# Holistic Interstitial Lung Disease Detection using Deep Convolutional Neural Networks: Multi-label Learning and Unordered Pooling

Mingchen Gao, Ziyue Xu, Le Lu, Adam P. Harrison, Ronald M. Summers, Daniel J. Mollura<sup>1</sup>

#### Abstract

Accurately predicting and detecting interstitial lung disease (ILD) patterns given any computed tomography (CT) slice without any pre-processing pre-requisites, such as manually delineated regions of interest (ROIs), is a clinically desirable, yet challenging goal. The majority of existing work relies on manually-provided ILD ROIs to extract sampled 2D image patches from CT slices and, from there, performs patch-based ILD categorization. Acquiring manual ROIs is labor intensive and serves as a bottleneck towards fully-automated CT imaging ILD screening over large-scale populations. Furthermore, despite the considerable high frequency of more than one ILD pattern on a single CT slice, previous works are only designed to detect one ILD pattern per slice or patch.

To tackle these two critical challenges, we present multi-label deep convolutional neural networks (CNNs) for detecting ILDs from holistic CT slices (instead of ROIs or sub-images). Conventional single-labeled CNN models can be augmented to cope with the possible presence of multiple ILD pattern labels, via 1) continuous-valued deep regression based robust norm loss functions or 2) a categorical objective as the sum of element-wise binary logistic losses. Our methods are evaluated and validated using a publicly available database of 658 patient CT scans under five-fold cross-validation, achieving promising performance on detecting four major ILD patterns: Ground Glass, Reticular, Honeycomb, and Emphysema. We also investigate the effectiveness of a CNN activation-based deep-feature encoding scheme using Fisher vector encoding, which treats ILD detection as spatially-unordered deep texture classification.

*Keywords:* Interstitial Lung Disease Detection, Convolutional Neural Network, Multi-label Deep Regression, Unordered Pooling, Fisher Vector Encoding

<sup>&</sup>lt;sup>1</sup>Mingchen Gao, Ziyue Xu, Adam P. Harrison, and Daniel J. Mollura are with Center for Infectious Disease Imaging; Le Lu and Ronald M. Summers are with the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory and Clinical Image Processing Service. All authors are with the Radiology and Imaging Sciences Department, National Institutes of Health Clinical Center, Bethesda, MD 20892-1182, USA.

#### 1. Introduction

Interstitial lung disease (ILD) refers to a group of more than 150 chronic lung diseases characterized by progressive scarring or inflammation of lung tissues and eventual impairment of breathing. The gold standard imaging modality for ILD diagnosis is computed tomography (CT) [1, 2]. Figure 1 depicts several examples of some most typical ILD-related CT imaging visual patterns. Automated detection of ILD patterns from CT images would aid the diagnosis and treatment of this morbidity.

The majority of previous work on ILD pattern detection is on 2D image classification at the patch level, which attempts to classify relatively small image patches (e.g.,  $32\times32$  pixels) into one of the ILD pattern classes. These image patches are extracted or sampled from manually annotated polygon-like regions of interest (ROIs) on 2D axial slices, following the protocol of [1]. Recent notable approaches include restricted Boltzmann machines [3], convolutional neural networks (CNNs) [4, 5, 6], local binary patterns [7, 8] and multiple instance learning [9]. One prominent exception to the predominant patch-based approach is Gao *et al.*'s work [4] which assigns a *single* ILD class label directly upon whole axial CT slices, without any pre-processing to obtain ROIs.

When analyzing the Lung Tissue Research Consortium (LTRC) dataset [2], which is the most comprehensive lung disease image database with per-pixel annotated segmentation masks, a significant number of CT slices are observed as being associated with two or more ILD labels. Despite the importance of predicting multiple possible ILD pattern types given an input CT image, this challenge has not been addressed by previous studies [5, 3, 4, 7, 8]. ILD pattern detection is usually treated as a single-label classification problem from image patches [5, 3, 7, 8] or slices [4].

Detecting multiple possible ILD types on holistic CT slices simultaneously arguably causes more technical challenges, but it results in a fully automated and clinically more realistic ILD classification process, especially when considering the problem of pre-screening large populations. Without knowing the actual ILD locations and regions of spatial extents a priori (even lung segmentation), the methodological difficulties stem from several aspects, including 1) the tremendous amount of variation in ILD appearance, location and configuration; 2) the expense to obtain delicate pixel-level ILD annotations of large datasets for training and evaluation; and 3) the common occurrence of multiple ILD diseases coexisting on single CT slices. In this study, we target solving these three challenges at the same time.

One way to tackle the multi-label ILD recognition challenge is by replacing the softmax-based single-label loss [10] with a multi-label classification loss layer (Sec. 3.1). Our method works on a holistic CT slice as an input to directly provide multiple ILD patterns existing on that slice, which moves forward an important step to delivering better clinical relevance than previous work [11, 12]. Alternatively, partially inspired by the recent natural image

classification work [13], we explore an alternative method, which models this multi-label prediction problem using a continuously valued regression formulation (Sec. 3.2). Note that multi-label regression has also been used outside of ILD contexts to estimate heart chamber volume [14, 15].

We employ the end-to-end deep CNN regression model because of its simplicity and the fact that deep image features and final cost functions can be learned simultaneously [10, 16]. End-to-end deep neural network representations have shown significant performance superiority over the variants of "hand-crafted image features followed by a separate classifier", in recent studies [17, 18].

While CNNs are powerful image recognition models, its deep image feature learning and encoding representation is not invariant to the spatial locations and layouts of objects or texture patterns within a holistic visual scene (e.g., an input CT slice). As observed in [19, 20], this order-sensitive CNN feature encoding, reflecting the spatial layout of the local image descriptors, is effective in object and scene recognition but may not be beneficial, or can even be counter-productive, for texture classification. The default order-sensitive spatial encoding of CNN image descriptors can be removed through the schemes of unordered feature encoders, such as bag of visual words (BoVW), Fisher vectors (FV) [21], or aggregation of spatial pyramid matching (SPM) [20], etc. Previous work on image patch based approaches [5, 3, 7, 8], are equivalent to formulating ILD pattern recognition as texture classification since the gross image layout information is discarded. Therefore, given the above considerations, we attempt to answer the question whether ILD recognition is indeed a texture classification problem by performing spatially invariant feature encoding from image feature activations from the CNN regression architecture, followed by dimension reduction and multivariate linear regression (Sec. 3.3).

Our methods are validated using the publicly available LTRC ILD dataset [2], composed of 658 patients which are all the data in LTRC consisting of good ILD annotations. Our experiment protocol employs five-fold cross-validation (CV) to detect the most common ILD classes of Ground Glass, Reticular, Honeycomb, and Emphysema. Extensive quantitative experimental results show the promise of our approach in tackling the challenges of multi-class ILD classification given any input CT slice, without any manual pre-processing.

# 2. Related work

Detecting ILD patterns in CT imaging is commonly treated as a texture recognition and classification problem in many previous studies [22, 8, 23, 24, 25]. Moreover, texture based visual representation is adopted inside local image regions of interest (ROIs) or volumes of interest (VOIs) via extracting rectangular image patches, when a 2D and 3D CT imaging modality is used, respectively. In the sliding window manner, image classifiers can generate an ILD probability map within a pre-segmented lung region. Image feature extraction and machine learning based classification are two separate factors in building previous image recognition systems.

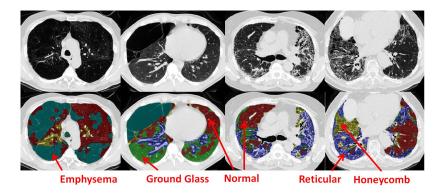


Figure 1: Examples of ILD patterns. Every voxel in the lung region is labeled as healthy or one of the four ILD patterns: Ground Glass, Reticular, Honeycomb, or Emphysema. The first row is the lung CT images. The second row refers to their corresponding labeling.

An early work on computer-aided ILD recognition is proposed to employ neural networks and expert rules to detect ground glass opacity (GGO) on CT images [26]. The follow-up work includes GGO detection and segmentation [27, 28]. Shyu et al. [29] describe a human-in-the-loop approach where the human annotator delineates the region of interest and anatomical landmarks in the images, followed by classification on image attributes related to variations in intensity, texture, shape descriptors and so on. Zheng et al. [30] analyze 3D ILD imaging regions that are combined from multiple candidates detected beforehand on 2D slices. Fukushima et al. [31] evaluate the diagnostic performance of an artificial neural network.

There are many types of hand-crafted image features that are adopted for ILD classification, such as filter banks[22, 8, 23], local binary patterns (LBPs) [24, 25], morphological operators followed by geometric measures, histogram of oriented gradients [8], texton based approaches [32], and wavelet and contourlet transforms [33, 34]. 2D texture features have also been extended into three dimensions [35, 36, 28]. Typical feature encoding scheme and classifiers include bag of words [37], support vector machines (SVMs) [24, 35, 34], random forest [23] and k-nearest neighbors (kNN) [22].

In contrast to separate hand-crafted features and classifier modeling, convolutional neural networks (CNN) can learn image features and the classifier simultaneously. Restricted Boltzmann machines (RBMs) have been used to learn unsupervised classification features within lung regions [38], whereas CNNs are used in a supervised formulation [39]. In [3], a convolutional classification RBM is trained combining a generative and a discriminative learning objective. [5] proposes a specially designed CNN architecture for the classification of ILD patterns. This network consists of five convolutional layers with  $2\times 2$  kernels and LeakyReLU activations, followed by average pooling and three fully connected layers. The size of the kernels in each layer is chosen to be minimal, which leads to deep networks, similar to VGG-net [40]. [6] articulates several important

approaches toward employing CNNs in medical imaging applications. The ILD pattern classification problem was explored and evaluated using different CNN architectures. In particular, transfer learning was studied using pre-trained ImageNet CNN models [10] to fine-tune on domain-specific tasks of medical imaging detection and diagnosis.

A preliminary version of this work appears in [11]. In this paper, we propose, extend and fully evaluate two different multi-label CNN classification architectures to address the phenomenon of multiple ILDs' co-occurrence on single CT images. Robust deep regression loss function under multi-label setting is also addressed. The improved algorithms are extensively validated with a more complete dataset, using comprehensive evaluation metrics, and by conducting comparable experiments against patch based ILD classification, which constitutes the majority of previous work. Superior quantitative performance in both detection accuracy and time efficiency is demonstrated.

#### 3. Methods

In this section, we propose three variations of multi-label deep convolutional neural network classification or regression models to address the multi-label ILD detection challenge. First, an end-to-end CNN network is trained using a multi-label image classification loss layer. Second, we outline a CNN network that uses a continuously-valued regression formulation, estimating either the actual pixel numbers occupied per ILD class per CT image or the binary [0,1] occurring status. Third, the convolutional image activation feature maps at different network depths are spatially aggregated and encoded through the orderless Fisher vector (FV) encoder [21]. This encoding scheme removes the spatial configurations/layouts of convolutional activations and turns them into location-invariant feature representations. This type of CNN is referred to as FV-CNN [19]. The formed orderless features are then trained with a multivariate linear regressor (Mvregress(\*) function in Matlab) to regress the ILD pixel numbers or binary labels.

There are several mainstream CNN architectures, such as AlexNet [10], VG-GNet [40], GoogLeNet [41], and deep residual networks [42]. Each network has its own advantages and is suitable for specific applications. Here we employ a variation of AlexNet, called CNN-F [43], for its good trade-off between efficiency and performance. Fully-annotated medical imaging datasets are usually of limited availability and can be much smaller than the popular computer vision ImageNet database [44]. The classical CNN-F contains five convolutional layers, followed by two fully-connected (FC) layers, and a last softmax layer for classification. We modify it to accommodate our application of detecting multiple ILD patterns in CT images, as shown in Fig. 2. Based on our empirical test using a much deeper CNN model of VGG-19, deeper models do not provide significantly noticeable quantitative performance boosts in ILD classification accuracy while at the same time they consume much more training and testing time.

Our three main deep learning algorithms, namely multi-label CNN classification, robust deep regression, and unordered pooling multivariate regression, are described in Sec. 3.1, Sec. 3.2 and Sec. 3.3, respectively. Two additional critical technical aspects are then addressed, i.e., balancing the distribution of different classes to achieve the performance boost (Sec. 3.4), and exploiting different CT attenuation scaling schemes to better capture the visual appearance of abnormal ILD patterns (Sec. 3.5).

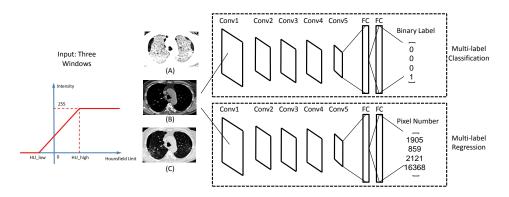


Figure 2: Multi-label CNN models: the input CT slices are transformed into three attenuation scales from the original CT value range that can highlight different anatomical tissues and lung disease patterns), before being fed into CNN for processing. Three CNN models are proposed to detect possibly multiple ILD diseases per CT image  $(512 \times 512 \text{ pixel CT}$  slice instead of smaller image ROIs or patches in most previous work). The soft-max layer of single-label multi-class CNN is replaced by either a multi-label classification CNN loss layer (Sec. 3.1) or real-valued regression loss layer (Sec. 3.2).

#### 3.1. Multi-label Classification Deep CNN

The multi-label classification problem has not been widely studied in ILD recognition because in previous image patch based representation, each patch belongs to a single ILD class making it unnecessary to add the additional complexity of multi-labeling. In the holistic CT slice based ILD prediction, we first tackle this task using a multi-label classification CNN loss, under a one-against-all formulation (Eq. 1), as studied in the recent computer vision literature [45, 46, 47]. The typical logistic regression loss in the soft-max CNN layer is only capable of predicting a single label for each instance, which is not suitable to our application. Our goal is to identify the existence of all occurring ILDs simultaneously where the input image may contain multiple ILD patterns. Our employed loss function is intuitive and effective in the form of a sum of C binary logistic regression losses, one for each of the C classes  $k \in \{1 \dots C\}$ ,

$$L(y, f(x)) = \sum_{k=1}^{C} log(1 + exp(-y_k f_k(x))),$$
 (1)

where  $y_k \in \{-1, 1\}$  is the label indicating the absence/presence of class k given input image x and  $f_k(x)$  is the output of the network.

In each loss computation per input image, the logistic loss and gradients according to all "positive" labels (on that image) are calculated and added within the network loss layer intrinsically, for the stochastic gradient back-propagation (BP) of neural network training. There is another possible design choice, which is to treat the multi-label classification problem as C independent multi-task classification problems. In contrast to Eq. 1, we need to have C separate binary loss layers to cover all ILD classes and the gradient BP process during training will be independently computed. Thus this "multi-head" multi-mask CNN can not model the intrinsic correlations among multiple ILD labels, which based on our empirical finding, causes the multi-mask network hard to converge.

#### 3.2. Robust Multi-label Regression Deep CNN

Next, we propose and investigate this multi-label problem via deep regression losses. Supposing that there are in total N images and C types of ILDs to be detected, the label vector for an image I is represented as a C-dimensional label vector  $\mathbf{y} = [y_1, \ldots, y_k, \ldots, y_C] \in \{0, 1\}^C, k \in \{1, \ldots, C\}$ , in which each entry can be 1 or 0, indicating whether a specific disease exists in the image. For the case of one image containing multiple diseases, there will several corresponding 1's in the label vector  $\mathbf{y}$ . An all-zero  $\mathbf{y}$  represents a healthy slice/instance, *i.e.*, no targeted ILD being found. Alternately, the actual number of pixels of each ILD pattern per image can be recorded in  $\mathbf{y}$ , replacing the binary value of 1 with an integer quantity. This multivariate label vector allows our algorithm to naturally preserve the frequent co-occurrence property of ILDs in CT imaging through (deep) regression. We summarize the different options below.

Loss Functions: Deep CNN regression loss is used to learn the presence or the spatial occupancy area (in terms of the pixel number) of ILD classes per image. The following loss function (Eq. 2) [13, 48] is adopted instead of the more widely used softmax loss for classification CNNs [4, 5, 10]. The loss cost function to be minimized is defined as

$$L(y, f(x)) = \sum_{k=1}^{C} L_{reg}(y_k - f(x)),$$
 (2)

where  $L_{reg}$  could be either  $L_2 = x^2$  loss or smooth  $L_1$  loss. The smooth  $L_1$  cost function [48] is defined as

$$\operatorname{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1, \\ |x| - 0.5 & \text{otherwise.} \end{cases}$$
 (3)

This robust smooth  $L_1$  loss is designed to be less numerically sensitive to outliers (extremely large targeted label values) than  $L_2$  loss. The use of Eq. 3 could eliminate the chances of exploding gradients, to which the  $L_2$  loss is subject.

Binary or Continuously Valued Regression: There are several options to form the regression labels for each image. One straightforward scheme is to

count the total number of pixels annotated per ILD disease, which represents its severity (Fig. 3 Left). The step function to represent the presence (1) or absence (0) of the disease (Fig. 3 Middle) is also possible. The binarizing threshold T may be defined using clinical knowledge: if the pixel number is  $\geq T$ , the label is set to be 1; otherwise as 0. A more sophisticated label transfer model is a piecewise linear function of the pixel counts with  $T_1, T_2$ , mapping pixel counts to the a range of [0,1] (Fig. 3 Right). Diseases with the number of total pixels  $\geq T_1$  but  $\leq T_2$  are linearly interpolated to between 0 and 1.

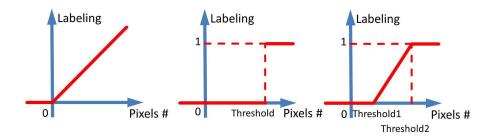


Figure 3: Three mapping functions transfer the pixel counts (per ILD class) to their label values for training the CNN regression losses.

# 3.3. Unordered Pooling via Fisher Vector Encoding for Multivariate Linear Regression

Both classification and regression CNN models (Sec. 3.1 and 3.2) can be seen as performing the spatially order-sensitive feature pooling through their use of fully-connected CNN layers. In this section, we investigate whether the spatial information captured inside CNN activation maps is beneficial for a task-specific image recognition problem. The typical representation of deep hierarchical CNN features inherits the gross image-activation spatial layouts. We are motivated by the observation that ILD patterns could happen anywhere inside the lung region, implying that the spatial layout may not be a strongly-correlated factor to ILD recognition. CNNs are designed to learn special feature layouts from the limited annotated ILD imaging data, which may be subject to over-fitting more easily. In our implementation, CNN activations are extracted from the convolutional layers at various depths of the CNN network and compiled using the Fisher vector (FV) feature encoding scheme [21, 19], allowing us to achieve a location-invariant deep texture description. ILD class labels can then be predicted via the simple multivariate linear regression.

The output of each k-th convolutional layer is a 3D descriptor or matrix  $X_k \in \mathbb{R}^{W_k \times H_k \times D_k}$ , where  $W_k$  and  $H_k$  are the width and height of the spatial reception field and  $D_k$  is the number of feature channels. In this sense, the specific deep feature activation map is represented by  $W_k \times H_k$  feature vectors and each feature vector is  $D_k$  dimension. We invoke the FV encoding to remove the spatial configurations of total  $W_k \times H_k$  vectors (denoted as the set

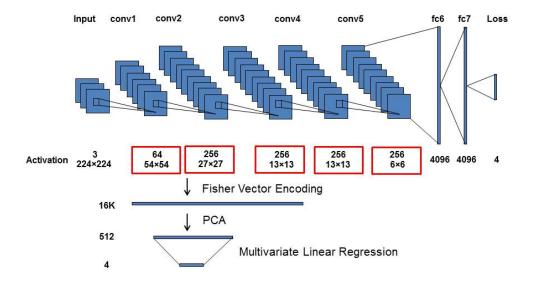


Figure 4: The overall architecture of the CNN model and unordered polling using FV encoding and multivariate linear regression.

 $X_k$ ) for each activation map. Following [21], each feature descriptor  $x_i \in X_k$  is soft-quantized using a Gaussian mixture model. The first- and second-order differences  $(u_{i,m}^T, v_{i,m}^T)$  between any descriptor  $x_i$  and each of the Gaussian cluster mean vectors  $\{\mu_m\}, m=1,2,...,M$  are accumulated into a  $2MD_k$ -dimensional image representation:

$$\boldsymbol{f_{i}^{FV}} = [u_{i,1}^{T}, v_{i,1}^{T}, ..., u_{i,M}^{T}, v_{i,M}^{T}]. \tag{4}$$

FV feature encoding produces a very high dimensionality of  $2MD_k$  from the deep feature activations of  $X_k$  for each image (e.g., M=32 and  $D_k=256$ ). For computational and memory efficiency, we adopt principal component analysis (PCA) to project the  $f_i^{FV}$  vectors to a lower-dimensional parameter subspace. Using their ground-truth ILD label vectors  $y_i$ , the multivariate linear regression Mvregress(\*) function in Matlab is called to predict the presence or non-presence of ILDs from the low-dimensional projected image feature vectors  $PCA(f_i^{FV})$ .

### 3.4. Class Balancing

Most computer-aided detection datasets have very biased distributions for instances of different classes. In our setting, some ILD types may appear more often than others. We utilize a simple and effective strategy to automatically balance the loss among different classes [49]. A class-balancing weight  $\beta_k$  for each class k is added into the classification CNN loss.

$$L(y, f(x)) = \sum_{k=1}^{C} \beta_k \log(1 + \exp(-y_k f_k(x))),$$
 (5)

Similarly, the weighted multi-label regression loss is

$$L(y, f(x)) = \sum_{k=1}^{C} \beta_k L_{\text{reg}}(y_k - f(x)), \tag{6}$$

where  $\beta_k = \frac{1 - |Y_k|/|Y|}{C}$ ,  $|Y_k|$  denotes the cardinality of the dataset of class k according to the ground truth labels, and  $|Y| = \sum_{k=1}^{C} |Y_k|$ .

#### 3.5. CT Attenuation Rescaling

To better capture the abnormal ILD patterns in CT images, we select three CT attenuation ranges or windows and rescale them to [0, 255] for CNN input. This is inspired by the fact that radiologists adjust the CT contrast window to optimize the visualization effects for certain tissues or pathologies on CT scans. For example, "Lung" window visualizes details in lung tissue that are not apparent on the "Bone" window, whereas much of the information in bone and soft tissue will be lost on the "Lung" window. In this work, we use three CT attenuation scales to highlight different lung disease patterns. As demonstrated in Fig. 2(A), this process is designated to preserve the attenuation values between HU\_low and HU\_high via a linear transformation. The intensity values outside the specific attenuation window are set as 0 or 255. In details, the low attenuation range (Fig. 2(B)) is used to capture ILD patterns with lower intensities, such as emphysema; the normal range (Fig. 2(C)) to represent normal appearance of lung regions; and high attenuation range for highlighting patterns with higher intensities, for example, consolidation and nodules. In our experiments, the low attenuation window is set as  $HU_low = -1400$  and  $HU_high = -950$ ; for normal range, HU\_low= -1400 and HU\_high= 200; for high attenuation scale,  $HU_{low} = -160$  and  $HU_{high} = 240$ .

#### 4. Experiments and Discussion

#### 4.1. Data

There are two main publicly available datasets for CT imaging based ILD classification [1, 2]. The LTRC [2] dataset provides complete ILD labeling at the per-voxel or per-pixel level. In contrast, missing labels are common in [1]. Additionally, not all ILD regions of interest per-slice are delineated by annotators and often only one prominent disease region is annotated on a slice, as studied in [12, 1]. As a result, we use the LTRC dataset for our method validation and performance evaluation. Every voxel inside the CT lung region is labeled as healthy or one of the four ILD types: Ground Glass, Reticular, Honeycomb or Emphysema. Our goal is, given any input axial CT slice, to predict ILD labels where multiple diseases could co-occur. The number of classes C is set as 4 to represent the four ILD types.

For ease of comparison, 2D axial CT slices or images are evaluated independently, without taking successive slices into consideration. Many CT scans of ILD study can have large inter-slice distances, for example, 10mm in [1] between

successive axial slices, making direct 3D volumetric analysis implausible. Utilizing only 2D axial image information makes the algorithm more generalizable to low-dose CT imaging based ILD screening protocols.

In total, there are 658 patients in the LTRC dataset for ILD classification and detection. Some ILD patients, which cannot find matched annotations with CT images, are eliminated. The original resolution of the 2D axial slices is  $512 \times 512$  pixels. All images are resized to the uniform size of  $224 \times 224$  pixels. Five-fold cross-validation (split at patient level) is conducted for the quantitative experimental evaluation. There are  $\sim\!240\rm k$  CT slices in total from the 658 patients for cross-validation (CV). CNN training is performed in Matlab using MatConvNet [16] and run on a PC with 3.1GHz CPU, 32 GB memory, and an Nvidia Tesla K40 GPU.

#### 4.2. Multi-label Classification Evaluation

Single label ILD classification can be quantitatively evaluated using recall, precision, and F-score metrics, respectively for each disease. Multi-label classification needs different performance metrics than those used in the single-label scenario [50]. Let T denote the ground truth set of labels; and S be the predicted set. Accuracy is measured by the Hamming score which is symmetrical measurement of how close T is to S, as illustrated in Eq. 7. Similar formulations are applied to calculate precision and recall under multi-label classification evaluation (Eq. 8 and 9). F-score, which is the harmonic mean of precision and recall, keeps the same for both single and multi-label classification evaluations (Eq. 10). In our experiments, we evaluate the overall multi-label ILD prediction performance and report the results for each individual ILD as well.

$$Accuracy(T,S) = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap S_i|}{|T_i \cup S_i|},$$
(7)

$$Precision(T, S) = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap S_i|}{|S_i|},$$
(8)

$$\operatorname{Recall}(T, S) = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap S_i|}{|T_i|}, \tag{9}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall},\tag{10}$$

## 4.3. Results on Multi-label Classification CNN

To conduct holistic CT slice based ILD classification, we first convert the pixel-wise annotated masks in LTRC [2] into slice-level labels. Without loss of generality, we set the pathology threshold T=6000 or T=4000 pixels to differentiate the presence (if  $\geq T$ ) or absence (if  $\leq T$ ) of ILDs. The number of slices containing each ILD pattern is outlined in Table 1 when the pathology presence threshold is set as T=6000. There are many CT slices or instances with multiple ILDs co-existing on the same slice, as shown in Table 2 .

Table 1: Statistics on the 658 patients and 240k CT slices from LTRC [2]. Without loss of generality, threshold T=6000 pixels is used to differentiate the presence or absence of ILD patterns.

| ILD pattern  | Positive | Negative |
|--------------|----------|----------|
| Healthy      | 226675   | 15362    |
| Ground Glass | 41194    | 200843   |
| Reticular    | 20560    | 221477   |
| Honeycomb    | 17392    | 224645   |
| Emphysema    | 36328    | 205709   |

Table 2: The number of slices with multiple ILD patterns coexisting on the same slice. Pathology presence threshold T=6000 pixels is used.

| Healthy | One Disease | Two Diseases | Three Diseases | Four Diseases |
|---------|-------------|--------------|----------------|---------------|
| 149950  | 70339       | 20127        | 1603           | 18            |

The classification results are shown using the ROC curves in Fig. 5, and F-scores are presented in Table 3 and Table 4 while setting the pathology presence thresholds to be T=6000 and T=4000, respectively. The overall F-score is calculated based on the multi-label classification evaluation mentioned in Sec. 4.2. We obtain good results using the setting of T=4000 but the quantitative results by setting T=6000 are further improved, indicating that our algorithm may be robust to detect smaller ILD patterns and can tolerate some pixel-level annotation errors in LTRC. In our setting, pixel-level ILD annotations are not essentially required. Therefore the medical experts can simply provide the holistic CT slice-level labels on any lung CT image to indicate if there are ILD presences worth reporting, without annotating particular ILD image ROIs. It would considerably save the labeling time for experts to annotate the training dataset.

# 4.4. Results on Multi-label Regression CNN

We can treat the continuously-valued output vector, either in the form of pixel number counts or binary presence status, as the "classification confidence scores" after the multi-label regression CNN processes an input CT image. These regressed confidence scores can be compared against the ground truth binary ILD labels obtained by thresholding on T as in Sec. 4.3. In this manner,

Table 3: F scores of multi-label classification and regression CNNs, with the setting T=6000 pixels.

|                              | F-score      |           |           |           |         |  |
|------------------------------|--------------|-----------|-----------|-----------|---------|--|
| Disease                      | Ground Glass | Reticular | Honeycomb | Emphysema | Overall |  |
| classification               | 0.8642       | 0.7686    | 0.4602    | 0.9468    | 0.7959  |  |
| regression $L_2$ loss        | 0.8343       | 0.4764    | 0.2314    | 0.8042    | 0.6882  |  |
| regression smooth $L_1$ loss | 0.9102       | 0.7095    | 0.3385    | 0.8991    | 0.8028  |  |

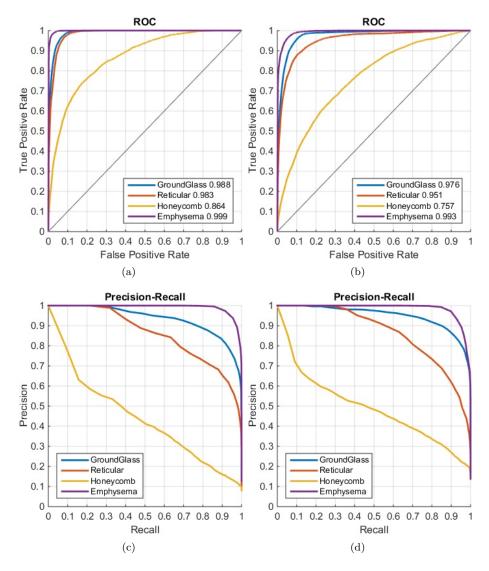


Figure 5: ILD classification results shown in ROC and Precision-Recall curves using the threshold to 6000 pixels in (a)(c) and 4000 pixels in (b)(d).

ILD classification receiver operating characteristic (ROC) curves are generated. Our experiments are conducted via the three label converting functions or plots in Fig. 3. Two variations of CNN regression outputs to match the ILD occupied pixel numbers per-slice, or the binary ILD presence labels produce similar quantitative ILD classification results. The piecewise linear transformation (Fig. 3 Right) yields slightly inferior results.

Table 3 and Table 4 show the multi-label regression CNN results where the

Table 4: F scores of multi-label classification and regression CNNs, with the setting T=4000 pixels.

|                              | F-score      |           |           |           |         |  |
|------------------------------|--------------|-----------|-----------|-----------|---------|--|
| Disease                      | Ground Glass | Reticular | Honeycomb | Emphysema | Overall |  |
| classification               | 0.8825       | 0.7667    | 0.5068    | 0.9361    | 0.7656  |  |
| regression $L_2$ loss        | 0.8385       | 0.5868    | 0.3102    | 0.8008    | 0.6487  |  |
| regression smooth $L_1$ loss | 0.9079       | 0.7190    | 0.4092    | 0.9152    | 0.7774  |  |

Table 5: AUC values among different CNN layers by FV encoding and linear regression. Both CNN and multi-variant linear regression regress to ILD pixel numbers.

|              |       | AUC   |       |       |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|
| Disease      | conv1 | conv2 | conv3 | conv4 | conv5 | fc6   | CNN   |
| Ground Glass | 0.979 | 0.978 | 0.984 | 0.985 | 0.984 | 0.970 | 0.990 |
| Reticular    | 0.951 | 0.953 | 0.955 | 0.957 | 0.950 | 0.900 | 0.964 |
| Honeycomb    | 0.765 | 0.770 | 0.780 | 0.744 | 0.753 | 0.743 | 0.809 |
| Emphysema    | 0.985 | 0.990 | 0.987 | 0.989 | 0.988 | 0.985 | 0.995 |

trained model regresses to the number of diseased pixels in each image. The use of a smooth  $L_1$  cost function greatly improves the performance and constantly outperforms the  $L_2$  cost function in all experiments by noticeably large margins. Fig. 6 illustrates some visual examples of successful or misclassified results. The first four examples are successfully detected cases, with multiple ILD patterns coexisting on the same slice. The last two are failure cases. Note that the first misclassified case is marked with two detected labels of "emphysema" and "ground glass". Both emphysema and ground glass co-occur on this image but the pixel count of ground glass occupied spatial region does not meet the pathology threshold of  $\geq T=6000$ . These qualitative results visually confirm the high performance demonstrated by our quantitative evaluation.

The overall performance of multi-label regression CNN when the smooth  $L_1$  loss is employed is generally comparable with the multi-label classification CNN (Sec. 4.3). From Table 4 and 3, the smooth  $L_1$  regression CNN performs slightly better overall and particularly for the ground glass class, but the multi-label classification CNN outperforms in the categories of reticular, honeycomb, and emphysema with moderate margins.

# 4.5. Unordered Pooling via Fisher Vector Encoding

When constructing the FV-encoded features,  $f_i^{FV}$ , the local convolutional image descriptors are pooled into 32 Gaussian components, producing a dimensionality as high as 16K [21]. We further reduce the FV features to 512 dimensions using PCA. The performance is empirically found to be insensitive to the number of Gaussian kernels and the dimensions after PCA. We compare the ILD classification performance with FV encoding, on the features pooled from different CNN layers, using area-under-the-curve (AUC) values (in Table 5) and F-scores (in Table 6), respectively.

Table 6: F-scores between different layers. Both CNN and multi-variant linear regression regress to ILD pixel numbers.

|              |       | F-score |       |       |       |       |       |
|--------------|-------|---------|-------|-------|-------|-------|-------|
| Disease      | conv1 | conv2   | conv3 | conv4 | conv5 | fc6   | CNN   |
| Ground Glass | 0.871 | 0.875   | 0.883 | 0.885 | 0.889 | 0.868 | 0.908 |
| Reticular    | 0.699 | 0.711   | 0.709 | 0.714 | 0.698 | 0.629 | 0.719 |
| Honeycomb    | 0.349 | 0.384   | 0.386 | 0.356 | 0.384 | 0.344 | 0.409 |
| Emphysema    | 0.879 | 0.895   | 0.887 | 0.897 | 0.893 | 0.877 | 0.915 |
| Overall      | 0.713 | 0.750   | 0.750 | 0.735 | 0.754 | 0.716 | 0.777 |

Table 7: Comparing the AUC values between different layers using a smaller dataset of 18k slices. Both CNN and multi-variant linear regression regress to ILD pixel numbers

|              |       | AUC   |       |       |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|
| Disease      | conv1 | conv2 | conv3 | conv4 | conv5 | fc6   | CNN   |
| Ground Glass | 0.984 | 0.955 | 0.953 | 0.948 | 0.948 | 0.930 | 0.943 |
| Reticular    | 0.976 | 0.958 | 0.954 | 0.951 | 0.950 | 0.939 | 0.917 |
| Honeycomb    | 0.898 | 0.826 | 0.828 | 0.823 | 0.806 | 0.773 | 0.698 |
| Emphysema    | 0.988 | 0.975 | 0.967 | 0.966 | 0.967 | 0.985 | 0.988 |

When evaluated using a smaller ILD dataset, the same as the one used in [11, 1] of 18k CT slices, FV order-less encoding is effective as demonstrated in Table 7. The unordered pooling operating on the first CNN convolutional layer conv1 produces the overall best quantitative results, especially for honeycomb. Despite residing in the first layer, the filters and activations on conv1 are still the integrated parts of a deep network since they are learned through backpropagation from deeper layers. From Table 7, FV encoding with deeplylearned conv1 filter activations produces the best ILD classification against FV encoding on other layers and without FV. Nevertheless, for the much larger dataset of 240k CT images under 5-fold CV, the computational complexity of FV encoding becomes the performance bottleneck. It could take an undesirably long time and huge memory requirement to calculate the FV Gaussian components and perform the feature encoding. In our experiments, we randomly select a smaller subset of deep activation features ( $\sim 1/3$ ) to calculate FV encoding, which may limit the FV encoding performance. In this setting of sufficiently large amount of data, CNN models without FV encoding perform better.

#### 4.6. Class Balancing

Even though data is not highly unbalanced, as shown in Table 1, using the strategy to balance the classes helps promote the performance considerably. With all the other settings fixed, the integration of the class balancing factors  $\beta$  into the CNN loss function (Eq. 5) improves the ILD classification performances on every disease class. The overall F-score also increases by  $\sim 5\%$ , as shown in Table 8. We observe similar a performance boost when Eq. 6 is employed.

Table 8: F-scores of multi-label classification, T=4000 pixels, comparing the results with and without class balancing.

|               | F-score      |           |           |           |         |  |  |
|---------------|--------------|-----------|-----------|-----------|---------|--|--|
| Disease       | Ground Glass | Reticular | Honeycomb | Emphysema | Overall |  |  |
| w/o balancing | 0.8507       | 0.7397    | 0.4308    | 0.8895    | 0.7146  |  |  |
| w balancing   | 0.8825       | 0.7667    | 0.5068    | 0.9361    | 0.7656  |  |  |

Table 9: Running time (seconds) comparison for the image patch-based classification methods and our proposed holistic approach.

|      | Patch-based Min | Patch-based Max | Patch-based Mean | Holistic Detection |
|------|-----------------|-----------------|------------------|--------------------|
| Time | 4.11            | 32.15           | 22.64            | 0.01               |

#### 4.7. Patch-based Classification Baseline and Processing Time

The image patch-based classification was the state-of-the-art ILD recognition paradigm. To set up a baseline, we have implemented a standard patch-based algorithm for comparison. The patch-based ILD detection is evaluated on user-defined  $32 \times 32$  pixel image patches and predicts a single ILD label for each patch. The image patch-based ILD classification is an easier problem compared to holistic slice-based recognition [4]. However, it is not suitable to efficiently predict multiple ILD patterns simultaneously that co-occur on a single CT image. As shown in Table 9, the running time is the heaviest burden of the patch-based methods. We adopt the most commonly used sliding window method to detect all the ILDs within an image. Labels are predicted for image patches sampled at a spatial interval of 10 pixels. The running time for each slice ranges from 4.11 to 32.15 seconds, with a mean of 22.64 seconds, depending on the actual area of lung region on that slice. On the other hand, our proposed holistic method takes less than 0.01 seconds to process a CT slice. It finishes evaluating 50k test slices in < 8 minutes.

Furthermore, most patch-based methods require the pixel-level lung segmentation as preprocessing. Although the healthy lung segmentation problem is relatively easy to solve, pathological lung segmentation remains an obstacle. The benefit of patch-based classification is that it can explicitly present the location of the ILDs on CT slices. However it is possible to adapt the slice-level CNN models to localize of the underlying diseased regions under a weakly-supervised learning fashion [51], which provides a way to overcome this drawback with high computational efficiency. We leave this as future work.

#### 5. Conclusion

In this paper, we present three multi-label deep CNN classification and regression models to accurately recognize potential multiple ILD co-occurrence on an input lung CT slice, in a holistic manner. In contrast to previous image patch based approaches where manual ILD ROIs are given as prerequisites [7, 5, 3],

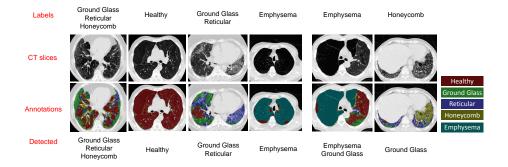


Figure 6: Examples of successfully detected and misclassified ILD slices. The left four are the correctly labeled cases, and the right two are failed cases.

our method performs the task of multi-label, multi-class ILD detection simultaneously, with no image preprocessing on CT slices. Moreover, we also investigate the effectiveness of exploiting the unordered reformation or pooling from deep CNN convolutional activation features via a FV encoding scheme. The proposed algorithms are validated on a publicly available dataset of 658 patients under five-fold cross-validation, achieving high AUC values and F-scores for detecting four main types of ILDs. Our method can be readily adapted to other CAD problems that face similar large spatial and appearance variations. Future work includes performing cross-dataset transfer learning and incorporating weakly-supervised deep CNN approaches to provide ILD localization information on the slices. Last but not least, the flexible holistic CT slice based deep ILD recognition protocol represents a significant step toward clinically useful automated image analyses.

#### Acknowledgments

This research is supported by the NIH Intramural Research Program, the Center for Infectious Disease Imaging, the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, the National Institute of Allergy and Infectious Diseases and the NIH Clinical Center. We also thank Nvidia for the donation of a Tesla K40 GPU.

#### References

#### References

[1] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, H. Müller, Building a reference multimedia database for interstitial lung diseases, Computerized medical imaging and graphics 36 (3) (2012) 227–238.

- [2] B. Bartholmai, R. Karwoski, V. Zavaletta, R. Robb, D. Holmes, The lung tissue research consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease.
- [3] G. van Tulder, M. de Bruijne, Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines, IEEE Trans. on medical imaging 35 (5) (2016) 1262–1272.
- [4] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, et al., Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2016) 1–6.
- [5] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, IEEE Trans. on medical imaging 35 (5) (2016) 1207–1216.
- [6] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, IEEE Trans. on medical imaging 35 (5) (2016) 1285–1298.
- [7] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, M. Chen, Large margin local estimate with applications to medical image classification, IEEE Trans. on medical imaging 34 (6) (2015) 1362–1377.
- [8] Y. Song, W. Cai, Y. Zhou, D. D. Feng, Feature-based image patch approximation for lung tissue classification, IEEE Trans. on medical imaging 32 (4) (2013) 797–808.
- [9] J. Hofmanninger, G. Langs, Mapping visual features to semantic profiles for retrieval in medical imaging, in: IEEE Conf. on CVPR, 2015, pp. 457–465.
- [10] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012, pp. 1097–1105.
- [11] M. Gao, Z. Xu, L. Lu, A. P. Harrison, R. M. Summers, D. J. Mollura, Multi-label deep regression and unordered pooling for holistic interstitial lung disease pattern detection, in: Machine Learning in Medical Imaging, Springer, 2016, pp. 147–155.
- [12] M. Gao, Z. Xu, L. Lu, A. Wu, I. Nogues, R. M. Summers, D. J. Mollura, Segmentation label propagation using deep convolutional neural networks and dense conditional random field, in: IEEE ISBI, 2016, pp. 1265–1268.
- [13] Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, S. Yan, CNN: Single-label to multi-label, arXiv preprint arXiv:1406.5726.

- [14] X. Zhen, A. Islam, M. Bhaduri, I. Chan, S. Li, Direct and simultaneous four-chamber volume estimation by multi-output regression, in: MICCAI, Springer, 2015, pp. 669–676.
- [15] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, S. Li, Direct estimation of cardiac bi-ventricular volumes with regression forests, in: MICCAI, Springer, 2014, pp. 586–593.
- [16] A. Vedaldi, K. Lenc, Matconvnet: Convolutional neural networks for matlab, in: ACM Conf. on Multimedia Conference, 2015, pp. 689–692.
- [17] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology detection using deep learning with non-medical training, IEEE ISBI (2015) 294–297.
- [18] B. van Ginneken, A. Setio, C. Jacobs, F. Ciompi, Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans, IEEE ISBI (2015) 286–289.
- [19] M. Cimpoi, S. Maji, I. Kokkinos, A. Vedaldi, Deep filter banks for texture recognition, description, and segmentation, Int. J. of Computer Vision 118 (1) (2016) 65–94.
- [20] Y. Gong, L. Wang, R. Guo, S. Lazebnik, Multi-scale orderless pooling of deep convolutional activation features, in: European Conference on Computer Vision, Springer, 2014, pp. 392–407.
- [21] F. Perronnin, J. Sánchez, T. Mensink, Improving the fisher kernel for large-scale image classification, in: European conference on computer vision, Springer, 2010, pp. 143–156.
- [22] I. C. Sluimer, P. F. van Waes, M. A. Viergever, B. van Ginneken, Computeraided diagnosis in high resolution CT of the lungs, Medical physics 30 (12) (2003) 3081–3090.
- [23] M. Anthimopoulos, S. Christodoulidis, A. Christe, S. Mougiakakou, Classification of interstitial lung disease patterns using local dct features and random forest, in: IEEE EMBS, 2014, pp. 6040–6043.
- [24] L. D. Bagesteiro, L. F. Oliveira, D. Weingaertner, Blockwise classification of lung patterns in unsegmented ct images, in: IEEE Symposium on Computer-Based Medical Systems, 2015, pp. 177–182.
- [25] L. Sorensen, S. B. Shaker, M. De Bruijne, Quantitative analysis of pulmonary emphysema using local binary patterns, IEEE Trans. on medical imaging 29 (2) (2010) 559–569.
- [26] K. Heitmann, H.-U. Kauczor, P. Mildenberger, T. Uthmann, J. Perl, M. Thelen, Automatic detection of ground glass opacities on lung HRCT using multiple neural networks, European radiology 7 (9) (1997) 1463–1472.

- [27] J. Zhou, S. Chang, D. N. Metaxas, B. Zhao, L. H. Schwartz, M. S. Ginsberg, Automatic detection and segmentation of ground glass opacity nodules, in: MICCAI, Springer, 2006, pp. 784–791.
- [28] Y. Tao, L. Lu, M. Dewan, et al., Multi-level ground glass nodule detection and segmentation in ct lung images, in: MICCAI, 2009, pp. (1): 715–723.
- [29] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, L. S. Broderick, Assert: a physician-in-the-loop content-based retrieval system for HRCT image databases, Computer Vision and Image Understanding 75 (1) (1999) 111–132.
- [30] B. Zheng, J. K. Leader, C. R. Fuhrman, F. C. Sciurba, D. Gur, Automated detection and classification of interstitial lung diseases from low-dose CT images, in: SPIE Medical Imaging, 2004, pp. 849–856.
- [31] A. Fukushima, K. Ashizawa, T. Yamaguchi, N. Matsuyama, H. Hayashi, I. Kida, Y. Imafuku, A. Egawa, S. Kimura, K. Nagaoki, et al., Application of an artificial neural network to high-resolution CT: usefulness in differential diagnosis of diffuse lung disease, American Journal of Roentgenology 183 (2) (2004) 297–305.
- [32] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. De Bruijne, M. Loog, A texton-based approach for the classification of lung parenchyma in CT images, in: MICCAI, Springer, 2010, pp. 595–602.
- [33] K. T. Vo, A. Sowmya, Multiple kernel learning for classification of diffuse lung disease using HRCT lung images, in: IEEE Engineering in Medicine and Biology, 2010, pp. 3085–3088.
- [34] A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, H. Muller, Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames, IEEE Trans. on Information Technology in Biomedicine 16 (4) (2012) 665–675.
- [35] Y. Xu, E. J. van Beek, Y. Hwanjo, J. Guo, G. McLennan, E. A. Hoffman, Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM), Academic radiology 13 (8) (2006) 969–978.
- [36] A. Depeursinge, P. Pad, A. S. Chin, A. N. Leung, D. L. Rubin, H. Müller, M. Unser, Optimized steerable wavelets for texture analysis of lung tissue in 3-D CT: Classification of usual interstitial pneumonia, in: IEEE ISBI, 2015, pp. 403–406.
- [37] R. Xu, Y. Hirano, R. Tachibana, S. Kido, Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach, in: MICCAI, Springer, 2011, pp. 183–190.

- [38] Q. Li, W. Cai, D. D. Feng, Lung image patch classification with automatic feature learning, in: IEEE EMBC, 2013, pp. 6079–6082.
- [39] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, M. Chen, Medical image classification with convolutional neural network, in: IEEE ICARCV, 2014, pp. 844–848.
- [40] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556.
- [41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: IEEE Conf. on CVPR, 2015, pp. 1–9.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, arXiv:1512.03385.
- [43] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv:1405.3531.
- [44] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. of Computer Vision 115 (3) (2015) 211–252.
- [45] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Is object localization for free?weakly-supervised learning with convolutional neural networks, in: IEEE Conf. on CVPR, 2015, pp. 685–694.
- [46] T. Durand, N. Thome, M. Cord, Weldon: Weakly supervised learning of deep convolutional neural networks, in: IEEE Conf. on CVPR, 2016.
- [47] D. Li, J.-B. Huang, Y. Li, S. Wang, M.-H. Yang, Weakly supervised object localization with progressive domain adaptation, in: IEEE Conf. on CVPR, 2016.
- [48] R. Girshick, Fast R-CNN, in: IEEE ICCV, 2015, pp. 1440–1448.
- [49] S. Xie, Z. Tu, Holistically-nested edge detection, in: IEEE ICCV, 2015, pp. 1395–1403.
- [50] G. Tsoumakas, I. Katakis, Multi-label classification: An overview, Int. J. Data Warehousing and Mining 2007 (2007) 1–13.
- [51] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: IEEE Conf. on CVPR, 2016, pp. 2921–2929.