Original Investigation

Novel Logistic Regression Model of Chest CT Attenuation Coefficient Distributions for the Automated Detection of Abnormal (Emphysema or ILD) Versus Normal Lung

Kung-Sik Chan, PhD, Feiran Jiao, MS, Marek A. Mikulski, MD, PhD, Alicia Gerke, MD, Junfeng Guo, PhD, John D. Newell Jr, MD, Eric A. Hoffman, PhD, Brad Thompson, MD, Chang Hyun Lee, MD, PhD, Laurence J. Fuortes, MD

Rationale and Objectives: We evaluated the role of automated quantitative computed tomography (CT) scan interpretation algorithm in detecting interstitial lung disease (ILD) and/or emphysema in a sample of elderly subjects with mild lung disease. We hypothesized that the quantification and distributions of CT attenuation values on lung CT, over a subset of Hounsfield units (HUs) range (–1000 HU, 0 HU), can differentiate early or mild disease from normal lung.

Materials and Methods: We compared the results of quantitative spiral rapid end-exhalation (functional residual capacity, FRC) and end-inhalation (total lung capacity, TLC) CT scan analyses of 52 subjects with radiographic evidence of mild fibrotic lung disease to the results of 17 normal subjects. Several CT value distributions were explored, including (1) that from the peripheral lung taken at TLC (with peels at 15 or 65 mm), (2) the ratio of (1) to that from the core of lung, and (3) the ratio of (2) to its FRC counterpart. We developed a fused-lasso logistic regression model that can automatically identify sub-intervals of –1000 HU and 0 HU over which a CT value distribution provides optimal discrimination between abnormal and normal scans.

Results: The fused-lasso logistic regression model based on (2) with 15-mm peel identified the relative frequency of CT values of over –1000 HU and –900 and those over –450 HU and –200 HU as a means of discriminating abnormal versus normal lung, resulting in a zero out-sample false-positive rate, and 15% false-negative rate of that was lowered to 12% by pooling information.

Conclusions: We demonstrated the potential usefulness of this novel quantitative imaging analysis method in discriminating ILD and/or emphysema from normal lungs.

Key Words: Fused lasso; interstitial lung disease; quantitative Computed Tomography; computer assisted.

Interstitial lung disease (ILD) is increasing in importance in part because of the aging population but also detection and/or incidence appear to be increasing. Data from the National Center for Health Statistics indicate that the age-adjusted mortality rate from pulmonary fibrosis has increased by 28.4% in men and by 41.3% in women between 1992 and 2014 (1).

Inter- and intrareader variability in the interpretation of radiographs for pulmonary fibrosis and pneumoconiosis has been long recognized as a potential issue for screening programs, epidemiologic studies, and medico-legal evaluations (2–4). Studies have found variable degrees of agreement for both parenchymal and pleural fibroses dependent on the extent of abnormalities and the training and medical specialty of the chest X-ray (CXR) readers (5–8). The National Institute for Occupational Safety and Health recommends using multiple, International Labour Organization radiographic pneumoconiosis classification system (ILO system)-trained readers and median profusion scores as the preferred reconciliation protocol to increase accuracy and precision in PA film classification (9,10). Clinical studies suggest that computed tomography (CT) imaging technology may become a

Acad Radiol 2016; 23:304–314

From the Department of Statistics and Actuarial Science, University of Iowa, Schaeffer Hall 241, Iowa City, IA 52242 (K.-S.C.); Department of Occupational and Environmental Health (F.J., M.A.M.); Department of Internal Medicine (A.G.); Department of Radiology and Biomedical Engineering (J.G., J.D.N.); Departments of Radiology, Internal Medicine and Biomedical Engineering (E.A.H.); Department of Radiology, University of Iowa, USA (B.T.); Department of Radiology, Seoul National University Hospital, Seoul, South Korea (C.H.L.); Department of Occupational and Environmental Health and Department of Epidemiology, University of Iowa, USA (L.J.F.). Received April 4, 2015; revised November 7, 2015; accepted November 16, 2015. **Address correspondence to:** K.-S.C. e-mail: kung-sik-chan@uiowa.edu

http://dx.doi.org/10.1016/j.acra.2015.11.013

gold standard for the evaluation of obstructive airways disease (11), but the literature is relatively limited on CT's use in quantifying and objectively characterizing patterns of subtle interstitial fibrosis.

High-resolution CT scanning detects finer anatomic detail than conventional CXR, and its superior sensitivity in diagnosing ILD, as well as lower potential for inter-reader variability, has been established in multiple studies (12–17). The present study utilizes recent advances in CT scanning, the spiral rapid CT with multidetector volumetric CT scanning, a time and cost-effective alternative to a single-detector row CT (18,19). Quantitative data analysis systems such as the Apollo (VIDA) software (20) allow for analysis of density histogram characteristics by lobe and region, similar to the ILO scheme.

Our approach leverages on the fact that scarring in the periphery of the lung results in measurable if not dramatic changes in the range, distribution, and relative frequency of HU summed over sets of voxels. Similarly, changes in the distribution and relative frequencies of voxels in differing ranges of HU are useful in detecting and quantifying other lung diseases such as emphysema. Our approach allows the identification of ranges of CT values, which lead to the best discrimination between fibrotic and normal lung. It is well known that the distribution of HU distributions and frequencies could be affected by potential confounders including gender, age, and body mass index (BMI). Therefore, we investigated that these confounding effects can be effectively mitigated by using the ratio function of the CT value distribution over the periphery of the lung to that over the core, including the mediastinal structures and hilum. The purpose of this study was to assess the potential of this novel CT technology in identifying and characterizing patterns of subtle interstitial changes especially in evaluation of normal aging lung and ILD.

MATERIALS AND METHODS

Data

The data for this study consisted of lung CT images from 17 subjects with no radiographic and functional abnormalities and from 52 subjects with ILD diagnosis confirmed by ILO review of CXR radiographs and functional testing. Table 1 summarizes demographic and clinical characteristics of the study subjects. The subjects with ILD were significantly different from normal subjects in the following demographics at 5% level: predominantly male (92% vs. 47% in the normal group), older (77.45 ± 8.41 years old vs. 40.35 ± 16.66), smokers (20% never smokers vs. 100%), presence of pleural plaques (38% vs. 0%) and majority having CT evidence of emphysema (88%) and bronchiectasis (83%). Table 2 contrasts several pulmonary test functions between the two groups. Compared to the normal subjects, subjects with ILD had lower total lung capacity (TLC) and residual volume (RV), but the differences were nonsignificant at 5% level. The subjects with ILD also had significantly lower functional vital capacity (FVC), forced expiratory volume in 1 s (FEV_1), diffusion capacity for the lungs of carbon monoxide (DLCO), and FEV_1/FVC, all of which, except FEV_1/FVC, were adjusted for age, gender, and body composition (and additionally hemoglobin level for DLCO). CT imaging was performed on a Siemens SOMATOM Definition FLASH (Siemens Healthcare, Erlangen, Germany). The CT scanning protocol consisted of obtaining multidetector CT (MDCT) images at TLC and at functional residual capacity (FRC) supine, and one that is TLC prone. All results reported previously are based on data-measured supine. Unique breathing instructions were required to obtain appropriate lung volume images. The MDCT acquisition parameters

TABLE 1. Demographic and Clinical Characteristics of Study Populations

	ILD	Normal	*P* Value
N	52	17	
Sex			0.0002
Male	48 (92%)	8 (47%)	
Female	4 (8%)	9 (53%)	
Age, year	77.35 ± 8.41	40.35 ± 16.66	<0.0001
Race			0.16
African American	0 (0%)	1 (6%)	
Caucasian	50 (96%)	16 (94%)	
Hispanic	2 (4%)	0 (0%)	
BMI	28.81 ± 5.25	25.65 ± 3.69	0.025
Smoking history			0.00006
Never	20 (38%)	17 (100%)	
Former	25 (48%)	0 (0%)	
Current	7 (14%)	0 (0%)	
Pack-years	42.39 ± 34.47	0 ± 0	0.003
Plaques			<0.0001
No	32 (62%)	17 (100%)	
Yes	20 (38%)	0 (0%)	
Emphysema			<0.0001
No	6 (12%)	17 (100%)	
Yes	46 (88%)	0 (0%)	
Bronchiectasis			<0.0001
No	9 (17%)	17 (100%)	
Yes	43 (83%)	0 (0%)	

BMI, body mass index; ILD, interstitial lung disease.

TABLE 2. Pulmonary Function in Patients with ILD Compared to Normal Subjects

	ILD	Normal	*P* Value
N	52	17	
TLC	96.41 ± 22.20	103.41 ± 8.52	0.21
RV	96.08 ± 31.26	100.18 ± 17.44	0.61
FVC	82.71 ± 24.11	109.76 ± 19.62	<0.0001
FEV_1	78.40 ± 24.40	112.82 ± 20.66	<0.0001
DLCO	62.25 ± 19.91	126.71 ± 15.81	<0.0001
FEV_1/FVC	68.88 ± 14.14	82.94 ± 6.43	0.0002

DLCO, diffusion capacity for the lungs of carbon monoxide; FEV_1, forced expiratory volume in 1 s; FVC, functional vital capacity; RV, residual volume; TLC, total lung capacity.