the grant UNAM PAPIIT IN113611. R. Nava thanks Consejo Nacional de Ciencia y Tecnologia for the doctoral scholarship 167161.

REFERENCES

1. American Thoracic Society, "Definitions, epidemiology, pathophysiology, diagnosis, and staging," *Am J Respir Crit Care Med* **152**, S78–S121 (1995).
2. Taraseviciene-Stewart, L. and Voelkel, N.F., "Molecular pathogenesis of emphysema," *J. Clin. Invest.* **118**, 394–402 (2008).
3. Murray, C.J. and Lopez, A.D., "Alternative projections of mortality and disability by cause 1990-2020: global burden of disease study," *Lancet* **349**, 1498–1504 (1997).
4. Dunnil, M.S., "Quantitative methods in the study of pulmonary pathology," *Thorax* **17**, 320–328 (1962).
5. Parameswaran, H., Majumdar, A., Ito, S., Alencar, A. M. and Suki, B., "Quantitative characterization of airspace enlargement in emphysema," *J. Appl. Physiol.* **100**, 186–193 (2006).
6. Munoz-Barrutia, A., Ceresa, M., ArtaechevarrÃa, X., Montuenga, L. M. and Ortiz-de-Solorzano, C., "Automated focusing in bright-field microscopy for tuberculosis detection," *Int. J. Biomed. Imaging*, ID 734734 (2012).
7. Salembier, P. and Serra, J., "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Trans. Image Process.* **4**, 1153–1160 (1995).
8. Acton, S.T. and Mukherjee, D. P., "Scale space classification using area morphology," *IEEE Trans. Image Process.* **9**, 623–635 (2000).
9. Maragos, P., "Pattern spectrum and multiscale shape representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 701–716 (1989).
10. Fletcher, N. D. and Evans, A. N., "Texture segmentation using area morphology local granulometries," in [*Mathematical morphology: 40 years on*], C. Ronse et al (eds.), 367–376 (2005).
11. Ljungqvist, M. G., Nielsen, M. E., Ersbol, B. K. and Frosch, S., "Image analysis of pellet size for a control system in industrial feed production," *Plos One* **6**, e26492 (2011).
12. Ross, N. E., Pritchard, C. J., Rubin, D. M. and Dusé, A. G., "Automatic image processing method for the diagnosis and classification of malaria on thin blood smears," *Med. Biol. Eng. Comput.* **44**, 427–436 (2006).
13. Jobson, J. D., " *Applied multivariate data analysis. Volume II: categorical and multivariate methods*, Springer, New York (1991).
14. Bishop, C. M., *Neural Networks for pattern recognition*, Oxford University Press, Oxford (1995).
15. Lakatos, H. F., Burgess, H. A., Thatcher, T. H., Redonnet, M. R., Hernady, E., Williams, J. P. and Sime, P. J., "Oropharyngeal aspiration of a silica suspension produces a superior model of silicosis in the mouse when compared to intratracheal instillation," *Exp. Lung Res.* **32**, 181–199 (2006).
16. Otsu, N., "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man Cybern.* **9**, 62–66 (1979).
17. Preibisch, S., Saalfeld, S. and Tomancak, P., "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics* **25**, 1463–1465 (2009).
18. González, R. C. and Wintz, P., " *Digital image processing*, Addison-Wesley, Reading, MA (1977).
19. Nabney, I.T., " *NETLAB: Algorithms for pattern recognition*, Springer, Berlin (2002).
20. Zweig, M.H. and Campbell, G., "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clin. Chem.* **39**, 561–577 (1993).
21. Miniati, M., Coppini, G., Monti, S., Bottai, M., Paterni, M. and Ferdeghini, E. M., "Computer arded recognition of emphysema on digital Cheste radiography," *Eur. J. Radiol.* **80**, e169–e175 (2011).