0040

Content-based retrieval and analysis of HRCT images from patients with interstitial lung diseases: a comprehensive diagnostic aid framework

A. Depeursinge[1], A. Vargas[1], F. Gaillard[2], A. Platon[3], A. Geissbuhler[1], P.-A. Poletti[3], H. Müller[1,2]
[1]University Hospitals of Geneva, Medical Informatics Service, Geneva 14, Switzerland
[2]University of Applied Sciences Sierre, Business Information Systems, Sierre, Switzerland
[3]University Hospitals of Geneva, Service of Emergency Radiology, Geneva 14, Switzerland

Keywords: texture analysis, computer-aided diagnosis, content-based image retrieval, interstitial lung diseases, high-resolution CT

Purpose: The interpretation of high-resolution computed tomography (HRCT) of the chest from patients with interstitial lung diseases (ILD) is often challenging with numerous differential diagnoses. We created a multimedia database of ILD cases and developed an image analysis tool to assist the radiologist to the diagnosis workup of ILDs. The database constituted the basis for developing automatic detection and characterization of pulmonary tissue in HRCT images as well as content-based retrieval of similar cases. The automated categorization of lung tissue patterns and retrieval of similar cases are found to be complementary both on the algorithmic and on the user side. This combination yields a hybrid detection-retrieval-based computer-aided diagnosis (CAD) system. In a first step, the automated detection and categorization of the pathological lung tissue sorts reduces the risk of omission and ensures the reproducibility of the diagnosis by drawing the radiologists' attention on diagnostically interesting events in medical image series. Then, case-based retrieval delivers quick and precious information for diagnosis aid and treatment planning through examples of cases with confirmed diagnosis.

Methods: Cases associated with 13 frequent diagnoses of ILDs (see Figure 1) that underwent an HRCT examination were retrospectively collected at the University Hospitals of Geneva. Based on each histological diagnosis, the most discriminative clinical parameters for the establishment of the differential diagnostic were kept, resulting in 159 clinical attributes. For each case having a biopsy-based or equivalent (e.g. tuberculin skin test for TB, Kveim test for Sarcoidosis, ...) proven diagnosis, the HRCT image series were annotated by two radiologists with more than 15 years of experience. Then, the 159 clinical attributes were filled based on their availability in the electronic health record (EHR) within a time interval of two weeks around the date of the HRCT image series.

Blockwise affine-invariant texture analysis based on tailored wavelet transforms and support vector machines was used for the categorization of the lung tissue in HRCT. The considered classes of lung tissue are *healthy*, *emphysema*, *ground glass*, *fibrosis* and *micronodules* and were selected as being the most represented lung tissue sorts in the database as well as being encountered in the 13 histological diagnoses listed in Figure 1. The segmentation results for the five classes are color-coded to constitute a three-dimensional map of the lung tissue sorts as diagnostic aid (see Figure 2).

Based on the output of the detection-based CAD, a multimodal similarity measure is introduced to enable case-based retrieval, yielding a hybrid detection-retrieval-based CAD system. The multimodal inter-case measure used is a weighted Euclidean distance that combines clinical parameters of two importance levels as well as the respective volumes of the five lung tissue sorts.

Results: 128 cases were captured from which 108 have HRCT image series annotated with clinical data (see Figure 1). 1946 regions of interest (ROIs) were delineated manually by expert radiologists resulting in a total volume of 41.65 liters of normal and pathological lung tissue as reference data.

The five lung tissue sorts described above are represented by ROIs with a total volume of 35.69 litres.

In order to evaluate the performance of the texture analysis framework for classifying the lung tissue regions, a leave-one-patient-out (LOPO) cross-validation of 69 image series from patients affected with ILDs was performed. Global arithmetic and geometric means of the accuracies of 75.1% and 74.7% are obtained respectively. Confusions between *healthy* and *micronodules* patterns are observed as some of the bronchovascular structures are mixed with small nodules and inversely. The case-based retrieval precision was evaluated based on the diagnosis of the retrieved cases using the 7 most represented histological diagnoses. A mean average retrieval precision at rank 1 of 71% is obtained with a LOPO cross-validation with 69 cases.

Graphical user interfaces were built for database browsing, 3D categorization of the lung tissue as well as case-based retrieval to bring the full potential of the proposed CAD system to the end-users (see Figure 2). The 3D segmented regions are displayed to the clinician using a modified version of YaDiV, where tabs were added for the lung tissue categorization.

Conclusion: Image-based diagnostic aid tools for HRCT images of the chest of patients affected with ILDs are available. As images are now accessible to each actor of the medical staff through the EHR, computerized diagnostic aid tools are becoming important to provide quick and accurate information to non-specialists as well. Besides providing ground truth for the evaluation of the hybrid CAD system, the multimedia database of ILD cases can be used as a reference library for teaching purposes. The evaluation of both lung tissue categorization and case-based retrieval based on LOPO cross-validation is representative to actual clinical situations where the CAD system is beforehand trained with a given set of cases and the observed performance is assessed on the patient to be treated. A validation of the CAD system in a clinical environment is necessary to assess the performance of the radiologists with and without the system in order to investigate the clinical benefits of the proposed methods.

fig.1

histological diagnosis	patients	image series	age (mean ± std)	female (%)
Healthy	2	2	63.5 ± 9.5	100
Pulmonary fibrosis (PF)	40	39	71.4 ± 13.4	50
Hypersensitivity pneumonitis (HP)	24	19	65.3 ± 17.1	80.3
Tuberculosis (TB)	15	12	41.1 ± 17.6	40
Pneumocystis pneumonia (PCP)	8	4	59.6 ± 20.4	12.5
Cryptogenic organizing pneumonia (COP/BOOP)	8	3	45 ± 24.6	50
Eosinophilic pneumonia (EP)	1	1	33	100
Sarcoidosis	20	18	48.5 ± 17.3	30
Acute interstitial pneumonia (AIP)	4	4	65.3 ± 5.5	50
Desquamative interstitial pneumonia (DIP)	1	1	46	100
Respiratory bronchiolitis associated ILD (RB–ILD)	1	1	54	100
Non–specific interstitial pneumonia (NSIP)	2	2	61.5 ± 12.5	50
Langerhans cell histiocytosis (LCH)	1	1	24	0
Lymphocytic interstitial pneumonia (LIP)	1	1	32	0
total	128	108	59 ± 20.2	36.7

Figure 1. Distribution, mean age and gender statistics of the diagnoses of the case library.

fig.2

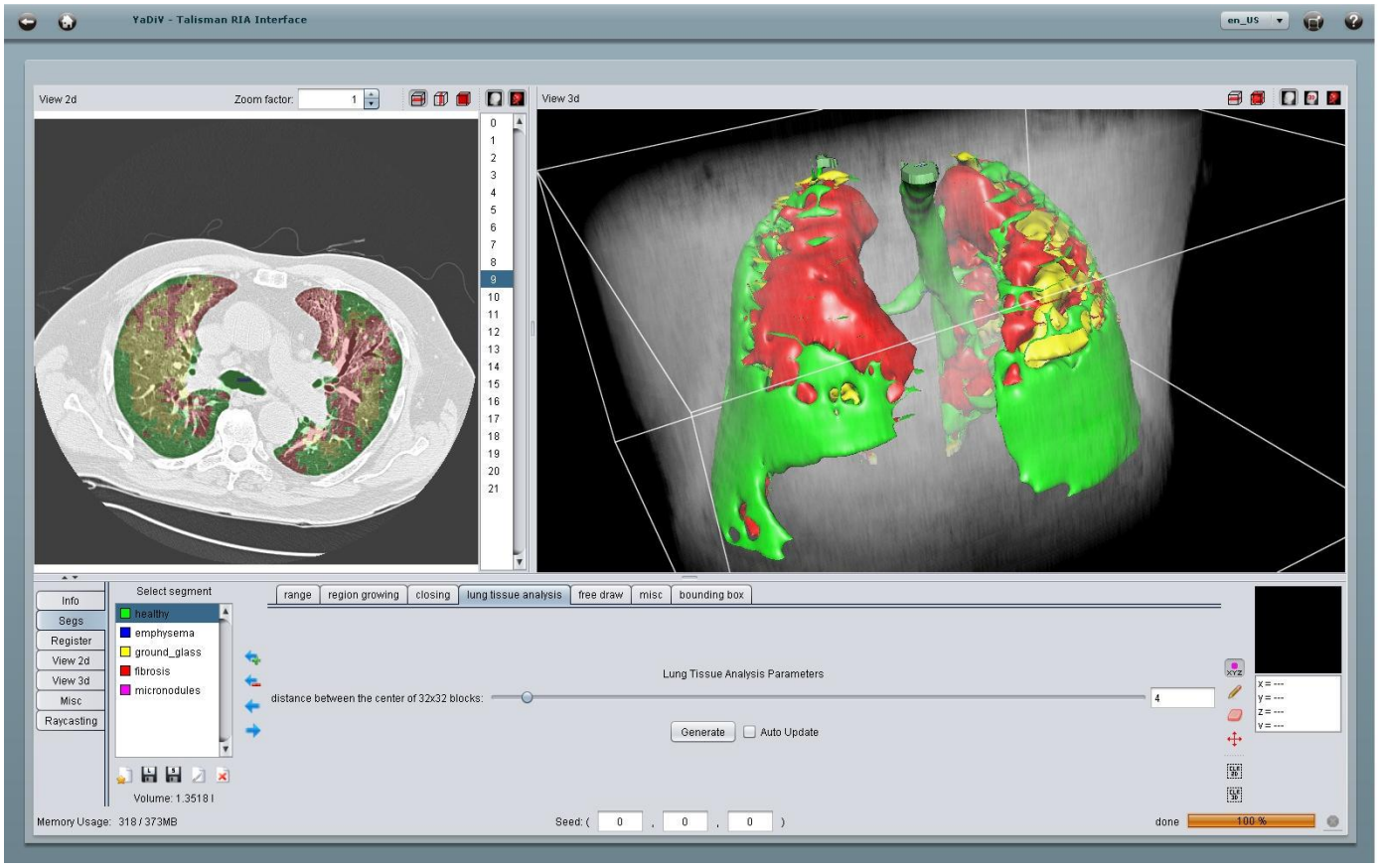

Figure 2. Segmentation of the lung tissue using YaDiV. A "lung tissue analysis" tab was created to run the blockwise feature extraction and classification.