Preliminary results on automatic quantification of histological studies in allergic asthma

Mónica Abella[1], José Manuel Zubeldia[2], Laura Conejero[2], Norberto Malpica[3], Juan José Vaquero[1] and Manuel Desco[1]

[1] *Laboratorio de Imagen, Unidad de Medicina Experimental. Hospital Gregorio Marañón, Madrid, Spain*
[2] *Servicio de Alergia, Unidad de Medicina Experimental, Hospital Gregorio Marañón, Madrid, Spain*
[3] *Laboratorio de Análisis de Imagen Médica, Universidad Rey Juan Carlos, Madrid, Spain*

***Abstract*— The evaluation of new therapies to treat allergic asthma makes frequent use of histological studies. Some of these studies are based on the microscope observation of stained paraffin lung sections to quantify cellular infiltration, an effect directly related to allergic processes. To our knowledge, there is no software tool for doing this quantification automatically nowadays. This paper presents a method for the quantification of cellular infiltrate of lung tissue images in a mouse model of allergic asthma. Each image is divided into regions of equal size that are classified by means of a segmentation algorithm based on texture analysis. The classification uses three discriminant functions, built from parameters derived from the histogram and the co-occurrence matrix and calculated by performing an initial stepwise discriminant analysis on 79 samples from a training set. Results provide a correct classification of 96.8 % on an independent test set of 251 samples labeled manually.**

***Index Terms*— Asthma, automated microscopy, lung tissue, co-occurrence matrix, image analysis, segmentation, texture.**

I. Introduction

Asthma is a chronic inflammatory disease of the lung, characterized by airway hyperresponsiveness to a variety of stimuli, eosinophilic inflammation of the airways, mucus hypersecretion, and elevated serum IgE levels [1]. Mortality of asthma has increased worldwide, despite the use of currently available medications, underlining the need for the development of novel therapies [2-3]. Only classic immunotherapy (vaccine with allergen based on olive proteins - Olea europaea) has proven to have effect on the allergic reaction mechanisms, although showing a limited efficacy.

Research carried out by the Allergy Department of Hospital Gregorio Marañón studies the preventive effect on allergic inflammation of the simultaneous administration of immunostimulatory sequences of DNA oligodeoxynucleotides (ISS-ODNs) together with olive pollen proteins (Olea europaea) in a mouse model of allergic asthma [4,5]. Results show that simultaneous administration of ISS-ODN and olea prior to sensitization prevents the development of an allergic phenotype in their mouse model of bronchial asthma. Approaches of this type may provide a novel strategy for the prevention of allergic diseases.

Three types of studies are commonly performed in this type of experiments to assess the effects: immunologic parameters (immunoglobulines and cytokines), pulmonary function (bronchial hyperactivity) and histological studies (cellular infiltrate and bronchial mucus secretion). The latter are assessed nowadays by visual inspection by an expert, as there is no automatic analysis tool available. Such a tool would speed up the process, also providing a more repeatable quantification.

In order to quantify the histological studies it is necessary to segment the image so that the infiltration area can be identified [6-11]. This could be done by trying to identify each nucleus of an infiltrating cell, as they are clearly identified in the blue component of the color image. This method is not recommended considering that cells in bronchial wall are stained the same way and have the same size as infiltrate cells and that cell aggregation may hide edges between cells.

It can be seen, however, that the infiltrate area has a texture pattern different from other regions in the image. This leads us to apply region identification based on texture analysis [12].

This paper presents a method for automatic quantification of cellular infiltrate in this kind of histological studies.

II. Materials and Methods

Segmentation is performed by classifying each region of the image into one of three a priori classes, on the basis of a vector of texture parameters. The optimal texture parameters were obtained from the luminance component of the images in a training data set by means of stepwise discriminant analysis that also provided the corresponding discriminant (Fisher) functions.

Corresponding autor: Manuel Desco, email: desco@mce.hggm.es

A. *Sample preparation and working dataset*

For histological analysis, lungs are fixed in 4% paraformaldehyde-PBS and tissue blocks are embedded in paraffin. Ten-µm sections are stained with hematoxilyn and eosin (H&E) [5]. Images are acquired with an Olimpus IX70 microscope at 10x magnification, and captured with a Sony DXC 151P color CCD camera. Images of size 1712x1368 are transferred to a PC for analysis.

To assess the potential of texture analysis approach, we made a first attempt defining five classes matching the number of found tissue types by examining the images (Fig. 1).

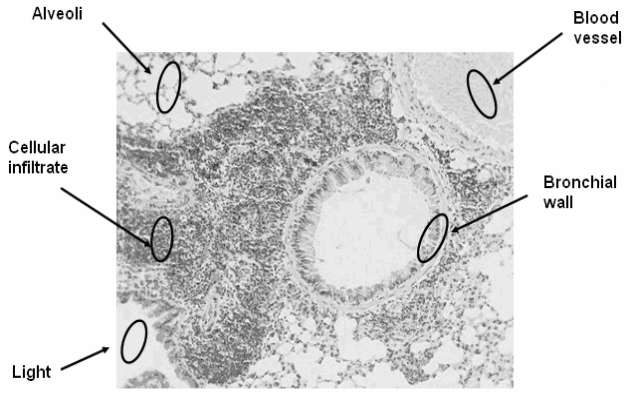

Fig. 1. Five textures identified in lung studies images.

We obtained 330 samples from 9 images with different percentages of infiltrate area to build the working dataset: 79 samples used for training and 251 samples for testing the classifier. Each of these samples was manually classified by an expert.

Texture features extraction is done over the luminance component as it has the most representative information for the textures selected. A median filter with a 7-point mask is applied to eliminate impulsive noise. Sample size (denoted as "analysis window") was set to 85x85 pixel (1/20 of the image), which appeared to be enough to separate alveoli from bronchial light (the most critical issue for a small sample size) while allowing a reduction in processing time and a gain of accuracy in segmentation.

B. *Texture parameters*

Texture parameters used in this application can be classified into first-order statistics, computed from the normalized histogram, and second-order statistics, computed using the Gray Level Coocurrence matrix (GLCM) [13, 14].

The normalized histogram is the probability density function for the grey levels of a specific region. If we denote by $p(z_i)$ the probability of each grey value z_i, the following parameters can be defined:

$$\text{Mean: } \mu = \sum_{z_i=0}^{N-1} z_i p(z_i)$$

$$\text{Variance: } \sigma^2 = \sum_{z_i=0}^{N-1} (z_i - \mu)^2 p(z_i)$$

$$\text{Skewness: } \mu_3 = \sigma^{-3} \sum_{z_i=0}^{N-1} (z_i - \mu)^3 p(z_i)$$

$$\text{Kurtosis: } \mu_4 = \sigma^{-4} \sum_{z_i=0}^{N-1} (z_i - \mu)^4 p(z_i) - 3$$

$$\text{Percentile X = T: } \quad X = \sum_{z_i=0}^{T} p(z_i)$$

The histogram considers the grey level of each pixel separately and no spatial information is conveyed in these parameters. To incorporate spatial distribution of the grey levels we make use of the grey-level co-occurrence matrix (GLCM). Any GLCM element $P_d(i,j)$ reflects the distribution of the probability of occurrence of a pair of grey levels (i,j) separated by a given distance *d*. The GLCM is computed by mapping the grey-level co-occurrence probabilities based on spatial relations of pixels in different angular directions θ.

As with the histogram, a normalized version of the co-occurrence matrix can be computed, dividing each element by the total number of neighbors for each *d* and *θ*. These values depend on the texture. From the co-occurrence matrix, the following parameters were derived:

Second order angular moment:

$$AngScMom = \sum_{z_i=0}^{N-1} \sum_{z_j=0}^{N-1} [p(z_i, z_j)]^2$$

Contrast

$$Contrast = \sum_{z_i=0}^{N-1} \sum_{z_j=0}^{N-1} (z_i - z_j)^2 p(z_i, z_j)$$

Correlation:

$$Correlat = \sum_{z_i=0}^{N-1} \sum_{z_j=0}^{N-1} \frac{z_i z_j (z_i - z_j) - \mu_x \mu_y}{\sigma_i \sigma_j}$$

Sum of Squares:

$$SumOfSqs = \sum_{z_i=0}^{N-1} \sum_{z_j=0}^{N-1} (z_i - \mu_x)^2 p(z_i, z_j)$$

Corresponding autor: Manuel Desco, email: desco@mce.hggm.es

Difference Variante

$$DifVarnc = \sum_{i=0}^{2(N-1)} (i - \mu_{x-y})^2 \, p_{x-y}(i)$$

where $p_{x-y}(k) = \frac{1}{R} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j)$

Inverse difference moment:

$$InvDfMom = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p(i,j)}{1+(i-j)^2}$$

Entropy:

$$Entropy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j) \log[p(i,j)]$$

Entropy of difference

$$DifEntrop = -\sum_{i=0}^{2(N-1)} p_{x+y}(i) \log(p_{x-y}(i))$$

Sum Average:

$$SumAverage = \sum_{i=0}^{2(N-1)} i \, p_{x+y}(i)$$

$$p_{x+y}(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j) \text{ where}$$

$$i+j = 0,1,\ldots,2(N-1)$$

Sum Entropy

$$SumEntrop = -\sum_{i=0}^{N-1} p_{x+y}(i) \log(p_{x+y}(i))$$

Sum Variance:

$$SumVariance = \sum_{i=0}^{2(N-1)} i - SumAveg) \, p_{x+y}(i)$$

In these expressions, N is the number of grey levels, z_i are the different grey levels, $p(z_i,z_j)$ is the value of the GLCM at point (i,j), μ_x and μ_y are the mean values of GLCM values accumulated in the x and y directions and μ_{x-y} is the mean value of the distribution p_{x-y}. To improve computation speed, advantage can be taken from the fact that the co-occurrence matrix is symmetric. On the other side, the size of the GLCM depends on the grey level resolution of the image. Texture parameters have shown to be reasonably invariant to grey-level quantization. In this work, images were quantized to 4 bits/pixel before the matrix was calculated, to increase computational speed, and GLCM was computed for d = 1, 2, 3, 4, 5 and θ = 0°, 45°, 90°, 135°.

C. *Training: discriminant analysis*

Texture parameters have been extracted using *MaZda*, a software tool developed at the Institute of Electronics, Technical University of Lodz, Poland [15]. As a result, we obtained 229 parameters (9 from the normalized histogram and 220 of the 20 coocurrence matrix computed) for each sample. The final result for the 79 samples was a table of 79x229 parameter values. Given this amount of data, a reduction of dimensionality of the features vector was clearly advisable.

This reduction was made by means of a feature selection of the most discriminant ones with a stepwise method based on Fisher criterion. At each step, the feature that better contributed to class separation, following the criterion of maximization of the change in Wilk's Lambda, was included [16]. An inclusion condition, based on an F-test, was used to evaluate if the change in discrimination was significant. The F-to-enter and F-to-remove values used were 3.84 and 2.71, respectively. This analysis was performed with *SPSS* for Windows, Rel. 11.5.1. (SPSS Inc., Chicago).

D. *Reduction of the number of classes*

Resulting clusters for the five classes defined as a function of the two main functions are presented in Fig. 2. It can be noticed that there are three main groups, what suggests a possible clustering regrouping. Provided that the target is just the infiltrate area, it is not necessary to identify separately all the five classes.

As a result of this preliminary study, we decided to reduce the number of classes to three: cellular infiltrate, light and other tissues (blood vessels, alveoli, bronchial walls).

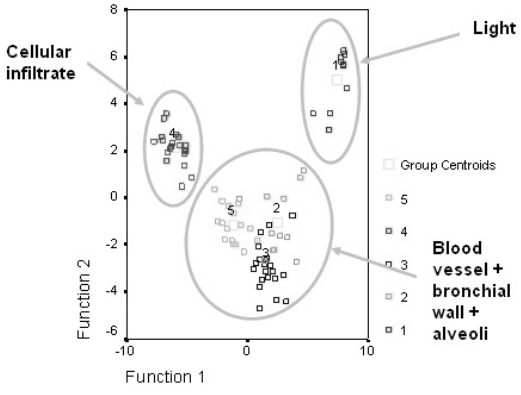

Fig. 2. Five textures identified in lung studies images

Corresponding autor: Manuel Desco, email: desco@mce.hggm.es

Samples from the classes in the 'other tissues' class where re-labeled and discriminant analysis was repeated. Stepwise discriminant analysis resulted in the selection of 12 parameters (Table I). Resulting clusters are shown in Fig. 3.

E. Classification tool

A classification tool was implemented following a procedure similar to the one described in [17]. This tool tessellates the images, calculates texture parameters for each region and applies the discriminant functions obtained in the training phase. Images are divided into regions of size 55x55 pixels, denoted as 'classification window'. Each region is assigned to a class on the basis of texture parameters computed on the previously defined region of 85x85 pixels, denoted as 'analysis window'.

A tissue area mask is generated by thresholding the hue component of the image. Final result is expressed as the percentage of cellular infiltrate area in the total tissue area.

F. Validation

A leave-one-out method was used to obtain an initial estimate of the correct classification rate. This method involves leaving out each case in turn, calculating the function based on the remaining cases, and then classifying the left-out case. A more accurate estimate of the correct classification rate was derived by classifying a test set of 251 samples not included in the training set.

III. Results

Stepwise discriminant analysis on the data obtained with MaZda resulted in the selection of 12 parameters. Parameter names and their weighting coefficients for the three linear discriminant functions are shown in Table I. Resulting clusters as a function of the two main characteristics selected are presented in Fig. 3.

The leave-one-out method yielded a classification success rate of 97,5 % on the training set. Results on the test set (251 samples) showed a 96,8 % of classification success.

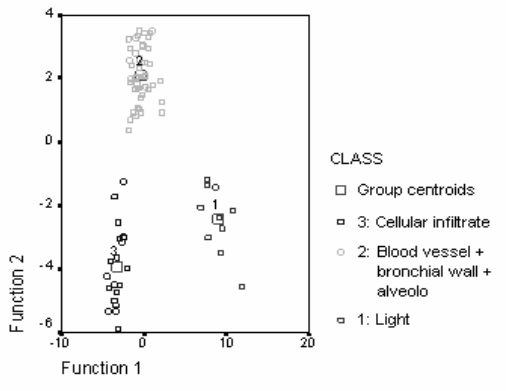

Fig. 3. Clusters obtained for 79 samples.

TABLE I
COEFICIENTS FOR THE THREE LINEAR FISHER FUNCTIONS

	Infiltrates	Light	Others
Percentile 01%	14.812	16.323	15.033
Percentile 90%	-9.592	-9.770	-9.048
Percentile 99%	-1.371	-2.426	-2.145
AngMom(1,0)	3360.578	3797.413	3531.605
DifEntrop(1,0)	11601.510	11928.582	11897.886
Correlation(0,1)	-475.031	-320.065	-482.230
Correlation(1,1)	277.963	72.031	268.837
SumEntrop(1,-1)	9.493	10.974	9.317
InvDfMom(2,0)	9546.252	9880.968	9970.419
AngMom(5,0)	-3002.056	-2963.440	-3390.924
SumEntrop(5,5)	4303.205	4736.825	4448.329
AngMom(5,-5)	3889.051	3851.420	4209.196
(Constant)	-6994.747	-7742.933	-7494.682

The tool has been tested on an Intel Pentium, 3Ghz with 1GB RAM. For images of size 1712x1368, average classification time is about 76 seconds, 65 % of the time dedicated to the computation of the co-ocurrence matrix. An example of the result obtained is shown in Fig. 4.

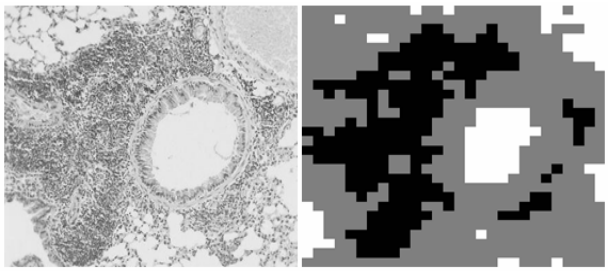

Fig. 4. Original image and segmentation result; black pixels define infiltration area.

IV. Discussion

This work presents an approximation to the automatic quantification of microscopy images of histological lung studies. These studies, relevant for research in allergic asthma and other pulmonary diseases, are currently performed by simple visual inspection of the images. Classification results are promising (96.8 hits) in a reasonable time (76 secs. for images of size 1712x1368). A further validation of the quantification tool in terms of results for complete images compared to those obtained from observation by an expert seems advisable.

In the present study, both the 'analysis window' and 'classification window' were of a fixed size, determined empirically. 'Analysis window' must be large enough to have a sufficient number of pixels for parameter computing. The size of the 'classification window' determines the resolution of the classification. Using a smaller 'classification window' increases the resolution (minimizing partial volume effect) but also the computational cost.

Corresponding autor: Manuel Desco, email: desco@mce.hggm.es

The method uses only histogram and grey-level co-occurrence matrix parameters.

Regarding processing time, the slowest stage in the quantification process is the texture feature extraction algorithm. Although no attempt was made to speed up the algorithms, the processing time would only be a concern if the study subject included a high number of images.

On the other hand, we derived discriminant functions from 12 parameters, which might be too many. A study of the marginal discrimination power provided by the additional parameters would be advisable.

The system developed could be used to segment other histological studies, although the parameter selection and discriminant function calculation would have to be repeated for the new image types.

A reliable and easy to implement method has been developed, providing an automatic way of quantifying microscopy images of lung histological studies. Preliminary results show similar accuracy to that provided by an expert, while allowing analyzing a much larger number of fields in a repeatable way.

ACKNOWLEDGMENTS

The authors thank the Institute of Electronics, Technical University of Lodz (Poland), for providing the MaZda texture analysis software, in the framework of the EU COST B11 project.

REFERENCES

[1] Tortora G. J. , Grabowski, S., Principles of Anatomy and Physiology, John Wiley & Sons, 2002.

[2] D'Amato G, Liccardi G, D'Amato M, Cazzola M: Outdoor air pollution, climatic changes and allergic bronchial asthma. *Eur Respir J* 2002, 20(3):763-776.

[3] Holt PG, Sly PD, Martinez FD, et al. Drug development strategies for asthma: in search of a new paradigm. *Nat Immunol* 2004;**5**(7)**:**695-8.

[4] Horner, A.A., et al., DNA-based immunotherapeutics for the treatment of allergic disease. Immunol Rev, 2001. 179: p. 102-18.

[5] Laura Conejero Hall, Yoko Higaki, María Fernández-Bohorquez, Isabel Varela Nieto, María Luisa Baeza de Ocáriz, José Manuel Zubeldia Ortuño, Preventive use of the co-administration of immunostimulatory DNA sequences (ISS-ODN) with olive pollen proteins in a mouse model of allergic asthma., Journal of Allergy and Clinical Immunology 2005 Feb; 115(2): S263.

[6] J.S. Weszka, C.R.D , A. Rosenfeld, A Comparative Study of Texture Measures for Terrain Classification. IEEE Trans. Systems, Man, Cybernetics. vol.6: p. pp.269-285.

[7] Trygve Randen, J.H.H., Filtering for Texture Classification: A Comparative Study. IEEE Trans. On Pattern Analysis and Machine Intelligence, April 1999. vol.21, no.4: p. pp.291-309.

[8] J.M.H. du Buf, M.K., M. Spann, 'Texture Feature Performance for Image Segmentation', Pattern Recognition. 1990. vol.23, no.3/4: p. pp.291-309.

[9] J. Strand, T.T., 'Local Frecuency Features for Texture Classification', Pattern Recognition. Vol. vol.27, no.10. 1994. pp.1397-1406.

[10] A.H. Bhalerao and N.M. Rajpoot. Discriminant Feature Selection for Texture Classification. In Proc. British Machine Vision Conference (BMVC'2003), Sep. 2003.

[11] K. Valkealathi, E.O., Reduced Multidimensional Co-Ocurrence Histograms in Texture Classification. IEEE Trans. Pattern Analysis and Machine Intelligence, 1998. vo.20, no.1: p. pp.90-94.

[12] M. Tuceryan, A.K.J., Texture Analysis, in Handbook Pattern Recognition and Computer Vision. 1993, C.H. Chen, L.F. Pau, and P.S.P. Wang eds.: Singapore: World Scientific. p. ch.2, pp. 235-276.

[13] Julesz, B., Experiments in the visual perception of texture. Sci Am, 1975. 232(4): p. 34-43.

[14] M. Sonka, V.H., R. Boyle, Image processing, analysis and machine vision. 1993, Great Britain, University Press, Cambridge: Chapman & Hall.

[15] Strzelecki, A.M.a.M., Texture Analysis Methods - A Review. 1998, Technical University of Lodz, Poland. p. COST B11 report (presented and distributed at MC meeting and workshop in Brussels, June 1998).

[16] William R. Dillon, M.G , Multivariate análisis. Methods and applications. 1984: John Wiley & Sons.

[17] N. Malpica, Automatic quantification of viability in epithelial cell cultures by texture analysis, Journal of Microscopy, Vol. 209, Pt 1 January 2003, pp. 34–40

Corresponding autor: Manuel Desco, email: desco@mce.hggm.es