Automated Classification of Usual Interstitial Pneumonia Using Regional Volumetric Texture Analysis in High-Resolution Computed Tomography

Adrien Depeursinge, PhD,*† Anne S. Chin, MD,* Ann N. Leung, MD,* Donato Terrone, MD,* Michael Bristow, MD,* Glenn Rosen, MD,* and Daniel L. Rubin, MD, MS*‡

**Objective:** We propose a novel computational approach for the automated classification of classic versus atypical usual interstitial pneumonia (UIP).

**Materials and Methods:** Thirty-three patients with UIP were enrolled in this study. They were classified as classic versus atypical UIP by a consensus of 2 thoracic radiologists with more than 15 years of experience using the American Thoracic Society evidence-based guidelines for computed tomography diagnosis of UIP. Two cardiothoracic fellows with 1 year of subspecialty training provided independent readings. The system is based on regional characterization of the morphological tissue properties of lung using volumetric texture analysis of multiple-detector computed tomography images. A simple digital atlas with 36 lung subregions is used to locate texture properties, from which the responses of multidirectional Riesz wavelets are obtained. Machine learning is used to aggregate and to map the regional texture attributes to a simple score that can be used to stratify patients with UIP into classic and atypical subtypes.

**Results:** We compared the predictions on the basis of regional volumetric texture analysis with the ground truth established by expert consensus. The area under the receiver operating characteristic curve of the proposed score was estimated to be 0.81 using a leave-one-patient-out cross-validation, with high specificity for classic UIP. The performance of our automated method was found to be similar to that of the 2 fellows and to the agreement between experienced chest radiologists reported in the literature. However, the errors of our method and the fellows occurred on different cases, which suggests that combining human and computerized evaluations may be synergistic.

**Conclusions:** Our results are encouraging and suggest that an automated system may be useful in routine clinical practice as a diagnostic aid for identifying patients with complex lung disease such as classic UIP, obviating the need for invasive surgical lung biopsy and its associated risks.

**Key Words:** computer-assisted image processing, usual interstitial pneumonia, multidetector computed tomography, computer-assisted diagnosis

Received for publication June 7, 2014; and accepted for publication, after revision, October 8, 2014.

From the *Department of Radiology, School of Medicine, Stanford University, Stanford, CA; †Department of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland; and ‡Section on Biomedical Informatics Research, Department of Medicine, School of Medicine, Stanford University, Stanford, CA.

Supported by the Swiss National Science Foundation (P300P2_142283, PZ00P2_154891, and 205320_141300/1) and the National Cancer Institute, National Institutes of Health (U01-CA-142555).

Conflicts of interest and sources of funding: none declared.

Supplemental digital contents are available for this article: Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s Web site (www.investigativeradiology.com).

Reprints: Adrien Depeursinge, PhD, Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), TechnoArk 3, 3960 Sierre, Switzerland. E-mail: adrien.depeursinge@hevs.ch.

Copyright © 2014 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0020-9996/15/0000-0000

Investigative Radiology • Volume 00, Number 00, Month 2015 www.investigativeradiology.com
There were 15
www.investigativeradiology.com
Investigative Radiology
using cubic spline inter-

if the summed value of its own neigh-

/C2

Fig. 2).

X

as well as 3 diagonals

Y

YZ

and

The user initiated a region growing al-

© 2014 Wolters Kluwer Health, Inc. All rights reserved.

guidelines for CT diagnosis of IPF/UIP listed in Table 1. The

2

classical UIP on the basis of the American Thoracic Society evidence worked in consensus to clas-

ic UIP, the latter requiring further evaluation (eg, biopsy) to confirm the diagnosis of IPF. To the best of our knowledge, our work is the first attempt to automatically differentiate the UIP subtypes using computational methods.

In this article, we used a basic anatomic atlas to define and extract morphological tissue properties (ie, 3D texture) from 36 anatomical subregions of the lung parenchyma. The latter were aggregated to create computational models of the radiological phenotypes for classic and atypical UIP. We hypothesized that the characterization of these radiologic phenotypes differentiates the UIP subtypes, specifically classic versus atypical UIP. To produce an automated system for this classification task, machine learning was used to derive a score from the regional texture attributes. This score identified patients with typical versus atypical UIP, the latter requiring further evaluation (eg, biopsy) to confirm the diagnosis of IPF. To the best of our knowledge, our work is the first attempt to automatically differentiate the UIP subtypes using computational methods.

MATERIALS AND METHODS

Data set

Computed tomographic examinations of 33 patients with biopsy-proven UIP/IPF from Stanford Hospital and Clinics were retrospectively reviewed. The examinations were performed in compliance with national legislation and Declaration of Helsinki guidelines, with institutional review board waiver of informed consent for this retrospective analysis. Two thoracic radiologists with more than 15 years of experience worked in consensus to classify each patient as classic versus atypical UIP on the basis of the American Thoracic Society evidence–based guidelines for CT diagnosis of IPF/UIP listed in Table 1. There were 15 patients categorized as having a classic UIP and 18 patients with an atypical UIP appearance on CT. Additional readings were performed by 2 cardiothoracic fellows, each having 1-year specialty training to establish an optimal human performance. A volumetric CT scan was available for each patient, which was acquired within the year of the biopsy-proven diagnosis. The standard routine protocols yielded a computed tomographic dose index of 19.51 ± 8.81 mGy and a dose-length product of 622.11 ± 269.97 mGy·cm. The CT scans were reconstructed with slice thicknesses of 0.625 mm (1), 1 mm (15), 1.25 mm (14), and 2 mm (3). The spacings between the slices are 0.6 mm (3), 1 mm (13), 1.25 mm (14), as well as 2 mm (3), and the pixel spacings are in the range of 0.59 to 0.82 mm. All volumes were resampled to have isotropic cubic voxels of 0.59 × 0.59 × 0.59 mm³ using cubic spline interpolation. This ensures that the physical dimensions (ie, image scales and directions) are comparable between the patients.

3D Anatomical Atlas of the Lungs

A simple anatomical atlas of the lung served as a 3D reference to localize the various lung tissue types. The locations were chosen according to the predominant pattern sites of classic UIP (Table 1). First, a semiautomated segmentation of the lung volumes was carried out with a graphical user interface. The user initiated a region growing algorithm from a seed point placed inside the lungs. The growing 3D region mimics the propagation of air in lungs, where each neighboring voxel is added to the region if the summed value of its own neighbors differs by less than a threshold value defined by the user. The lungs are divided vertically into apical, middle lung, and basal regions on the basis of the center of mass of . The axial division results into central, intermediate, and peripheral lung regions. Left, right, anterior, and posterior regions are defined on the basis of the intersections of the coronal and sagittal planes. The intersections of all 10 lung divisions result in 36 subregions.

3D Texture Quantification Using Riesz Wavelets

Three-dimensional multiscale Riesz filter banks were used to characterize the morphological properties of the lung parenchyma in volumetric CT. These filters are advantageous for texture characterization because they quantify the local amount of directional image patterns at multiple scales. Second-order Riesz wavelets were used, yielding 6 filters per image scale that are oriented along the main image directions X, Y, and Z as well as 3 diagonals XY, XZ, and YZ (Fig. 2). Likewise, Riesz wavelets allow for a complete coverage of image scales and directions. We hypothesized that the local morphological tissue properties of the normal lung, ground glass, reticulation, and honeycombing can be expressed as the combinations of the responses of the oriented filters. The filters were implemented at multiple scales to analyze both fine morphological structures and coarser reticulations of the lung parenchyma.

| Peripheral, basal predominance | Upper or midlung predominance |
| Reticular abnormality | Peribronchovascular predominance |
| Honeycombing with or without traction bronchiectasis | Extensive ground-glass abnormality (extent > reticular abnormality) |
| Absence of features listed as inconsistent with UIP pattern | Profuse micronodules (bilateral, predominantly upper lobes) |
| | Discrete cysts (multiple, bilateral, away from areas of honeycombing) |
| | Diffuse mosaic attenuation/air trapping (bilateral, in 3 or more lobes) |
| | Consolidation in bronchopulmonary segment(s)/lobe(s) |

TABLE 1. Radiological Criteria for UIP

<table>
<thead>
<tr>
<th>Classic UIP (All Required)</th>
<th>Inconsistent With UIP (Any)</th>
</tr>
</thead>
</table>
| Peripheral, basal predomi-
| nance                    | Upper or midlung predomi-
| nance                    | Peribronchovascular predomi-
| nance                    | Extensive ground-glass abnor-
| mality (extent > reticular |
| abnormality)             | Profuse micronodules (bi-
| (bilateral, predominantly |
| upper lobes)             | Discrete cysts (multiple, |
| (bilateral, away from a-
| reas of honeycombing)    | Diffuse mosaic attenuation/|
|                          | air trapping (bilateral, |
|                          | in 3 or more lobes)       |
|                          | Consolidation in broncho-
|                          | pulmonar y segment(s)/lo-
|                          | be(s)                     |

UIP indicates usual interstitial pneumonia.

FIGURE 1. Common parenchymal appearances of UIP in CT.
The prototype regional distributions of the morphological tissue properties of classic versus atypical UIPs were learned using support vector machines (SVMs). The SVM is a supervised machine learning algorithm that can learn the complex relationship between a group of variables (ie, the vector $v$) and the presence or absence of a class from an ensemble of examples called the training set. Once the SVM model has been built from the example cases, it can predict the class of an unseen case with a confidence score (called computer score thereafter). The group of variables feeding SVMs consisted of the responses (ie, energies) of the multiscale Riesz filters in each of the 36 anatomical regions of the lungs (Fig. 3). The size of the vector $v$ regrouping the responses of the 6 Riesz filters at 4 scales from the 36 regions was equal to 864.

To compare Riesz wavelets with other features that could capture the radiological phenotype of diffuse lung disease, 2 different feature groups were extracted for each region to provide a baseline performance:

1. Histogram bins of the gray levels in the extended lung window $[-1000; 600]$ Hounsfield units (HU) and 3D gray-level co-occurrence matrices (GLCM). Statistical measures from GLCMs are popular texture attributes that were used by several studies in the literature to characterize the morphological properties of lung tissue associated with interstitial lung diseases. They consist in counting the co-occurrence of voxels with identical gray level values that are separated by a distance $d$, which results in a co-occurrence matrix. Eleven statistics were extracted from these matrices as texture attributes. The choices of $d$ and the number of gray levels were optimized by considering values in $\{-3; 3\}$ and $\{8, 16, 32\}$, respectively. The size of the vector of attributes $v$ was 540 for the gray-level histogram attributes (called HU thereafter) and 396 for the GLCM attributes.

RESULTS

A leave-one-patient-out cross-validation evaluation was used to estimate the performance of the proposed approach. The leave-one-patient-out cross-validation consisted of using all patients but 1 to train the SVM model and to measure the prediction performance on the remaining test patient. The prediction performance was then averaged over all possible combinations of training and test patients. Receiver operating characteristic (ROC) curves of the system's performance in classifying between classic and atypical UIP are shown in Figure 4 for different feature groups and their combinations. The ROC curves were obtained by varying the decision threshold between the minimum and

<table>
<thead>
<tr>
<th>Vertical</th>
<th>Axial</th>
<th>Coronal</th>
<th>Sagittal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apical, central, basal</td>
<td>Peripheral, middle, axial</td>
<td>Left, right</td>
<td>Anterior, posterior</td>
</tr>
</tbody>
</table>

The lungs are split perpendicularly to 4 axes. Images can be viewed online in color at www.investigativeradiology.com.
the maximum of the score provided by the SVMs. A maximum area under the ROC curve (AUC) of 0.81 was obtained with the regional Riesz attributes, which suggests that prediction was correct for more than 4 of 5 patients. The performance of HU and GLCM attributes was close to random (0.54 and 0.6 for HU and GLCMs, respectively). On the other hand, predictive SVM models based on the responses of the Riesz filters, averaged over the entire lungs, had an AUC of 0.72.

Our system's performance was also compared with the interpretations of 2 fellowship-trained cardiothoracic fellows, each having 1 year of experience. Interobserver agreement was assessed with the Cohen \( \kappa \) statistics and the percentage of agreement (ie, number of times the 2 observer agreed). The comparisons are detailed in Tables 3 and 4. The operating points of the 2 independent observers are reported in Figure 4 (top right). A detailed analysis of the 6 cases that were misclassified by our system is shown in the Supplcumentary Table, Supplemental Digital Content 1, http://links.lww.com/RLI/A189.

Overall, 7 incorrect predictions were made by the fellows and 6 incorrect predictions by the computer. The fellows and the computer made only 2 common errors (cases 1 and 13).

DISCUSSION

We developed a novel computational method for the automated classification of classic versus atypical UIP based on regional volumetric texture analysis. This constitutes, to the best of our knowledge, a first attempt to automatically differentiate the UIP subtypes with computational methods. An SVM classifier yielded a score that predicts if the UIP is classic or atypical. The classifier was based on a group of attributes that characterize the radiological phenotype of the lung parenchyma, specifically the morphological properties (ie, texture) of the parenchyma. Because diffuse lung diseases can vary in the distribution and severity of abnormalities throughout the lungs, we extracted our quantitative image features from 36 anatomical regions of the lung. To our knowledge, adding this spatial characterization to the computational model is also innovative, and it is particularly relevant for assessing diffuse lung disease.
System Performance and Comparison With Chest Specialists

The proposed approach successfully predicted UIP subtypes for more than 4 of 5 patients (AUC, 0.81), with high specificity for classic UIPs. This performance is encouraging and suggests that an automated system may be useful in routine clinical practice for radiologists with limited experience in interpretation of diffuse lung diseases. The detailed case analyses (see Supplemental Table, Supplemental Digital Content 1, http://links.lww.com/RLI/A189) reveal that, with the exception of 1 case (case 20), the computer score was able to sort and grade the patients from highly atypical (eg, patients 2 and 3) to classic (eg, patient 33). The observer agreement between the computer and the consensus was associated with a κ value of 0.63 and a percentage of agreement of 81.82, which is considered as being substantial in the literature. The performance of the computer was found to be comparable with that of the cardiothoracic fellows with 1 year of specialty training (Fig. 4, top right). The agreement between each of the 2 fellows and the computer was moderate (ie, 0.51 and 0.57 κ values and 75.8% and 78.8% agreement) when compared with their excellent interreader agreement (ie, κ value and percentage of agreement of 0.94 and 97%, respectively). However, the agreement between each of the fellows with the consensus classification (our criterion standard) is considerably lower (ie, 0.63 and 0.57 κ values and 81.8% and 78.8% agreement). Interestingly, the agreement between the computer and the consensus was comparable with the agreement between the fellows and the consensus, suggesting that, although the algorithm can still be improved in terms of agreement with the consensus, the current method performs at least as well as cardiothoracic fellows. The performance was also consistent with the interreader agreement between the 2 expert radiologists (ie, 4 and 8 years of experience, κ value of 0.67) reported by Assayag et al in their study. Because the target application for this algorithm is for general radiologists, the proposed system could considerably improve their performance, a study which we plan to undertake.

Detailed Case Analysis of the System’s Performance

Our cohort had a much higher proportion of atypical UIPs accounting for more than half the cases, which likely reflects a selection bias. In current practice, only cases that are atypical have histologic/pathologic confirmation, which was required for this cohort analysis. Many of the cases had subtle findings that required adhering very strictly to the guidelines for classification by the consensus, which may account for 3 of 6 erroneous predictions of atypical UIP by the computer (see Table 5 and the Supplemental Table, Supplemental Digital Content 1, http://links.lww.com/RLI/A189). We believe that the algorithm would perform better in a real-world setting, which does not reflect unusual cases referred to a quaternary academic institution.

Errorneous predictions of classic UIP by the computer occurred in 2 patients (patients 24 and 27) where bronchiectasis was mistaken for honeycombing (Table 5). This can be a difficult task even for experienced chest radiologists, but it will constitute a target for further improvement of the system. One patient with diffuse disease without definite basilar predominance was also misclassified as classic UIP by the computer.

Overall, 7 incorrect predictions were made by the cardiothoracic fellows and 6 incorrect predictions by the computer. Combining these predictions results in only 2 incorrect predictions of 33 (ie, 94% correct predictions), demonstrating the potential benefit of computer-assisted diagnosis in the setting of UIP.

Comparison of the Quantitative Image Features

The gray-level histograms (ie, HU) failed to predict the UIP subtypes, and 3D GLCMs provided a slightly better specificity for classical UIPs. The 3D Riesz wavelets provided an AUC of 0.81, which was found to be higher than that of 3D GLCMs and HU. The combination of Riesz and HU did not lead to a performance improvement. This can be explained by the fact that the distinction between tissue types associated with UIP requires a fine characterization of the volumetric morphologic tissue properties (ie, 3D texture) that only 3D Riesz wavelets were able to extract. The prototype gray-level distributions of ground glass, reticular, and honeycombing may not be sufficiently distinct to accurately discriminate among them, as shown in Figure 1. The comparison between global and regional texture analysis (Fig. 4, right) highlights the importance of localizing tissue texture properties, which is consistent with the medical knowledge.

Overall, we believe that our automated system may be useful in general radiology practices, accounting for the majority of sites where patients with IPF are initially examined, to accurately identify patients with classic UIP for which an unnecessary surgical biopsy can be avoided. The high-resolution imaging of lung disease afforded by modern-day CT provides accurate 3D anatomic/pathologic fidelity in patients with classic UIP features. A confident CT diagnosis of classic UIP can eliminate the need for an invasive surgical biopsy to confirm this diagnosis; this underscores the importance of an accurate and reliable interpretation of the CT examination. This task is often challenging, particularly in less-specialized practice centers without access to

Table 3. Interobserver Agreement (Cohen κ Statistics)

<table>
<thead>
<tr>
<th></th>
<th>Consensus</th>
<th>Fellow 1</th>
<th>Fellow 2</th>
<th>Computer (Regional Riesz Texture Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fellow 1</td>
<td>0.629</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fellow 2</td>
<td>0.569</td>
<td>0.9374</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Computer (regional Riesz)</td>
<td>0.633</td>
<td>0.506</td>
<td>0.569</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4. Interobserver Agreement (Percentage of Times That the Observers Agreed)

<table>
<thead>
<tr>
<th>% Agreement</th>
<th>Consensus</th>
<th>Fellow 1</th>
<th>Fellow 2</th>
<th>Computer (Regional Riesz Texture Analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>100</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fellow 1</td>
<td>81.82</td>
<td>100</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fellow 2</td>
<td>78.79</td>
<td>96.97</td>
<td>100</td>
<td>—</td>
</tr>
<tr>
<td>Computer (regional Riesz)</td>
<td>81.82</td>
<td>75.76</td>
<td>78.79</td>
<td>100</td>
</tr>
<tr>
<td>Case</td>
<td>Computer Score</td>
<td>Consensus</td>
<td>Fellow 1</td>
<td>Fellow 2</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>1</td>
<td>-1.13</td>
<td>classic UIP</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>-0.82</td>
<td>classic UIP</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>-0.34</td>
<td>classic UIP</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>20</td>
<td>0.01</td>
<td>atypical UIP</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>24</td>
<td>0.1</td>
<td>atypical UIP</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>27</td>
<td>0.24</td>
<td>atypical UIP</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Coronal and axial views of the lung parenchyma in computed tomographic scans are shown with a window level of ~500 HU and a width of 1400 HU. UIP indicates usual interstitial pneumonia. Images can be viewed online in color at www.investigativeradiology.com.
experts experienced in interstitial lung disease. Nevertheless, reduction of needless invasive biopsies is desired when possible to reduce costs and, more importantly, to avoid unnecessary morbidity/mortality in patients with UIP. We recognize several limitations of the current work, including the small number of cases included in the study and the use of a very simple digital atlas of the lung anatomy. Future work will include validating our results in an independent data set and performing regional learning of 3D rotation-covariant texture models of normal, ground glass, reticular, and honeycombing.27,31

REFERENCES
24. Zrimec T, Wong JSJ. Improving computer aided disease detection using knowledge of disease appearance. Paper presented at: 12th World Congress on Health (Medical) Informatics; August 20–24, 2007; Brisbane, Australia.