A framework for diagnosis of 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], 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*

Summary

In this paper we describe the goals and the latest outcomes of the TALISMAN project, the purpose of which is to furnish image-based diagnostic assistance 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 computerised analysis of HRCT images with integration of the clinical context is ready to be used for computer-aided diagnosis of ILDs. As future tasks, implementation of multimodal retrieval of ILD cases and clinical evaluation of the software are planned to meet clinical needs.

Introduction

Interstitial lung diseases (ILDs) are a heterogeneous group of some 150 diseases, many forms of which are rare and thus fall largely outside the experience of many radiologists. Hence differential diagnosis of ILDs is considered difficult [1]. ILDs are often diagnosed on a collaborative basis between clinicians, radiologists and pathologists. Images play an important role and patients may not require surgical lung biopsy if the clinical and radiographic impression is consistent with safe diagnosis of ILDs [2]. The first imaging procedure used is the chest radiograph, because of its low cost and weak radiation dose. It also provides a rapid overview of the whole chest. Where chest x–ray does not carry enough elements to finalise the diagnosis, high-resolution computed tomography (HRCT) provides an accurate assessment of lung tissue patterns [3]. Original three-dimensional HRCT data avoid superposition of anatomical organs and provide an accurate assessment of the pattern and distribution of lung tissue. It is becoming 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. Correct interpretation of the three-dimensional form requires significant reading time, effort and experience [4]. Moreover, the context is fundamental for correct interpretation: healthy tissue, for example, may have different visual aspects depending on the patient's age or smoking history. Computerised HRCT analysis with integration of the clinical context can provide rapid and invaluable information for emergency radiologists and other non-chest specialists [5, 6]. The computer-aided detection (CAD) system should be used as first reader to improve the radiologist's productivity and reduce reading fatigue [7, 8].
One approach to building image-based computerised diagnostic aid for ILDs is to imitate the radiologists' human vision system. The latter can be schematised into two main parts. First, the eyes act as captors to extract relevant features from the observed scene [9]. The cerebral cortex then takes decisions based on the preprocessed 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 (ie classification of image regions).
This paper details the latest outcomes of the TALISMAN (Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce) project, whose purpose is to provide image-based diagnostic assistance for ILDs with secondary data integration.

Methods

The methods of building the image-based diagnostic aid tool consist of 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 departments of the University Hospitals of Geneva (HUG), 15 diseases described as the most frequent causes of lung parenchymal disorders were selected. In relation to these 15 diseases, a library of annotated pulmonary CT cases with clinical data integration

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/

was created (see Section 3.2). Sections 3.3 and 3.4 detail the steps for detection and classification of lung tissue patterns in HRCT data, with integration of the clinical context of the image series.

Results

Global description

Since ILDs comprise more than 150 illnesses of interstitial lung tissue which in most cases are comorbid, building a computerised system to provide a final diagnosis with high precision is not conceivable. A computerised diagnostic aid built on content-based image retrieval (CBIR), along with secondary data integration such as the relevant clinical parameters related to ILDs, can provide less experienced radiologists and non-chest experts with rapid and invaluable information. The information system provides results in two steps (see fig. 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. The relevant patterns along with secondary data (ie age, smoking history, laboratory tests, etc) can then be used as query for automatic retrieval of similar cases from an associated multimedia database of typical HRCT scans accompanied by the patient's corresponding clinical parameters [8]. The radiologist must regard the system as a second opinion in addition to the experience gained with similar cases for the purpose of differential diagnosis.

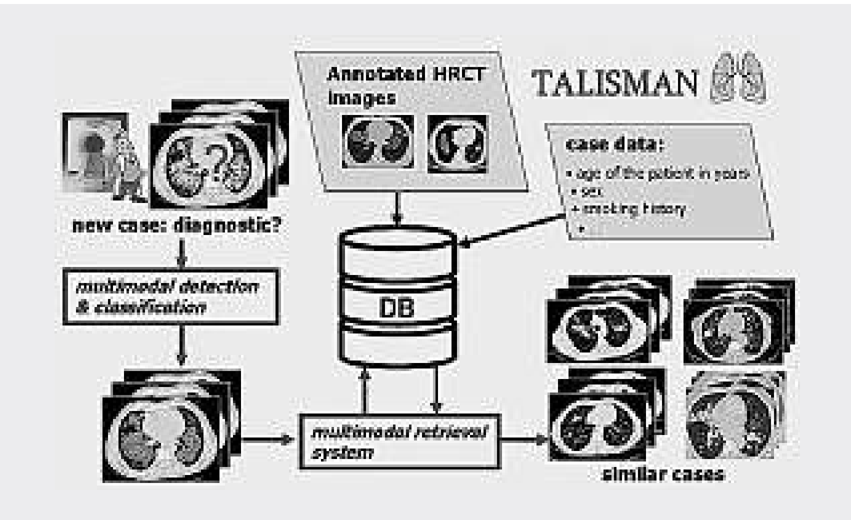

Figure 1. *Global block diagram of the diagnostic aid system for ILDs. The user can submit an unknown HRCT image for analysis and then retrieve visually and/or clinically similar cases from the multimedia database.*

This differs radically from automated diagnosis aimed 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 fig. 2). The radiologist can browse the three-dimensional stack of images and delineate regions of interest (ROIs) in 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.

Multimedia library of ILD cases

Building a high-quality multimedia database is an essential step in implementing a computerised diagnostic aid for ILDs, HRCT image series from control cases being 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 also be used for teaching purposes. Building of the database of cases was organised as follows: taking each of the 15 most frequent pathologies as a basis, the 99 most discriminative clinical parameters for differential diagnosis 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. To establish ground truth for lung tissue classification, experienced radiologists delineated typical pathological 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 1 mm without contrast medium. 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 1,163 ROIs were drawn, constituting six classes of well represented lung tissue, including healthy tissue.

Image processing: visual feature extraction

To extract measures from HRCT images as discriminating features for classification of lung tissue, several image processing steps are performed. 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 characterised using grey-level histograms. Full resolution (12-bit grey values) HRCT images contain values in Hounsfield

Units (H.U.) in the interval [-1500; 1500]. These values correspond unequivocally to densities of the anatomical organs and thus allow identification of lung tissue components. Secondly, the spatial organisation of the pixels is studied through the coefficients of isotropic polyharmonic B-spline wavelet frames using a quincunx subsampling scheme. [11, 12 for more detail].

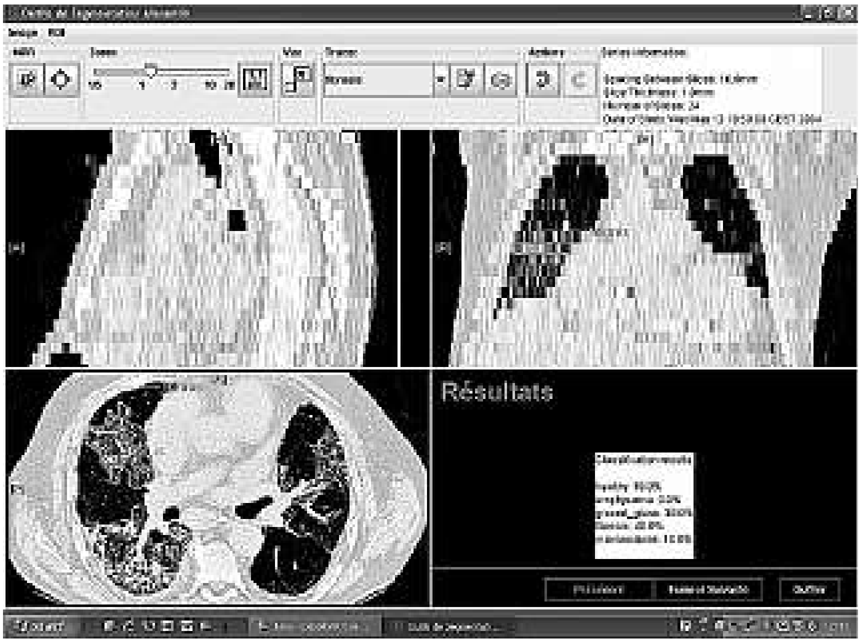

Figure 2. *A screenshot of the DICOM viewer for the classification of image regions.*

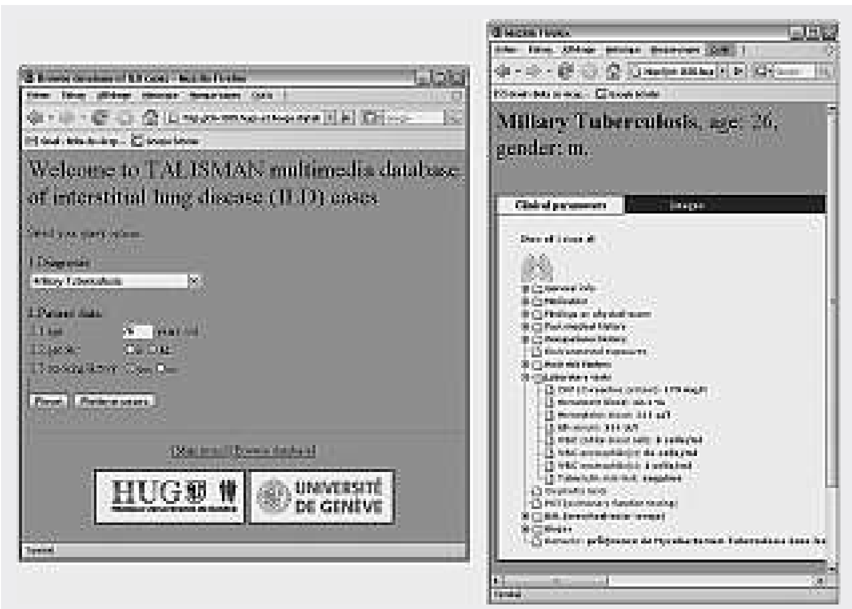

Figure 3. *Web-based interface for the retrieval of ILD cases in the multimedia database.*

Supervised machine learning: contextual image analysis

Classification algorithms are required to find the boundaries between the distinct classes of lung tissue represented in the feature space. Five common classifier families with optimised parameters were compared for their ability to categorise six lung tissue patterns [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 integrating 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 purely visually-based categorisation. Integration of the clinical parameters in the feature vector must be performed with caution, since bad synergies between visual and clinical attributes may occur.

Discussion and conclusions

A high-quality multimedia database of ILD cases is now available to train and test 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, as is required in order to establish "normality" of the values. The pattern recognition task is challenging, being partly based on the radiologists' experience. Nevertheless, the high rate of correct prediction of the lung tissue classes and the improvements made by integration of the images' clinical context suggest that the visual feature set with optimised SVMs is ready to be used for computer-aided diagnosis of ILDs. Classification accuracy values are trustworthy for further usage in clinical routine, since the six classes of lung tissue pattern tested are realistically distributed and make it possible to diagnose a wide variety of ILDs [3].

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 to integrate the clinical context of HRCT images for the classification of ROIs, 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 characterisation of textures to allow for automatic detection of pathological tissues in the entire HRCT volume. Finally, improvement and clinical evaluation of the user interface are planned for optimal adaptation of the computer tools in clinical routine.

Acknowledgements

This work was supported by the Swiss National Science Foundation (FNS) with grant 200020–118638/1, the equalisation fund of Geneva University Hospitals and the University of Geneva (grant 05-9-II) and a grant from the Swiss Confederation for the work of Dr. Herizo Andriambololoniaina.

References

1 King TE. Approach to the adult with interstitial lung disease, Uptodate, August 2004.
2 Flaherty KR, King TE, Ganesh Raghu J, Lynch III JP, Colby TV, Travis WD, et al. Idiopathic interstitial pneumonia: What is the effect of a multidisciplinary approach to diagnosis? Am Resp Crit Care Med. 2004;170:904–10.
3 Stark P. High resolution computed tomography of the lungs, Uptodate, September 2007.
4 Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol. 2005;78: 3–19.
5 Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment. Radiology. 2003;228:265–70.
6 Shyu CR, Brodley C, Kak A, Kosaka A. ASSERT: A Physician-in-the-loop content-based retrieval system for HRCT image databases. Computer Vision and Image Understanding 1999;75(1-2):111–32.
7 Nishikawa RM. Current status and future directions of computer-aided diagnosis in mammography. Comput Med Imaging Graph. 2007;31:224–35.
8 Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medicine – clinical benefits and future directions. Int J Med Inform. 2004;73:1–23.
9 Biedermann I. Recognition-by-components: A theory of human image understanding. Psychol Rev. 1987;94(No 2):115–47.
10 Sasso G, Marsiglia HR, Pigatto F, Basilicata A, Gargiulo M, Abate AF, et al. A visual query-by-example image database for chest CT images: Potential role as a decision and educational support tool for radiologists. J Digit Imaging. 2005;18:78–84.
11 Depeursinge A, Sage D, Hidki A, Platon A, Poletti P-A, Unser M, Müller H. 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 Depeursinge A, Van De Ville D, Unser M, Müller H. Lung tissue analysis using isotropic polyharmonic B-spline wavelets, *submitted to* Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008.
13 Depeursinge A, Iavindrasana J, Hidki A, Cohen G, Geissbuhler A, Platon A, et al. A classification framework for lung tissue categorization. In *SPIE Medical Imaging*, San Diego, CA, USA, February 2008 – *to appear*.
14 Depeursinge A, Iavindrasana J, Cohen G, Platon A, Poletti P-A, Müller H. 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*.