

HIGH-RESOLUTION COMPUTED TOMOGRAPHY TO DIFFERENTIATE CHRONIC DIFFUSE INFILTRATIVE LUNG DISEASES WITH CHRONIC MULTIFOCAL CONSOLIDATION PATTERNS USING LOGICAL ANALYSIS OF DATA

Constance de Margerie-Mellon^{1*}, Geneviève Dion^{2*}, Julien Darlay³, Imene Ridené⁴,
Marianne Kambouchner⁵, Nadia Brauner³, Michel Brauner⁶, Dominique Valeyre⁷, Pierre-Yves Brillet⁶

¹Department of Radiology, Hôpital Saint-Louis, Université Paris 7 Denis Diderot, Sorbonne Paris-Cité, Paris, France; ²Department of Pneumology, Centre de Recherche de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, Québec, Canada; ³Laboratoire G-SCOP, Université Joseph Fourier, Grenoble, France; ⁴Department of Radiology, Hôpital Abderrahmane Mami, Ariana, Tunisie; ⁵Department of Pathology, EA2363 Laboratory, Hôpital Avicenne, Université Paris 13, Sorbonne Paris Cité, Bobigny, France; ⁶Department of Radiology, EA2363 Laboratory, Hôpital Avicenne, Université Paris 13, Sorbonne Paris Cité, Bobigny, France; ⁷Department of Pneumology, EA2363 Laboratory, Hôpital Avicenne, Université Paris 13, Sorbonne Paris Cité, Bobigny, France

ABSTRACT. *Background:* Chronic lung consolidation has a limited number of differential diagnoses requiring distinct managements. The aim of the study was to investigate how logical analysis of data (LAD) can support their diagnosis at HRCT (high-resolution computed tomography). *Methods:* One hundred twenty-four patients were retrospectively included and classified into 8 diagnosis categories: sarcoidosis (n=35), connective tissue disease (n=21), adenocarcinoma (n=17), lymphoma (n=13), cryptogenic organizing pneumonia (n=11), drug-induced lung disease (n=9), chronic eosinophilic pneumonia (n=7) and miscellaneous (n=11). First, we investigated the patterns and models (association of patterns characterizing a disease) built-up by the LAD from combinations of HRCT attributes (n=51). Second, data were recomputed by adding simple clinical attributes (n=14) to the analysis. Third, cluster analysis was performed to explain LAD failures. *Results:* HRCT models reached a sensitivity >80% and a specificity >90% for adenocarcinoma and chronic eosinophilic pneumonia. The same thresholds were obtained for sarcoidosis, connective tissue disease, and drug-induced lung diseases when clinical attributes were added to HRCT. LAD failed to provide a satisfactory model for lymphoma and cryptogenic organizing pneumonia, with overlap between both diseases shown on cluster analysis. *Conclusion:* LAD provides relevant models that can be used as a diagnosis support for the radiologist. It highlights the need to add clinical data in the analysis due to frequent overlap between diseases at HRCT. (*Sarcoidosis Vasc Diffuse Lung Dis* 2016; 33: 355-371)

KEY WORDS: interstitial lung disease, computed tomography, medical informatics

INTRODUCTION

Consolidation appears on high-resolution computed tomography (HRCT) as a homogeneous increase in pulmonary parenchymal attenuation that obscures the margins of vessels and airway walls (1). Chronic consolidations may be due to diagnoses requiring distinct managements, especially regarding

Received: 7 January 2016

Accepted after revision: 14 March 2016

Correspondence: Constance de Margerie-Mellon,
Department of Radiology, Hôpital Saint-Louis,
Assistance Publique-Hôpitaux de Paris,
Université Paris 7 Denis Diderot,
Sorbonne Paris-Cité, Paris, France

*These authors contributed equally to this work

malignancy. HRCT plays a key role in the diagnostic approach along with clinical data (2, 3). Consolidations can be defined by their morphological characteristics (contours, density, and bronchogram), distribution, or their association with other features. Some HRCT features are considered highly indicative of a disease. For instance, a bulging fissure sign favors adenocarcinoma (4), a superior and peripheral distribution advocates chronic eosinophilic pneumonia, and an association with a nodular galaxy sign suggests sarcoidosis (5). However, the diseases often share similar HRCT features, and the specificity of isolated signs remains to be determined. Moreover, in the daily practice, diagnosis is usually made by associating features, and the determination of pertinent combinations needs to be investigated.

The development of data mining provides innovating tools to extract relevant information from a database. Among the various available methods, logical analysis of data (LAD) allows, from a large set of data, the identification of informative combinations of features (named attributes) suggestive of a specific diagnosis (6). LAD has already been applied in a series of medical studies (7-10), particularly for the diagnosis of chronic diffuse infiltrative lung diseases presenting with predominant ground-glass opacity (11). Results provided by the LAD constitute a diagnostic decision support for non-expert radiologists (11).

Our study aimed to investigate how logical analysis of data (LAD) may improve the diagnosis of chronic pulmonary consolidations at HRCT.

METHODS

The present study was retrospective and monocentric. It was validated by a local ethics committee, and informed consent was waived. First, we worked on the clinical and HRCT attributes. We calculated the interobserver agreement (Cohen's Kappa test using the following κ ranges: 0.21-0.40=poor, 0.41-0.60=fair, 0.61-0.80=moderate, 0.81-1.00=good), sensitivity and specificity of each HRCT attribute. Second, we defined HRCT and HRCT + clinical models for the main diseases by combining attributes, and we calculated their sensitivity and specificity. Our endpoint was to fulfill a specificity >90% (first optimal scenario) or >80% (second scenario) with the highest sensitivity as possible. The last step

consisted of understanding the overlaps between diseases by analyzing misclassified or non-classified cases by LAD during the agglomeration of models step and by using cluster analysis. Our aim was to obtain for each disease a model with the highest sensitivity and specificity.

Patient selection

Three hundred one cases of possible chronic diffuse infiltrative lung diseases with multifocal consolidation on lung HRCT that were referred at a tertiary care hospital (Hôpital Avicenne, Bobigny, France) from January 1988 to August 2009 were retrospectively reviewed. Cases with predominant or exclusive multifocal consolidation on lung HRCT, a definitive diagnosis of chronic diffuse infiltrative lung diseases and clinical symptoms evolving for at least 2 months were included, after optimal work-up and antibiotic treatment leading to exclusion of chronic infectious diseases. Cases with the following criteria were excluded: lone lung consolidation (n=12), absence of lung consolidation at presentation (n=15), absence of a definite diagnosis (no histological proof (n=47), incomplete files (n=48), other (n=8)), prior comorbidity known to be confusing (i.e., cardiac failure, known lung neoplasia, other chronic infiltrative lung disease or lung infection) (n=16), immunodeficiency (HIV/AIDS, organ transplantation, immunosuppressive drug or corticosteroids use) (n=5) and clinical symptoms evolving for less than 2 months (n=28). Eventually, 124 cases were included and analyzed for this study (Appendix A).

Clinical data collection

Clinical and epidemiological data were collected retrospectively from clinical records using a standardized data sheet by a pulmonologist with expertise in infiltrative lung diseases (GD). For each case, the final diagnosis was confirmed in consensus by a pulmonologist (DV), a radiologist (MB) and a pathologist (MK) specialized and experienced (> 20 years of experience) in infiltrative lung disease according to strict criteria (Appendix B). Subsequently, cases were classified into eight categories: sarcoidosis (n=35), connective tissue disease (n=21), adenocarcinoma (n=17), lymphoma (n=13), cryptogenic organizing pneumonia (n=11), drug-induced lung disease (n=9),

chronic eosinophilic pneumonia (n=7) and miscellaneous (n=11, Appendix B).

The following clinical and epidemiological attributes were included (n=14, Table 1): age, gender, exposure to drugs known to cause lung disease, smoking history, Caucasian ethnicity, cough, dyspnea, weight loss, asthenia, fever, extrathoracic visceral involvement, rheumatologic symptoms, crackles and Raynaud syndrome or myalgias. Each attribute was binary coded (presence=1 or absence=0) for the LAD analysis except for age (in years).

HRCT Protocol

HRCT was performed using a Toshiba X-Press unit (Toshiba, Tokyo, Japan) or a CE 10 000 unit (Thomson CGR, Buc, France) between 1988 and 1999 and an MX 8000 unit (Philips Healthcare, Best, the Netherlands) since 1999. HRCT was acquired using standard lung protocols (with $120 < kVp < 140$ and standard mAs). All images were obtained during inspiration, with patients in the supine position. Images were reconstructed using a high spatial frequency algorithm (lung algorithm). A series of 1.5- or 1-mm-thick sections with 10-mm intersection spacing was available in all cases. The matrix size was 512×512 pixels, and the field of view was adapted to the patient size. All images were examined at standard pulmonary and mediastinal window settings, either on film (n=85) or on a PACS workstation (n=39).

HRCT analysis

HRCT were retrospectively and independently reviewed by two specialized and experienced radiologists (> 10 years of experience) in infiltrative lung disease (MB, PYB), blinded to the clinical data. In the case of disagreement, a consensus was reached thereafter. Retained HRCT attributes (n=51), including morphological characteristics of the consolidation, distribution, and other associated features are listed in Table 1. Specific attributes describing consolidation were cavitation, a reversed halo sign (central ground-glass opacity surrounded by denser consolidation of a crescentic or a ring shape), a stretched or absent bronchogram, a bulging fissure and an abnormal density. Distribution was evaluated in the cranio-caudal, axial and antero-posterior planes and was described as symmetrical, peribron-

chovascular, lobar or having no predominant distribution. The cranio-caudal distribution was evaluated in three zones: the upper zone above the level of the carina, middle zone between the level of the carina and level of the inferior pulmonary veins, and lower zone below the level of the inferior pulmonary veins. The axial distribution was also evaluated in three zones: the central zone situated in the inner half of the lung, peripheral zone referring to the outer half of the lung and peripheral and subpleural zones situated under the pleura. The antero-posterior distribution was evaluated in three equal zones (anterior, middle and posterior). Thirty-two additional HRCT features describing other pleuro-pulmonary and hilo-mediastinal abnormalities were considered (Table 1). Each attribute was binary coded (presence=1 or absence=0) for the LAD analysis.

Logical analysis of data

The analysis was performed using the data mining method LAD previously described by Martin et al (11) for radiological applications. The principle of a data mining process is to extract information from a data set and transform it into an understandable structure for further use. This information extraction can be performed using supervised learning models (if the aim is to classify data in previously established groups) such as the nearest neighbor method, neural networks, support vector machine, or unsupervised techniques (if the aim is to find a hidden structure in unlabeled data - for instance clustering). Logical analysis of data is a supervised data mining method that combines ideas and concept from optimization, combinatorics and Boolean functions (6, 12). Its aim is to classify new observations in a way consistent with past classifications. It consists of extracting, from a set of instances sharing a common property, one or more logical relations called "patterns" satisfied by a large number of those instances such that these logical relations are satisfied by only a small number of instances not having the property. An important feature of LAD methodology is the possibility of using patterns to explain the results of the classification to human experts by standard formal reasoning.

Logical analysis of data has been applied to data analysis problems in different domains, including biology and medicine (13), particularly for the diag-

Table 1. Details of the attributes evaluated by logical analysis of data for the 7 main categories

	Sarcoidosis n=35	CTD n=21	ADC n=17	Lymphoma n=13	COP n=11	DILD n=9	CEP n=7	Total n=124
Clinical data								
Age* [†]	38	50	66	59	57	68	49	52
Gender (females/males)*	12/23	16/5	9/8	7/6	8/3	5/4	7/0	68/56
Exposure to drugs*	6	8	6	5	3	9 [§]	2	39
Smoker*	21	2	8	5	3	6	1	46
Caucasian ethnic group*	20	7	9	10	6	2	3	57
Cough	24	15	15	11	7	5	7	92
Dyspnea	29	18	6	5	5	3	5	58
Weight loss	16	12	6	5	5	3	5	58
Asthenia	17	12	7	10	5	3	4	65
Fever*	2	8	2	4	5	6	6 [§]	33
Extrathoracic visceral involvement*	18	18	2	6	1	3	2	50
Rheumatologic symptoms*	6	9	1	3	0	1	1	21
Crackles*	7	18	12	6	7	6	2	58
Raynaud syndrome or myalgias*	4	8	0	0	0	1	0	13
Characteristics of alveolar consolidations								
Cavitation*	3	0	10	1	0	0	0	14
Halo sign	16	10	13	9	9	7	5	69
Reversed halo sign*	7	7	5	1	3	0	5 [§]	28
Stretched or absent bronchogram*	11	0	15 [§]	5	1	5	0	37
Bulging fissure*	3	0	8	2	0	1	0	14
Abnormal density [‡] *	0	0	1	0	0	5	0	6
Distribution								
- <i>Asymmetrical</i> *	5	4	11	3	4	3	3	36
- <i>Peribronchovascular</i> *	8	13	0	5	3	1	2	32
- <i>Lobar</i> *	3	1	13 [§]	3	2	3	0	25
Cranio-caudal distribution								
- <i>Upper</i> *	9	2	2	1	2	0	5 [§]	21
- <i>Middle</i> *	15	5	8	8	6	6	2	50
- <i>Lower</i> *	12	15	10	7	7	6	0	57
Axial distribution								
- <i>Central</i>	6	0	0	0	0	1	0	7
- <i>Peripheral</i>	7	7	2	0	3	1	1	21
- <i>Peripheral and subpleural</i> *	10	10	4	3	2	4	4	37
Antero-posterior distribution								
- <i>Anterior</i> *	2	0	2	4	0	0	0	8
- <i>Middle</i>	12	2	4	2	2	3	1	26
- <i>Posterior</i> *	13	13	7	2	7	1	1	44
No predominant distribution*	5	3	3	3	0	0	2	16
Ground-glass pattern								
As associated predominant pattern*	1	4	5	1	3	2	1	17
Well-defined margins*	0	0	0	0	2	0	0	2
Centrilobular*	0	0	0	0	0	2	0	2
Nodular pattern								
As associated predominant pattern*	12	0	0	2	0	1	0	15
Nodules*	29	12	16	10	7	7	3	84
Centrilobular micronodules*	0	1	4	0	0	4	1	10
Tree in bud*	0	0	0	0	0	1	0	1
Galaxy sign*	11	1	0	0	0	0	0	12
Perilymphatic micronodules*	27 [§]	1	2	3	0	0	0	33
Cavitated nodule*	1	0	5	1	0	0	0	7
Linear pattern								
Bands*	3	5	0	0	5	2	1	16
Intralobular reticulations*	0	4	0	1	0	1	0	6
Smooth septal thickening*	2	1	3	3	1	0	0	10
Nodular septal thickening*	2	0	1	2	0	0	0	5

(continued)

Table 1 (continued). Details of the attributes evaluated by logical analysis of data for the 7 main categories

	Sarcoidosis n=35	CTD n=21	ADC n=17	Lymphoma n=13	COP n=11	DILD n=9	CEP n=7	Total n=124
Hilo peripheral lines*	2	5	0	0	3	0	1	11
Subpleural curvilinear lines*	0	4	0	0	2	2	1	9
Peripheral honeycombing*	1	1	1	0	0	0	0	3
Other lines*	20	10	2	4	6	5	3	50
Bronchial disease								
Proximal traction bronchiectasis*	0	0	2	0	0	1	0	3
Distal traction bronchiectasis*	2	5	1	2	1	2	1	14
Bronchiectasis (except traction)*	0	4	1	1	0	0	0	6
Peribronchovascular thickening*	15	6	1	5	0	3	2	32
Lymph nodes								
Hilar*	23	2	1	5	1	0	2	34
Mediastinal*	25 [‡]	2	5	5	0	0	2	39
Pleura								
Thickening*	5	0	4	2	0	0	0	11
Effusion*	5	1	4	1	1	0	2	14
Other HRCT signs								
Fissural distortion*	6	1	0	2	0	1	0	10
Lung collapse*	7	2	6	3	1	0	1	20
Centrilobular emphysema*	0	0	2	0	0	2	0	4
Thin-walled cyst*	0	1	2	0	0	0	0	3
Thick-walled cyst*	0	0	1	0	0	0	0	1
Esophageal dilatation*	0	2	1	0	0	0	0	3

CTD: connective tissue disease, ADC: adenocarcinoma, COP: cryptogenic organizing pneumonia, DILD: drug-induced lung disease, CEP: chronic eosinophilic pneumonia

Total column includes the 7 previous columns and miscellaneous category

* Informative data used by the logical analysis of data to construct the different models

† Data are presented as the number of patients except for age (mean number of years)

‡ Hypodense or hyperdense compared to vessel density

§ Sensitivity > 70% and specificity >80% for the considered diseases

nosis of chronic diffuse interstitial lung diseases presenting with predominant ground-glass opacity (11).

Based on our data set, LAD defined combinations of attributes (called patterns) and models (compilations of patterns) characterizing each disease. Next, the agglomeration of models allowed patients with or without a given disease to be classified and the sensitivity of the models to be calculated. The analysis was based on HRCT data alone and on combined HRCT and clinical data. Validation was performed according to the leave-one out cross-validation method. Leave-one-out cross-validation is a simply n-fold cross-validation, where n is the number of patients in the dataset. (14).

Selection of relevant attribute combinations and definition of models

Each pattern in our study included a maximum of 3 HRCT and/or clinical attributes which could be

observed in homogeneous subsets of patients. This maximum of 3 attributes per pattern was chosen to limit the number of generated patterns and to keep readable models. The method identified positive patterns, mainly covering patients with the disease, and negative patterns, mainly covering patients without the disease. The selection of combinations used in the models was performed by a radiologist (MB) and a pulmonologist (DV) according to relevancy.

Understanding overlaps

Failures of LAD were investigated by the agglomeration of models and cluster analysis. Both approaches aimed to help to understand the HRCT overlap between diagnoses.

During the agglomeration of models, each patient in the data set was tested for each model. Therefore, a patient could be classified as follows: correct classification (the correct diagnosis was given by the

correct model), partially correct classification (if two positive diagnoses were proposed including the correct one), misclassification (if the patient fitted with one or several models, but none was the correct diagnosis) and unclassified (if the patient was not classified as positive by any model). Our analysis focused on partially correct and misclassified patients.

Cluster analysis consists of classifying diseases into subsets (clusters) in such a way that HRCT and HRCT + clinical attributes in the same cluster are more similar to each other than to those in other clusters. The aim is to obtain higher homogeneity within a cluster and, conversely, higher heterogeneity among the different clusters. Cluster analysis is a general task that can be performed using different methods. The expected maximization algorithm using the Weka software package (<http://www.cs.waikato.ac.nz/ml/weka/>) was chosen in this study to isolate our clusters. Our analysis focused on overlaps among the different diseases.

RESULTS

Analysis of HRCT attributes alone

Table 1 shows the frequency of clinical and HRCT attributes for each category of disease. Cohen's kappa agreement measure for the most frequent signs (observed by at least one observer in more than 10% of the cases) was good for 21/39 attributes. A lower level of agreement was noted for hilo-peripheral lines (poor), peribronchovascular thickening (fair), a stretched or an absent bronchogram (fair), smooth septal lines (fair), the association of ground-glass opacity as a predominant finding (fair) and lung collapse (moderate).

Among all HRCT attributes, some were specific, few were sensitive, but none of them had both a sensitivity and specificity >80% (Figure 1A -C). Particularly, regarding the standard signs of chronic consolidation, some were highly specific of a given disease but had a low (<50%) sensitivity: for example the "galaxy" sign in sarcoidosis or the bulging fissure sign in adenocarcinoma had a specificity of 99% and 94% but a sensitivity of 31% and 47%, respectively. Eventually, some signs were neither specific nor sensitive (for instance, the halo sign).

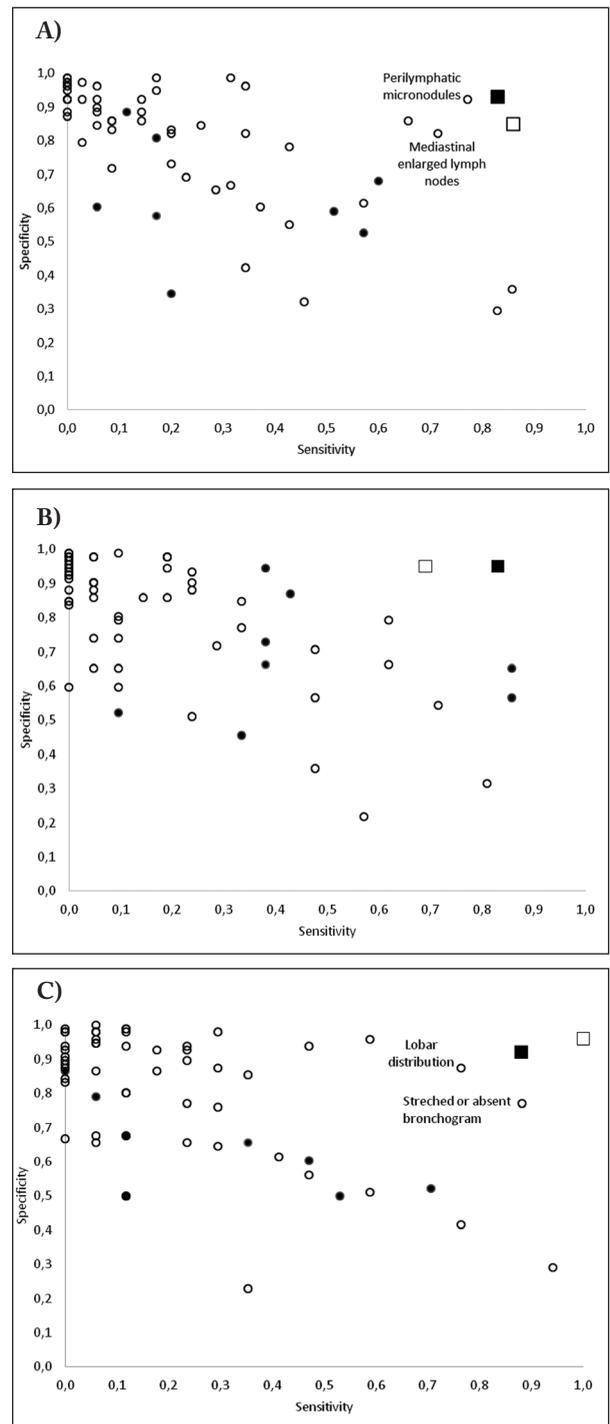


Fig. 1. Scatter plots of the sensitivity and specificity of isolated HRCT and clinical features, and of HRCT and HRCT + clinical models for sarcoidosis (1A), connective tissue disease (1B), adenocarcinoma (1C). White circles represent isolated HRCT attributes, black circles represent clinical attributes. The attribute name is mentioned if the sensitivity is >70%, and the specificity is >80%. White squares represent HRCT models, and black squares represent HRCT + clinical models

In sarcoidosis, the best attributes were perilymphatic micronodules and mediastinal nodes, which had a good sensitivity and were both highly specific (sensitivity of 77% and 71% and specificity of 92% and 82%, respectively). For adenocarcinoma, the best attributes were lobar distribution of consolidation and a stretched or an absent bronchogram (sensitivity of 76% and 88% and specificity of 87% and 77%, respectively). As expected, chronic eosinophilic pneumonia had a trend for upper lobe predominance (sensitivity of 71% and specificity of 85%).

Definition of models for HRCT and HRCT+ clinical attributes

Almost all attributes were used for the definition of models (46/50 for HRCT and 10/14 for clinical attributes) (Tables 1 and 2a-n).

The HRCT models had a sensitivity and specificity >80% for the diagnosis of sarcoidosis (sensitivity and specificity of 86% and 85%, respectively) and reached a higher specificity (>90%) for adenocarcinoma (sensitivity and specificity of 100% and 96%, respectively) and chronic eosinophilic pneumonia (86% and 96%). The specificity remained >90% but with a lower sensitivity for the diagnosis of connective tissue disease (69%) and drug-induced lung disease (75%). Conversely, we could not obtain accurate models for the diagnosis of lymphoma and cryptogenic organizing pneumonia (sensitivity <80%, specificity ≤50%).

Adding clinical data to HRCT data improved the model specificity (>90%) without decreasing the sensitivity for sarcoidosis (sensitivity and specificity of 83% and 95%, respectively), connective tissue disease (83% and 95%), and drug-induced lung disease

Table 2 (A-B). Models obtained using LAD for sarcoidosis

A. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with sarcoidosis (n= 35)	Patients without sarcoidosis (n=89)
Positive	Hilar lymph nodes	No nodular septal thickening	No band	21	8
Positive	No lobar distribution (AC)	Perilymphatic Micronodules	No smooth septal thickening	24	4
Positive	Perilymphatic micronodules	No nodular septal thickening	No bronchiectasis	25	4
Negative	No perilymphatic micronodules	No fissural distortion	No hilar lymph nodes	3	67
Negative	No perilymphatic micronodules	No peripheral honeycombing	No hilar lymph nodes	3	63
Negative	No galaxy sign	No peribroncho-vascular thickening	No hilar lymph nodes	3	59

B. HRCT + clinical model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with sarcoidosis (n= 35)	Patients without sarcoidosis (n=89)
Positive	Age ≤ 60	No fever	Smoker	20	3
Positive	Age ≤ 55	No fever	No crackles	25	7
Positive	Age ≤ 40	No fever	No cavitated nodule	23	4
Positive	Perilymphatic micronodules	No nodular septal thickening	No band	24	3
Negative	Age ≥ 40	No fissural distortion	No hilar lymph nodes	2	63
Negative	Age ≥ 40	No perilymphatic micronodules		3	70
Negative	No hilar lymph nodes	No fissural distortion	No perilymphatic micronodules	3	67
Negative	No peribroncho-vascular thickening	No peripheral honeycombing	No perilymphatic micronodules	3	63

Table 2 (C-D). Models obtained using LAD for connective tissue diseases

C. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with CTD (n= 21)	Patients without CTD (n=103)
Positive	Peribronchovascular distribution (AC)	Posterior distribution (AC)	No band	9	1
Positive	Peribronchovascular distribution (AC)	Lower distribution (AC)	No pleural thickening	10	5
Positive	Peribronchovascular distribution (AC)	No middle (cranio-caudal) distribution (AC)	No mediastinal lymph nodes	10	5
Negative	No peribroncho-vascular distribution (AC)	No intralobular reticulations	No subpleural curvilinear lines	4	74
Negative	No posterior distribution (AC)	No intralobular reticulations	No bronchiectasis	2	65
Negative	No reversed halo sign (AC)	Nodule	No bronchiectasis	3	61

D. HRCT + clinical model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with CTD (n= 21)	Patients without CTD (n=103)
Positive	Extrathoracic visceral involvement	Crackles	No mediastinal lymph nodes	16	9
Positive	Age ≤ 55	Crackles	No hilar lymph nodes	13	8
Positive	Extrathoracic visceral involvement	Peribronchovascular distribution (AC)	No mediastinal lymph nodes	12	4
Negative	No extrathoracic visceral involvement	No Raynaud/ myalgias	No intralobular reticulations	2	66
Negative	No rheumatologic symptoms	No peribroncho-vascular distribution (AC)	No intralobular reticulations	2	66
Negative	No extrathoracic visceral involvement	No Raynaud/ myalgias	No esophageal dilatation	2	66
Negative	No posterior distribution (AC)	No intralobular reticulations	No bronchiectasis	65	2
Negative	No crackles	No intralobular reticulations	No bronchiectasis	59	1
Negative	No crackles	No thin wall cyst	No bronchiectasis	59	1

(100% and 94%). The model specificity remained above this threshold for chronic eosinophilic pneumonia (sensitivity and specificity of 100% and 97%, respectively) and adenocarcinoma (88% and 92%). Regarding lymphoma and cryptogenic organizing pneumonia, the specificity of the models increased (>90% for both), but the sensitivity showed a significant decrease (≤55%). Sensitivity and specificity results obtained with the leave-one out cross-validation method provided slightly lower sensitivities (Appendix C). This result is probably due to overfitting, which occurs when a model is overly complex, especially when the number of attributes is too high relative to the number of observations.

Comprehensive analysis of overlaps

Results of the model agglomeration and cluster analysis are shown in Table 3. In most cases (68/124, 55% based on HRCT and 70/124, 56% based on HRCT and clinical attributes), the patients fulfilled only one of the different models and were correctly classified. The diagnosis was correct in all cases of chronic eosinophilic pneumonia. The diagnosis was also correct for all cases of adenocarcinoma based on HRCT attributes and for sarcoidosis when adding clinical data. In a limited number of cases, patients fulfilled several models simultaneously, including being diagnosed correctly (partially correct classifi-

Table 2 (E-F). Models obtained using LAD for adenocarcinoma

E. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with adenocarcinoma (n=17)	Patients without adenocarcinoma (n=107)
Positive	Lobar distribution (AC)	Asymmetrical distribution (AC)	Nodules	9	1
Positive	No predominant nodular pattern	Cavitation (AC)	No galaxy sign	10	2
Positive	No predominant nodular pattern	Lobar distribution (AC)	Asymmetrical distribution (AC)	10	3
Negative	No predominant ground glass pattern	No cavitation (AC)	No stretched nor absent bronchogram (AC)	1	82
Negative	Symmetrical distribution (AC)	NO cavitation (AC)	No thick-walled cyst	0	76
Negative	No lung collapse	No bulging fissure (AC)	No cavitated nodule	1	83

F. HRCT + clinical model				Number of patients fitting the pattern	
Type of pattern	Pattern description (clinical + HRCT data)			Patients with adenocarcinoma (n=17)	Patients without adenocarcinoma (n=107)
Positive	Age ≥ 50	No fever	Lobar distribution (AC)	11	2
Positive	Age ≥ 50	Lobar distribution (AC)	No peribroncho-vascular thickening	13	4
Negative	No predominant ground-glass pattern	No stretched nor absent bronchogram (AC)	No cavitated nodule	1	82
Negative	No cavitation (AC)	Symmetrical distribution (AC)		0	76

Table 2 (G-H). Models obtained using LAD for lymphoma

G. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with lymphoma (n=13)	Patients without lymphoma (n=111)
Positive	Anterior distribution (AC)	No Reversed halo sign (AC)	No stretched or absent bronchogram (AC)	4	0
Positive	Middle (cranio-caudal) distribution (AC)	Peribronchovascular thickening	No other lines	4	1

H. HRCT + clinical model				Number of patients fitting the pattern	
Type of pattern	Pattern description (clinical + HRCT data)			Patients with lymphoma (n=13)	Patients without lymphoma (n=111)
Positive	Stretched or absent bronchogram (AC)	No reversed halo sign (AC)	Anterior distribution (AC)	4	0
Positive	No peribroncho-vascular distribution (AC)	Middle (cranio-caudal) distribution (AC)	No other lines	4	1
Negative	Age ≤ 75	No anterior distribution (AC)	No distal traction bronchiectasis	6	93
Negative	No predominant distribution (AC)	No anterior distribution (AC)		6	92
Negative	No smooth septal thickening	No nodular septal thickening	No fissural distortion	6	90

Table 2 (I-J). Models obtained using LAD for cryptogenic organizing pneumonia

I. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with COP (n=11)	Patients without COP (n=113)
Positive	No posterior distribution (<i>AC</i>)	Band	No pleural effusion	4	2
Positive	No middle (cranio-caudal) distribution (<i>AC</i>)	Band	No pleural effusion	5	4
J. HRCT + clinical model				Number of patients fitting the pattern	
Type of pattern	Pattern description (clinical + HRCT data)			Patients with COP (n=11)	Patients without COP (n=113)
Positive	Caucasian	Fever	No middle (cranio-caudal) distribution (<i>AC</i>)	4	2
Negative	No predominant ground-glass pattern	No upper distribution (<i>AC</i>)	No band	2	67
Negative	Age ≤ 55	No ground glass with well-defined margins	No subpleural curvilinear lines	2	66

Table 2 (K-L). Models obtained using LAD for drug-induced lung disease (DILD)

K. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with DILD (n=8)	Patients without DILD (n=116)
Positive	Abnormal density (<i>AC</i>)	No subpleural curvilinear lines	No pleural effusion	5	0
Negative	No abnormal density (<i>AC</i>)	No centrilobular emphysema	No proximal traction bronchiectasis	1	113
L. HRCT + clinical mode				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with DILD (n=8)	Patients without DILD (n=116)
Positive	Exposure to drugs	No Caucasian	Abnormal density (<i>AC</i>)	5	0
Positive	Age ≥ 70	No posterior distribution (<i>AC</i>)	No lung collapse	4	2
Positive	Exposure to drugs	Fever	No reversed halo sign (<i>AC</i>)	6	4
Negative	No fever	No tree in bud	No centrilobular ground glass	0	84
Negative	Age ≤ 60	No abnormal density (<i>AC</i>)	No proximal traction bronchiectasis	0	82

cation for 12/124, 10% based on HRCT and 8/124, 6% based on HRCT and clinical attributes) or being misclassified (6/124, 5% based on HRCT and 5/124, 4% based on HRCT and clinical attributes). Mismatches were mainly noted between sarcoidosis and lymphoma and between connective tissue disease and cryptogenic organizing pneumonia. Finally

31% of the studied population remained unclassified, especially patients included in the miscellaneous category, as this category was not described by a model.

The cluster analysis individualized three clusters based on HRCT alone or with clinical data (Table 4). Cluster A grouped most of the sarcoidosis cases

Table 2 (M-N). Models obtained using LAD for chronic eosinophilic pneumonia

M. HRCT model				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with CEP (n=8)	Patients without CEP (n=116)
Positive	Upper distribution (AC)	Peripheral and subpleural distribution (AC)	Reversed halo sign (AC)	4	1
Positive	Upper distribution (AC)	No lobar distribution (AC)	No perilymphatic Micronodules	5	4
Negative	No upper distribution (AC)	Predominant distribution ≠ NO (AC)		0	85
Negative	No upper distribution (AC)	Nodule		0	74
Negative	No reversed halo sign (AC)	No peribronchovascular distribution (AC)		0	63

N. HRCT + clinical mode				Number of patients fitting the pattern	
Type of pattern	Pattern description (HRCT)			Patients with CEP (n=8)	Patients without CEP (n=116)
Positive	Female gender	No lower distribution (AC)	No nodule	4	3
Positive	Age ≤ 60	Fever	Upper distribution (AC)	5	0
Negative	NO upper distribution (AC)	Nodule		0	74
Negative	No upper distribution (AC)	Symmetrical distribution (AC)	No centrilobular micronodule	0	69
Negative	Age ≥ 30	No reversed halo sign (AC)		0	68
Negative	No upper distribution (AC)	No peribronchovascular distribution (AC)	No hilo peripheral lines	0	68

AC = refers to characteristic of alveolar consolidations

(71 and 80% for HRCT alone or with clinical data, respectively), cluster B grouped most of the connective tissue disease cases (95 and 90% for HRCT alone or with clinical data, respectively), and cluster C grouped most of the adenocarcinoma cases (82 and 82% for HRCT alone or with clinical data, respectively). No specific cluster was recognized for the other diagnoses, with cases of lymphoma distributed among the 3 clusters. Finally, many cases of cryptogenic organizing pneumonia were observed in cluster B based on HRCT data (72 and 55% for HRCT alone or with clinical data, respectively).

DISCUSSION

By providing models based on the combinations of attributes, LAD supports the diagnosis of chronic pulmonary consolidations. Indeed, LAD allows the

attainment of a sensitivity >80% and a specificity >90% based on HRCT data for adenocarcinoma and chronic eosinophilic pneumonia, respectively. Secondly, LAD highlights the need to combine HRCT and clinical data to obtain the best results for sarcoidosis, connective tissue disease, and drug-induced lung diseases, with models reaching 90% specificity. Third, LAD failed to provide a satisfactory model for lymphoma and cryptogenic organizing pneumonia, with both diseases having a wide HRCT polymorphism spectrum as evidenced by cluster analysis.

Data mining techniques are a new field of research in radiology. LAD method reproduces daily practice, where features are combined to propose a diagnosis. From a patient’s point of view, LAD should be considered as support for the physician as it provides a justification of the results based on the list of positive and negative patterns. Each pattern comprises attributes that suggest or exclude a given

Table 3 – Agglomeration of models generating the results for the classification of patients by LAD

Diagnosis	Number (n)	Correct classification n (%)		Partially correct classification n (%) Alternative diagnoses (number)		Misclassification n (%) -Evoked diagnosis (number)		Unclassified n (%)	
		HRCT attributes	HRCT and clinical attributes	HRCT attributes	HRCT and clinical attributes	HRCT attributes	HRCT and clinical attributes	HRCT attributes	HRCT and clinical attributes
Sarcoidosis	35	26 (76)	27 (77)	2 (6) <i>-lymphoma (1)</i> <i>-COP (1)</i>	0 (0)	0 (0)	0 (0)	7 (20)	8 (23)
CTD	21	10 (48)	14 (67)	3 (14) <i>-sarcoidosis (1)</i> <i>-COP (2)</i>	2 (10) <i>-CEP (1)</i> <i>-COP (1)</i>	1 (5) <i>-CEP (1)</i>	0 (0)	7 (33)	5 (24)
ADC	17	17 (100)	11 (65)	0(0)	2 (12) <i>-CTD (1)</i> <i>-DILD (1)</i>	0 (0)	1 (6) <i>-sarcoidosis (1)</i>	0 (0)	3 (18)
Lymphoma	13	4 (31)	3 (23)	3 (23) <i>-sarcoidosis (3)</i>	1 (8) <i>-ADC (1)</i>	1 (8) <i>-sarcoidosis (1)</i>	2 (15) <i>-sarcoidosis (1)</i> <i>-CTD (1)</i>	5 (38)	7 (54)
COP	11	3 (27)	1 (9)	2 (18) <i>-CTD (2)</i>	2 (18) <i>-DILD (1)</i> <i>-ADC (1)</i>	0 (0)	1 (9) <i>-CTD (1)</i>	6 (55)	7 (64)
DILD	9	3 (33)	7 (78)	2 (22) <i>-ADC (1)</i> <i>-COP (1)</i>	1 (11) <i>-ADC (1)</i>	0 (0)	1 (11)	4 (44)	0 (0)
CEP	7	5 (71)	7 (100)	0 (0)	0 (0)	0 (0)	0 (0)	2 (29)	0 (0)
Miscellaneous	11	NA	NA	NA	NA	4 (36)	0 (0)	7 (64)	11 (100)
Total	124	68 (55%)	70 (56%)	12 (10%)	8 (6%)	6 (5%)	5 (4%)	38 (31%)	41 (33%)

CTD: connective tissue disease; ADC: adenocarcinoma; COP: cryptogenic organizing pneumonia; DILD: drug-induced lung disease; CEP: chronic eosinophilic pneumonia

diagnosis. For instance, perilymphatic micronodules and enlarged hilar lymph nodes are the cornerstone of sarcoidosis on HRCT. Consequently, they contribute to the diagnosis of sarcoidosis in positive patterns, and their absence tends to block this diagnosis in negative patterns.

Our results confirm previously sparse data from the literature showing various degrees of overlap between benign conditions (15, 16). Concerning drug-induced lung disease and connective tissue disease, clinical attributes have to be considered in addition to HRCT attributes to reach both a high sensitivity and specificity. As expected, exposure to drugs known to cause lung diseases, age, crackles and the presence of extrathoracic signs were combined on HRCT in the different patterns. Regarding crypto-

genic organizing pneumonia, we could not obtain an accurate model, and the diagnosis should be considered by exclusion of others. In the agglomeration of models, we noticed several misclassifications or partially correct classifications involving connective tissue disease and cryptogenic organizing pneumonia, with the two diagnoses well represented in cluster B. This result was expected because we know from the pathological literature that these diagnoses are both included in the wide spectrum of organizing pneumonia (17). Future studies should focus on the distribution of consolidation, which seem to be more frequently peribronchovascular and peripheral in the case of connective tissue disease.

For malignant diseases, the interpretation of results differed between adenocarcinoma and lympho-

ma, the first one being easily diagnosed, while the second one likely to be misinterpreted as sarcoidosis on HRCT images. For adenocarcinoma, our results corroborate the former descriptions of bronchioloalveolar carcinomas (18, 19, 20). Adenocarcinomas display a pneumonic growth by the filling of alveolar spaces, expansion to the lobe leading to bulging fissures and narrowing bronchi, and frequent cavitations (4, 18, 20). They finally disseminate through airways to produce distant nodules. Conversely, for lymphoma, no satisfactory model could be identified, and the diagnosis should be advocated when LAD generates several diagnoses, including sarcoidosis, or remains unclassified. Indeed, in the cluster analysis, we noticed the polymorphism of lymphoma, which shares similarities with other diseases (21–23). Lymphatic involvement in lymphoma, such as in sarcoidosis (24), most likely explains the occurrence of the latter disease as an alternative diagnosis or a false diagnosis for lymphoma in the LAD models. The confusion urges caution, and the diagnosis of sarcoidosis should be retained in the case of typical cases with perilymphatic micronodules or nodules, hilar nodes and subtle signs of fibrosis (fissural distortion). In other cases, further histologic examinations seem to be required, including transparietal biopsy (25).

Our study has several limitations. First, we could not obtain models for rare diseases (miscellaneous category), and a significant number of patients remained unclassified by LAD. Indeed, the study was monocentric and retrospective, and strict diagnostic criteria were used. This resulted in a limited number of included cases ($n=124$), with a lack of representation of rare diseases and the need to group cases (particularly in the drug-induced lung disease category). Second, the limited number of patients, as regard to the complexity of models, is also responsible for a certain degree of overfitting. Overfitting occurs when the LAD “memorized” rather than “learn” from the training set. It induces a decreased of sensitivities during the leave-one-out cross-validation, as it can exaggerate minor fluctuations in the data. Third the study included HRCT performed between 1988 and 2009, and most of them were realized with sequential scanning. Attribute detection may be more powerful with helical scanning.

In conclusion, there was recently a paradigm shift in the diagnostic imaging approach of infil-

trative lung disease that was initially based on predominant lesions, often leading to numerous possible diagnoses. LAD, which is a mathematical approach that can combine numerous data, seems relevant for the diagnosis of chronic consolidation. Additionally, LAD highlights the need to combine clinical data in most of cases due to frequent overlaps at HRCT. Finally, LAD is a powerful support for diagnosis and provides imaging and clinical evidence to the physician. It may improve the performance of less experienced radiologists who cannot easily integrate the imaging and clinical findings.

REFERENCES

1. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697–722.
2. Hansell DM, Lynch DA, McAdams HP, Bankier AA. Basic patterns in lung disease. In: Hansell DM, Lynch DA, McAdams HP, Bankier AA editors. *Imaging of Diseases of the Chest*, 5th ed, St. Louis, Mo: Mosby-Year Book; 2009.
3. Webb WR, Muller NL, Naidich DP. HRCT findings of lung disease. In: Webb WR, Muller NL, Naidich DP editors. *High-Resolution CT of the Lung*, 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2009.
4. Patsios D, Roberts HC, Paul NS, et al. Pictorial review of the many faces of bronchioloalveolar cell carcinoma. *Br J Radiol* 2007; 80: 1015–1023.
5. Nakatsu M, Hatabu H, Morikawa K, et al. Large Coalescent Parenchymal Nodules in Pulmonary Sarcoidosis: “Sarcoid Galaxy” Sign. *Am J Roentgenol* 2002; 178: 1389–1393.
6. Crama Y, Hammer PL, Ibaraki T. Cause-effect relationships and partially defined boolean functions. *Ann Oper Res* 1988; 16: 299–326.
7. Alexe G, Alexe S, Liotta LA, et al. Ovarian cancer detection by logical analysis of proteomic data. *Proteomics* 2004; 4: 766–783.
8. Alexe G, Alexe S, Axelrod DE, et al. Logical analysis of diffuse large B-cell lymphomas. *Artif Intell Med* 2005; 34: 235–267.
9. Alexe S, Blackstone E, Hammer PL, et al. Coronary Risk Prediction by Logical Analysis of Data. *Ann Operations Res* 2002; 119: 15–42.
10. Reddy A, Wang H, Yu H, et al. (2008) Logical Analysis of Data (LAD) model for the early diagnosis of acute ischemic stroke. *BMC Med Inform Decis Mak* 2008; 8: 30.
11. Martin SG, Kronek L-P, Valeyre D, et al. High-resolution computed tomography to differentiate chronic diffuse interstitial lung diseases with predominant ground-glass pattern using logical analysis of data. *Eur Radiol* 2010; 20: 1297–1310.
12. Boros E, Hammer PL, Ibaraki T, et al. An implementation of logical analysis of data. *IEEE Trans Knowl Data Eng* 2000; 12: 292–306.
13. Hammer PL, Bonates TO. Logical analysis of data, an overview: From combinatorial optimization to medical applications. *Ann Oper Res* 2006; 148: 203–225.
14. Witten IH, Frank E, Hall MA (2011) *Credibility*. In: Witten IH, Frank E, Hall MA (eds) *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Morgan Kaufmann Publishers, San Francisco; 2001.
15. Arakawa H, Kurihara Y, Niimi H, et al. Bronchiolitis Obliterans with Organizing Pneumonia Versus Chronic Eosinophilic Pneumonia. *Am J Roentgenol* 2001; 176: 1053–1058.

16. Marchiori E, Zanetti G, Hochegger B, et al. Reversed halo sign on computed tomography: state-of-the-art review. *Lung* 2012; 190: 389-394.
17. Robertson BJ, Hansell DM. Organizing pneumonia: a kaleidoscope of concepts and morphologies. *Eur Radiol* 2011; 21: 2244-2254.
18. Lee KS, Kim Y, Han J, et al. Bronchioloalveolar carcinoma: clinical, histopathologic, and radiologic findings. *Radiogr Rev Publ Radiol Soc N Am Inc* 1997; 17: 1345-1357.
19. Akira M, Atagi S, Kawahara M, et al. High-resolution CT findings of diffuse bronchioloalveolar carcinoma in 38 patients. *Am J Roentgenol* 1999; 173: 1623-1629.
20. Trigaux JP, Gevenois PA, Goncette L, et al. Bronchioloalveolar carcinoma: computed tomography findings. *Eur Respir J* 1996; 9: 11-16.
21. Wislez M, Cadranet J, Antoine M, et al. Lymphoma of pulmonary mucosa-associated lymphoid tissue: CT scan findings and pathological correlations. *Eur Respir J* 1999; 14: 423-429.
22. Hare SS, Souza CA, Bain G, et al. The radiological spectrum of pulmonary lymphoproliferative disease. *Br J Radiol* 2012; 85: 848-864.
23. Okada F, Ando Y, Kondo Y, et al. Thoracic CT Findings of Adult T-Cell Leukemia or Lymphoma. *Am J Roentgenol* 2004; 182: 761-767.
24. Honda O, Johkoh T, Ichikado K, et al. Comparison of high resolution CT findings of sarcoidosis, lymphoma, and lymphangitic carcinoma: is there any difference of involved interstitium? *J Comput Assist Tomogr* 1999; 23: 374-379.
25. Ferretti GR, Jankowski A, Rodière M, et al. CT-guided biopsy of nonresolving focal air space consolidation. *J Thorac Imaging* 2008; 23: 7-12.
26. Costabel U, Hunninghake GW. ATS/ERS/WASOG statement on sarcoidosis. Sarcoidosis Statement Committee. American Thoracic Society. European Respiratory Society. World Association for Sarcoidosis and Other Granulomatous Disorders. *Eur Respir J* 1999; 14: 735-7.
27. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). *N Engl J Med* 1975; 292: 344-7.
28. Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *An Rheum Dis* 2002; 61: 554-8.
29. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 31: 315-24.
30. Mosca M, Neri R, Bombardieri S. Undifferentiated connective tissue diseases (UCTD): a review of the literature and a proposal for preliminary classification criteria. *Clin Exp Rheumatol* 1999; 17: 615-20.
31. Spickard A 3rd, Hirschmann JV. Exogenous lipoid pneumonia. *Arch Intern Med* 1994; 154: 686-92.
32. Masi AT, Hunder GG, Lie JT, et al. The American College of Rheumatology 1990 criteria for the classification of Churg-Strauss syndrome (allergic granulomatosis and angiitis). *Arthritis Rheum* 1990; 33: 1094-100.
33. Leavitt RY, Fauci AS, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of Wegener's granulomatosis. *Arthritis Rheum* 1990; 33: 1101-7.
34. Raghu G, Collard HR, Egan JJ, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *Am J Respir Crit Care Med* 2011; 183:788-824

APPENDICES

APPENDIX A

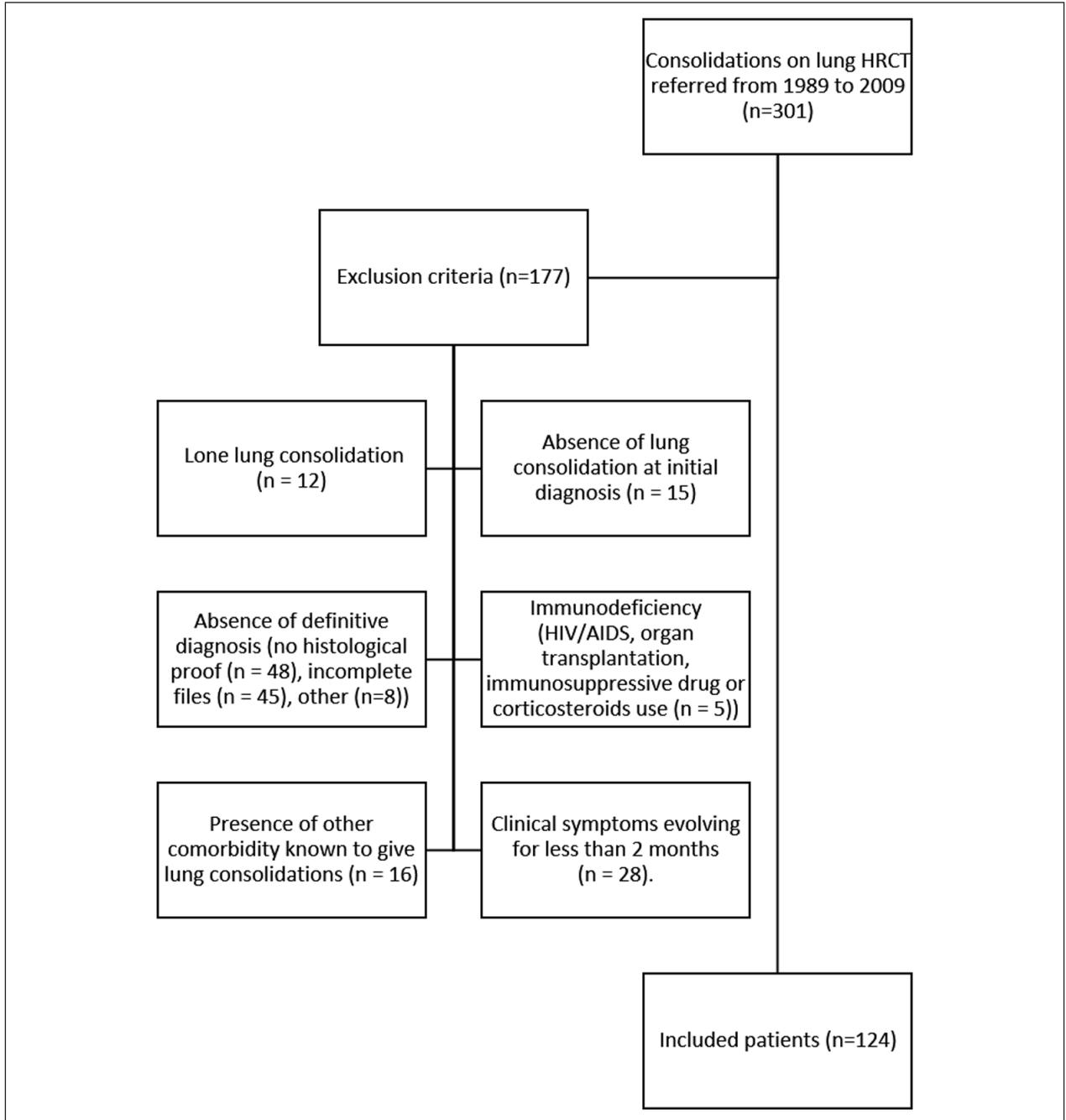


Fig. A.1. Study Flow Chart

APPENDIX B

Table B.1 Diagnostic criteria of 22 diseases (8 categories) with chronic diffuse interstitial lung disease and predominant alveolar consolidation

CDILD	N	Diagnostic criteria
Sarcoidosis	35	(i) Compatible clinical history and thoracic imaging, (ii) Histologic demonstration of noncaseating granulomas and (iii) Exclusion of an alternative diagnosis of granulomatosis (26)
Connective tissue disease	21	(i) Diagnosis of connective tissue disease according to specific diagnosis criteria, (ii) Compatible clinical history and thoracic imaging, (iii) Exclusion of an alternative diagnosis
Polymyositis and dermatomyositis	13	Peter and Bohan criteria (27)
Sjögren's syndrome	5	American-European Consensus Group criteria (28)
Rheumatoid arthritis	2	American Rheumatoid Association criteria (29)
Undifferentiated connectivitis	1	Mosca criteria (30)
Adenocarcinoma with lepidic growth	17	(i) Compatible clinical history and thoracic imaging, (ii) adenocarcinoma with lepidic growth on lung cytology or histology
Lymphoma	13	(i) Compatible clinical history and thoracic imaging, (ii) Lymphoma on lung biopsy
Cryptogenic organizing pneumonia	11	(i) Compatible clinical history, (ii) Mobile lung infiltrations on thoracic imaging, (iii) Absence of eosinophilia on BAL, (iv) Exclusion of alternative diagnosis <u>or</u> (i) Compatible clinical history and thoracic imaging, (ii) Organizing pneumonia on lung biopsy (iii) Exclusion of an alternative diagnosis of organizing pneumonia
Drug-induced lung disease	9	
Amiodarone	4	(i) Exposure to a drug known to be responsible for CDILD, (ii) Compatible clinical history and thoracic imaging, (iii) BAL compatible, (iv) Exclusion of an alternative diagnosis, (v) Clinical resolution with drug withdrawal
Bicalutamide	1	
Exogenous lipid pneumonia	4	(i) Compatible clinical history and thoracic imaging, (ii) Presence of intrapulmonary lipids, (iii) Exogenous origin of the lipid pneumonia (31)
Chronic eosinophilic pneumonia	7	(i) Compatible clinical history and thoracic imaging, (ii) Blood eosinophilia > 1000 <u>or</u> BAL eosinophilia > 25% <u>or</u> eosinophil infiltration on lung biopsy, (iii) Exclusion of an alternative diagnosis
Miscellaneous (n ≤ 4)	11	
Vasculitis	4	(i) Diagnosis of vasculitis according to their specific diagnosis criteria, (ii) Compatible clinical history and thoracic imaging, (iii) Exclusion of alternative diagnosis
<i>Churg and Strauss</i>	3	American College of Rheumatology criteria (32)
<i>Wegener</i>	1	American College of Rheumatology criteria (33)
Idiopathic pulmonary fibrosis	1	(i) Possible UIP on thoracic imaging (34) (ii) UIP on lung biopsy, (iii) Exclusion of an alternative diagnosis of UIP
Chronic beryllium disease	1	(i) A history of beryllium exposure, (ii) A positive beryllium-specific lymphocyte proliferation test, (iii) Non-necrotizing granulomas on biopsy of lung/affected tissue
Silicosis	1	(i) A history of crystalline silica exposure, (ii) Compatible clinical history and thoracic imaging, (iii) Silicotic granulomas or birefringent body on lung biopsy
Crohn's disease	1	(i) Diagnosis of Crohn's disease according to specific diagnosis criteria, (ii) Compatible clinical history and thoracic imaging, (iii) Organizing pneumonia on lung biopsy
Lung amyloidosis	2	(i) Compatible clinical history and thoracic imaging, (ii) Amyloidosis on lung biopsy
Radiation pneumonitis	1	(i) History of radiation therapy, (ii) Compatible history and thoracic imaging, (iii) Exclusion of an alternative diagnosis

BAL: bronchioloalveolar lavage

CDILD: chronic diffuse interstitial lung disease

UIP: usual interstitial pneumoni

APPENDIX C

Comparison of sensitivity and specificity obtained with LAD and with leave-one out cross-validation

Table C.1 HRCT models

Disease category	Sensitivity (LAD models)	Sensitivity (leave-one out cross-validation method)	Specificity (LAD models)	Specificity (leave-one out cross-validation method)
Sarcoidosis	0.86	0.71	0.85	0.83
Connective tissue disease	0.69	0.55	0.95	0.84
Adenocarcinoma	1.00	0.59	0.96	0.95
Lymphoma	0.77	0.04	0.50	0.98
Cryptogenic organizing pneumonia	0.73	0.09	0.48	0.95
Drug-induced lung disease	0.75	0.5	0.99	0.93
Chronic eosinophilic pneumonia	0.86	0.44	0.96	0.98

Table C.2 HRCT + clinical models

Disease category	Sensitivity (LAD models)	Sensitivity (leave-one out cross-validation method)	Specificity (LAD models)	Specificity (leave-one out cross-validation method)
Sarcoidosis	0.83	0.74	0.93	0.92
Connective tissue disease	0.83	0.6	0.95	0.91
Adenocarcinoma	0.88	0.53	0.92	0.96
Lymphoma	0.38	0.08	1.00	0.95
Cryptogenic organizing pneumonia	0.55	0	0.91	0.97
Drug-induced lung disease	1.00	0.5	0.95	0.97
Chronic eosinophilic pneumonia	1.00	0.42	0.97	0.97