Holistic Classification of CT Attenuation Patterns for Interstitial Lung Diseases via Deep Convolutional Neural Networks

Mingchen Gao[1], Ulas Bagci[2], Le Lu[1], Aaron Wu[1], Mario Buty[1], Hoo-Chang Shin[1], Holger Roth[1], Georgios Z. Papadakis[1], Adrien Depeursinge[3], Ronald M. Summers[1], Ziyue Xu[1*], and Daniel J. Mollura[1]

[1] National Institutes of Health (NIH), Bethesda, MD 20892, US
[2] University of Central Florida (UCF), Orlando, FL 32816, US
[3] University of Applied Sciences Western Switzerland (HES-SO), Sierre 3960, Switzerland

Abstract. Interstitial lung diseases (ILD) involve several abnormal imaging patterns observed in computed tomography (CT) images. Accurate classification of these patterns plays a significant role in precise clinical decision making of the extent and nature of the diseases. Therefore it is important for developing automated pulmonary computer-aided detection (CAD) systems. Conventionally, this task relies on experts' manual identification of regions of interest (ROIs) as a prerequisite to diagnose potential diseases. This protocol is time consuming and inhibits fully automatic assessment. In this paper, we present a new method to classify ILD imaging patterns on CT images. The main difference is that the proposed algorithm uses the entire image as a holistic input. By circumventing the prerequisite of manually input ROIs, our problem setup is significantly more difficult than previous work but can better address the clinical workflow. Qualitative and quantitative results using a publicly available ILD database demonstrates state-of-the-art classification accuracy under the patch based classification and shows the potential of predicting the ILD type using holistic image.

Keywords: Interstitial Lung Disease, Convolutional Neural Network, Holistic Medical Image Classification

1 Introduction

The interstitial lung diseases cause progressive scarring of lung tissue, which would eventually affect the patients' ability to breathe and get enough oxygen into the bloodstream. High-resolution computed tomography (HRCT) is the standard in-vivo radiology imaging tool for visualizing normal/abnormal imaging patterns to identify the

specific type of ILD [16], and to develop appropriate therapy plans. Examples of these lung tissue patterns are shown in Fig. 1. Computer-aided detection/classification systems are needed for achieving higher recalls on ILD assessment [1]. In particular, the amounts and anatomical positions of abnormal imaging patterns (along with patient history) can help radiologists to optimize their diagnostic decisions, with better quantitative measurements.
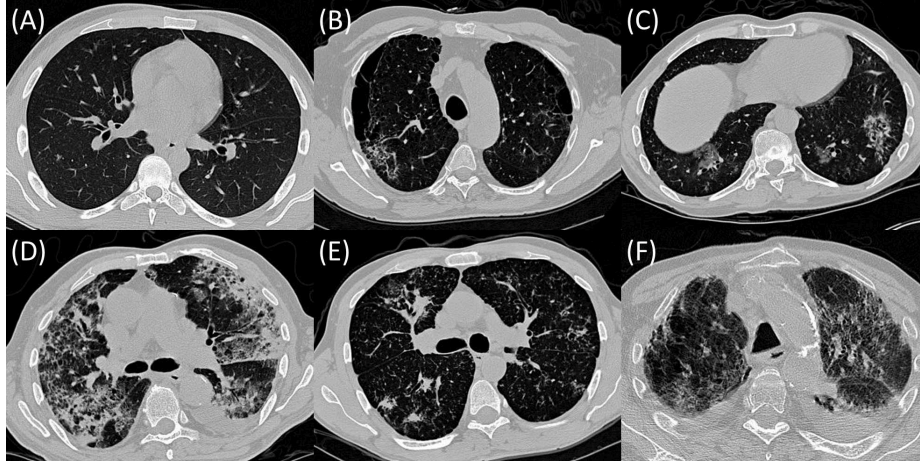

Fig. 1. Example images(segment of HRCT axial slices) for each of the six lung tissue types. (A) Normal (NM). (B) Emphysema (EM). (C) Ground Glass (GG). (D) Fibrosis (FB). (E) Micronodules (MN). (F) Consolidation (CD).

There are a vast amount of relevant literature on developing CAD systems of pulmonary diseases, but most of them focus on identifying and quantifying a single pattern such as consolidation or nodules [2]. For computer-aided ILD classification, all previous studies have employed a patch based image representation with the classification results of moderate success [4, 8, 13, 14]. There are two major drawbacks for the image patch based methods: 1), The image patch sizes or scales in studies [13, 14] are relatively small (31×31 pixels) where some visual details and spatial context may not be fully captured. The holistic CT slice holds a lot of details that may be overlooked in the patch based representation. 2), More importantly, the state-of-the-art methods assume the manual annotation as given. Image patches are consequently sampled within these ROIs. Image patch based approaches, which depend on the manual ROI inputs, are easier to solve, but unfortunately less clinically desirable. This human demanding process will become infeasible for the large scale medical image processing and analysis.

In this paper, we propose a new representation/approach to address this limitation. Our method classifies and labels ILD tags for holistic CT slices and can possibly be used to prescreen a large amount of radiology data. Additionally, the prescreened data can be used as feedbacks to enlarge the training dataset in a loop. This would be the essential component for a successful and practical medical image analysis tool at

a truly large scale. Different from [8, 13, 14], our CNN based method is formulated as a holistic image recognition task [12] that is also considered as a weakly supervised learning problem. Obtaining image tags alone is cost effective and can be obtained very efficiently. On the other hand, our new setup of using holistic images makes it significantly more challenging than the previous settings [8, 13, 14], since the manual ROIs are no longer required. Image patches as classification instances, which are extracted from the annotated ROIs, are well spatially-aligned or invariant to their absolute intra-slice CT coordinates. On the contrary, In our setup, only slice-level image labels or tags are needed and no precise contouring of ILD regions are necessary. This weakly supervised learning scheme can scale well with large scale image database. The experimental evaluation on the publicly available dataset demonstrates the state-of-the-art results under the same image patch based approaches and shows promising results under this new challenging protocol.

2 Methods

CNN has been successfully exploited in various image classification problems and achieved the state-of-the-art performances in image classification, detection and segmentation challenges such as MNIST, ImageNet, etc. [7, 10]. The typical image classification approach consists of two steps of feature extraction and classification. However, the most attractive characteristics of the CNN method is that it learns the end-to-end feature extraction and classification simultaneously. CNN also shows promise in medical image analysis applications, such as mitosis detection [3], lymph node detection [11] and knee cartilage segmentation [9]. In previous ILD classification work, hand-crafted local image descriptors (such as LBP, HOG) are used in [4, 13, 14] to capture the image patch appearance.

Our proposed framework of is illustrated in Fig. 2. Three attenuation scales with respect to lung abnormality patterns are captured by rescaling the original CT image in Hounsfield Units to 2-D inputs in training and testing. For this purpose, three different ranges are utilized: one focusing on patterns with lower attenuation, one on patterns with higher attenuation and one for normal lung attenuation. Using three attenuation ranges offers better visibility or visual separation among all six ILD disease categories. Another reason of using the three ranges is to accommodate the CNN architecture that we adapt from ImageNet [7] that uses RGB values of natural images. Finally, for each input 2-D slice, ten samples ("data augmentation") are cropped randomly from the original images and resized to 224×224 pixels via linear interpolation. This step generates more training data to reduce the overfitting. These inputs, together with their labels, are fed to CNN for training and classification. Each technical component is discussed in details as follows.

2.1 CNN Architecture

The architecture of our CNN is similar to the convolutional neural network proposed by Krizhevsky, et al. [7]. CNNs with shallow layers do not have enough discriminative power while too deep CNNs are computationally expensive to train and easy to be
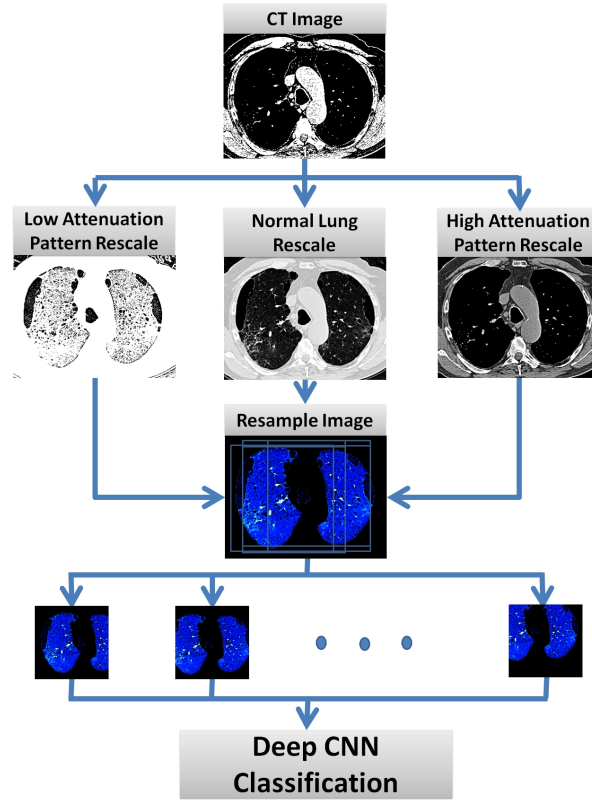
Fig. 2. Flowchart of the training framework.

overfitted. Our network contains multiple layers: first five layers are convolutional layers followed by three fully-connected (FC) layers and the final softmax classification layer, which is changed from 1000 classes to 6 classes in our application.

It is known from the computer vision community that supervised pre-training on a large auxiliary dataset, followed by the domain-specific fine-tuning on a small dataset, is an effective paradigm to boost the performance of CNN models (when the training data are limited [6]). In our experiments, the training convergence speed is much faster when using pre-trained model than using randomly initialized model. The use of three CT attenuation ranges also accommodate the CNN architecture of three input channels. The output of the last FC layer is formed into a six-way softmax to produces a distribution over the six class labels (with six neurons). We start the training via stochastic gradient descent (SGD) at a learning rate of 1/10th of the initial pre-training rate [7] expect for the output softmax layer. The adjusted learning rate allows appropriate fine-turning progresses without ruining the initialization. The output layer still needs a large learning rate for convergence to the new ILD classification categories.

2.2   CT Attenuation Rescale

To better capture the abnormal ILD patterns in CT images, we select three ranges of attenuation and rescaled them to $[0, 255]$ for CNN input. As shown in Fig. 3(A), this process is designed to select the attenuation value between HU_low and HU_high so that the value between can be highlighted to represent different visual patterns. A linear transformation is applied to rescale the intensities. Specifically, low attenuation range (Fig. 3(B)) is used to capture patterns with lower intensities, such as emphysema; normal range (Fig. 3(C)) to represent normal appearance of lung regions; and high attenuation range (Fig. 3(D)) for modeling patterns with higher intensities, such as consolidation and nodules. Specific HUs we chose in our experiments are: for low attenuation, HU_low= -1400 and HU_high= -950; for normal, HU_low= -1400 and HU_high= 200; for high attenuation, HU_low= -160 and HU_high= 240.
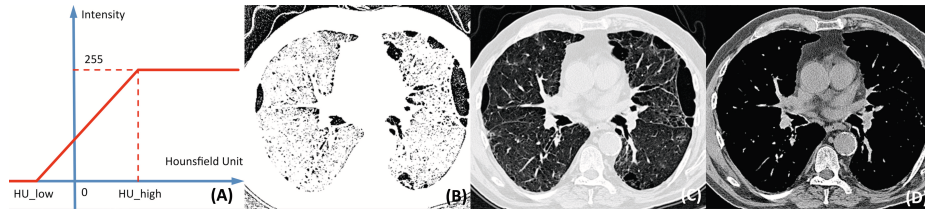

Fig. 3. (A) CT attenuation range rescale. (B) Low attenuation range. (C) Normal lung range. (D) High attenuation range.

2.3   Data Augmentation

The most common and effective way to reduce overfitting on image recognition training using CNN is to artificially enlarge or augment the original dataset by label-preserving geometric transformations. We generate new images by randomly jittering and cropping 10 subimages per original CT slice. Although the generated images are interdependent, the scheme would improve the training/testing performance by $\sim 5\%$ in classification accuracy. At test time, 10 jittered images are also generated and fed into the trained CNN model for any CT slice. Final per slice prediction is obtained by aggregating (e.g. majority voting, maximal pooling) over the CNN six-class softmax probabilities on 10 jittered images.

3   Experiments and Discussions

A publicly available ILD database has been released recently [5] to improve the detection and classification of a wide range of lung abnormal imaging patterns. This database contains 120 HRCT scans with 512×512 pixels per axial slice, where 17 types of lung tissues are annotated on marked regions (i.e., ROIs). Most existing classification methods [8, 13, 14] evaluated on the ILD dataset first extract many image patches from ROIs

and then only classify patches into five lung tissue classes: normal (NM), emphysema (EM), ground glass (GG), fibrosis (FB) and micronodules (MN). Here, consolidation (CD), as a highly prevalent type of ILD, is also included within our classification scheme. All of the six diseases are prevalent characteristics of ILD and identifying them is critical to determine their ILD types or healthy.
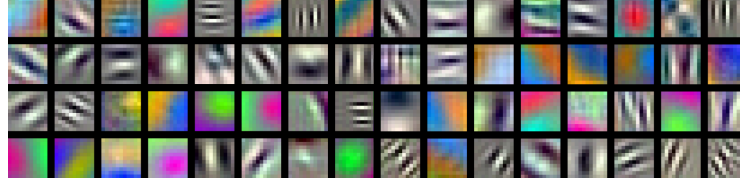

Fig. 4. Trained convolutional filters in the first layer.

The database contains 2084 ROIs labeled with specific type of ILD disease, out of 120 patients. All patients are randomly split into two subsets at the patient level for training (100 patients) and testing (20 patients). Training/testing data are separated at patient level, i.e., different slices from the same patient will not appear in both training and testing. All images containing the six types of diseases are selected, resulting 1689 images in total for training and testing. Note that previous work [8,13,14] report performance on patch classification only, rather than performance assessment for the whole image slices or at patient level, which are actually more clinically relevant.

Table 1. F-score of ILD classifications.

	EM	FB	GG	NM	MN	CD
[14]	0.753	0.841	0.782	0.840	0.857	-
[13]	0.768	**0.872**	0.795	0.877	0.888	-
[8]	0.5449	0.7624	0.7150	0.8395	**0.9096**	-
Ours	**1.0000**	0.8000	0.7500	0.4000	0.5600	**0.5000**
Ours Patch Setting	0.8940	0.8509	**0.8159**	**0.8844**	0.8950	-

For fair comparisons with previous work, we conduct experiments under two different settings. One is patch based classification, that is exact the same as in previous state-of-the-art work [13, 14]. An overall accuracy of 87.9% is achieved, comparing with 86.1% [14] accuracy of previous patch methods. The best F-scores are achieved in most classes as shown in Table 1. 31×31 patches are extracted from the ROI regions, and then resized to the size of 224×224 to accomodate the CNN architecture. Another experiment shows the holistic image classification results. The overall accuracy is 68.6%. Note that our per slice testing accuracy results are not strictly comparable to [8, 13, 14], reporting classification results only at the image patch level (a significantly less challenging protocol).

Table 2. Confusion matrix of ILD classification

| Ground | Prediction | | | | | |
Truth	EM	FB	GG	NM	MN	CD
EM	**1**	0	0	0	0	0
FB	0	**0.7111**	0.0889	0.0667	0	0.1333
GG	0	0	**0.9375**	0.0625	0	0
NM	0	0	0	**0.5**	0.5	0
MN	0	0	0	0.4615	**0.5385**	0
CD	0	0.2	0.3333	0	0	**0.4667**

Table 2 shows the confusion matrix of the classification results on holistic images. Majority voting based aggregation from jittered subimages is used. Emphysema is perfectly classified from other diseases. One of the three CT attenuation ranges is specifically designed to emphasize on the patterns with lower attenuation, which boosts the classification performance on emphysema significantly. Healthy images and micronodule patterns are difficult to be separated from, based on the confusion matrix result. Micronodule patterns are indeed visually challenging to be recognized from one single static CT slice [2]. 3D cross-slice image features may be needed. Majority voting performs slightly better ($\sim 2\%$) than choosing the highest value from 10 subimage CNN scores per ILD class, and assigning the CT slice into the class corresponding to the maximum of aggregated highest scores. Table 3 shows the confusion matrix of patch based classification.

Table 3. Confusion matrix of ILD patch classification

| Ground | Prediction | | | | |
Truth	EM	FB	GG	NM	MN
EM	**0.9142**	0.0078	0.0237	0.0047	0.0495
FB	0.0546	**0.8270**	0.0075	0.0464	0.0646
GG	0.0558	0.0025	**0.8151**	0.0930	0.0337
NM	0.0141	0.0108	0.0494	**0.8910**	0.0348
MN	0.0600	0.0070	0.0262	0.0268	**0.8799**

Our model is implemented in Matlab using MatConvNet package [15] for the CNN implementation, running on a PC with 3.10 GHz dual processors CPU and 32 GB memory. Training the CNN model consumes about $20 - 24$ hours, while classifying a new testing image takes only a few seconds.

4   Conclusion and Future Work

In this paper, we present a new representation and approach for interstitial lung disease classification. Our method with holistic images (i.e., CT slice) as input, is significantly different from previous image patch based algorithms. It addresses a more practical and

realistic clinical problem. Our preliminary experimental results have demonstrated the promising feasibility and advantages of the proposed approach.

There are several directions to be explored as future work. The image features learned from the deep convolutional network can be integrated into more sophisticated classification algorithms. There are some cases ($\sim 5\%$) with multiple disease tags on the same slice of CT image. Detection with multiple labels at a slice level would be interesting. Understanding the clinical meaning and value of the features learned from the network would also be a direction that we plan to pursue.

References

1. U. Bağcı, M. Bray, J. Caban, J. Yao, and D. J. Mollura. Computer-assisted detection of infectious lung diseases: a review. *CMIG*, 36(1):72–84, 2012.
2. U. Bagci, J. Yao, A. Wu, J. Caban, T. Palmore, A. Suffredini, O. Aras, and D. Mollura. Automatic detection and quantification of tree-in-bud (TIB) opacities from CT scans. *TBME*, 59(6):1620–1632, June 2012.
3. D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. *MICCAI*, pages 411–418, 2013.
4. A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Muller. Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4):665–675, 2012.
5. A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller. Building a reference multimedia database for interstitial lung diseases. *CMIG*, 36(3):227–238, 2012.
6. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, pages 580–587. IEEE, 2014.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.
8. Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. Medical image classification with convolutional neural network. In *ICCAR*. 2014.
9. A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *MICCAI*, pages 246–253. Springer, 2013.
10. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. In *CVPRW*, pages 512–519. IEEE, 2014.
11. H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers. A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations. In *MICCAI*, pages 520–527. Springer, 2014.
12. 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. *IJCV*, pages 1–42.
13. Y. Song, W. Cai, H. Huang, Y. Zhou, D. Feng, Y. Wang, M. Fulham, and M. Chen. Large margin local estimate with applications to medical image classification. *TMI*, 2015.
14. Y. Song, W. Cai, Y. Zhou, and D. D. Feng. Feature-based image patch approximation for lung tissue classification. *TMI*, 32(4):797–808, 2013.
15. A. Vedaldi and K. Lenc. Matconvnet-convolutional neural networks for matlab. *arXiv preprint arXiv:1412.4564*, 2014.
16. W. R. Webb, N. L. Muller, and D. P. Naidich. *High-resolution CT of the lung*. Lippincott Williams & Wilkins, 2014.