# Computer-aided diagnostic for interstitial lung diseases in HRCT: the TALISMAN project

Adrien Depeursinge[a], Jimison Iavindrasana[a], Herizo Andriambololoniaina[a], Pierre-Alexandre Poletti[b], Alexandra Platon[b], Antoine Geissbuhler[a] and Henning Müller[a,c]

[a] *Service of Medical Informatics, Geneva University Hospitals and University of Geneva (HUG), Switzerland*
[b] *Service of Emergency Radiology, HUG, Switzerland*
[c] *Business Information Systems, University of Applied Sciences Sierre, Switzerland*

**Abstract.**
**In this paper, we describe the goals and the latest outcomes of the TALISMAN project which aims to carry out image-based diagnostic aid for interstitial lung diseases (ILDs) with secondary data integration. Prototypes of the computer tools are implemented. High correct classification rates of lung tissue regions in high-resolution computed tomography (HRCT) based on a high-quality dataset built from clinical routine suggests that the computerized analysis of HRCT image with integration of the clinical context is ready to be used for computer-aided diagnosis of ILDs. As future work, implementation of multimodal retrieval of ILD cases and the clinical evaluation of the software are planned in order to cope with clinical needs.**
**Keywords:** high-resolution lung CT, interstitial lung diseases, CT library, computer-assisted image analysis, pattern recognition.

## 1. Introduction

Interstitial lung diseases (ILDs) are a heterogeneous group of around 150 illnesses of which many forms are rare and thus many radiologists have little experience with. Many of the diseases are rare and establishing the differential diagnosis for ILDs is considered difficult [1]. The diagnosis of ILDs is often established through the collaborations of the clinicians, radiologists, and pathologists. Images play an important role and patients may not require surgical lung biopsy when the clinical and radiographic impression is consistent with a safe diagnosis of ILDs [2]. The first imaging examination used is the chest radiograph because of its low cost and weak radiation dose. It also provides a quick overview of the whole chest. When the chest x–ray does not carry enough elements to finalize the diagnosis, high–resolution computed tomography (HRCT) is used to provide an accurate assessment of lung tissue patterns [3]. Indeed, original three–dimensional HRCT data avoid superposition of anatomic organs and provide an accurate assessment of the pattern and distribution of the lung tissue. It becomes an increasingly important method for the diagnosis of diffuse pulmonary parenchymal diseases. However, interpreting HRCT images of the chest showing patterns associated with ILDs is time–consuming and requires experience. The three–dimensional form requires significant reading time, effort, and experience for a correct interpretation [4]. Moreover, the context is fundamental for correct interpretation: healthy tissue, for example, may have different visual aspects depending on the age or the smoking history of the patient. Computerized HRCT analysis with integration of the clinical context can provide quick and precise information for emergency radiologists and other non–chest specialists [5,6]. The computer–aided detection (CAD) system should be used as first reader in order to improve the radiologist's productivity and reduce reading fatigue [7,8]. One approach for building image–based computerized diagnostic aid for ILDs is to imitate the radiologists' human vision system. The latter can be schematized into two main parts. First, the eyes act as captors and aim at extracting relevant features from the observed scene [9]. Then, the cerebral cortex takes decisions based on the pre-processed information provided by the eyes

and visual cortex as input, as well as the clinical context of the observed radiological image, and the knowledge and experience of the radiologist. In pattern recognition, these two tasks can be respectively identified as visual feature extraction (involving image processing) and supervised machine learning for decision-making (i.e. classification of image regions).

In this paper, the latest outcomes of the TALISMAN (Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce) project which aims at carrying out image-based diagnostic aid for ILDs with secondary data integration are detailed.

## 2. Methods

The methods for building the image–based diagnostic aid tool are constituted by several connected subtasks. A first step was to specify the scope of the diagnostic aid tool, determined by the selected subset of interstitial lung diseases included in the study. In collaboration with the pneumology and emergency radiology of the University Hospitals of Geneva (HUG), 15 diseases that are described as the most frequent causes of lung parenchymal disorders were selected. According to these 15 diseases, a library of annotated pulmonary CT cases with clinical data integration was created (see Section 3.2). In Sections 3.3 and 3.4, the steps for carrying out detection and classification of lung tissue patterns in HRCT data with integration of the clinical context of the image series are detailed.

## 3. Results

### 3.1. Global description

As ILDs are gathering more than 150 illnesses of the interstitial lung tissue, which are in most of the cases comorbid, building a computerized system to provide a final diagnostic with high precision is not conceivable. A computerized diagnostic aid built on content–based image retrieval (CBIR) along with secondary data integration such as the relevant clinical parameters related to ILDs can bring quick and precious information to less experienced radiologists and non–chest experts. The information system provides results in two steps (see Figure 1). First, the suspicious (abnormal) patterns in the new, non–interpreted HRCT are automatically detected and highlighted with a proposed tentative class of lung tissue disorders. Then, the relevant patterns along with secondary data (i.e. age, smoking history, laboratory tests, etc…) can be used as query for automatic retrieval of similar cases from an associated multimedia database of typical HRCT scans accompanied by corresponding clinical parameters of the patient [8]. The radiologist has to consider the system as a second opinion in addition to the experience gained with similar cases for providing a differential diagnosis. This differs distinctly from automated diagnosis aiming at replacing radiologists by computers [4].

A first prototype of a DICOM viewer for the detection and classification of image regions in HRCT data is implemented in Java (see Figure 2). The radiologist can browse the three-dimensional stack of images and delineate regions of interest (ROIs) of pulmonary tissue to be classified. Screenshots of the web-based prototype interface for the retrieval of cases in the multimedia library of ILD cases are shown in Figure 3.
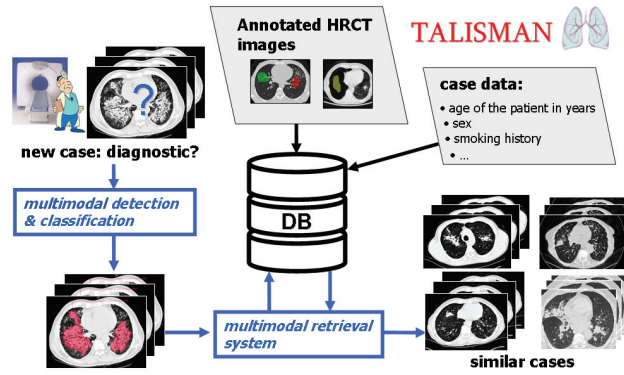
*Figure 1: Global block diagram of the diagnostic aid system for ILDs. The user can submit an unknown HRCT image to be analyzed and then retrieve visually and/or clinically similar cases from the multimedia database.*
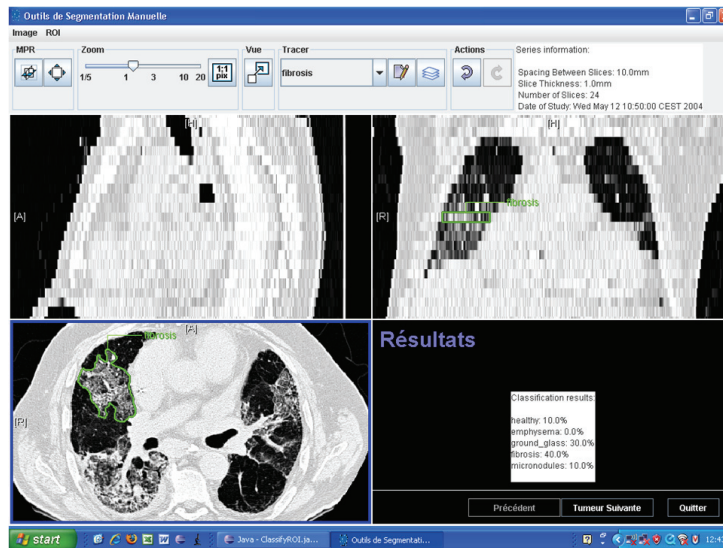


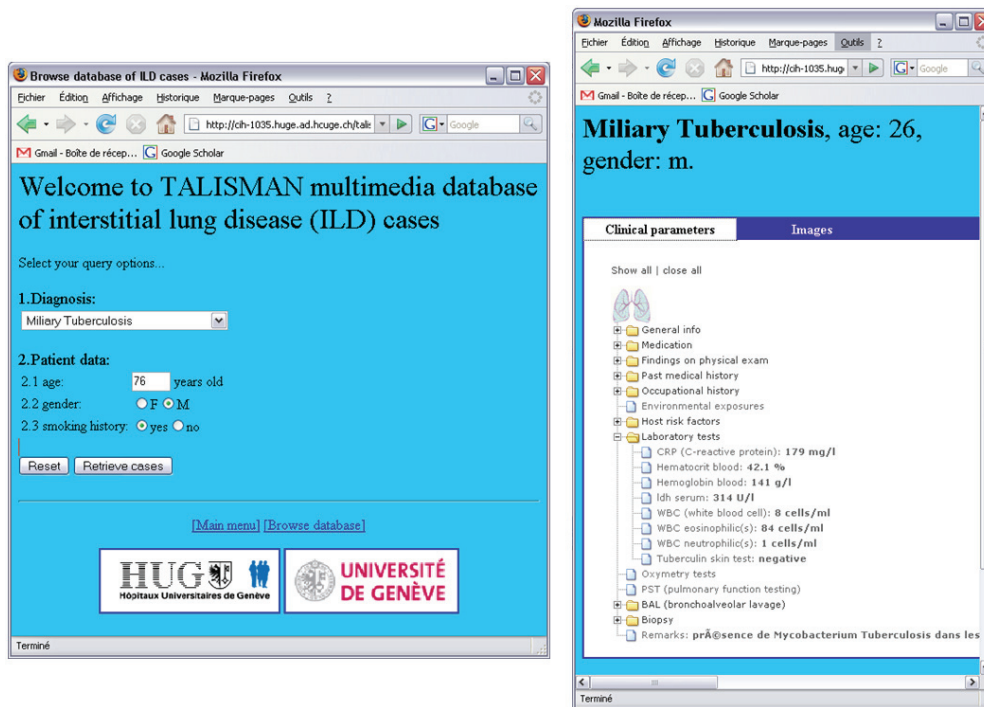*Figure 2: A screenshot of the DICOM viewer for the classification of image regions.*

*Figure 3: Web-based interface for the retrieval of ILD cases in the multimedia database.*

### 3.2. Multimedia library of ILD cases

Building a high-quality multimedia database is an essential step for carrying out computerized diagnostic aid for ILDs. Indeed, HRCT image series of control cases are required to train and test algorithms for detection and classification of image regions. Moreover, the database provides well-documented cases for content-based and multimodal retrieval of similar cases. The library itself can be used for teaching purposes as well. The building of the database of cases was organized as follows: Based on each of the 15 most frequent pathologies, the 99 most discriminative clinical parameters for the establishment of the differential diagnostic were kept. With the approval of the ethics commission, cases with HRCT images linked to one of the 15 ILDs were retrospectively collected in the database. In order to establish ground truth for lung tissue classification, experienced radiologists delineated typical pathologic regions in HRCT scans using custom-built software. The image series to be annotated were selected from cases with certified diagnoses and with slice thickness no larger than 1mm without contrast product. After 20 months of acquisition, 103 cases were captured (with a goal of 150 by the end of the project), a mean 65% of the clinical parameters were filled for each case and 1163 ROIs were drawn, which constitutes 6 classes of lung tissue well represented, including healthy tissue.

### 3.3. Image processing: visual feature extraction

In order to extract measures from HRCT images as discriminating features for classification of lung tissue, several steps of image processing are carried out. Image texture features are used as the taxonomy used by radiologists to interpret patterns in HRCT images often relating to texture properties. Two feature groups were used. First, pixel values are characterized using grey–level histograms. Full resolution (12–bit grey values) HRCT images are containing values in Hounsfield Units (H.U.) in the interval [-1500; 1500]. These values are corresponding univoquely to densities of the anatomic organs and thus allow the identification of lung tissue components. Secondly, the spatial organization of the pixels is studied through the coefficients of isotropic polyharmonic B-spline wavelet frames using a quincunx subsampling scheme. More details can be found in [11,12].

*3.4. Supervised machine learning: contextual image analysis*

Classification algorithms are required to find the boundaries among the distinct classes of lung tissue represented in the feature space. Five common classifier families with optimized parameters were compared in their ability to categorize six lung tissue patterns in [13]. Using a solid methodology for training and testing the classifiers, support vector machines (SVMs) with isotropic polyharmonic B-spline wavelet frames allowed a mean of 94.3% correct matches of 843 ROIs among the six lung tissue classes with high class-specific precision. The effect of the integration of the clinical context of HRCT images on tissue classification is studied in [14]. Along with appropriate selection of relevant clinical parameters, contextual image analysis allowed significant improvements in classification performances compared to a pure visually-based categorization. The integration of the clinical parameters in the feature vector has to be carried out with precautions as bad synergies between visual and clinical attributes can occur.

## 4. Discussion and Conclusions

A high-quality multimedia database of ILD cases is now available for training and testing the classification and retrieval algorithms as well as for teaching purposes. However, although containing ROIs of healthy tissue, it does not yet include healthy cases with a full set of "healthy" clinical parameters which is required in order to establish "normality" of the values. The pattern recognition task is challenging as it is partly based on experience of the radiologists. Nevertheless, the high rate of correct predictions of the lung tissue classes and the improvements gained by the integration of the clinical context of images suggest that the visual feature set with optimized SVMs is ready to be used for computer-aided diagnosis of ILDs. Indeed, classification accuracy values are trustworthy for further usage in clinical routine as the six classes of lung tissue pattern tested are realistically distributed and allow for diagnosing a wide variety of ILDs [3].

## 5. Future Work

As future work, the collection of ILD cases at the HUG (Hôpitaux Universitaires de Genève) will continue in order to increase the statistical validity of our evaluation results. The goal is to have at least 150 fully annotated cases with images and normal/abnormal ROIs marked by the end of the project. More investigations are required for the integration of the clinical context of HRCT images for the classification of ROIs in order to avoid bad synergies between clinical and visual attributes. A next step is to implement multimodal retrieval of the ILD cases along with three-dimensional characterization of textures to allow for automatic detection of pathological tissues in the entire HRCT volume. Finally, improvements and clinical evaluation of the user interface are planned for an optimal adaptation of the computer tools in clinical routine.

## 6. Acknowledgements

## 7. References

[1]   T E King, Approach to the adult with interstitial lung disease, Uptodate, August 2004.
[2]   K R Flaherty, T E King, J Ganesh Raghu, J P Lynch III, T V Colby, W D Travis, B H Gross, E A Kazerooni, G B Toews, Q Long, S Murray, V N Lama, S E Gay, and F J Martinez, Idiopathic interstitial pneumonia: What is the effect of a multidisciplinary approach to diagnosis?, American Journal of Respiratory and Critical Care Medicine 170, pp. 904–910, July 2004.
[3]   P Stark, High resolution computed tomography of the lungs, Uptodate, September 2007.
[4]   K Doi, Current status and future potential of computer–aided diagnosis in medical imaging, British Journal of Radiology 78, pp. 3–19, 2005.

[5]    A M Aisen, L S Broderick, H Winer-Muram, C E Brodley, A C Kak, C Pavlopoulou, J Dy, C R Shyu, A Marchiori, Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment, Radiology, 228, pp. 265-270, 2003.

[6]    C R Shyu, C Brodley, A Kak, A Kosaka, ASSERT: A Physician-in-the-loop content-based retrieval system for HRCT image databases, Computer Vision and Image Understanding 75 (1-2), pp. 111-132, 1999.

[7]    R M Nishikawa, Current status and future directions of computer-aided diagnosis in mammography, Computerized Medical Imaging and Graphics 31, pp. 224–235, June 2007.

[8]    H Müller, N Michoux, D Bandon, A Geissbuhler, A review of content-based image retrieval systems in medicine – clinical benefits and future directions, International Journal of Medical Informatics, 73, pp 1-23, 2004.

[9]    I Biedermann, Recognition–by–components: A theory of human image understanding, Psychological Review 94 No 2, pp. 115–147, 1987.

[10]   G Sasso, H R Marsiglia, F Pigatto, A Basilicata, M Gargiulo, A F Abate, M Nappi, J Pulley, and F S Sasso, A visual query–by–example image database for chest CT images: Potential role as a decision and educational support tool for radiologists, Journal of Digital Imaging 18, pp. 78–84, March 2005.

[11]   A Depeursinge, D Sage, A Hidki, A Platon, P-A Poletti, M Unser, and H Müller, Lung tissue classification using wavelet frames, in Engineering in Medicine and Biology Society, 2007, EMBS 2007, 29th Annual International Conference of the IEEE, (Lyon, France), August 2007.

[12]   A Depeursinge, D Van De Ville, M Unser and H Müller, Lung tissue analysis using isotropic polyharmonic B-spline wavelets, *submitted to* Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008.

[13]   A Depeursinge, J Iavindrasana, A Hidki, G Cohen, A Geissbuhler, A Platon, P-A Poletti and H Müller, A classification framework for lung tissue categorization. In *SPIE Medical Imaging*, San Diego, CA, USA, February 2008 - *to appear*.

[14]   A Depeursinge, J Iavindrasana, G Cohen, A Platon, P-A Poletti and H Müller, Lung tissue classification in HRCT data integrating the clinical context, in 21st IEEE Symposium on Computer-Based Medical Systems, CBMS 2008, Jyväskylä, Finland, June 2008 - *to appear*.

## 8. Address for correspondence

Adrien Depeursinge
Geneva University Hospitals and University of Geneva,
Service of Medical Informatics
24, rue Micheli-du-Crest,
CH-1211 Geneva 14, Switzerland
adrien.depeursinge@sim.hcuge.ch, http://www.sim.hcuge.ch/medgift/
Tel ++ 41 22 372 8875, fax ++ 41 22 372 8879