A Lung Graph Model for the Classification of Interstitial Lung Diseases on CT Images

Guillaume Vanoost[a,b], Yashin Dicente Cid[b,c], Daniel Rubin[e], and Adrien Depeursinge[b,d]

[a]INP-ENSEEIHT, Toulouse, France;
[b]University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland;
[c]University of Geneva, Geneva, Switzerland;
[d]Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital, Lausanne, Switzerland;
[e]Department of Biomedical Data Science, Radiology, and Medicine (Biomedical Informatics Research), Stanford University School of Medicine, Stanford, CA, USA;

ABSTRACT

Diagnosing Interstitial Lung Diseases (ILD) is a difficult task. It requires experienced chest radiologists that may not be available in less-specialized health centers. Moreover, a correct diagnosis is needed to decide for an appropriate treatment and prognostic. In this paper, we focus on the classification of 3 common subtypes of ILDs: Usual Interstitial Pneumonia (UIP), Non-Specific Interstitial Pneumonia (NSIP) and Chronic Hypersensitivity Pneumonitis (CHP). We propose a graph model of the lungs built from a large dataset. The structure of the graph is inspired from medical knowledge of disease predominance, where the nodes correspond to 24 distinct regions obtained from lateral, anterior-posterior and vertical splits of the images. The adjacency matrix is built from distances between intensity distributions of distinct regions. Graphs models are interpretable and were successfully used in neuroimaging. However, to the best of our knowledge, this is the first attempt to use a graph model of the lungs for classifying ILDs. In the particular case of ILDs, graph methods are relevant for the following reasons. In order to differentiate between the subtypes, not only the types of local patterns of the disease are important but also their anatomical location. Therefore, we hypothesize that the comparison between regional distributions of Hounsfield Unit (HU) values is relevant to discriminate between the considered ILD subtypes. For instance, typical UIP shows a spatial predominance of reticular abnormalities and honeycombing in the peripheral regions of the lung bases. Therefore, we expect a marked difference of HU distributions between the central and peripheral regions of the lung bases. Moreover, the construction of the graph leads to an interpretable patient descriptor. The descriptor led to encouraging area under the Receiver Operating Characteristic (ROC) curve in 0.6-0.8 for one-versus-one classification configurations, which also showed to outperform feature sets based on a simple concatenation of regional HU distributions.

Keywords: Computer-Assisted Diagnosis, Interstitial Lung Diseases, Graph-Models

1. INTRODUCTION

Interstitial Lung Diseases (ILD) is a group of more than 150 disorders of the lung tissue of varying origin causing texture and intensity changes of the lung parenchyma with a characteristic distribution within the lung anatomy. Thus CT diagnosis of most ILDs is based on the presence/absence of characteristic parenchymal alterations (e.g. ground glass, reticulation, honeycombing, micro-nodules) in specific regions of the lungs.[1] These findings are summarized in Table 1.

Several studies highlighted the difficulty to correctly establish the diagnosis of ILDs based on Computed Tomography (CT) scans alone as the same patterns are commonly found in different diseases.[5] While most studies focused on the detection and classification of these patterns,[6] their distributions in the lungs (upper vs. lower lungs, central vs peripheral) are known additional discriminating features that were captured by only a

Further author information: Send correspondence to Guillaume Vanoost Email: vanoost-guillaume@hotmail.fr

Table (1) Radiological CT findings for Usual Interstitial Pneumonia (UIP), Non-Specific Interstitial Pneumonia (NSIP) and Chronic Hypersensitivity Pneumonitis (CHP).[1–4]

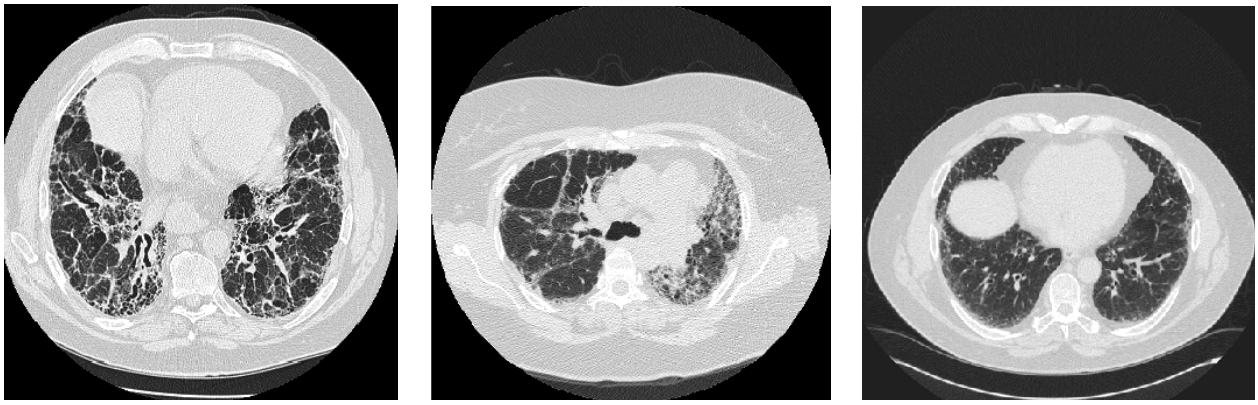
Typical UIP	Typical NSIP	Typical CHP
Peripheral, Basal. Reticular abnormality. Honeycombing.	Subpleural, Basal, Symmetric. Ground glass and reticular opacity. Irregular lines.	Predominance in the middle lung fields. Ground-glass and reticular opacity. Possible Honeycombing. Mosaic attenuation and air trapping.

small number of prior quantitative studies. Depeursinge et al.[7] demonstrated the wealth of a regional approach for the computed diagnosis using a geometric atlas of the lungs and regional texture features. More recently, a study from Jun et al.[8] highlighted the discriminative power of a texture comparison between upper-lower, anterior-posterior and central-peripheral parts of the lung. This paper builds on this previous work with a dataset of 450 labeled cases and with 3 classes of ILDs, Usual Interstitial Pneumonia (UIP), Non-Specific Interstitial Pneumonia (NSIP) and Chronic Hypersensitivity Pneumonitis (CHP). Instead of studying the features of the regions of the lung, we focus on the definition of a graph model of the lungs, defined in Figure 3. This type of graph was introduced by Dicente et al.,[9] to study the relations between features extracted in lung subregions. This graph was already tested on patients with pulmonary circulatory diseases[10] and with Tuberculosis.[11]

2. METHODS

2.1 Dataset

In our experiments we used the Lung Tissue Research Consortium (LTRC) public dataset*. It consists of 462 CT scans of ILD patients. Seven patients were excluded in our experiments due to inconsistencies in the DICOM headers, such as non-coherent number of acquisitions. The final dataset used consisted of 184 UIP patients, 59 CHP patients and 208 NSIP patients. Figure 1 shows one axial CT slice from each considered ILD class.



UIP CHP NSIP

Figure (1) Examples of axial views of CT scans from patients with each of the three diseases considered.

2.2 Lung Graph Model Construction

The purpose of encapsulating regional distributions of Hounsfield Unit (HU) values into a graph structure is to leverage inter-regional comparisons of voxel intensities. These comparisons are relevant to distinguish between the ILD subtypes according to their respective spatial predominance (see Table 1). For instance, typical UIP shows a spatial predominance of reticular abnormalities and honeycombing in the peripheral regions of the lung bases. Therefore, we expect a marked difference of HU distributions between the central and peripheral regions of the lung bases.

*https://ltrcpublic.com/, as of January 21 2019.

We modified the fully automated pipeline introduced by Dicente et al.[10] to obtain specific graph models of the lungs for each disease. Our pipeline is composed of five steps, which includes (i) segmentation of the lung fields to create a mask of the lungs, (ii) subdivision of this mask into regions to create an atlas, (iii) extraction of the features for each region, (iv) creation of the distance matrix encoding the comparison between the regional features, and (v) creation of the graph model from a statistical test for all elements of the distance matrix to highlight the most relevant connections between regions.

We used the automatic lung segmentation algorithm developed by Dicente et al.[12] to extract the lung fields. We initially divided the lungs using a geometric atlas introduced in,[13] based on the 3D model of the lung developed by Zrimec et al.[14] This atlas includes 36 regions generated from 4 axis divisions: coronal (right/left), sagital (anterior/posterior), vertical (apical/central/basal), and axial (peripheral/middle/central). For a small number of patients with strongly affected anatomy, this atlas produced very small or even empty regions. Therefore, we reduced the number of regions to 24 by fusing the peripheral and middle regions (see Fig. 3). This procedure produced more balanced region volumes.

For each region of the atlas we computed a normalized 10-bin histogram of the HUs. We decided to include at least 97% of the pixels of every patient in the histogram, hence, we chose the histogram limits from -1024 to 296 HU (see Fig. 2). A total of 25 patients had at least one region with less than 100 pixels. A surrogate histogram was created for these missing regions by averaging each bin of the adjacent regions.

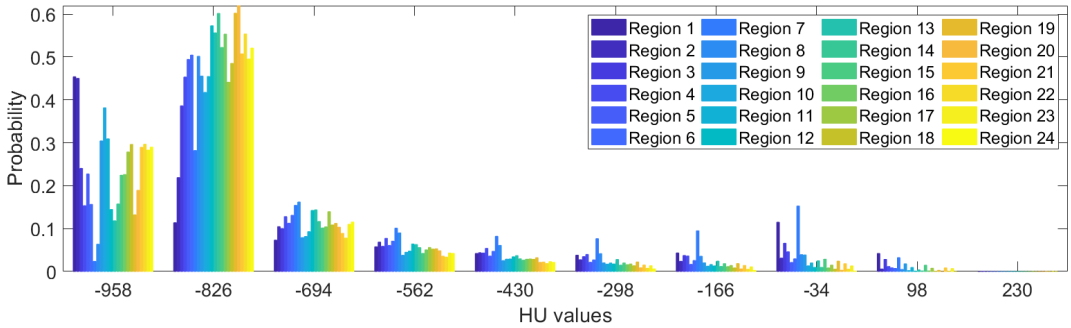


Figure (2) Example of a normalized histogram of the HU distribution in each of the 24 regions of the atlas for one patient with UIP.

For each patient $n = 1, \ldots, N$ we create the 24×24 distance matrix D_n:

$$D_n = \begin{pmatrix} d_{\mathrm{KS}_{1,1}} & d_{\mathrm{KS}_{1,2}} & d_{\mathrm{KS}_{1,3}} & \cdots & d_{\mathrm{KS}_{1,24}} \\ d_{\mathrm{KS}_{2,1}} & d_{\mathrm{KS}_{2,2}} & d_{\mathrm{KS}_{2,3}} & \cdots & d_{\mathrm{KS}_{2,24}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{\mathrm{KS}_{24,1}} & d_{\mathrm{KS}_{24,2}} & d_{\mathrm{KS}_{24,3}} & \cdots & d_{\mathrm{KS}_{24,24}} \end{pmatrix}$$

The matrix elements $d_{\mathrm{KS}_{i,j}}$ are obtained from the Kolmogorov-Smirnov (KS) distance between the HU histograms of regions i and j. This matrix is symmetric and only the distances in the upper triangular are kept. The vectorization of this upper triangle of D_n creates a 276-dimensional feature vector characterizing how similar (or dissimilar) the intensities of an anatomical region are to another. The feature vectors from specific ILD subtypes are then used to create ILD-specific graph models, where feature- (i.e. edge-) selection based on statistical testing is used to create the final graph adjacency matrix and associated feature vector $\boldsymbol{v}$ for its further use in ILD classification.

2.2.1 ILD-Specifc Graphs and Distance-Based Features

The next step is to create graph models that are specific to each ILD by selecting the edges that are most significantly different between the ILD subtypes. We expect a reduced graph that highlights only the relevant

edges for the classification of these three diseases. Therefore, we need to determine if there are edges for which the inter-class difference between the distribution of KS distances is statistically significant when considering the three ILD populations.

As there are edges for which the distribution of the KS distances in the same class is non-normal, we decided to use the non-parametric Kruskal-Wallis test on the means KS distance ($\mu_{i,j}$) for each edge $d_{\mathrm{KS}_{i,j}}$ with the following hypotheses (the indices i, j are omitted to simplify the notations):

$$H_0 : \mu_{\mathrm{UIP}} = \mu_{\mathrm{CHP}} = \mu_{\mathrm{NSIP}},$$
$$H_1 : \text{At least two of the population means are different.}$$

Edges for which H_0 was rejected with a significance level $\alpha = 0.05$ were kept. This resulted in a p-matrix constituting the adjacency matrix of the graph: a 24×24 matrix from the distance matrix of every patient. Each value is the p-value for a specific edge (see Fig. 4). This matrix reveals edges for which the distribution of the distance between one the two classes is dissimilar. Feature vectors $\boldsymbol{v}_{\mathrm{dist}}$ were created from the set of distances (i.e. edges) for which the test was significant.

Similarly, one-versus-one and one-versus-all graphs and feature sets were constructed for ILD subtypes pairs and single ILD subtypes using a Mann-Whitney U test instead of the Kruskal-Wallis.

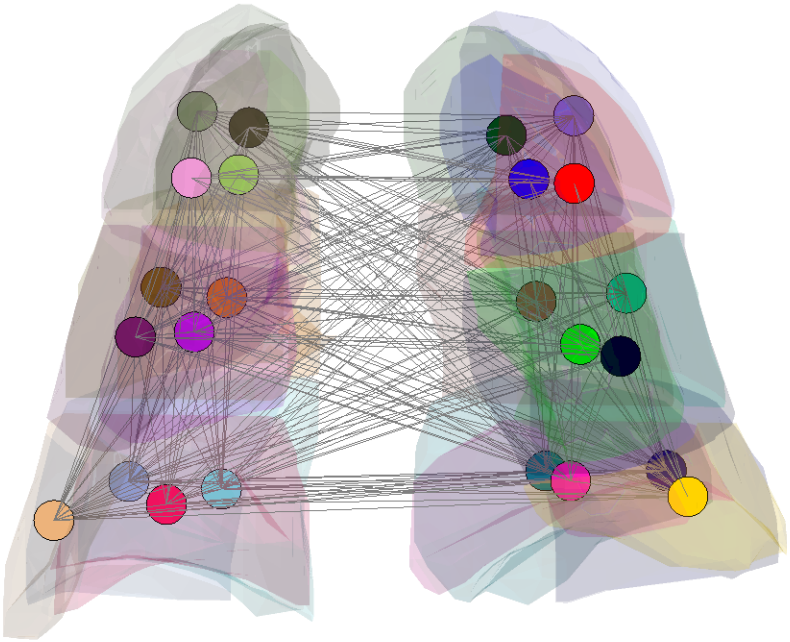


Figure (3) Example of a fully connected graph model of the lung based on 24 distinct regions (inspired from[9]).

3. EXPERIMENTAL SETUP

The relevance of the proposed graph features $\boldsymbol{v}_{\mathrm{dist}}$ is evaluated with one-versus-one classification configurations. A naive Bayes classifier was used, as it showed better results in preliminary experiments when compared to Support Vector Machines (SVM). We compared the performance of the proposed feature vector $\boldsymbol{v}_{\mathrm{dist}}$ with a feature vector based on the regional distributions of HU. In particular, it consisted of 10-bin histograms characterizing the distribution of HUs in each region (see Fig. 2), resulting in a 240-dimensional feature vector $\boldsymbol{v}_{\mathrm{HU}}$. The classification performance was estimated with a 10-fold Cross-Validation (CV). For each fold of the CV, the test (either Kruskal-Wallis or Mann-Whitney U) was carried out to select features and the naive Bayes model was built and evaluated with Receiver Operating Characteristic (ROC) curves. The Area Under the ROC Curves (AUC) were averaged over the folds and reported.

4. RESULTS

4.1 ILD graphs

First, the p-matrix was built when considering all three ILD subtypes and using the Kruskal-Wallis test (see Fig. 4). A subset of 12 regions were found to have at least one significant connection with another region.

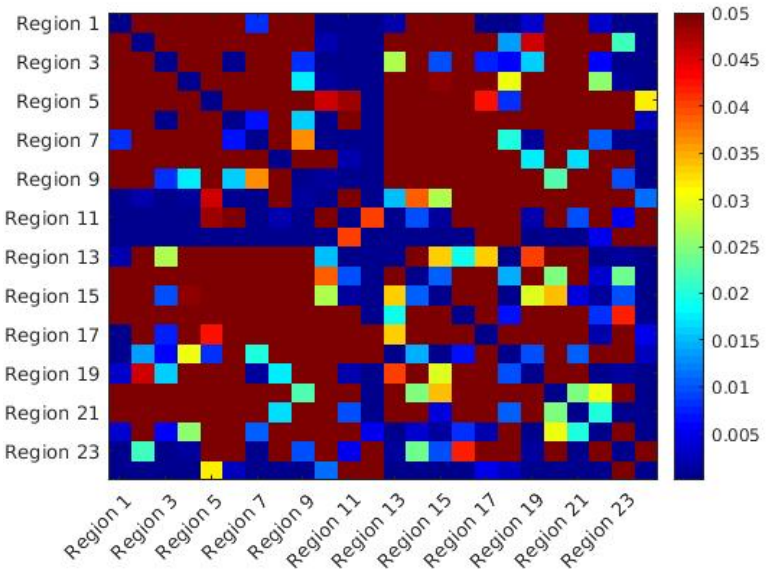


Figure (4) ILD-specific graph adjacency matrix resulting from Kruskal-Wallis tests (called p-matrix) when considering all three populations (UIP, CHP and NSIP). Every value above 0.05 was artificially set to 0.05 for visualization purposes to consider only significant p-values. Regions for which no edge were associated with a rejection of H_0 are omitted from the p-matrix.

In order to build ILD subtype-specific graph models, the Mann-Whitney U test was used in one-versus-all configurations. The weight of the edges is defined as $1 - p$ for each disease constructed from the p-matrix of a class versus the 2 others. The graphs for UIP, CHP and NSIP are depicted in Figures 5, 6 and 7, respectively. Only edges with weights above 0.99 are shown. Notice that the significant edges are not the same for each disease.

4.2 ILD Classification

The classification performance for the feature vectors $\boldsymbol{v}_{\text{dist}}$ and $\boldsymbol{v}_{\text{HU}}$ are compared in Fig. 8. A Student-t test assuming non-identical variance was used to determine whether the average AUC was significantly different between $\boldsymbol{v}_{\text{dist}}$ and $\boldsymbol{v}_{\text{HU}}$. The test revealed statistically different mean AUCs only in the case of CHP versus NSIP ($p = 0.0435$), whereas UIP versus CHP and UIP versus NSIP were not significant with $p = 0.7105$ and $p = 0.0953$, respectively.

5. DISCUSSIONS AND CONCLUSIONS

The goal of this study was to investigate the relevance of graph-based inter-regional comparisons of HU distributions in the lung to characterize ILD subtypes. The results suggest that adjacency between regions of the lungs is a relevant descriptor for UIP, CHP and NSIP, which are the three most common ILD subtypes, where classification performance with AUCs in 0.6-0.8 were observed. The performance was higher when using the graph features when compared to a baseline concatenation of regional HU histograms with around four times less features (see Figure 8). The image analysis pipeline is fully automatic.

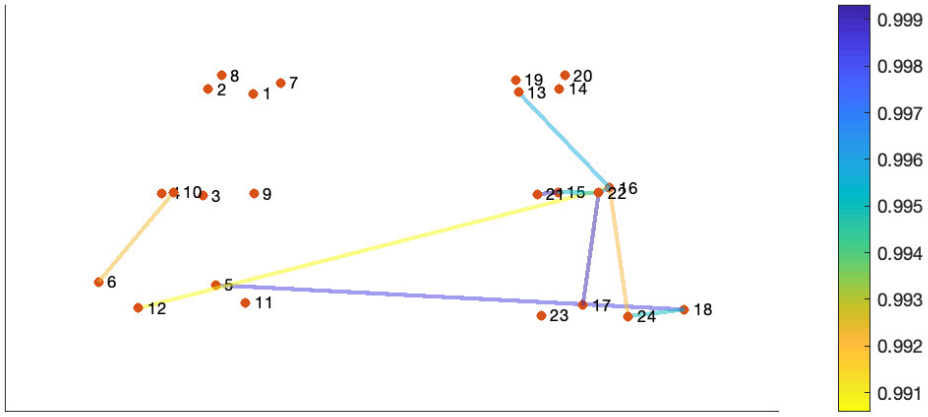


Figure (5) UIP-specific graph visualization.

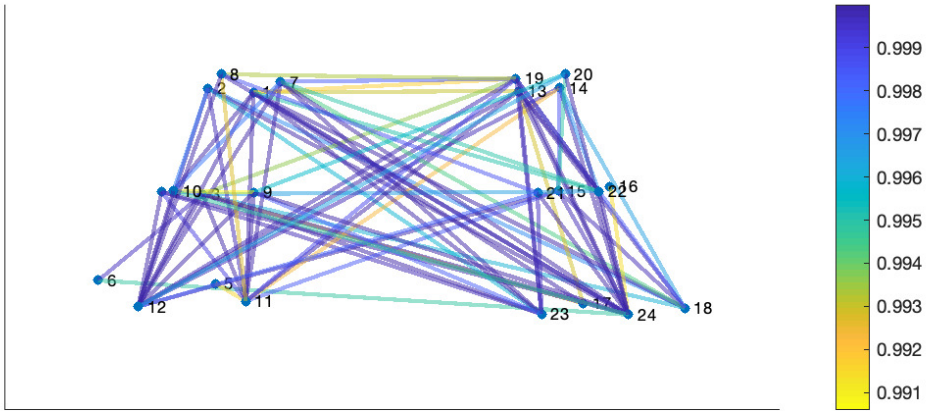


Figure (6) CHP-specific graph visualization.

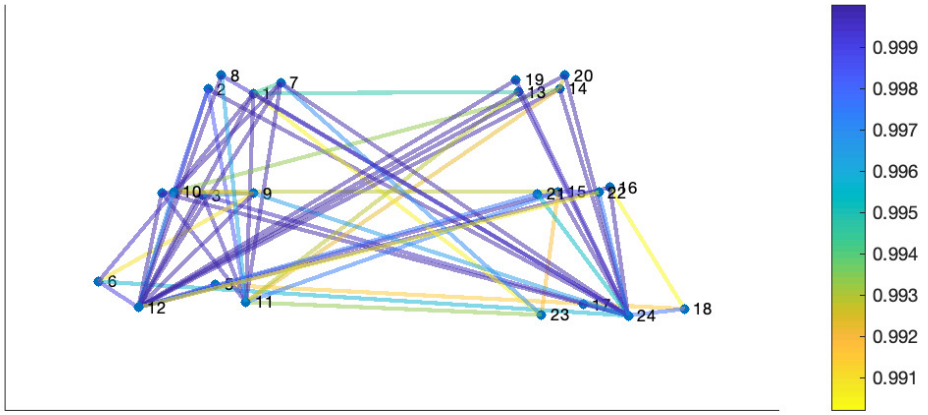


Figure (7) NSIP-specific graph visualization.

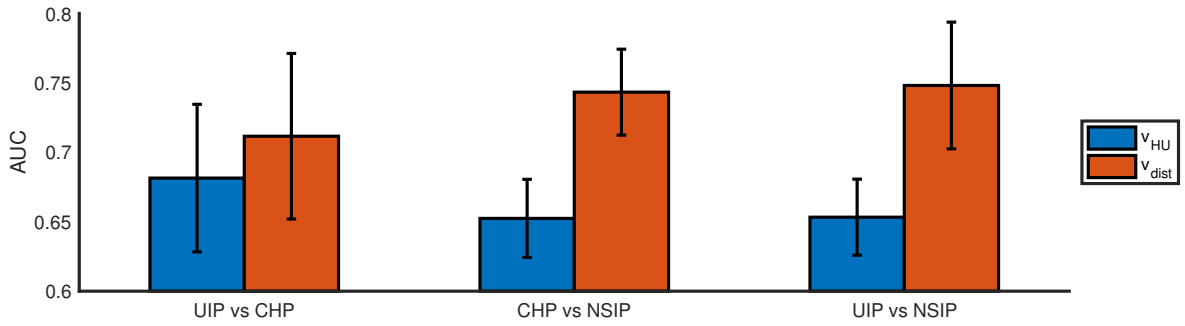


Figure (8) Classification performances for all three one-versus-one configurations. The two feature vectors $\boldsymbol{v}_{\mathrm{HU}}$ and the proposed $\boldsymbol{v}_{\mathrm{dist}}$ are compared. Error bars correspond to the standard error. While $\boldsymbol{v}_{\mathrm{HU}}$ has a fixed dimensionality of 240, the average dimensionalities of $\boldsymbol{v}_{\mathrm{dist}}$ are 56.4 ± 6, 82 ± 3.8 and 30.6 ± 6.9 for UIP versus CHP, CHP versus NSIP and UIP versus NSIP, respectively.

The proposed graph approach allows gaining insights on the relevant regional comparison in the context of a particular ILD subtype. We observed from the global p-matrix (see Fig. 4) that most of the edges coming from the posterior basal middle-peripheral are relevant. This is consistent with the reported CT findings of UIP and NSIP in the literature,[1] that points at these specific regions (see Table 1). In addition, we can also see relevant edges from the posterior central middle-peripheral region. This may be associated with typical CHP findings. The interpretation of subtype-specific graphs shown in Figures 5, 6 and 7 is difficult. One striking observation is the small number of significant edges in the UIP-specific graph (see Fig. 5). This suggests that inter-regional HU comparisons are less relevant than for NSIP and CHP. This is also confirmed by the classification performances reported in Fig. 8, where only CHP versus NSIP had a significantly improved AUC when using $\boldsymbol{v}_{\mathrm{dist}}$ as compared to using concatenated HU histogram features $\boldsymbol{v}_{\mathrm{HU}}$.

In future work, we plan to construct the graphs by extracting texture features that may be highly relevant due to the nature of the diseases.[7] Their combination with HU measurements are expected to allow optimal ILD classification performance.

ACKNOWLEDGMENTS

This work was partly supported by the Swiss National Science Foundation in the context of the PH4D project (grant agreement 320030_146804) and the VISIBLE project (grant agreement 205320_179069). This study utilized biological specimens [and/or] data provided by the Lung Tissue Research Consortium (LTRC) supported by the National Heart, Lung, and Blood Institute (NHLBI).

REFERENCES

[1] Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J.-F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., Hansell, D. M., Johkoh, T., Kim, D. S., King, T. E., Kondoh, Y., Myers, J., Müller, N. L., Nicholson, A. G., Richeldi, L., Selman, M., Dudden, R. F., Griss, B. S., Protzko, S. L., and Schünemann, H. J., "An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence–based guidelines for diagnosis and management," *American Journal of Respiratory and Critical Care Medicine* **183**, 788–824 (Feb. 2011).

[2] European, R. S., Society, A. T., et al., "American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001.," *American Journal of Respiratory and Critical Care Medicine* **165**(2), 277 (2002).

[3] Torres, P. P. T., Moreira, M. A. R., Silva, D. G. S. T., Gama, R. R. M. d., Sugita, D. M., Moreira, M. A. d. C., et al., "High-resolution computed tomography and histopathological findings in hypersensitivity pneumonitis: a pictorial essay," *Radiologia brasileira* **49**(2), 112–116 (2016).

[4] Torres, P. P. T., Rabahi, M. F., Moreira, M. A. C., Meirelles, G. d. S. P., Marchiori, E., et al., "Usual interstitial pneumonia: typical, possible, and inconsistent patterns," *Jornal Brasileiro de Pneumologia* **43**(5), 393–398 (2017).

[5] Chung, J. H., Montner, S. M., Adegunsoye, A., Oldham, J. M., Husain, A. N., Vij, R., Noth, I., and Strek, M. E., "CT findings associated with survival in chronic hypersensitivity pneumonitis," *European radiology* **27**(12), 5127–5135 (2017).

[6] Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.-A., and Müller, H., "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics* **36**, 227–238 (Apr. 2012).

[7] Depeursinge, A., Chin, A. C., Leung, A. N., Terrone, D., Bristow, M., Rosen, G., and Rubin, D. L., "Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high–resolution CT," *Investigative Radiology* **50**, 261–267 (Apr. 2015).

[8] Jun, S., Park, B., Seo, J. B., Lee, S., and Kim, N., "Development of a computer-aided differential diagnosis system to distinguish between usual interstitial pneumonia and non-specific interstitial pneumonia using texture- and shape-based hierarchical classifiers on hrct images," *Journal of Digital Imaging* **31**, 235–244 (Apr 2018).

[9] Dicente Cid, Y., Müller, H., Platon, A., Janssens, J.-P., Frédéric, L., Poletti, P.-A., and Depeursinge, A., "A lung graph-model for pulmonary hypertension and pulmonary embolism detection on DECT images," in [*MICCAI Workshop on Medical Computer Vision: Algorithms for Big Data*], *MICCAI-MCV*, 58–68 (Oct. 2016).

[10] Dicente Cid, Y., Jimenez-del-Toro, O., Platon, A., Müller, H., and Poletti, P.-A., "From local to global: A holistic lung graph model," in [*Medical Image Computing and Computer–Assisted Intervention – MICCAI 2018*], (2018).

[11] Dicente Cid, Y., Batmanghelich, K., and Müller, H., "Textured graph–based model of the lungs: Application on tuberculosis type classification and multi–drug resistance prediction," in [*CLEF 2018*], *Springer LNCS* (2018).

[12] Dicente Cid, Y., Jimenez-del-Toro, O., Depeursinge, A., and Müller, H., "Efficient and fully automatic segmentation of the lungs in CT volumes," in [*Proceedings of the VISCERAL Challenge at ISBI*], Orcun Goksel, Jimenez-del-Toro, O., Foncubierta-Rodriguez, A., and Müller, H., eds., *CEUR Workshop Proceedings*, 31–35 (Apr 2015).

[13] Depeursinge, A., Zrimec, T., Busayarat, S., and Müller, H., "3D lung image retrieval using localized features," in [*Medical Imaging 2011: Computer-Aided Diagnosis*], **7963**, 79632E, SPIE (Feb. 2011).

[14] Zrimec, T., Busayarat, S., and Wilson, P., "A 3D model of the human lung," in [*Proceedings of MICCAI 2004*], LNCS, S., ed., **3217**, 1074–1075 (Oct. 2004).