Study Lung Tool: A Way to Understand HRTC Lung Parenchyma

Verónica Vasconcelos[1,3], José Silvestre Silva [2,3], and João Barroso[4]

[1] Instituto Superior de Engenharia, Instituto Politécnico de Coimbra, Portugal
[2] Departamento de Física, Faculdade de Ciências e Tecnologia, Universidade de Coimbra, Portugal
[3] Centro de Instrumentação, Faculdade de Ciências e Tecnologia, Universidade de Coimbra, Portugal
[4] Universidade de Trás-os-Montes e Alto Douro, Portugal

Abstract—**The purpose of the described system is to aid radiologists on their daily routine in the task of analyzing HRCT lung images and to contribute to a more accurate and fast diagnosis. We developed a framework -Study Lung Tool- with the objective of gather information from radiologists, in a systematic way. Using Study Lung Tool framework, the radiologist analyzes HRCT scans, outlines regions of typical pattern and characterizes the patterns. A database of typical patterns associated with common pulmonary diseases was created. The information gathered can be a valuable teaching tool to every one that intends to understand HRCT lung parenchyma. The proposed system discriminates between normal and abnormal patterns of lung parenchyma based on statistical texture analysis extracted from HRCT lung scans. An overall accuracy of 89,2%, a sensitivity of 92,7% and a specificity of 83,6% were achieved.**

Keywords—**High-resolution computed tomography, pulmonary parenchyma, computer-aided diagnosis, texture analysis.**

I.   INTRODUCTION

In clinical practice, Computed Tomography (CT) is an important imaging modality for the diagnosis, treatment, and follow-up of lung diseases. Nowadays, the image quality achieved by High Resolution Computed Tomography (HRCT) scanners provides high morphologic detail of lung parenchyma and has already become the accepted imaging golden standard for lung study [1].

However, the HRCT analysis and interpretation of lung parenchyma patterns is a complex task. Several lung changes are difficult to understand because patterns are often mixed, can be present in different diseases and also may change during different stages of the disease. Many computer-aided diagnosis (CAD) systems have been developed during the last several years [2 - 7] showing the importance of this issue.

Due to intra-observer and inter-observer differences in analysis and interpretation of HRCT scans and the difficulty of establish a reliable reference standard we were motivated to develop a framework - Study Lung Tool - that collects information of normal and abnormal lung parenchyma from HRCT exams acquired at Coimbra University Hospital [8].

In this paper, we present an automated system capable of detect abnormal lung tissue in HRCT scans. Based on a database of patterns the system was trained and tested using pattern recognition approach. Each region of interest (ROI) is characterized through a set of 97 statistical texture descriptors. The performance of the system was evaluated by overall accuracy, sensitivity and specificity.

II.   DATA AQUISITION

Usually the diagnosis of lung disease on a thoracic CT is based on three elements [9]: recognition and classification of abnormal pattern in a category, determination of location and distribution of the abnormalities in the lung, careful analysis of the patient data that are available at the time the CT scan is done.

In order to gather information from radiologists we developed an independent, intuitive and user-friendly framework - Study Lung Tool. This software has multiple applications:

- Allows the visualization of complete CT scans, in an automatic mode or image by image mode. Details can be observed by magnifying a region or whole CT slice through zoom in/out option;
- After the characterization of diseases/patterns in the exam, by radiologist, the software can be used as a learning tool. Users of this framework can understand lung anatomy and pathology in HRTC images, understand the CT changes in patients with lung disease, recognize the appearance and distribution of lung patterns in different stages of the disease and to use these patterns to assist in the diagnosis;
- Radiologist can review complete HRCT scans, select and put a label on the outlined ROI, typifying tissue patterns. The Figure 1 illustrates some software options and some outlined ROIs. These ROIs were split in small ROIs of 32x 32 pixels and used as a training and test set in the classification process.

Fig. 1 Some outlined ROIS done by a radiologist using Study Lung Tool framework

Figure 2 shows some examples of the ROIs highlighted by radiologists, such as normal lung tissue and several abnormal patterns. The data were collected from the regular practice from Radiology Department of Coimbra University Hospital. The images were acquired on a LightSpeed VCT 64 multidetector spiral scanner from General Electrics Healthcare. The images were stored as individual DICOM 16-bit grayscale files, with the pixel intensity proportional to tissue density. Images were displayed using a lung window with a centre of -700 Hounsfield Units (HU) and a width of 1500 HU. Each ROI was selected from three predefined anatomic levels: apex to carina, carina to the lower lung veins and lower lung veins to the diaphragm.

III. TEXTURE ANALYSIS

Texture plays an important role in medical images analysis. In statistical texture methods, texture features are computed from statistics distribution of observed combinations of intensities positions relative to each other in the image. The results presented is this paper are based on statistical methods including first-order gray level statistics, gray level run length (GLRL) method and gray level co-occurrence (GLCM) method.

The spatial dependence of gray levels distributions were computed from a set of texture descriptors extracted from a set of co-occurrence matrices. The co-occurrence matrix $P(i,j|d,\theta)$ corresponds to the number of pairs of gray levels encountered in the neighborhood of each pixel [10]. The six texture descriptors used in our method, are presented in equations (1) to (6), where L is the number of gray levels, P is the normalized symmetric GLCM of dimension LxL and $P(i,j)$ is the (i, j)th element of P. The direction $\theta=0º$, $\theta=45º$, $\theta=90º$ and $\theta=135º$ with a range of $d=1$ and $d=2$ was used in this work, in a total of eight GLCM.
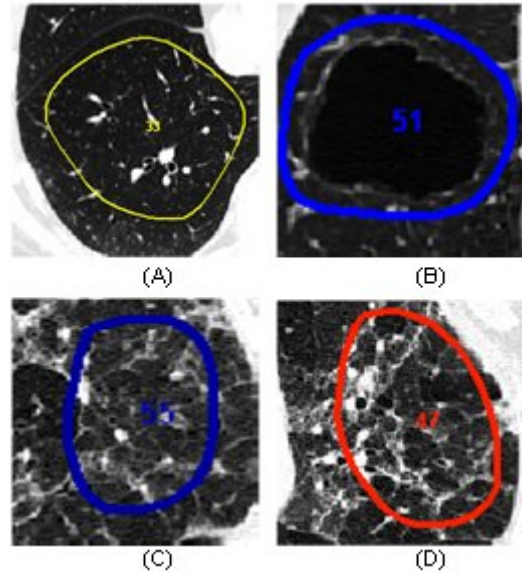

Fig. 2 Several examples of tissue patterns as outlined by radiologist. On each image, a region of interest (ROI) highlights a typical pattern. (A) Normal lung tissue. (B) Bullous emphysema. (C) Ground-glass opacity. (D) Honeycombing

1. Angular Second Moment (ASM) or Energy

$$ASM = \sum_{i=1}^{L}\sum_{j=1}^{L} P^2(i,j) \tag{1}$$

2. Entropy (ENT)

$$ENT = -\sum_{i=1}^{L}\sum_{j=1}^{L} P(i,j)\log_2\big(P(i,j)\big) \tag{2}$$

3. Maximum Probability (MaxProb)

$$MaxProb = Max\left(\sum_{i=1}^{L}\sum_{j=1}^{L} P(i,j)\right) \tag{3}$$

4. Inverse Difference Moment (IDM)

$$IDM = \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{P(i,j)}{1+(i-j)^2} \tag{4}$$

5. Contrast (CON)

$$CON = \sum_{i=1}^{L}\sum_{j=1}^{L} P(i,j)(i-j)^2 \tag{5}$$

6. Correlation (COR)

$$COR = \sum_{i=1}^{L}\sum_{j=1}^{L} P(i,j)\left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i\sigma_j}\right] \tag{6}$$

where μ_i is the Mean and σ_i the Standard Deviation

$$\mu_i = \sum_{i=1}^{L} \sum_{j=1}^{L} P(i,j)i \tag{7}$$

$$\sigma_i = \sqrt{\sum_{i=1}^{L} \sum_{j=1}^{L} P(i,j)(i-\mu_i)^2} \tag{8}$$

The shape aspects of the gray level primitives of each ROI were described through RLM features. The run length features were extracted from run length matrix M. A gray level run primitive is defined as a set of consecutive and collinear pixels which have the same gray level intensity, the run length is the number of pixels in the run and the run length value is the number of times such a run occurs in the image. Fine textures tend to contain more short runs with similar gray level intensities, while coarse textures have more long runs with significantly different gray level intensities [11].

Each element of a run length matrix $M(a,r|\theta)$ represents the number of runs with pixels of gray level intensity a and length r along the orientation θ. The typical orientations of 0°, 45°, 90°, and 135° were used in this work. Calculating the run length encoding for each direction will produce four run length matrices. The size of each matrix M is LxN_r, where L is the number of gray levels and N_r is the longest run length in the ROI.

Each ROI was quantized into 16 gray levels scale. A set of four run length matrices were computed, one for each direction θ and eleven scalar measures were extracted as proposed by Galloway [12], Chu et al. [13] and Dasarathy and Holder [14]:

1. Short Run Emphasis (SRE):

$$SRE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} \frac{M(a,r)}{r^2} \tag{9}$$

where n_r is the total number of observed runs

$$n_r = \sum_{a=1}^{L} \sum_{r=1}^{Nr} M(a,r) \tag{10}$$

2. Long Run Emphasis (LRE):

$$LRE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} M(a,r)r^2 \tag{11}$$

3. Gray Level Non-uniformity (GLNU):

$$GLNU = \frac{1}{n_r} \sum_{a=1}^{L} \left(\sum_{r=1}^{Nr} M(a,r) \right)^2 \tag{12}$$

4. Run Length Non-uniformity (RLNU):

$$RLNU = \frac{1}{n_r} \sum_{r=1}^{Nr} \left(\sum_{a=1}^{L} M(a,r) \right)^2 \tag{13}$$

5. Run Percentage (RP)

$$RP = \frac{n_r}{n_p} \tag{14}$$

where n_r is the total number of pixels.

6. Low Gray Level Run Emphasis (LGLRE)

$$LGLRE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} \frac{M(a,r)}{a^2} \tag{15}$$

7. High Gray Level Run Emphasis (HGLRE)

$$HGLRE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} M(a,r)a^2 \tag{16}$$

8. Short Run Low Gray Level Emphasis (SRLGLE)

$$SRLGLE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} \frac{M(a,r)}{a^2 r^2} \tag{17}$$

9. Short Run High Gray Level Emphasis (SRHGLE)

$$SRHGLE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} \frac{M(a,r)a^2}{r^2} \tag{18}$$

10. Long Run Low Gray Level Emphasis (SRLGLE)

$$LRLGLE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} \frac{M(a,r)r^2}{a^2} \tag{19}$$

11. Long Run High Gray Level Emphasis (SRHGLE)

$$LRHGLE = \frac{1}{n_r} \sum_{a=1}^{L} \sum_{r=1}^{Nr} M(a,r)a^2 r^2 \tag{20}$$

In order to describe the HRCT attenuation characteristics of each ROI, a set of textural features were obtained in the form of a 97-dimensional vector **v**. These features included mean, standard deviation, skewness, and kurtosis first-order features, 48 co-occurrence features, and 44 run length features.

IV. PERFORMANCE ASSESSMENT

To perform automated discrimination of the samples, between normal and abnormal lung tissue, we used a Support Vector Machine (SVM) classifier. SVM is a machine-learning method, introduced by Vapnik in 1995 [15], based on the principle of structural risk minimization. The mapping from original feature space to a higher dimension feature space is done using a set of mathematical functions known as kernels. As the feature extraction is implicit in the training procedure we did not performed feature selection [16]. The implementation of the SVM classifier was done through the Bioinformatics Toolbox

available in Matlab R2007b and a Gaussian radial basis function kernel was used.

From a set of 241 scans, radiologists outlined several ROIs. Each one was split into smaller ROIs of 32x32 pixels. To make optimal use of the available dataset a four-fold cross validation method was performed, randomly dividing the dataset in four groups. On average there were 630 ROIs in the train set and 210 ROIs in the test set.

The system was evaluated using sensitivity, specificity and overall accuracy. A positive sample is a ROI where lung pathology is present, and a negative sample is a ROI where lung tissue is normal. Sensitivity is defined as the number of samples correctly classified as positive divided by total number of positive samples. Specificity is defined as the number of samples correctly classified as negative divided by total number of negative samples. Overall accuracy is defined as the total number of correctly classified samples divided by the total number of samples. The results achieved were an accuracy of 89,2%, a sensitivity of 92,7%, and a specificity of 83,6%.

V. Conclusions

We described a CAD system that automatically distinguishes normal from pathologic tissue in HRCT lung scans. The system was based on a pattern recognition approach, in which a set of statistical texture descriptors were used to characterize each ROI.

Classification was done based on a database of typical samples from different pathologies, acquired from Coimbra University Hospital daily routine. This database was acquired through Study Lung Tool framework that showed to be very practical and also a valuable teaching tool.

Due to the large amount of extracted features for each ROI, a feature selection algorithm should be included to reduce computational complexity. We also intend to evaluate the ROI size in overall performance of our system.

Acknowledgment

References

1. Verschakelen J A, Wever W D (2007) Computed Tomography of the Lung - A Pattern Approach. Springer Berlin Heidelberg Edition
2. Sluimer I, Waes P F, Viergever M A (2003) Computer-aided diagnosis in high resolution CT of the lungs. Med Phys 30:3081–3090
3. Uppaluri R, Hoffman E A et al (1999) Computer Recognition of Regional Lung Disease Patterns. Am J Respir Crit Care Med 160:648-654
4. Xu Y, Sonka M, McLennan G et al (2006) MDCT-Based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies, IEEE Transactions on Medical Imaging vol. 25, nº 4, 2006, pp 464-474
5. Uchiyama Y, Katsuragawa S et al (2003) Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. Med Phys 30:2440-2454
6. Sluimer I, Prokop M et al (2006) Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung. Med. Phys 33:2610-2620
7. Zrimec T, Busayarat S (2007) A System for Computer Aided Detection of Diseases Patterns in High Resolution CT images of the Lungs, Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007, pp 41-46
8. Sluimer I, Schilham A, et al. (2006) Computer Analysis of Computed Tomography Scans of the Lung: A Survey, IEEE Transactions on Medical Imaging vol.25(4), 2006, pp 385-405
9. Webb W R, Muller N L, Naidich D P (1993) Standardized terms for high-resolution computed tomography of the lung: a proposed glossary. J Thoracic Imaging 8:167–185
10. Haralick R M (1979), Textural Features for Image Classification, Proc. of the IEEE vol.67(5), 1979, pp 786-804
11. Albregtsen F (1995) Statistical texture measures computed from gray level run-length matrices. University of Oslo
12. Galloway M M (1975) Texture analysis using gray level run lengths. Comput. Graphics Image Process 4:172–179
13. Chu A, Sehgal C M, Greenleaf J F (1990) Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Lett 11:415–420
14. Dasarathy B R, Holder E B (1991) Image characterizations based on joint gray-level run-length distributions. Pattern Recognition Lett 12:497–502
15. Vapnik V (1995) The Nature of Statistical Learning Theory. New York, Springer-Verlag
16. Kim H, Howland P, Park H (2006) Dimension Reduction in Text Classification with Support Vector Machines. J Machine Learning Research 6:37–53

Author:　Verónica Maria Vasconcelos
Institute: Instituto Superior de Engenharia de Coimbra
Street:　　Rua Pedro Nunes – Quinta da Nora
City:　　　Coimbra
Country: Portugal
Email:　　veronica@isec.pt