Putting the image into perspective: The need for domain knowledge when performing image-based diagnostic aid

Asmaa Hidki[1], Henning Müller[1], Adrien Depeursinge[1], Pierre-Alexandre Poletti[2],

Christian Lovis[1], Antoine Geissbuhler[1]

[1]Service for medical informatics, University and Hospitals of Geneva (HUG

[2]Emergency Radiology, HUG

Abstract:

The term interstitial lung disease includes around 150 pathologies of multiple and often idiopathic etiologies. The interstitial syndrome is of complex radiological interpretation and differential diagnosis is often difficult. Some criteria such as age, gender or environmental exposures have a high importance depending on the disease, and need to be collected for every case. Most common imaging method is the chest x-ray but for more complex cases computed tomographies (CT) of the lung are performed that contain more information on the lung tissue. Some projects on diagnostic aid based on lung CTs have been proposed and shown to improve diagnostics but often without taking into account other clinical data on the patient.
The goal of the project described in this paper is the creation of a database of cases containing images and other metadata on the patient. This paper describes the definition of these metadata necessary for a diagnostic aid system for interstitial lung disease. The most important data for 15 frequent lung diseases are discussed and these data will subsequently be taken into account for the creation of a digital library of lung CTs destined for diagnostic aid and teaching.

1. Introduction

The term interstitial lung disease groups more than 150 pathologies of which the clinical symptoms are often little specific. Parameters such as age and the environmental factors have to be taken into account to reach a diagnosis [4]. The chest radiography is the first imaging method used for investigation because it is inexpensive and low in radiation dose. On the other hand this method is little sensitive for many pulmonary lesions and a High Resolution Computed Tomography (HRCT) imposes itself as a complement. This method is more expensive but contains more of information. Due to the variety of interstitial lung diseases (ILD) many other parameters need to be taken into account. Some of the pathologies are very rare and radiologists who are not chest specialist can often refer to little experience. The consequence is that the diagnostic precision on lung HRCT is often not optimal. This precision can be enhanced by giving access to similar cases; reference books are frequently used, but the search for good information can take long. The creation of a database of reference CT images that are easily accessible permits to reduce the time of analysis and enhance diagnostic quality. Some databases already exist for the processing and the analysis of pulmonary nodules, which refer to annotated reference databases. However, no database of this type exists at the moment for interstitial lung illnesses [3, 4]. Image analysis based on visual similarity is in general a very active research domain. In the medical domain, several research groups work on the image analysis and indexing. In pneumology, these investigations are promising but difficult [1].
The goal of the project described in this paper is the creation of a database containing lung CT images and demographic data on the patients, which is complementary and indispensable for diagnostics. The dataset to be created will contain regions in images selected to well represent certain pathologies. The necessary tools to get the information directly from the patient record and the image archive will be developed. An annotation tool to mark regions of interests will also be developed. On the basis of the literature, a radiologist will establish precisely what the most applicable data are than the system should research. Objective is the accessibility of a reference database with marked regions containing representative examples for several ILD. It will allow the radiologist, and particularly the emergency radiologist to use an electronic system to find similar cases rather than looking them up in a book. The project will help diagnostics by allowing to search for similar cases by submitting a current case under observation.

2. Methods:

The project contains three parts building upon each other and on complementary knowledge:

- Part 1: Literature research to get a list as exhaustive as possible of the important data permitting to characterize each of the interstitial pulmonary diseases other than images. For our research we will collect this data and eliminate all data permitting to identify the patient. The goal is to constitute a diagnostic help for the radiologist and a tool for teaching. Such a library will give access to a large number of rare cases, exceeding personal experience.
- Part 2: Development of computer tools to acquire image datasets in DICOM format allowing the annotation of regions in these images. Developments will be done in Java, allowing access within the hospital via a web browser. A database will be created to store the images, annotations and acquired metadata.
Part 3: Retrospective selection of cases from clinical routine. The cases will be annotated according to the criteria described in this article. The quality of chosen cases and relevance of region annotations are extremely important to assure a good final result.
In this article, we focus in the first part which consist in the extraction by exhaustive review of literature of the main criteria characterized each of the interstitial lung diseases [4].

2. Results

Below, we describe the main criteria permitting to characterize the most frequent interstitial pulmonary diseases, listed in the order of their frequency.

2.1. Most frequent diseases and data of influence

Miliary Tuberculosis affects 33% of the world population, occurs at the extremes of ages. In the EU 80% are over 50 years of age and in Africa 80% are under 50 years of age. Risk factors are human immunodeficiency virus (HIV) infection, immunosuppression (especially in association with solid organ transplantation), pregnancy, malignancy, diabetes, renal failure, alcohol abuse and malnutrition [6]. The tuberculin test can be a supportive test if positive, but a negative skin test should not be interpreted to exclude the diagnosis [6, 7].
Pneumocystis carinii pneumonia (PCP) officially renamed" Pneumocystis jiroveci pneumonia": Most common opportunistic infection (OI) in patients infected with the human immunodeficiency virus (HIV) with 30% of HIV-infected patients after 75 years of ages in the EU and in many cases reports black subjects had 1/3 the risk of PCP compared to white patients. PCP is frequently presenting as the first manifestation of the acquired immunodeficiency syndrome (AIDS) and is generally gradual in onset [8, 9]. HIV transmission category, age, smoking history, and use of antiretroviral therapy did not predict development of PCP [9].
Lymphangitic carcinomatosis: The most likely form of the disease to present with diffuse infiltrative lung disease representing 3 percent of all ILD. The most common sites of primary cancer are the stomach, breast, bronchus, pancreas, and prostate. It is often difficult to diagnose lymphangitic carcinomatosis because symptoms referable to the primary neoplasm are often absent [10]. Recognition of the suggestive history and chest radiographic findings can accomplish an early diagnosis, especially when the patient has a known underlying malignancy [10].
Pneumoconiosis, **Silicosis** is the most frequent of all the pneumoconiosis, its incidence has neatly decreased in France, it is a professional illness and affects male between 21-70 years of ages [11]. Silicosis can occur in many industries and work settings including mining, quarrying, sandblasting, masonry, founding, and ceramics [11].
Hypersensitivity pneumonitis (**HP**) also called "extrinsic allergic alveolitis". In France 20 percent of the population are infected but only a small proportion of exposed individuals develop clinically significant HP and genetic factors play a major role in determining an individual's risk of disease. Cigarette smoking is associated with a decreased risk of HP. Over 300 etiologies of HP have been reported [12] ex:Farming, vegetable, or dairy cattle workers-Ventilation and water-related contamination - Bird and poultry handling -Veterinary work and animal handling-Grain and flour processing and loading-Lumber milling, construction, wood stripping, paper and wallboard manufacture - Plastic manufacture, painting, electronics industry, other chemicals -Textile workers[12].
Drug-induced pulmonary toxicity like **Amiodarone i**s highly effective in suppressing ventricular and supraventricular tachyarrhythmias. Because of the importance of total cumulative dose pulmonary toxicity occurs several months to as late as several years after the initiation of amiodarone therapy [13].
Sarcoidosis: The disease appears to vary in incidence among geographical regions and can also aggregate in families and specific races. For example, blacks tend to be affected more acutely and with more severe disease than whites, who tend to present with asymptomatic and chronic disease. In addition, the patient's immunogenetic background may play a role in the clinical manifestations of sarcoidosis and could underlie the heterogeneity of the illness because significant heterogeneity in disease presentation and severity occurs among different ethnic racial groups [14].
Community-acquired pneumonia (CAP) in adults is a common and serious illness with considerable morbidity and mortality, as approximately 20 percent of episodes result in hospitalization [15]. There is

seasonal variation, with more cases occurring during the winter months. The rates of pneumonia are higher for men than for women and for black persons compared with Caucasians. In addition to microbial virulence factors, diseases and conditions in the host may lead to an impairment of pulmonary defense and increased risk of CAP: alterations in the level of consciousness - alcohol consumption - hypoxemia - acidosis - malnutrition- administration of immunosuppressive agents [15]. Clinical features and radiographic changes are not helpful in identifying the etiologic pathogen of CAP.

Cryptogenic organizing pneumonitis (**COP**) also called idiopathic BOOP. The incidence and prevalence of COP are unknown but though a cumulative incidence of six to seven per 100,000 hospital admissions was found at a major teaching hospital and the disease onset is typically in the fifth or sixth decades of life, with men and women affected equally [16]. COP is a distinct clinical entity with features of pneumonia, rather than a primary airway disorder thus the clinical presentation of COP often mimics that of community-acquired pneumonia [16].

Acute decompensated heart failure is a common and potentially fatal cause of acute respiratory distress. It results from severe left atrial outflow impairment or left ventricular systolic or diastolic dysfunction. The clinical examination, laboratory analysis and in some cases, direct measurement of pulmonary capillary wedge pressure can be used to diagnose cardiogenic pulmonary edema [17].

Idiopathic pulmonary hemosiderosis (IPH) is a diffuse alveolar hemorrhage. Eighty percent of cases of IPH occur in children, generally manifesting before 10 years of age. In adults, most cases are recognized before 30 years of age [18]. Familial clustering has been described in several reports, suggesting that hereditary factors play a role in the development of IPH and/or that environmental factors might precipitate the disease in genetically predisposed individuals. The exact incidence and prevalence of IPH are largely unknown and there is no feature that is pathognomonic of IPH [18].

Pulmonary alveolar proteinosis (PAP) is a diffuse lung disease. The typical age at presentation of the patient with PAP is 30 to 50 years[19]. There is a male to female ratio of 2:1. The clinical presentation is usually insidious; approximately one-third of affected patients are asymptomatic despite infiltration of the alveolar air spaces [19]. The most risk factors found: patients with acute silicosis ("silicoproteinosis"), aluminum or titanium dust exposure, infection with Pneumocystis jiroveci, hematologic malignancies, and immunosuppressive disorders [19].

Idiopathic pulmonary fibrosis (IPF) represents 60% of all ILD. More males have been reported with IPF than females and patients with IPF are often middle aged, usually between 40 and 70 years of age. Approximately two-thirds of patients with IPF are over the age of 60 years at the time of presentation, with a mean age at diagnosis of 66 years. Potential risks are: Cigarette smoking - Exposure to commonly prescribed drugs - Chronic aspiration - Environmental factors (Metal dust and wood dust exposure, exposure to solvents) - Infectious agents (Epstein - Barr virus (EBV); influenza, cytomegalovirus (CMV), and hepatitis C; etc…) - Genetic predisposition to IPF [20]. The etiology of IPF is unknown [20], the onset is insidious of otherwise unexplained dyspnea on exertion and the duration of illness is less than or equal to three months [20].

A clinical lung manifestation of mixed connective tissue disease (MCTD) is much more common in women than in men (ratio of 16 to 1) [4, 21]. Most patients present in the second or third decades of life [21]. Unlike systemic lupus erythematosus SLE, however, sun exposure is not a precipitating factor. The only environmental agents associated with MCTD have been vinyl chloride and silica, while drug-induced MCTD has not been described. It is unusual for patients to exhibit overlapping features in the early stages of MCTD [21].

Lymphocytic interstitial pneumonia (LIP) is a rare form of interstitial lung disease in adults. An increased incidence of LIP has been seen in the pediatric population, especially in children with AIDS [22]. In adults, LIP occurs most commonly in women (2 to 1 female to male ratio), usually between the ages of 40 and 70 years (median age 56 years) and fewer than 5 percent of patients are asymptomatic at presentation [22].

Pulmonary Langerhans cell histiocytosis is an uncommon interstitial lung disease that primarily affects young adults. The true incidence and prevalence are unknown and no occupational or geographic predisposition has been reported but all affected individuals have a history of current or prior cigarette smoking. The signs and symptoms of pulmonary Langerhans cell histiocytosis are nonspecific and often point to other, more common pulmonary diagnoses [23]. However, a history of recurrent pneumothorax, diabetes insipidus, or bone pain can be helpful in suggesting the diagnosis [5, 23].

2.2. Data to be collected for the lung CT library

The most important data to be collected in our case are:

- age of the patient in years,
- gender,
- smoking history (duration, amount),
- duration of the disease in weeks/months,
- prior medication or drug use (quantity, which drugs),
- family history (genetic susceptibility),

- geographic factors (place of birth, stay and duration in places of risk such as cities),
- occupational history (local practices and customs, season),
- ethnic or specific races (blacks / white subjects),
- environmental exposures (industries and work settings).

All the before mentioned data will be collected for all cases of the lung CT library. Depending on new requirements we might add other data to this list.

3. Discussion and conclusion

The radiological image does not constitute the only way of diagnosing patients with ILD, although it is an important part of diagnostics. Several decision support systems have been developed concentrating on the image, only. An evaluation showed that diagnostic quality can be improved by purely image-based solutions. Still, in clinical practice diagnostics remains very closely linked to the integration of all patient data with the interpretation of images. ILD are pathologies of multiple and often idiopathic etiologies and of complex radiological interpretation. Specific criteria such as age, sex, and risk factors are extremely important for diagnostics. The literature review for criteria characterizing all pathologies permitted to underline the importance of these patient data. It is therefore the aim to create a decision support system capable to integrate patient data with images for teaching and as diagnostic aid. This is not only expected to help in fields such as emergency radiology but particularly for teaching. Still, much work will be necessary to reach the goal and a quality of retrieval really sufficient for diagnostics.

4. Acknowledgements

This work has been supported by the Swiss National Science Foundation through grant (205321-109304-1), by the University Hospitals of Geneva (subside PRD 05-9-II) and by the Swiss Confederation through a scholarship of the first author.

5. References

[1] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler. A review of content-based image retrieval systems in medicine - clinical benefits and future directions. Intern Jour of Medical Informatics. 2004; 73 :1-23.
[2] H. Müller, S. Marquis, G. Cohen, PA. Poletti, C. Lovis, A. Geissbuhler. Automatic abnormal region detection in lung CT images for visual retrieval. Swiss Medical Informatics. 2005; 57:2-6.
[3] A. Aisen, L. Broderick, H. Muram. Automated storage and retrieval of thin – section CT images to assist diagnosis: System descrition and preliminary assessment. Radiology. 2003; 228:265-270
[4] JR. Talmadge E. King. Approach to the adult with interstitial lung disease. Uptodate 2006.
[5] P. Stark. High resolution computed tomography of the lungs. Uptodate 2006.
[6] T. Matsushima. Miliary tuberculosis or disseminated tuberculosis. Intern Med. 2005 Jul; 44(7):687.
[7] I. Miyoshi, M. Daibata, N. Kuroda, H. Taguchi and H. Enzan. Miliary Tuberculosis not affecting the Lungs but Complicated by Acute Respiratory Distress Syndrome. Internal Medicine. 2005; 44(6):622-624.
[8] A. Knauer, AK. Das, S. Tansuphasawadikul, W. Supanaranond, P. Pitisuttithum, WH. Wernsdorfer. Clinical features, aetiology and short-term outcome of interstitial pneumonitis in HIV/AIDS patients at Bamrasnaradura Hospital, Nonthaburi, Thailand.Wien Klin Wochenschr. 2005; 117 (4):49-55.
[9] E. Castaner, X. Gallardo, Y. Pallardo, J. Branera, MA. Cabezuelo, JM. Mata. Diseases affecting the peribronchovascular interstitium: CT findings and pathologic correlation. Curr Probl Diagn Radiol. 2005 Mar-Apr; 34(2):63-75.
[10] KA. Keogh, AH. Limper. Characterization of lymphocyte populations in nonspecific interstitial pneumonia. Respir Res. 2005; 6(1): 137.
[11] ET. Nij, D. Heederik. Risk assessment of silicosis and lung cancer among construction workers exposed to respirable quartz. Scand J Work Environ Health. 2005; 31(2):49-56.
[12] Y. Cormier. Hypersensitivity pneumonitis: restrictive diagnostic criteria or a different disease?. Ann Allergy Asthma Immunol. 2005 Aug; 95(2):99.
[13] JL. Myers, AH. Limper, J. Swensen. Drug-induced lung disease: a pragmatic classification incorporating HRCT appearances. Semin Respir Crit Care Med. 2003 Aug; 24(4):445-54.
[14] RC. Richie. Sarcoidosis: a review. J Insur Med. 2005; 37(4):283-94.
[15] HF. Arancibia, PO. Diaz. Severe community-acquired pneumonia in adults. Rev Chilena Infectol. 2005; (22):46-51. Epub 2005 Sep 13.
[16] R. Lazor. Cryptogenic organising pneumonia. Rev Pneumol Clin. 2005 Jun; 61(3):193-202.
[17] S. Wilson Colucci. Acute decompensated heart failure (cardiogenic pulmonary edema). Uptodate 2006.
[18] J. Sharma, M. Nitsure. Idiopathic pulmonary hemosiderosis. Indian Pediatr. 2005 Oct; 42(10):1056-7.
[19] S. Pissarra, G. Rocha, I. Acevedo, H. Guimaraes. Congenital pulmonary alveolar proteinosis: a case report. Acta Med Port. 2005 Mar-Apr; 18(2):163-8. Epub 2005 Apr 28.
[20] MG. Patti, P. Tedesco, J. Golden, S. Hays, C. Hoopes, A. Meneghetti, T. Damani, LW. Way. Idiopathic pulmonary fibrosis: how often is it really idiopathic?. J Gastrointest Surg. 2005 Nov; 9(8):1053-6.

[21] RM. Bennett. Clinical manifestations of mixed connective tissue disease. Uptodate 2006.
[22] KA. Fagan. JR. Talmadge E King. Lymphocytic interstitial pneumonia in adults. Uptodate 2006.
[23] R. Vassallo, AH. Limper. Pulmonary Langerhans' cell histiocytosis.Semin Respir Crit Care Med. 2002 Apr; 23(2):93-101.

6. Address for correspondence

Asmâa Hidki
University and Hospitals of Geneva, Service of Medical Informatics
24, rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland.
Asmaa.hidki@sim.hcuge.ch.
Tél : +41 22 372 62 01
Fax : +41 22 372 62 55.