OPEN

Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images

Andreas Christe, MD, Alan A. Peters, MD,* Dionysios Drakopoulos, MD,* Johannes T. Heverhagen, PhD,* Thomas Geiser, MD,† Thomai Stathopoulou, PhD,‡ Stergios Christodoulidis, PhD,‡ Marios Anthimopoulos, PhD,‡ Stavroula G. Mougiakakou, PhD,‡ and Lukas Ebner, MD**

Objectives: The objective of this study is to assess the performance of a computer-aided diagnosis (CAD) system (INTACT system) for the automatic classification of high-resolution computed tomography images into 4 radiological diagnostic categories and to compare this with the performance of radiologists on the same task.

Materials and Methods: For the comparison, a total of 105 cases of pulmonary fibrosis were studied (54 cases of nonspecific interstitial pneumonia and 51 cases of usual interstitial pneumonia). All diagnoses were interstitial lung disease board consensus diagnoses (radiologically or histologically proven cases) and were retrospectively selected from our database. Two subspecialized chest radiologists made a consensual ground truth radiological diagnosis, according to the Fleischner Society recommendations. A comparison analysis was performed between the INTACT system and 2 other radiologists with different years of experience (readers 1 and 2). The INTACT system consists of a sequential pipeline in which first the anatomical structures of the lung are segmented, then the various types of pathological lung tissue are identified and characterized, and this information is then fed to a random forest classifier able to recommend a radiological diagnosis.

Results: Reader 1, reader 2, and INTACT achieved similar accuracy for classifying pulmonary fibrosis into the original 4 categories: 0.6, 0.54, and 0.56, respectively, with $P > 0.45$. The INTACT system achieved an F-score (harmonic mean for precision and recall) of 0.56, whereas the 2 readers, on average, achieved 0.57 ($P = 0.991$). For the pooled classification (2 groups, with and without the need for biopsy), reader 1, reader 2, and CAD had similar accuracies of 0.81, 0.70, and 0.81, respectively. The F-score was again similar for the CAD system and the radiologists. The CAD system and the average reader reached F-scores of 0.80 and 0.79 ($P = 0.898$).

Conclusions: We found that a computer-aided detection algorithm based on machine learning was able to classify idiopathic pulmonary fibrosis with similar accuracy to a human reader.

Key Words: idiopathic pulmonary fibrosis, computed tomography, nonspecific interstitial pneumonia, usual interstitial pneumonia, interstitial lung diseases, artificial intelligence, machine learning, computer-assisted diagnosis

(*Invest Radiol* 2019;54: 627–632)

Received for publication February 15, 2019; and accepted for publication, after revision, March 19, 2019.

From the Departments of *Diagnostic, Interventional, and Pediatric Radiology, and †Pulmonology, Inselspital, Bern University Hospital, University of Bern; and ‡ARTORG Center for Biomedical Engineering Research, Bern University, Bern, Switzerland.

Conflicts of interest and sources of funding: This study was funded by the Swiss National Science Foundation, Roche Pharma Ltd, and the Lindenhof Foundation.

The authors declare no conflicts of interest.

Correspondence to: Andreas Christe, MD, Department of Diagnostic, Interventional, and Pediatric Radiology, Inselspital, Bern University Hospital, Freiburgstrasse 10, 3010 Bern, Switzerland. E-mail: andreas.christe@insel.ch.

ISSN: 0020-9996/19/5410–0627
DOI: 10.1097/RLI.0000000000000574

The differential diagnosis of idiopathic interstitial pneumonias is largely based on a number of uniform criteria and guidelines that have been proposed by the American Thoracic Society and the European Respiratory Society in Wells[1] and Travis et al,[2] and have been recently updated by the white paper recommendations of the Fleischner Society.[3] Table 1 portrays the radiological patterns and how these associate with the idiopathic pulmonary fibrosis (IPF) diagnosis. Typically, radiologists screen the patient's high-resolution computed tomography (HRCT) for these usual interstitial pneumonia (UIP) patterns (ie, typical UIP, probable UIP, indeterminate for UIP, and non-IPF), and along with a clinical board of pneumonologists and histopathologists, they decide on the diagnosis. In most of the cases, if any other than typical or probable UIP pattern is identified, additional invasive procedures, such as transbronchial or surgical biopsy, are required for the final diagnosis.[4] Early diagnosis is crucial for making fundamental treatment decisions, particularly for IPF or nonspecific interstitial pneumonia (NSIP), whereas a misdiagnosis may lead to life-threatening complications. It is well understood that biopsy-proven NSIP cases improve under the treatment of steroids and that pirfenidone increases the survival rate of IPF patients.[5,6]

Correct pattern identification on computed tomography (CT) images plays a central role in the diagnosis and further treatment of patients with interstitial lung disease (ILD), even more so in conjunction with the recently proposed Fleischner Society recommendations. These guidelines aim to expand the role of CT for the diagnosis of IPFs. However, pattern recognition on CT can be challenging, even for subspecialized chest radiologists. Radiologists find it difficult to identify such CT patterns in HRCTs, and interreader variability is high.[7] Performing a surgical biopsy exposes the patient to the potential risks of general anesthesia and requires at least a short hospital stay, which increases common health care costs. Moreover, lung biopsies have a potential risk of bleeding or even death and are not always diagnostic. Thus, supporting tools, such as computer-aided diagnosis (CAD) systems for CT pattern detection and disease classification, have the potential to improve the radiological diagnosis of pulmonary fibrosis. In particular, the combination of machine learning approaches, advanced segmentation algorithms, and clinical data (sex, age, etc) can provide the opportunity for the development of a lung fibrosis CAD system.[8–11] Computer-aided diagnosis systems have been thoroughly investigated for a variety of clinical applications, and previous investigators have reported their usefulness in various scenarios.[12–19]

Despite the extensive research that has been undertaken on the quantification of pathological lung tissue on CT images,[20–24] no pulmonary diagnosis pipeline or system has yet been proposed that is able to suggest automatically a final diagnostic decision for a case. Van Ginneken et al[25] proposed an automatic method for the segmentation of lung fields into 42 regions, followed by a classification step, which assigns to each region a confidence value for being abnormal. The product of the individual confidence values provides a global diagnosis on the abnormality of the whole lung. Fukushima et al[26] proposed the use of an artificial neural network that combines 10 clinical parameters with 23 HRCT features to provide a final differential diagnosis.

TABLE 1. Diagnostic Categories of UIP, Based on CT Patterns[3]

	Features	Characteristics
Typical UIP CT pattern	Reticulation, honeycombing, absence of features suggesting a non-IPF diagnosis	Basal and subpleural
Probable UIP CT pattern	Reticulation, absence of features suggesting a non-IPF diagnosis	Basal and subpleural
CT pattern indeterminate for UIP	Reticulation with inconspicuous features suggestive of a non-UIP pattern	Variable or diffuse
CT features most consistent with a non-IPF diagnosis	Any of the following: predominant consolidation, extensive pure ground glass opacity (without acute exacerbation), extensive mosaic attenuation with extensive sharply defined lobular air trapping on expiration, diffuse nodules or cysts	Upper, mid-lung, or peribronchovascular predominant or subpleural sparing

UIP indicates usual interstitial pneumonia; CT, computed tomography.

However, the HRCT features used were not computed automatically but rated manually by radiologists.

The purpose of this study was to assess the performance of the INTACT system—a CAD designed for the automatic classification of IPF cases into radiological diagnostic CT patterns, as based on HRCT chest images and clinical markers.

MATERIALS AND METHODS

Databases

For the purposes of this study, multiple databases were used for the training and evaluation of the different components of the system. More details about each database are contained in the following paragraphs.

Lung Tissue Research Consortium Database

The Lung Tissue Research Consortium database (LTRC-DB; https://ltrcpublic.com) is a resource program of the National Heart, Lung, and Blood Institute that provides CT scans, as well as biospecimens to qualified investigators for use in their research. The LTRC was originally created in 2005 by the National Institutes of Health and is composed of 4 clinical centers from around the United States: Mayo Clinic Rochester, University of Michigan–Ann Arbor, University of Pittsburgh, and Temple University. Each center contributes to the recruitment and enrollment of protocol-eligible participants, as well as to the procurement of the data. This library contains cases with different lung diseases along with annotations of the lung parenchyma, the airways, and pathological tissue. There are more than 100 cases with proven ILD diagnosis.

Multimedia Database of Interstitial Lung Diseases

The multimedia database for ILDs (MD-ILD; http://medgift.hevs.ch/wordpress/databases/ild-database/)[27] was developed within the framework of the Talisman project at the University Hospital of Geneva and is made publicly available. The database consists of HRCT image series of 10-mm slice spacing with annotated regions of pathological lung tissue and the lung parenchyma along with clinical parameters from patients with pathologically proven diagnoses of ILDs. The library contains cases from 128 patients affected with 1 of the 13 histological diagnoses of ILDs, 108 image series with 1946 delineated polygons (more than 3 million pixels) of annotated lung tissue patterns, as well as a comprehensive set of 99 clinical parameters related to ILDs.

Inselspital Interstitial Lung Diseases Database

The Inselspital ILD database (INSEL-DB) can be split into 2 parts based on the accompanying annotations. The first part consists of 60 unique HRCT lung scans with annotated tissue from 2 radiologists (INSEL-DB-Seg), and the second consists of 105 HRCT lung scans provided by the ILD board of Bern University Hospital, in which pneumologists, radiologists, and pathologists have diagnosed the patients according to international guidelines[3] (INSEL-DB-Diag: 54 NSIP and 51 IPF cases). The first database (INSEL-DB-Seg) was used for the training of the pathological tissue segmentation, whereas the second (INSEL-DB-Diag) was used for the diagnosis. Computed tomography scans were retrospectively collected with irreversible data anonymization from October 2015 to June 2017. Images were acquired on a third-generation dual-source CT (Somatom Definition Flash; Siemens Healthineers, Forchheim, Germany). Computed tomography scans were performed during the end-inspiratory phase using the breath-hold technique with patients in the supine position, from the apex of the lung to the costodiaphragmatic recess with a slice thickness of 1 mm. A tube voltage from 100 to 120 kVp and a reference mAs from 100 to 120 were applied. On the 128-detector scanner, a collimation of 128 × 0.6 mm was used, with a pitch of 0.6. A slice thickness of 1 mm was reconstructed, with a SAFIRE (Sinogram Affirmed Iterative Reconstruction) level 3 and with a hard lung kernel of I70f. Moreover, any associated clinical and biochemical data (eg, sex, age, smoking history, duration of illness, lung function tests, results of blood tests) were gathered to investigate whether they correlate with the actual diagnosis of each case, so they can provide information additional to the radiological data. Institutional board approval for the diagnosis was waived due to the retrospective collection of the patients.

The median age in the UIP group was 70 years (range, 49–84 years), and the group consisted of 11 female and 40 male patients. This group included 38 IPF cases, 8 cases of rheumatoid arthritis, and 5 connective tissue disease patients, with accompanying pulmonary fibrosis and UIP patterns. In the NSIP group, the median age was 64 years (range, 38–83 years), with 20 female and 34 male patients, consisting of 5 idiopathic NSIP patients and 49 cases with the known etiology of NSIP: 24 hypersensitivity pneumonitis, 7 antisynthetase syndrome, 6 medication related, 4 rheumatoid arthritis, 3 systemic sclerosis, 3 sarcoidosis, and 2 Sjögren syndrome patients.

Radiologists HRCT Readout

Two chest radiology specialists classified the cases into the 4 UIP CT patterns of Table 1 through consensus, according to the Fleischner Society recommendations,[3] to establish the ground truth. The radiologists first reviewed and classified all cases independently and then met to discuss the cases without agreement to determine the classification through consensus. This radiological consensus represented the ground truth for further calculations. Moreover, a radiologist with 10 years (reader 1) of experience in chest imaging and a chest fellow with 4 years (reader 2) of experience read the images on a Picture Archiving and Communication System (PACS Sectra, Linköping,

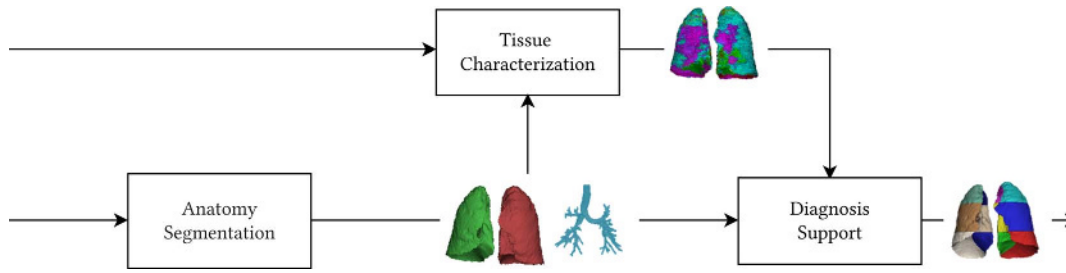

FIGURE 1. CAD (INTACT system) overall scheme.

Sweden). Lung windows settings were used to read the hard kernel reconstructions (I70f). Both radiologists were blinded to the ground truth and had to classify the cases into the 4 categories.

CAD System HRCT Readout

The INTACT system was implemented for the diagnosis support of IPF cases using HRCT images. INTACT was designed by biomedical engineers and trained by chest radiologists and pulmonologists from our University. Such a CAD system typically consists of 3 stages: (1) lung anatomy segmentation, (2) lung tissue characterization, and (3) diagnosis. In Figure 1, an overview of the overall pipeline is presented. In the following sections, a description of each component is presented.

Anatomy Segmentation

For the anatomy segmentation, a well-established algorithm from the literature was used to segment the airways and the lung parenchyma. This algorithm has been implemented in house, and its hyperparameters were defined using the LTRC-DB. This algorithm does not include a training component, and it is based on simple region growing, thresholding, and morphological operations. The pipeline is based on publications[28–31] and consists of the following steps: (1) extraction of large airways, (2) segmentation of lung regions; (3) separation of the left and right lungs; and (4) morphological 3-dimensional smoothing.

Tissue Characterization

After the segmentation of the lung parenchyma has been completed, a convolutional neural network (CNN) is used for the characterization of the pathological lung tissue. For the implementation, training, and evaluation of this step, the pathological tissue segmentations from MD-ILD and INSEL-DB-Seg were used. Examples of the ground truth annotation polygons of the pathological tissue are presented in Figure 2 (top row).

Interstitial lung diseases typically consist of an admixture of the following basic tissue pathologies: reticulation, honeycombing, ground glass opacity (GGO), consolidation, micronodules, and normal lung. The proposed system for the pathological tissue segmentation uses texture classification schemes to detect, classify, and calculate the extent of pathological tissue on CT images. The suggested quantification scheme for lung disease takes as input a section of a 2-dimensional CT slice of interest and uses a purely CNN scheme to calculate a corresponding label map with a single tissue class for each pixel. The proposed architecture is designed in such a way that the pixels are the training samples instead of the CT images. Thus, the number of training samples is of the order of 10,[6] and therefore, training of such a deep network is possible. We adopted a 5-fold cross-validation scheme to ensure the validity of the results stratified on a patient level. The balanced accuracy of the proposed CNN averaged over all folds was 81.8%. A few examples are presented in Figure 2. More details about the design and training of the CNN are presented in Anthimopoulos et al.[32]

Diagnosis

The diagnosis support module is the final step of the pipeline, in which all previous outputs are aggregated to achieve a final diagnosis. The diagnosis labels from INSEL-DB-Diag along with the associated clinical parameters (age, sex, etc) were used for the evaluation of this step.

To calculate the distribution of the different pathological tissue types in the different areas of the lung, an additional step was used that divided each lung into 6 regions (upper-central/peripheral, middle-central/peripheral, lower-central/peripheral). For this step, a volume-based split was used for the upper, middle, and lower segmentation, whereas for the central and peripheral segments, k-means clustering was applied

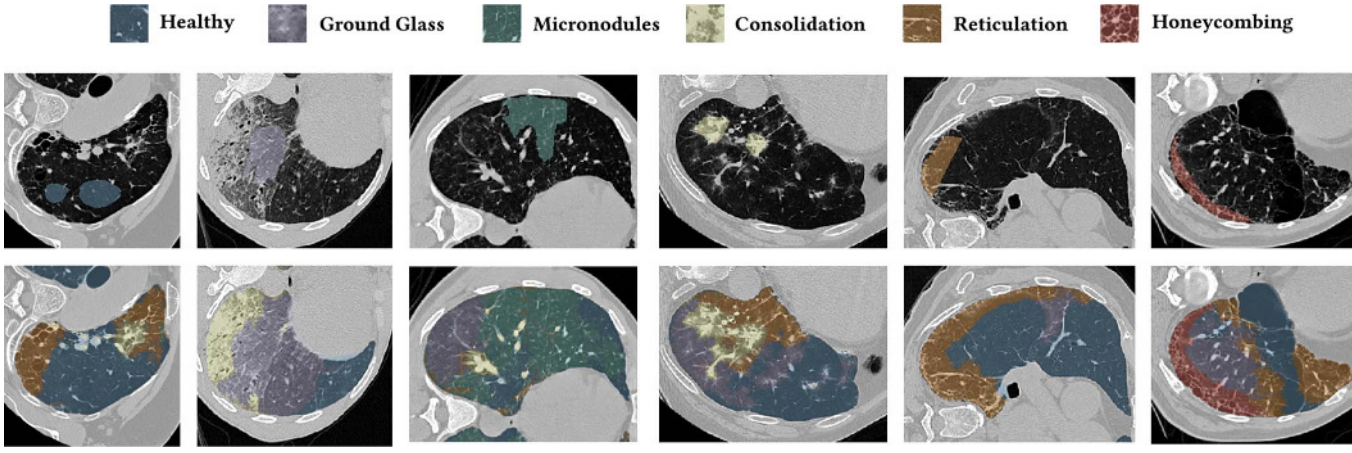

FIGURE 2. Output examples of the segmentation model for tissue characterization. Top row, ground truth; bottom row, model output. Each example has a different pattern annotated. From left to right: healthy (blue), GGO (purple), micronodules (green), consolidation (yellow), reticulation (orange) and honeycombing (red).

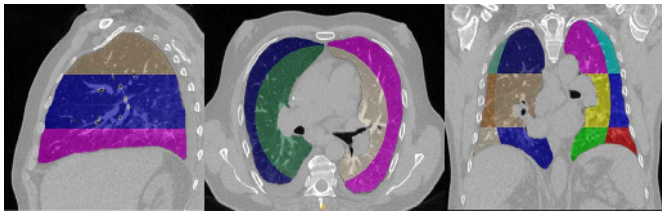

FIGURE 3. An example of the lung partitioning in 12 regions. From left to right: craniocaudal splitting, central-peripheral splitting, and their intersection.

and was calculated for the distances from the center of mass and the border. Particularly, the center of mass was calculated in terms of the average of the *x*, *y*, and *z* coordinates of all the pixels that reside within the lung parenchyma, then the distance of each of the pixels within the lung and the center of mass was calculated. Moreover, the distance of each pixel from the closest border pixel was calculated using the fast marching method (scikit-fmm). Each of the pixels was described by these 2 values and clustered using a k-means algorithm into 2 clusters that split the lung in peripheral and central. The intersection of the aforementioned segments produces a total of 12 segments in both lungs; an example is presented in Figure 3. The distribution of pathological tissue as estimated by the tissue characterization CNN is calculated over each segment, and these are used as features together with the clinical data (age, sex, etc) to train multiple one-versus-all random forest classifiers to classify the lung fibrosis for each case into (1) a typical UIP CT pattern, (2) a probable UIP CT pattern, (3) a CT pattern indeterminate for UIP, and (4) CT features that are most consistent with a non-IPF diagnosis (Table 1, Lynch et al[3]). In total, 90 features were used, which were composed of clinical parameters, the distribution of the pathological tissue in the 12 regions, as well as some extra features that were calculated over the total extend of the lung (eg, total extend of pathological tissue, total extend of GGO, etc). The random forest was configured to use 200 decision trees and no more than 5 features per leaf.

The different steps of the CAD system required approximately 6 minutes per case for the scan to be properly processed and for a diagnosis to be available. In Figure 4, 2 example cases are presented (Typical UIP, non-IPF) along with the output of the pathological tissue segmentation step on a single slice and a radial histogram visualization of the distribution of pathological tissue in the whole lung. For all calculations, we used a CPU implementation running on an Intel Core i7-5960X CPU, except for the tissue characterization, for which we used a NVIDIA GeForce GTX TITAN GPU. The steps needed the following mean times:

- Lung and airways segmentation, 47.1 seconds,
- ILD pathology quantification, 192.4 seconds, and
- Diagnosis support, 124.4 seconds.

Experimental Setup

Sensitivity, specificity, accuracy, and positive predictive values were calculated for the readers and the proposed CAD system, using the independent chest radiology experts' consensus classification (1–4) as the ground truth. Positive predictive value (precision) and sensitivity (recall) were used to calculate the F-score (harmonic mean for precision and recall).

McNemar test was used to compare the sensitivity, specificity, and accuracy between the readers and the CAD system. Comparison of proportions was used to compare the F-scores. The significance level was set to 0.05. MedCalc version 15.0 (MedCalc Software, Ostend, Belgium) was used. The 4 categories were first analyzed unpooled for all entries. Then the 4 groups were pooled into 2 categories: (*a*) cases needing a biopsy for further diagnostic testing, according to the Fleischner Society White Paper[3] (groups 3 and 4), and (*b*) cases without further need for diagnostic testing (groups 1 and 2). A separate analysis of the correct classification into these 2 groups was performed for the findings of the readers and the machine. Furthermore, interobserver agreement between the radiologists and between the radiologists and

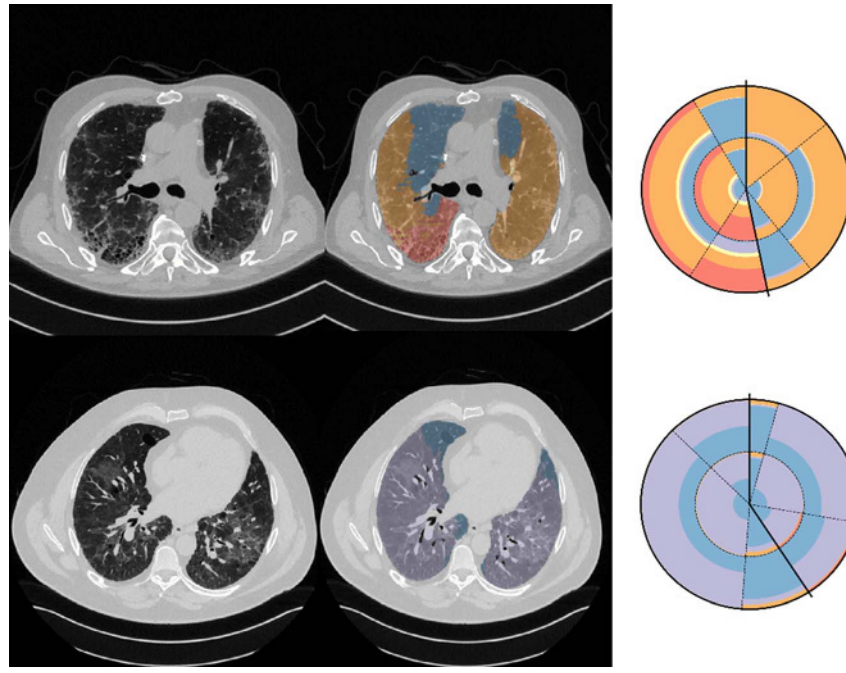

FIGURE 4. Interstitial lung disease pathological lung tissue quantification (middle column) and visualization using radial histograms (right column) similar to that used in.[31] Each sector denotes a region of the lung and is split into 2 parts, one for the central (inner) and one for peripheral (outer). The color scheme denotes the pathological tissue similar to Figure 2. Solid lines denote the division of left and right lungs. Top case is a typical UIP and bottom a non-IPF.

the machine was calculated individually from the weighted Kappa, as follows: slight (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), substantial (0.61–0.8), and almost perfect agreement (0.81–1).

RESULTS

Fleischner Classification

Reader 1, reader 2, and CAD demonstrated similar (no significant difference) accuracy for classifying the pulmonary fibrosis, according to the Fleischner Society Guidelines[3]: 0.6, 0.54, and 0.56, respectively, with $P > 0.45$. The CAD system achieved an F-score (harmonic mean for precision and recall) of 0.56, whereas the 2 readers on average achieved 0.57 ($P = 0.991$).

Fleischner Pooled Classification

When the 4 classification groups were pooled into the group requiring further workup (groups 3 and 4) and into the group for which the diagnosis was clear without biopsy (UIP group 1 and 2), the accuracy increased. Reader 1, reader 2, and the CAD again scored similar accuracies: 0.81, 0.70, and 0.81, respectively. Reader 2 was slightly inferior to reader 1 (without reaching the level of statistical significance) and CAD, with P values of 0.059 and 0.189, respectively. The sensitivities for choosing the cases that needed further workups were similar for reader 1, reader 2, and CAD, at 0.86, 0.84, and 0.79, respectively, and $P > 0.39$. The CAD system demonstrated the best specificity; the specificities for reader 1, reader 2, and CAD were 0.63, 0.44, and 0.67, respectively. The CAD system performed significantly better than reader 2 ($P = 0$ .012) and just as well as reader 1 ($P = 0.773$). The specificity of reader 1 was significantly higher than that of reader 2 ($P = 0.037$). The F-score was again similar for the CAD system and the readers. The CAD and the mean reader F-scores were 0.80 and 0.79 ($P = 0.898$), respectively.

Interobserver Agreement

Reader 1 versus reader 2 demonstrated fair interreader agreements, with a weighted kappa (±standard error) of 0.30 ± 0.08 for group 4 (unpooled) and 0.3 ± 0.1 for group 2 (pooled) classifications. The CAD system versus reader 1 demonstrated moderate classification agreement, with a weighted kappa of 0.47 ± 0.08 and 0.54 ± 0.1, respectively, for the unpooled and pooled classifications. Compared with reader 2, the agreement was only fair: 0.33 ± 0.08 and 0.3 ± 0.09, respectively. The interreader agreement of the 2 chest experts, who set the ground truth, was substantial: weighted kappa was 0.63 ± 0.08 and 0.66 ± 0.1 for the unpooled and pooled classifications, respectively.

DISCUSSION

The accuracy of the proposed CAD system for dichotomous classification into the group needing further intervention and the group without the need for further workups was as good as that of the radiologists. The automated system even outperformed the inexperienced radiologist, in terms of the specificity for patient identification requiring subsequent intervention. Demonstrating the best specificity means having the lowest false-positive rate in the group undergoing further workups. This rate is important, if we consider the high mortality and morbidity of surgical wedge resection of the lung on the one hand and the low operability of this group on the other.[33] Our results support the beneficial implementation of a computer-aided diagnostic algorithm. As dedicated chest radiologists are scarce, and IPF is almost considered to be an orphan disease, it is difficult to build the required expertise in this field. Under these circumstances, the importance of CAD solutions becomes evident, which highlights the importance of such a CAD system.

There has been much discussion of whether the Fleischner Society White Paper on classifying pulmonary fibrosis into 4 groups and recommending interventions in only 2 of these groups should be accepted. Some even feel that biopsy is needed for 3 groups (including the group with a probable UIP CT pattern).[4,34] The authors believe that the number of biopsies can be even further reduced in a proper setting of multidisciplinary ILD boards, with follow-up of these patients. The pattern recognition method is certainly a promising approach for classifying the probability of IPF today; however, with increasing precision and recall of CAD for ILD, it may be possible to skip this classification and then call the disease by its name.

The accuracy of all readers (CAD inclusive) between 0.5 and 0.6 for the unpooled classification is not particularly high, but is still substantially above the level of chance, as 4 classes were considered. In practice, differentiation between groups 1 and 2 is less important than differentiation between group 2 and 3, as the diseases are labeled IPF for both groups 1 and 2. The positive predictive value for IPF is 80% in group 2, so a biopsy is not recommended for the first 2 groups.[3] For all other cases (groups 3 and 4), a biopsy would be needed. Therefore, it is important to differentiate between those patients with and without the need for biopsy. There is an ongoing debate whether a surgical lung biopsy is essential in probable UIP cases.[34] Although, there is mild evidence against the biopsy,[3] this issue needs further validation in future investigations. To appreciate the performance of the readers and the CAD system, one has to examine the results for the dichotomous differentiation into the group with and without the need for biopsy; the accuracy rate of 0.81 more appropriately represents the performance of both the radiologists and the machine. This accuracy is comparable with the results published by Depeursinge et al.[35] The relatively low accuracy for classifying pulmonary fibrosis is a recognized phenomenon in chest imaging in general. Watadani et al[36] reported a disagreement in identifying simple honeycombing of the lungs in 29% among experts. As radiologists and the CAD system did not have the same false-positive and false-negative cases, there is also the potential of CAD to help radiologists classify the UIP pattern. Further analyses evaluating the use of this CAD as second reader and the potential reduction of the interobserver variability are part of an ongoing study. For the next iteration of the INTACT-CAD, we are aiming to implement a 3-dimensional texture analysis on a new batch of annotated CT images of volume scans. This strategy will allow for a substantial increase in accuracy to help in minimizing the total number of required lung biopsies.

It should be noted that our study suffers from a few limitations. One of the limitations is that the number of cases was low. For more significant results, we are currently conducting a larger-scale study using an improved version of the CAD. However, given the low prevalence of interstitial lung fibrosis, these results are of considerable impact. Another limitation of the study is introduced by a number of algorithmic failures, such as undersegmentation of the lung field or a failure to split the left/right lungs. Such failures are mostly due to radiological or anatomical characteristics, such as high-intensity pathologies (eg, extensive consolidation near the lung edges) or touching lungs. These failures are then propagated in the following steps, dropping the overall performance of the system. In the next version of the system, the manual intervention of the radiologists will be investigated to reduce such failures. Furthermore, only cases of pulmonary fibrosis were covered in this study, due to the fine-tuning of our CAD toward fibrosis; a pattern extension for the algorithm will be implemented in later versions. Furthermore, we included the idiopathic forms of pulmonary fibrosis, and we also allowed for diseases with associated pulmonary fibrosis to be included, and this may have confounded these results.

The diagnosis of ILDs is a challenging task; therefore, the formation of interdisciplinary ILD boards in university centers proved to be very useful.[3,4] In our study, the interrater agreement of the 2 radiologists who set the ground truth was between moderate and perfect. To overcome this downside, the radiologists met again for a second consensus read out of the cases without agreement. This study stayed on the level of the radiological diagnosis (4 UIP classifications), because the

radiological diagnosis of typical UIP pattern replaces the tissue sampling. Our approach for a robust ground truth and robust CAD may improve when using more image date for machine learning and more radiologist with a consensus annotation. Future CADs should also provide a diagnosis on the level of the ILD-board, and for training of such a system, all cases must be used: radiologically or histopathologically approved diagnosis.

To conclude, in this study, we present a preliminary evaluation of an integrated pipeline for the automatic classification of IPF. A multistep approach is used for the segmentation of anatomical structures of the lung cavity, the segmentation of lung pathological tissue, and finally the classification into a radiological diagnostic CT pattern. In summary, we found that a machine learning-supported computer-aided detection algorithm was able to classify IPF with similar accuracy to a human reader. Moreover, the computer algorithm delivered results comparable to those of radiologists when grouping fibrosis patterns according to the Fleischner Society's newest recommendation.

ACKNOWLEDGMENTS

The authors would like to thank the Swiss National Science Foundation, Roche Pharma Ltd, and the Lindenhof Foundation for funding this study.

REFERENCES

1. Wells AU. Managing diagnostic procedures in idiopathic pulmonary fibrosis. *Eur Respir Rev*. 2013;22:158–162.
2. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2013;188:733–748.
3. Lynch DA, Sverzellati N, Travis WD, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med*. 2018;6: 138–153.
4. 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.
5. Lee JY, Jin SM, Lee BJ, et al. Treatment response and long term follow-up results of nonspecific interstitial pneumonia. *J Korean Med Sci*. 2012;27:661–667.
6. Fisher M, Nathan SD, Hill C, et al. Predicting life expectancy for pirfenidone in idiopathic pulmonary fibrosis. *J Manag Care Spec Pharm*. 2017;23(suppl 3-b): S17–S24.
7. Sluimer I, Schilham A, Prokop M, et al. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Trans Med Imaging*. 2006;25:385–405.
8. Maldonado F, Moua T, Rajagopalan S, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J*. 2014;43:204–212.
9. Kim HG, Tashkin DP, Clements PJ, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol*. 2010;28(5 suppl 62):S26–S35.
10. Jacob J, Bartholmai BJ, Rajagopalan S, et al. Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med*. 2018;198:767–776.
11. Salaffi F, Carotti M, Di Donato E, et al. Computer-aided tomographic analysis of interstitial lung disease (ILD) in patients with systemic sclerosis (SSc). Correlation with pulmonary physiologic tests and patient-centred measures of perceived dyspnea and functional disability. *PLoS One*. 2016;11:e0149240 eCollection 2016 11: e0149240.
12. Christe A, Szucs-Farkas Z, Huber A, et al. Optimal dose levels in screening chest CT for unimpaired detection and volumetry of lung nodules, with and without computer assisted detection at minimal patient radiation. *PLoS One*. 2013;8: e82919 e Collection 2013 8: e82919.
13. Silva M, Schaefer-Prokop CM, Jacobs C, et al. Detection of subsolid nodules in lung cancer screening: complementary sensitivity of visual reading and computer-aided diagnosis. *Invest Radiol*. 2018;53:441–449.
14. Jacobs C, van Rikxoort EM, Scholten ET, et al. Solid, part-solid, or non-solid?: classification of pulmonary nodules in low-dose chest computed tomography by a computer-aided diagnosis system. *Invest Radiol*. 2015;50:168–173.
15. Wagner AK, Hapich A, Psychogios MN, et al. Computer-aided detection of pulmonary nodules in computed tomography using ClearReadCT. *J Med Syst*. 2019;43:58.
16. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol*. 2017;52:281–287.
17. Godoy MC, Kim TJ, White CS, et al. Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin- and thick-section CT. *Am J Roentgenol*. 2013;200:74–83.
18. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–1350.
19. Ciompi F, Chung K, Van Riel SJ, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*. 2017;7:46479.
20. Anthimopoulos M, Christodoulidis S, Christe A, et al. Classification of interstitial lung disease patterns using local DCT features and random forest. *Conf Proc IEEE Eng Med Biol Soc*. 2014;2014:6040–6043.
21. Sorensen L, Shaker SB, De Bruijne M. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans Med Imaging*. 2010;29: 559–569.
22. Gangeh MJ, Sørensen L, Shaker SB, et al. A texton-based approach for the classification of lung parenchyma in CT images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg, Germany: Springer; 2010:595–602.
23. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016;35:1207–1216.
24. Christodoulidis S, Anthimopoulos M, Ebner L, et al. Multisource transfer learning with convolutional neural networks for lung pattern analysis. *IEEE J Biomed Health Inform*. 2017;21:76–84.
25. Van Ginneken B, Katsuragawa S, ter Haar Romeny BM, et al. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Trans Med Imaging*. 2002;21:139–149.
26. Fukushima A, Ashizawa K, Yamaguchi T, et al. Application of an artificial neural network to high-resolution ct: usefulness in differential diagnosis of diffuse lung disease. *AJR Am J Roentgenol*. 2004;183:297–305.
27. Depeursinge A, Vargas A, Platon A, et al. Building a reference multimedia database for interstitial lung diseases. *Comput Med Imaging Graph*. 2012;36: 227–238.
28. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*. 2001; 20:490–498.
29. Sluimer I, Prokop M, Van Ginneken B. Toward automated segmentation of the pathological lung in CT. *IEEE Trans Med Imaging*. 2005;24:1025–1038.
30. van Rikxoort EM, de Hoop B, Viergever MA, et al. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Med Phys*. 2009;36:2934–2947.
31. Mori K, Hasegawa JI, Toriwaki JI, et al. Recognition of bronchus in three-dimensional X-ray CT images with applications to virtualized bronchoscopy system. *Pattern Recognition*. Proceedings of the 13th International Conference IEEE. 1996;528–532.
32. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Semantic segmentation of pathological lung tissue with dilated fully convolutional networks. *IEEE J Biomed Health Inform*. 2019;23:714–722.
33. National Lung Screening Trial Research TeamAberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
34. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*. 2018;198:e44–e68.
35. Depeursinge A, Chin AS, Leung AN, et al. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. *Invest Radiol*. 2015;50:261–267.
36. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology*. 2013;266:936–944.