Multi-Scale Rotation-Invariant Convolutional Neural Networks for Lung Texture Classification

Qiangchang Wang, Yuanjie Zheng, Gongping Yang, Weidong Jin, Xinjian Chen*, Yilong Yin*

***Abstract*—We propose a new Multi-scale Rotation-invariant Convolutional Neural Network (MRCNN) model for classifying various lung tissue types on high-resolution computed tomography (HRCT). MRCNN employs Gabor-local binary pattern (Gabor-LBP) which introduces a good property in image analysis - invariance to image scales and rotations. In addition, we offer an approach to deal with the problems caused by imbalanced number of samples between different classes in most of the existing works, accomplished by changing the overlapping size between the adjacent patches. Experimental results on a public Interstitial Lung Disease (ILD) database show a superior performance of the proposed method to state-of-the-art.**

***Index Terms*—Interstitial Lung Disease (ILD) classification, lung classification, convolutional neural network (CNN), Gabor filter, local binary pattern (LBP).**

I. Introduction

INTERSTITIAL lung disease (ILD) is a clinical umbrella term referring to a diverse range of lung disorders that involve inflammation and fibrosis of interstitium [1]. The most common symptoms of ILD include labored breathing, dry cough and shortness of breath, which produce seriously adverse impact on our lives. High-resolution computed tomography (HRCT) scan of chest is being widely used for visualizing the subtle texture changes of different ILD lung tissues. HRCT image analysis algorithms are potentially capable of providing invaluable assists on clinical radiologists in detecting these differences [2]. Besides, in emergency radiology, for example, radiologists have to process a wide range of image modalities applied to different organs. They need to provide initial diagnosis shortly without being specialized in a single domain. This further aggravates the confusion between diverse lung tissue types [3]. Due to the inherent complexity of interpretation of HRCT images, a computerized approach for differentiating the lung tissues to provide helpful suggestions would be appealing for lung specialists.

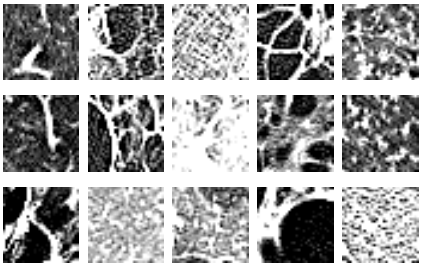


Fig. 1. Three example images of each ILD type. From left to right: healthy, emphysema, ground glass, fibrosis, micronodules.

However, distinguishing between lung tissue patterns (e.g. healthy, emphysema, fibrosis, ground glass and micronodules) on HRCT images is of a challenging problem. Lung tissue exhibits similar appearances between different tissue categories but also great variations between different subjects for a same tissue class (e.g. the emphysema), as illustrated in Fig. 1. This can be reflected in sense of not only the image gray values but also the geometric structure composed by the tissues. How to extract effective features with which intra-class patterns are more similar and inter-class patterns are more distinctive has been found to be a critical but difficult task.

A. Related Works

There exist extensive lung tissue representation approaches and they can be basically classified as hand-crafted features and learned ones. For the learning based approaches, almost all previous works are plagued with inferior performances caused by imbalanced data distribution.

1) Hand-crafted features: Hand-crafted features are used to capture the subtle variations of different lung tissue types, furthermore, in order to describe the lung tissues at different scales and handle the rotational variations in lung tissue patterns, a group of multi-scale or rotation-invariant features are proposed.

The popular texture features include, gray-level histogram [4]–[10], filters [7], [11], run length (RLE) [8], [12], gray-level cooccurrence matrices (GLCM) [8], [12], [13], local binary pattern (LBP) [14], [15].

While these feature descriptors have generated promising performance, some works focus on learning multi-scale or rotation-invariant features, multi-scale features like Riesz

Manuscript received August 29, 2016; revised October 17, 2016; accepted March 13, 2017.

The work is supported by NSFC Joint Fund with Guangdong under Key Project No. U1201258, National Natural Science Foundation of China under Grant No. 61573219, 61472226, and the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.

Q. Wang is with the School of Computer Science and Technology, Shandong University, Jinan 250101, China (e-mail: shdyn2000@mail.sdu.edu.cn).

Y. Zheng is with the School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China (e-mail: zhengyuanjie@gmail.com).

G. Yang is with the School of Computer Science and Technology, Shandong University, Jinan 250101, China (e-mail: gpyang@sdu.edu.cn).

W. Jin is with the School of Economics, Economic School of Shandong University of Finance and Economics, Jinan 250014, China (e-mail: jinweidong@sdufe.edu.cn).

X. Chen is with the School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: xjchen@suda.edu.cn).

Y. Yin is with the School of Computer Science and Technology, Shandong University, Jinan 250101, China (e-mail: ylyin@sdu.edu.cn).

*X. Chen and Y. Yin are the corresponding authors.

transform [16], [17], Gabor filters [4], learning-based filters [9] are used to describe the different image details at various scales, rotation-invariant features such as LBP [4], [18], wavelet transform [6], [10], Gabor filters [4], histogram of oriented gradients (HOG) [4] can eliminate the rotational variances since no dominant orientations associated with all lung patterns that contain complex structures.

After the feature descriptors are derived, the next stage is to employ machine learning to assign each patch a label. Commonly used classifiers include sparse representation [4], [19]–[22], support vector machine (SVM) [6], [10], [11], [16], [17], random forest (RF) [7], Bayesian classifiers [8], [23], k-nearest neighbor (kNN) [18], linear discriminant classifier [14], standard Gaussian Mixture Model (GMM) [24], Boosted decision tree (BDT) [25]. For clinical experts, they can utilize prior knowledge to extract representative information, but as the domain background varies, another group of features are needed to accommodate the new domain variations, this would be inaccessible when there are lots of clinical domains to be analyzed.

2) Feature learning: Feature learning is a set of techniques that can exploit more descriptive features from raw data and avoid the domain-specific problem.

Among the various ways of feature learning, it has been recently shown that approaches based on deep learning have achieved good performance in many computer vision and machine learning problems [26]–[29]. For lung tissue classification, several works have used deep learning to automatically extract useful information from data: multi-scale features produced by Gaussian restricted Boltzmann machine (GRBM) are used for SVM classification [30]; the convolutional classification restricted Boltzmann machine (RBM), a combination of convolutional RBM and classification RBM, is introduced for discriminative feature learning [31], [32].

If provided with sufficient annotated samples, CNNs are capable of learning more informative features, this is because that CNNs concatenate the feature extraction stage and classification stage as a integral pipeline and create synergy through close interactions, so both stages are optimized jointly toward the higher classification performance. CNNs have popularized the topic of deep learning in computer vision research. Through the use of CNN, great achievements have been made in the classification of natural images [33], [34]. In addition, CNNs have been used in many medical applications, such as mitosis detection in breast histology images [35], [36], magnetic resonance imaging (MRI) based knee cartilage segmentation [37], infant brain image segmentation [38], AD and MCI diagnosis [39], characterization of lung nodule malignancy [40], lung tissue pattern classification [41]–[45]. Despite the fact that CNNs are capable of learning high-level features, they cannot learn multi-scale and rotation-invariant representation of lung tissue patterns, which are significantly beneficial for classifying different lung categories.

3) Multi-scale rotation-invariant CNN: to combine the multi-scale and rotation-invariant superiorities in hand-crafted features with the advantage of supervised high-level features in CNNs, we propose to use multi-scale and rotation-invariant convolutional neural network (MRCNN) to learn image representations, this is mainly based on the following considerations: during CT scanning, slight pose variations of the patients may cause the lung CT images to be rotationally biased. Besides because the lung tissue patterns contain complex structures, no dominant direction could be associated with all lung tissue patterns, thus the rotation-invariant features are expected; however, CNNs are not actually invariant to large rotations of the input data [46], [47]. On the other hand, since breathing has a large effect on lung volume size, different lung details are displayed at various scales on computed tomography (CT) imaging, so multi-scale analysis is needed to get comprehensive description. Considering the facts stated above, lung tissue patterns may be better represented by multi-scale rotation-invariant features rather than gray-level images. Gabor LBP images are one of such features. First, LBP feature is able to describe the spatial structure of local lung image texture. Second, multi-scale Gabor filters decompose gray-level images into several sub-images that preserve valuable information at different scales. Third, in order to reduce rotation variations, rotation-invariant Gabor filter and LBP characterize lung tissue patterns. These Gabor LBP images are concatenated to multiple channels of the inputs in CNN to learn multi-scale high-level features. Due to the fact that the Gabor LBP images are rotation-invariant, thus the learned CNN features are also rotation-invariant.

4) Imbalanced data distribution: Besides, we notice a common problem that emphysema and ground glass tissues based on the same database [48] have fewer samples, which leads to their underrepresentation in most models and relatively suboptimal performance. For example kNN would have the lower classification performance of the minority class. This is because the majority class would probably dominate the neighborhood of a test sample even considering distance measurement; SVM and CNN's objective functions are to maximize the classification accuracy which would naturally focus on the majority class. SVM [4], [6], [16], [30], kNN [4], and CNN [41] on the same database prove this conclusion.

There exist data-level methods which can handle imbalanced data distribution by resampling the data space, like oversampling or synthetic minority over-sampling technique (SMOTE) [49]. Unfortunately, they would cause the model overfitting [50] or over generalization [51].

B. Our Contributions

In this paper, we present a multi-scale and rotation-invariant convolutional neural network (MRCNN) which is capable of classifying the lung texture pattern more accurately. We also propose a technique to deal with the imbalanced data problem, which is accomplished by changing the overlapping size of adjacent patches. Our major contributions are detailed as below.

First, each image patch is classified by a MRCNN model in order to achieve a better classification accuracy based on several considerations. First, motivated by the success of Gabor LBP representations in lung tissue classification, we investigate them into an end-to-end system that employs a hierachical architecture for transforming the low-level representations of input images into the high-level structural information.

Second, it can detect complementary features on the basis of multi-scale images which could provide a rich and effective representation of lung tissue patterns. Finally, lung tissue patterns contain complex structures that no dominant direction could be associated with all lung patches. Meanwhile, CNN is vulnerable to rotation variations, which presumably reduces its predictive power. In our approach, we use rotation-invariant images as the inputs of CNN, and therefore, the learned CNN features are also rotation-invariant.

Second, we notice a common problem among most of the previous works: different ILD tissue categories have imbalanced number of samples. Any classifier being sensitive to this data imbalance would inevitably perform poorly. We solve this problem in the image patch preparation step by changing the overlapping size of adjacent patches flexibly, increasing the overlapping size to get more samples of minority classes and decreasing the overlapping size to get fewer samples of majority classes.

The rest of paper is organized as follows. Section II describes the proposed tissue pattern classification MRCNN model. Section III discusses the experimental setup and parameter settings. Section IV reports the experimental results and discussions. Conclusions and future works are presented in Section V.

II. Methods

The overall framework for learning discriminative features by CNN based on Gabor LBP images is illustrated in Fig. 2. As Gabor LBP features have shown promising results in face recognition [52], [53], lung classification [4], we believe multi-scale rotation-invariant Gabor LBP images are a better choice to act as the inputs of CNN rather than gray-level images. CNN is an end-to-end model that optimizes the values of a huge number of parameters in a supervised way. Gabor filter, LBP and CNN are detailed in the following.

A. Gabor filter

Gabor filter is biologically inspired by the receptive fields of simple cells in the primary visual cortex [54]. With good characteristics in providing the optimized resolution in space and frequency domains, Gabor filter has been widely used in lung tissue patterns for image enhancement [55], feature extraction [56], multi-scale representations [57], [58].

It is defined as follows

$$g_{s,r}(x,y) = \frac{k_{s,r}^2}{\sigma^2} exp\left(-\frac{k_{s,r}^2 z^2}{2\sigma^2}\right)\left[exp(ik_{s,r}z) - exp(-\frac{\sigma^2}{2})\right] \quad (1)$$

where r, s define orientation and scale of the Gabor filters, z=(x, y) and the wave vector $k_{s,r}$ is defined as

$$k_{s,r} = k_s exp(i\phi_r) \quad (2)$$

with $k_s = k_{max}/f^s$, k_{max}=1, f=2, ϕ_u=$(r\pi)/4$, s=(0, 1, …, S-1), r=(0, 1, …, R-1), where S is the number of scales and R is the number of orientations.

As implied in Eq. 2, we can get a group of images at various scales by convolving with multiple sets of Gabor filters at different scales. The image under investigation is decomposed into sub-images at several scales, i.e., different image details are presented at different scales.

However, being rotation-variant, Gabor filter may be not directly appropriate to describe the complex lung tissue. To overcome the rotation-variant shortcomings encountered in traditional Gabor representations, all orientations of a certain scale s are summed together [59]

$$g_s^R(x,y) = \sum_{r=0}^{R-1} g_{s,r}(x,y) \quad (3)$$

Hence, the extracted Gabor images could describe lung tissue patterns without any assumption on prevailing orientations.

B. Local Binary Pattern

LBP [60] measures the differences between a pixel and its surrounding pixels and captures the spatial structure of a local image texture, this is especially useful to describe the various types of lung tissue patterns.

Formally, given a pixel, its LBP code is obtained by thresholding neighboring pixels with its intensity as follows

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^P, s(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x <= 0 \end{cases} \quad (4)$$

where g_c represents the gray value of the center pixel and g_p (p=0, …, P-1) denotes the gray value of the surrounding pixels in the circle neighborhood with a radius R and the total number of neighbor pixels P.

The operator $LBP_{P,R}$ could produce 2^P different values formed by P surrounding pixels around the center pixel. Once the image is rotated, the neighbor pixels will move correspondingly along the perimeter of the circle around the center pixel, so except for patterns comprised of only 0s (or 1s), the rotation will result in a different $LBP_{P,R}$ value. In order to assign a unique value to the rotation-variant LBP, a new type of LBP is proposed

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}i)|i = 0, 1, \ldots, P-1\} \quad (5)$$

where ROR(x, i) performs a circular bit-wise right shift on the P-bit number x, i times. The operator $LBP_{P,R}^{ri}$ can be considered as a feature detector to preserve certain micro-features in the image corresponding to the rotation-invariant patterns [60].

C. Convolutional Neural Network

Recently, the availability of enough labeled data and the accessibility of the high parallel computation resources provided by graphics processing units (GPUs) have make it possible to train large CNNs. CNNs were first presented by Lecun et al. for handwritten digit recognition [61] and have achieved remarkable results in a wide array of computer vision tasks [33], [34], [62]–[64]. Generally speaking, CNNs can be seen as a variant of multilayer perceptron that can learn hierarchical features by building high-level features from low-level represenations with alternating convolutional layers and subsampling layers.

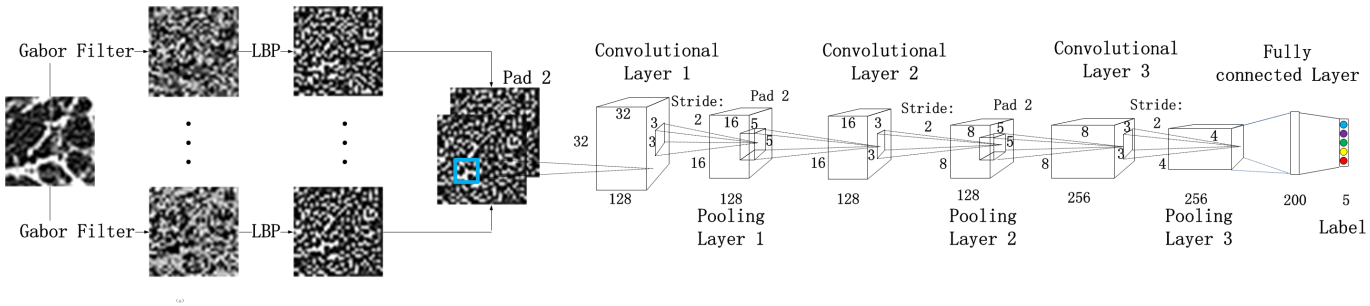


Fig. 2. Topology of MRCNN. (a) Lung tissue image is decomposed into rotation-invariant Gabor filtered images of different scales. (b) Rotation-invariant LBP is used to further characterize the Gabor filtered images. (c) All Gabor LBP images are concatenated together as the multiple channels of CNN to learn high-level features.

Convolutional layer is the key component of CNNs. It exploits the stationary nature of images by convolving with patches sampled from a image. Unlike the standard, fully connected hidden layer, it has two important characteristics. First, each convolutional unit receives input only from a local area of the input. This means that each unit represents some features of a local region of the input. Second, the units of the convolution layer can be organized into a number of feature maps, where all units in the same feature map share the same weights but receive input from different locations of the lower layer.

The output of a convolutional layer is defined as

$$d_j^{l+1} = f(\sum_i d_i^l * w_{ij}^{l+1} + b_j^{l+1}) \tag{6}$$

where * denotes the convolution operation, $d_j^{(l)}$ denotes j-th feature map of the l-th layer, let d_0 be the 2D input image patch, w_{ij}^l is the connecting weight from the i-th feature map in the layer l-1 to the j-th feature map in the l-th layer, $b_j^{(l)}$ is a bias added to the j-th feature map, and f(x) represents the non-linear activation function applied to the each element of x.

In order to reduce the number of features, subsampling layers compute the max or mean value of a particular feature over a region of the image. This will decrease computational complexity, introduce invariance properties and also reduce the chances to be overfitting. Subsampling operation will not reduce the number of feature maps of layer l.

For mean pooling, the subsampling factor s, the output of the element (x, y) of j-th feature map of layer l+1 is computed as

$$d_j^{(l+1)}(x,y) = \frac{\sum_{m=0}^{s-1}\sum_{n=0}^{s-1} d_j^{(l)}(s\times x+m, s\times y+n)}{s^2} \tag{7}$$

For max pooling, the output is computed as

$$d_j^{(l+1)}(x,y) = max\left(d_j^{(l)}(s\times x+m, s\times y+n)\right) \tag{8}$$

where m, n=0, ..., s-1.

Rectified linear units (ReLU) is a layer of neurons that use the non-saturating activation function $f(x) = max(0, x)$ Compared to other functions that are used to increase nonlinearity, such as sigmoid function $f(x) = 1/(1+exp(-x))$ and hyperbolic tangent function $f(x) = tanh(x)$, ReLU expedites convergence of the training procedure and leads to better solutions [33].

When training a large neural network on a small training set, it is prone to be overfitting. Large networks are also slow to train, making it difficult to reduce test errors by combining predictions of several different networks. Dropout is a simple way to address these problems by randomly dropping neurons from neural network on each training iteration [65]. These dropped neurons do not contribute to the forward pass or participate in the back-propagation. So every iteration, different architectures are presented, this significantly reduces overfitting. Besides, it is an effective regularizer to force each unit to learn useful information by itself, and prevent the coadaptation between units. At test time, all neurons are turned on but multiply their outputs by the probability of average turn on rate during training.

III. Experimental Setup

A. Data

The data comes from the publicly available database of ILD cases [48] which contains HRCT images with a slice thickness of 1mm. 2062 2-D regions of interest (ROIs) that were manually annotated in 113 sets of HRCT by two radiologists with 15 and 20 years of experience at the University Hospitals of Geneva (HUG). Following the previous works [4], [6], [19], [22], [48], we also focus on the classification of five lung tissue patterns-healthy (H), emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). The hand-drawn ROIs are divided to 2D patches of 32*32 pixels and each patch must contain at least 70% of its pixels falling inside of the annotated ROI.

We notice a common phenomenon: if the adjacent patches overlap with size 16*32, this will lead to imbalanced number of patches between different lung tissue patterns-emphysema and ground glass patterns have relative fewer patches and would cause low classification accuracy if the classifiers such as SVM, kNN are sensitive to the imbalanced data. Unlike other methods that attempt to design a cost-sensitive classifier or resample the data space, we propose to change the overlapping size between the adjacent patches, namely, for emphysema and ground glass patterns, to increase the overlapping size to get more samples, here we adopt 23*32; for micronodules

TABLE I
DATA DISTRIBUTION OF EACH TISSUE CATEGORY

Tissue Category	# Patients	# Patches
Healthy (H)	14	4195
Emphysema (E)	9	5385
Ground Glass (G)	31	5567
Fibrosis (F)	32	3819
Micronodules (M)	15	5139

patterns, to decrease the overlapping size to get fewer samples, here 9*32. As shown in Table I, five types of lung tissue patterns have relatively balanced number of patches.

B. Performance evaluation

All samples are randomly partitioned to 5 equally sized groups on patch-level. Of the 5 groups, a single group is retained as the testing data, and the remaining 4 groups are used as training data. The process is repeated 5 times so that every group used exactly once as the testing data. The five results are averaged to produce the final result.

The quantitative performance of our MRCNN model and different models are analyzed using the metrics in Eq. 9-11.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F - measure = \frac{2 * Recall * Precision}{Precision + Recall} \tag{11}$$

C. Parameter setting

The detailed architecture is shown in Fig. 2. We train the network from scratch. In our experiments, 8 orientations of Gabor filters at a single scale are summed together to form a rotation-invariant Gabor filter, we derive Gabor representations by convolving gray-level images with the rotation-invariant Gabor filters at 6 scales. For every pixel p(x, y), we have 6 scales of rotation-invariant texture features, this means that we can obtain 6 Gabor filtered images from a single gray-level image. Next, rotation-invariant LBP to encode the Gabor filtered images at every scale.

After that, 6 Gabor LBP images of every sample are concatenated to form multiple channels of the inputs in CNN, that is, the CNN architecture contains 6 input maps with size 32*32 corresponding to Gabor LBP images at 6 scales. Then we apply three convolutional and pooling layers, the remaining two are fully-connected. ReLU function is applied to the output of every convolutional and fully-connected layer. The three convolutional layers contain 128, 128 and 256 feature maps respectively. Each of the feature maps is connected to all of the feature maps in the previous layer through filters of size 5*5 with a stride of one pixel. All convolutional layers are padded with size 2*2 to keep the same shape as previous layer, so the input units near the border can also be analyzed. Pooling layers are followed after each convolutional layer, they are max pooling, mean pooling and mean pooling with kernels of size 3*3, stride 2. The two fully-connected layers have 200-5 neurons. We use dropout rate of 0.5 in the first fully-connected layer. The output of the last fully-connected layer is fed into a 5-way softmax activation function which produces the distribution of five lung tissue labels.

The weights in the network are initialized randomly from a zero-mean Gaussian distribution with standard deviation 0.01. We update weights by stochastic gradient descent (SGD) algorithm with a batch size of 100 examples, a momentum 0.9 and weight decay 0.004. The leaning rate is set to $3*10^{-4}$ for the first $3*10^4$ iterations, $3*10^{-5}$ for the following 10^4 iterations and $1*10^{-5}$ for the last $5*10^3$ iterations. All these learning parameters are chosen based on experimental tests.

All the experiments were carried on a workstation (Intel Core(TM) 3.4 GHz processor with 64 GB of RAM) and a NVIDIA GTX 980 GPU. We implement our CNN model using Caffe package [66].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Patch-level classification

Table III shows the confusion matrix of the classification results. All the tissue categories achieve higher than 85% classification rates. 9.5% of emphysema images are misclassified as fibrosis while 8.6% of fibrosis images are predicted as emphysema. This is because that emphysema and fibrosis images have very similar patterns as shown in Fig. 1. For other three lung tissue categories-healthy, ground glass and micronodules, the classification accuracy can be up to 90% or higher.

Table IV summarizes the classification recall, precision and F-measure of each lung category. Overall, the results show relatively high performance among the different tissue classes. Note that the relatively lower fibrosis precision is mainly because the similarities between emphysema and fibrosis patterns and the smaller number of fibrosis patches.

We also implement the classification task on patient-level split, as shown in the Table II. The results came out with the classification accuracy of 79%, which is comparable to other feature learning methods. However, a little bit worse than the results based on hand-crafted features [4]. This may because the deep model could work not well if training samples are not enough, and our MRCNN model is limited by the population size. If the dataset is large enough, the deep model should have good generalization ability either clear separation between the training and testing or not.

We visualize the learned features as shown in Fig. 3. Although there are no recognizable structures, they are advantageous for classifying lung texture.

B. CNN configurations

Table V compares different CNN architectures in terms of classification performance. The proposed network configuration is presented in bold. In order to compare the running

TABLE II
COMPARISON OF CLASSIFICATION ACCURACY ON BOTH PATIENT-LEVEL AND PATCH-LEVEL

Patient-level	79%
Patch-level	90.1%

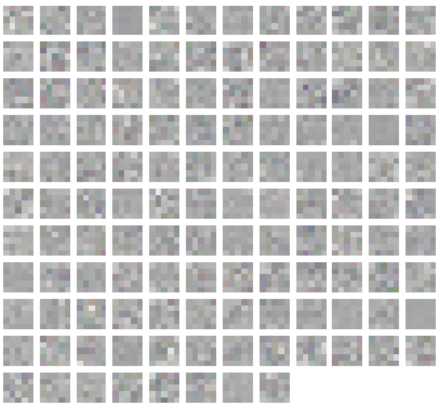


Fig. 3. Visualization of the learned features.

TABLE III
CONFUSION MATRIX OF PATCH-LEVEL CLASSIFICATION

Ground Truth	Prediction (%)				
	H	E	G	F	M
H	89.4	2.9	3	1.1	3.6
E	2.6	86.1	0.2	9.6	1.5
G	2.3	0.2	93.4	3.3	0.8
F	1.2	10.5	3.6	83.5	1.2
M	1.9	0.6	1.1	0.6	95.8

time of one iteration, we use SGD with a mini-batch of 100 examples and every experiment is repeated 100 times.

First, we note that using zero padding improves classification accuracy (A in comparison with B), so we employ zero padding in other models. Zero padding the input enables us to set the kernel width and the output size independently; therefore networks can be defined deeper to enhance their expressive power. Besides, the input pixels near the border can also be well analyzed and discriminative information may be learned.

Second, we observe that classification error decreases with the increased CNN depth: from 2 layers in C to 3 layers in M; however, the error rate increases in D with 4 layers. This proves that the architecture saturates, but deeper models might be beneficial for larger databases. Decreasing the size of the

TABLE IV
THE RECALL, PRECISION AND F-MEASURE MEASURES OF THE PATCH-LEVEL CLASSIFICATION

	H	E	G	F	M
Recall (%)	89.4	86.1	93.4	83.5	95.8
Precision (%)	90.2	89	94.1	80.3	93.9
F-measure (%)	89.8	87.5	93.8	81.9	94.8

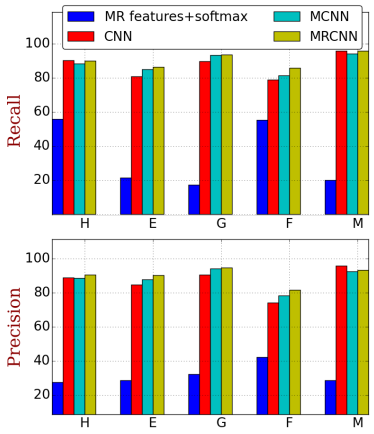


Fig. 4. The classification recall (top) and precision (down), comparing our MRCNN model with MR features+softmax, CNN and MCNN.

kernels to 3*3 in E reduces the performance by approximately 7% while increasing the size of kernels to 7*7 in F and 9*9 in G result in a rise of less than 1%, accompanied by a considerable increase in training time. This indicates that 5*5 is the optimal kernel size to analyze the local structures in lung tissues. To identify the optimal number of feature maps, we experiment in H and I, which show that the choice in M is the most appropriate one according to accuracy and efficiency.

Third, we set the dropout fraction to 0 in J, which leads to a decrease of 1.6% in classification accuracy. This verifies that dropout is an effective way to address overfitting problem. Besides, we change the number of hidden units in fully-connected layer, 100 in K and 400 in J respectively, yielding results that are inferior.

C. Evaluation of MRCNN model

The framework of the proposed MRCNN approach has three characteristics. First, multi-scale CNN features are learned based on Gabor LBP images at different scales which would provide a rich representation of lung tissue patterns. Second, in order to eliminate the rotational variations in lung tissues, rotation-invariant Gabor LBP images serve as the inputs of CNN. Finally, CNN transforms the low-level lung tissue representations into high-level structural features. Fig. 4 analyzes the contribution of every characteristic by comparing the classification performance using different combinations. MR features+softmax: the histogram of multi-scale and rotation-invariant Gabor LBP images is used for classification. The total number of possible rotation-invariant LBP values is 36, so the concatenated histogram of Gabor LBP images at 6 scales is 216 dimensional. Like the classification stage in CNN, we use softmax regression as classifier with L2 regularization term and inverse of regularization term 1.0; meanwhile, L-BFGS algorithm is used in the optimization process and 1000 iterations taken for the solver to converge. CNN: CNN directly learns from raw images instead of Gabor LBP images. MCNN: we treat the multi-scale Gabor LBP images as the inputs of CNN. MRCNN: multi-scale and rotation-invariant Gabor LBP images are severed as the inputs of CNN.

TABLE V
PERFORMANCE OF DIFFERENT CNN ARCHITECTURES. EVERY CONVOLUTIONAL LAYER IS FOLLOWED BY POOLING LAYER. THE RELU ACTIVATION FUNCTION IS NOT SHOWN FOR BREVITY.

Model	Input	Conv1	Conv2	Conv3	Conv4	FC	Dropout	Accuracy (%)	Time (ms)
A	32*32, 6	5*5, 128	5*5, 256	-	-	200	0.5	86.5	27.02
B	32*32, 6 (2)	5*5, 128 (2)	5*5, 256 (2)	-	-	200	0.5	88.2	45.67
C	32*32, 6 (2)	5*5, 128 (2)	5*5, 256 (2)	-	-	200	0.5	88.2	45.67
D	32*32, 6 (2)	5*5, 128 (2)	5*5, 128 (2)	5*5, 256 (2)	5*5, 256 (2)	200	0.5	89.4	44.62
E	32*32, 6 (1)	3*3, 128 (1)	3*3, 128 (1)	3*3, 256 (1)	-	200	0.5	83.2	17.10
F	32*32, 6 (3)	7*7, 128 (3)	7*7, 128 (3)	7*7, 256 (3)	-	200	0.5	90.8	70.81
G	32*32, 6 (4)	9*9, 128 (4)	9*9, 128 (4)	9*9, 256 (4)	-	200	0.5	90.9	119.53
H	32*32, 6 (2)	5*5, 64 (2)	5*5, 64 (2)	5*5, 128 (2)	-	200	0.5	87.7	13.53
I	32*32, 6 (2)	5*5, 256 (2)	5*5, 256 (2)	5*5, 512 (2)	-	200	0.5	90.8	119.95
J	32*32, 6 (2)	5*5, 128 (2)	5*5, 128 (2)	5*5, 256 (2)	-	200	0	88.4	36.40
K	32*32, 6 (2)	5*5, 128 (2)	5*5, 128 (2)	5*5, 256 (2)	-	100	0.5	88.9	36.19
L	32*32, 6 (2)	5*5, 128 (2)	5*5, 128 (2)	5*5, 256 (2)	-	400	0.5	89.8	36.79
M	**32*32, 6 (2)**	**5*5, 128 (2)**	**5*5, 128 (2)**	**5*5, 256 (2)**	**-**	**200**	**0.5**	**90.1**	**36.29**

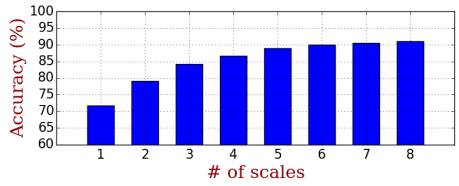


Fig. 5. The classification accuracy with various number of scales S for Gabor LBP images.

TABLE VI
THE CLASSIFICATION ACCURACY WITH DIFFERENT INPUTS OF CNNS

Input of CNN	Accuracy(%)
Gabor LBP	**90.1**
Gabor	89.7
LBP	86.7

The lower performance of MR features+softmax may because the low-level Gabor LBP histograms provide the detailed description about the individual components and are prone to be more susceptible to signal transformations, whereas CNN could capture high-level shape and edge interactions in lung tissues. When directly learned from gray-level images, CNN may be influenced by the rotational variances in all lung tissue patterns. Besides, images at a single scale may be not descriptive enough, thus these features are not discriminative enough. When MCNN learns from multi-scale Gabor LBP images, the classification performance is better than the results from images of a single scale, this suggests that a multi-scale approach is beneficial. When Gabor filters and LBP are combined with rotation-invariant property, these rotation-invariant Gabor LBP images are better representations of lung tissue patterns, as indicated in Fig. 4;

We measure the classification accuracy based on different number of scales S of Gabor LBP images. As shown in Fig. 5, the classification accuracy becomes higher and higher when S increases from 1 to 6, and is comparable when S=6, 7, 8. This indicates the benefit of having a multi-scale representations for high-level feature learning.

D. Comparison with different Multi-scale and Rotation-invariant CNNs

As shown in Table VI, we compare different inputs of CNN to learn multi-scale rotation-invariant features: proposed Multi-scale Rotation-invariant (MR) Gabor LBP images, MR Gabor images and MR LBP images. MR Gabor LBP images are able to enhance local intensity with the spatial information. This is the reason why the performance of MR Gabor LBP images is slightly better than that of MR Gabor images. Besides, MR LBP images are another alternative. However, their performance is inferior to the performance of proposed MR Gabor LBP images. This is because multi-scale Gabor filters can bring much more rich information.

To combine multiple scales of Gabor LBP images, we investigate three fusion schemes: input fusion, feature fusion and decision fusion. The input fusion directly concatenates 6 Gabor LBP images as multiple channels of the inputs in CNN. The feature fusion combines features from the first fully-connected layer in 6 networks to make a decision. The decision fusion scheme combines the probabilistic values of 6 separately trained networks. As shown in Fig. 6, input fusion works better than feature fusion or decision fusion, intuitively, input fusion is a natural way to guide the network to learn multi-scale features and reduces the chances of overfitting. Input fusion has 1,318,400 parameters, while the latter two methods have 7,810,560 parameters.

E. Evaluation of different balanced data distribution

We evaluate different methods to handle the imbalanced data distribution problem. Table VII indicates the training data distribution during five cross validation process. Original: like the previous works on the same database, if the neighbouring

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2017.2685586, IEEE Journal of Biomedical and Health Informatics

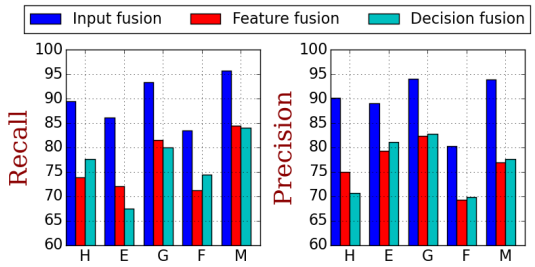


Fig. 6. The classification recall (left) and precision (right), comparing different fusion schemes in our model: input fusion, feature fusion and decision fusion.

patches overlap with size 16*32, we would get the imbalanced data distribution among the five tissue types, emphysema and ground glass patterns only have 1157, 1187 samples, whereas, the number of micronodules patterns is 9619. Resample: the number of micronodules overwhelms other four tissues types in original data, which would make the MRCNN model tend to classify the majority class, so we undersample 2800 micronodules patches during every training process. We resample extra 1672 emphysema and ground glass patches to alleviate the underrepresentation problem of minority classes. SMOTE: unlike the resampleing method, we increase the number of emphysema and ground glass pathces by SMOTE method [49]. SMOTE method resamples the small class through taking small class example and introduces synthetic examples along the line segments joining its small class nearest neighbors. Proposed: the proposed method to change the overlapping size of adjacent patches in image patch preparation step.

We adopt the MRCNN model to classify the five lung tissues and make a classification performance comparison between the imbalanced data and different balanced data in Table VII. As shown in Fig. 7, emphysema and ground glass patterns have much lower classification recall and precision values compared to the resampled data, this is because that emphysema and ground glass patterns are underrepresented in our model due to the relatively small number of samples. By resampling the minority classes and undersample majority class, we could get a balanced data distribution. MRCNN model could learn discriminative representation of every tissue type, this could account for the higher recall value of emphysema and ground glass. Although we only have 2800 micronodules instances, classification recall in micronodules is slightly affected and classification precision in micronodules is even higher, we can attribute the results to the relatively equal representation of every tissue type rather than the skewed prediction of being micronodules in the imbalanced data. SMOTE method is another oversampling method, but our result shows that a dramatic decreasing performance of all tissue types. This may be because that SMOTE method generates each minority example without considering the neighbouring examples, which increases the occurrences of overlapping between classes. On the other hand, by changing the overlapping size of adjacent patches, the emphysema and ground glass performance increases significantly, this verifies the benefits of our method to handle imbalanced data. It reduces the chances of being overfitting in resampling method or over generalizaiton in

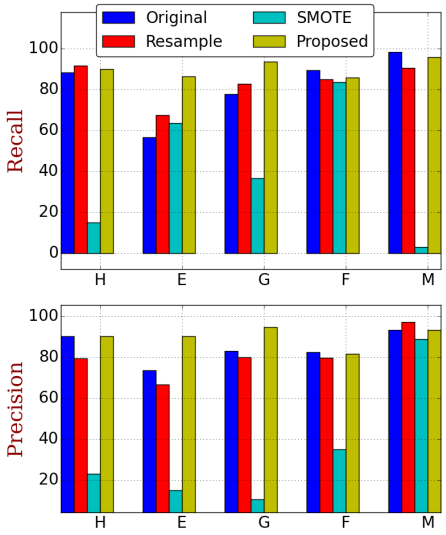


Fig. 7. Comparisons in recall (top) and precision (down) between imbalanced data distribution and various balanced data distributions.

TABLE VII
TRAINING DATA DISTRIBUTION OF DIFFERENT METHODS IN EVERY FIVE CROSS-VALIDATION PROCESS. H:HEALTHY, E:EMPHYSEMA, G:GROUND GLASS, F:FIBROSIS, M:MICRONODULES

METHODS	# Patches(H\E\G\F\M)
Original	2830\1157\1187\2690\9619
Resample	2830\2829\2859\2690\2800
SMOTE	2830\2829\2859\2690\2800
Proposed	3356\4308\4453\3055\4111

SMOTE method.

F. Comparisons with popular hand-crafted features

The performance of our MRCNN model is compared with the existing popular hand-crafted features: (i) LF [6], which combined wavelet-based textures and gray-level histogram, followed by SVM classifier; (ii) PASA [4], which used sparse representation with reference adaption; (iii) LSRE [22], which is based on sparse representation in a clustering-based subcluster generation model and is the state-of-the-art in patch-wise ILD tissue classification; (iv) MR-LBP+I [18], which combines the histogram of multi-resolution rotation-invariant LBP and histogram of intensity as features and classifies lung tissues by kNN. It is worth noting that PASA and LSRE used the same feature vector, which contains three types of information: multi-scale and rotation-invariant Gabor LBP features, intensity histogram and multi-coordinate HOG. The feature vector is specially designed for ILD representation and achieves good performance [4], [19], [20], [22]. We note that the results of $i - iii$ are obtained directly from [6], [4], [22], which used slightly different image patches from ours.

Fig. 8 shows that the comparisons in recall and precision between our MRCNN model on balanced data, MRCNN model on imbalanced data and the compared approaches. Under the similarly imbalanced data distribution like the compared

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2017.2685586, IEEE Journal of Biomedical and Health Informatics

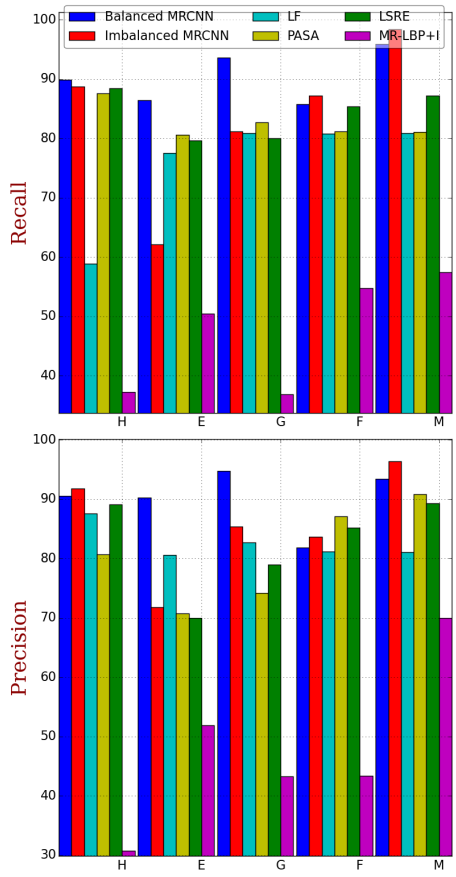


Fig. 8. Comparisons in recall (left) and precision (right) between our MRCNN model and other popular hand-crafted features.

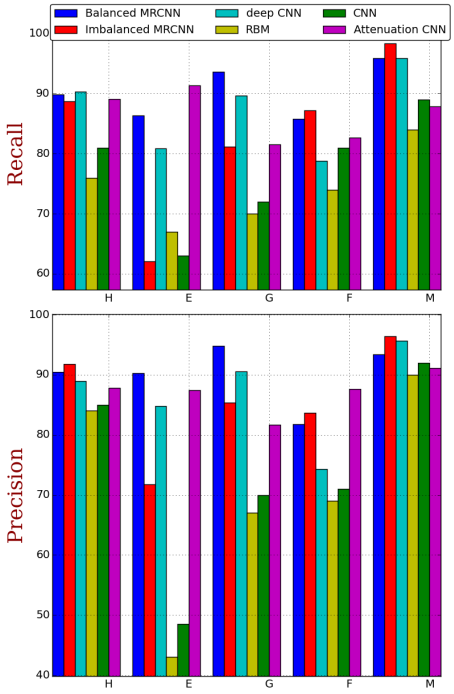


Fig. 9. Comparisons in recall (top) and precision (down) between our MRCNN model and the other feature learning methods.

methods, our MRCNN model exhibits comparative performance for healthy and fibrosis tissues and better performance for ground glass and micronodules categories. The higher recall of micronodules also contributes to the high precision of ground glass. However, the emphysema recall in our model is lower than the other compared approaches, this is mainly because misclassification of fibrosis category as emphysema tissue and the smaller number of emphysema patches; on the other hand, due to that sparse representation in [4] defines five dictionaries associated with each tissue category and labels every image patch according to the minimum feature reconstruction error, thus it is slightly affected by the minority data distribution and accounts for the relatively high emphysema recall. Although LSRE [22] uses a global sparse representation classifier, it assigns a higher fusion weight where the images in the subcluster belong to only a small number of classes, this would make minority class such as emphysema to be sufficiently represented. Recall that methods $i - iv$ define several complementary features and use SVM, kNN or sparse representation as classifier, the improvement of our MRCNN model demonstrates the advantage to simultaneously optimize the feature representation and classification parameters as an integral end-to-end pipeline toward better classification performance.

The cost-sensitive characteristic of MRCNN model and the skewed distribution of five lung tissue categories motivate us to prepare a balanced data distribution. By altering the overlapping size of adjacent patches in emphysema, ground glass and micronodules tissues, we could get the fairly balanced data distribution as indicated in Table I. Our MRCNN model based on balanced data leads to substantial improvements for all tissue categories particularly in emphysema, ground glass and micronodules patterns, which suggest the benefit of balanced data.

G. Comparisons with other feature learning methods

To evaluate our design of MRCNN model and the benefit of balanced data distribution, we compare the patch-level classification performance using the following approaches: (i) the proposed MRCNN approach on balanced data; (ii) the proposed MRCNN model on imbalanced data; (iii) deep CNN with gray-level images as inputs; (iv) the customized CNN with shallow convolutional layer [41]; (v) the multi-scale GRBM filters are convolved with image patches to produce the feature vectors for SVM classification [30]; (vi) fine-tuned CNN with three CT attenuation ranges as inputs [42]. Although most experiments in [42] do not require manual delineation of the ROIs, the patch based classification is also conducted in [42], that is exact the same as in previous state-of-the-art work for comparisons with previous work: 31 * 31 patches are extracted from the ROI regions.

Under the similarly imbalanced data distribution of ILD patches, our MRCNN model outperforms both the CNN and RBM considerably in the recall of the four tissue categories except for emphysema and all five lung tissues of precisions as well. The multi-scale GRBM features in [30] are data-adaptive, highly descriptive and could capture the intrinsic

image features without manual feature design; CNN in [41] uses supervised learning method and hence better classification result could be expected. However, both of them are shallow models that only contain one layer for high-level feature representation. We notice that the deep CNN uses gray-level images as inputs and has the same architecture of the proposed MRCNN model. The performance comparison between deep CNN and shallow CNN suggests the deep CNN could extract more informatively high-level features to characterize the complex lung tissue structures. On the other hand, the advantages over deep CNN imply the benefits of multi-scale and rotation-invariant properties of MRCNN. Furthermore, as highlighted in [42], by rescaling the original CT image in Hounsfield Units to three different attenuation scales, good visual distinctions among several ILD categories are captured. Since low attenuation range is specifically designed to describe the emphysema, the best emphysema performance can be achieved by attenuation CNN; however, the fibrosis and micronodules performance is marginally worse than our model. Finally, the MRCNN on balanced data could improve the classification performance especially for minority emphysema and ground glass tissues.

V. CONCLUSIONS AND FUTURE WORKS

A new multi-scale rotation-invariant (MRCNN) model is presented in this paper for classifying five lung tissue patterns. Motivated by the success of Gabor LBP representations in face recognition [52], [53] and lung classification [4], we propose to treat the multi-scale rotation-invariant Gabor LBP images as the inputs of CNN instead of the original image as in traditional methods. This improves the classification performance due to three reasons. First of all, unlike the hand-crafted methods that relying on human knowledge, we directly extract useful information from big data. The learned high-level features are able to capture highly complex structures and more consistently discernible. Moreover, different from the generative feature learning models, CNN could jointly optimize the feature transformations and classification. Second, the learned CNN features are from multi-scale analysis of the image and thereore provide er description of the lung tissue patterns. Third, to handle rotational biases discovered in the lung tissue patterns, we propose to use rotation-invariant Gabor LBP images as the inputs of CNN. Our method incorporate the rotation-invariant Gabor filter and LBP and therefore the learned CNN features are also robust to rotations.

Additionally, previous works using the ILD database have a common problem, i.e. the data used in training has imbalanced distribution between different lung tissue types. For example, emphysema and ground glass patterns have limited representations while micronodules patterns are abundant. However, most machine learning algorithms assume balanced distribution between different classes and could be defeated by this skewed distribution. This is caused by the fact that the classification power can be affected overwhelmingly by the major class while the minor classes are ignored. Unlike the existing methods that handle imbalanced data, such as the resampling method or SMOTE method, we show that changing overlapping size between neighbouring patches to get balanced distribution helps on ILD classification. By increasing the overlapping size to get more emphysema and ground glass instances and decreasing the overlapping size to get fewer micronodules instances, we get a relatively balanced data distribution which leads to significantly better results especially for emphysema and ground glass patterns.

Our experimental results show that learning multi-scale information is effective for lung tissue classification. As our future work, we plan to investigate other ways to learn multi-scale features, e.g. replacing Gabor filter with other multi-scale filters. Moreover, we found that CNN is important to the classification performance, and as alternative future work of ours, we plan to incorporate the recent progress of CNN into our work. Finally, we will apply our model to other medical image classification problems, e.g. breast cancer diagnosis.

REFERENCES

[1] W. R. Webb, N. L. Muller, and D. P. Naidich, *High-resolution CT of the lung*. Lippincott Williams & Wilkins, 2014.

[2] R. F. Wagner, M. F. Insana, D. G. Brown, B. S. Garra, and R. J. Jennings, "Texture discrimination: radiologist, machine and man," *Vision: Coding and efficiency*, vol. 558, pp. 310–318, 1990.

[3] R. Kakinuma, H. Ohmatsu, M. Kaneko, K. Eguchi, T. Naruke, K. Nagai, Y. Nishiwaki, A. Suzuki, and N. Moriyama, "Detection failures in spiral ct screening for lung cancer: Analysis of ct findings 1," *Radiology*, vol. 212, no. 1, pp. 61–66, 1999.

[4] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 4, pp. 797–808, 2013.

[5] M. Asherov, I. Diamant, and H. Greenspan, "Lung texture classification using bag of visual words," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2014, pp. 90 352K–90 352K.

[6] A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 4, pp. 665–675, 2012.

[7] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Classification of interstitial lung disease patterns using local dct features and random forest," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 6040–6043.

[8] Y. Xu, M. Sonka, G. McLennan, J. Guo, E. Hoffman *et al.*, "Mdct-based 3-d texture classification of emphysema and early smoking related lung pathologies," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 4, pp. 464–475, 2006.

[9] Y. Song, W. Cai, S. Huh, M. Chen, T. Kanade, Y. Zhou, and D. Feng, "Discriminative data transform for image feature extraction and classification," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 452–459.

[10] A. Depeursinge, D. Racoceanu, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, and H. Müller, "Fusing visual and clinical information for lung tissue classification in high-resolution computed tomography," *Artificial intelligence in medicine*, vol. 50, no. 1, pp. 13–21, 2010.

[11] M. Kale, S. Mukhopadhyay, J. K. Dash, M. Garg, and N. Khandelwal, "Differentiation of several interstitial lung disease patterns in hrct images using support vector machine: role of databases on performance," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 97 852Z–97 852Z.

[12] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb, "Nonlinear histogram binning for quantitative analysis of lung tissue fibrosis in high-resolution ct data," in *Medical Imaging*. International Society for Optics and Photonics, 2007, pp. 65 111Q–65 111Q.

[13] U. Bagci, J. Yao, A. Wu, J. Caban, T. N. Palmore, A. F. Suffredini, O. Aras, and D. J. Mollura, "Automatic detection and quantification of tree-in-bud (tib) opacities from ct scans," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1620–1632, 2012.

[14] C. Jacobs, C. I. Sánchez, S. C. Saur, T. Twellmann, P. A. de Jong, and B. van Ginneken, "Computer-aided detection of ground glass nodules in thoracic ct images using shape, intensity and context features," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer, 2011, pp. 207–214.

[15] X. Liu, L. Ma, L. Song, Y. Zhao, X. Zhao, and C. Zhou, "Recognizing common ct imaging signs of lung diseases through a new feature selection method based on fisher criterion and genetic optimization," *IEEE journal of biomedical and health informatics*, vol. 19, no. 2, pp. 635–647, 2015.

[16] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van de Ville, and H. Müller, "Lung texture classification using locally–oriented riesz components," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer, 2011, pp. 231–238.

[17] ——, "Multiscale lung texture signature learning using the riesz transform," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 517–524.

[18] L. Sørensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 2, pp. 559–569, 2010.

[19] Y. Song, W. Cai, H. Huang, Y. Zhou, Y. Wang, and D. Feng, "Boosted multifold sparse representation with application to ild classification," in *Proc. ISBI*, 2014, pp. 1023–1026.

[20] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, and M. Chen, "Large margin aggregation of local estimates for medical image classification," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer, 2014, pp. 196–203.

[21] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, and M. Chen, "Large margin local estimate with applications to medical image classification," *IEEE transactions on medical imaging*, vol. 34, no. 6, pp. 1362–1377, 2015.

[22] Y. Song, W. Cai, H. Huang, Y. Zhou, Y. Wang, and D. D. Feng, "Locality-constrained subcluster representation ensemble for lung image classification," *Medical image analysis*, vol. 22, no. 1, pp. 102–113, 2015.

[23] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *American Journal of Respiratory and Critical Care Medicine*, vol. 160, no. 2, pp. 648–654, 1999.

[24] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.

[25] U. Bagci, K. Jaster-Miller, K. N. Olivier, J. Yao, and D. J. Mollura, "Synergistic combination of clinical and imaging features predicts abnormal imaging patterns of pulmonary infections," *Computers in biology and medicine*, vol. 43, no. 9, pp. 1241–1251, 2013.

[26] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[27] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 609–616.

[28] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative *et al.*, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.

[29] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE Transactions on Medical Imaging*, vol. 35, 2016.

[30] Q. Li, W. Cai, and D. D. Feng, "Lung image patch classification with automatic feature learning," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 6079–6082.

[31] G. van Tulder and M. de Bruijne, "Learning features for tissue classification with the classification restricted boltzmann machine," in *Medical Computer Vision: Algorithms for Big Data*. Springer, 2014, pp. 47–58.

[32] ——, "Combining generative and discriminative representation learning for lung ct analysis with convolutional restricted boltzmann machines," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1262–1272, 2016.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv preprint arXiv:1409.4842*, 2014.

[35] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[36] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 411–418.

[37] 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 *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 246–253.

[38] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[39] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of ad/mci patients," *IEEE journal of biomedical and health informatics*, vol. 19, no. 5, pp. 1610–1616, 2015.

[40] M. Buty, Z. Xu, M. Gao, U. Bagci, A. Wu, and D. J. Mollura, "Characterization of lung nodule malignancy using hybrid shape and appearance features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 662–670.

[41] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 844–848.

[42] 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*, pp. 1–6, 2016.

[43] S. R. Tarando, C. Fetita, A. Faccinetto, and P. Yves, "Increasing cad system efficacy for lung texture analysis using a convolutional network," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 97 850Q–97 850Q.

[44] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[45] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[46] T. S. Cohen and M. Welling, "Transformation properties of learned visual representations," *arXiv preprint arXiv:1412.7659*, 2014.

[47] K. Lenc and A. Vedaldi, "Understanding image representations by measuring their equivariance and equivalence," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 991–999.

[48] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Computerized medical imaging and graphics*, vol. 36, no. 3, pp. 227–238, 2012.

[49] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2011.

[50] P. Cao, D. Zhao, and O. Zaiane, *An Optimized Cost-Sensitive SVM for Imbalanced Data Learning*. Springer Berlin Heidelberg, 2013.

[51] B. Wang and N. Japkowicz, "Imbalanced data set learning with synthetic samples," in *Proc. IRIS Machine Learning Workshop*, 2004, p. 19.

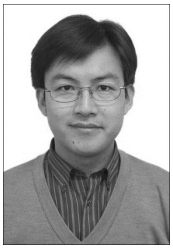
[52] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 786–791.

[53] Z. Lei, S. Liao, R. He, M. Pietikainen, and S. Z. Li, "Gabor volume based local binary pattern for face representation and recognition," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[54] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[55] M. S. AL-TARAWNEH, "Lung cancer detection using image processing techniques," *Leonardo Electronic Journal of Practices and Technologies*, vol. 11, no. 21, pp. 147–58, 2012.
[56] M. Kakar and D. R. Olsen, "Automatic segmentation and recognition of lungs and lesion from ct scans of thorax," *Computerized Medical Imaging and Graphics*, vol. 33, no. 1, pp. 72–82, 2009.
[57] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli, "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 4, pp. 344–357, 2003.
[58] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 55–73, 1990.
[59] J. Han and K.-K. Ma, "Rotation-invariant and scale-invariant gabor features for texture image retrieval," *Image and vision computing*, vol. 25, no. 9, pp. 1474–1481, 2007.
[60] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
[61] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[62] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," *arXiv preprint arXiv:1412.1265*, 2014.
[63] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 580–587.
[64] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *arXiv preprint arXiv:1411.4038*, 2014.
[65] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[66] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.

Qiangchang Wang received MS degree from Shandong University, Jinan, Shandong province, China , in 2016. His major research interests are in machine learning, computer vision, and medical image processing.



Yuanjie Zheng received his Ph.D. degree in Pattern Recognition and Intelligent Systems from Shanghai Jiao Tong University, Shanghai, China, in 2006. He is a professor and a vice dean in both the School of Information Science and Engineering and the Institute of Life Sciences of Shandong Normal University. He is also a Taishan Scholar of Shandong Provincial Government and a principal investigator directing the Key Laboratory of Intelligent Information Processing of Shandong Normal University. He used to be a senior research investigator in Perelman School of Medicine at the University of Pennsylvania and the primary contact of Image Analysis Core at the Penn Vision Research Center. His current major research interests are in computer vision, medical image analysis and clinical study.

Gongping Yang received his Ph.D. degree in computer software and theory from Shandong University, China, in 2007. Now he is a professor in the School of Computer Science and Technology, Shandong University. His research interests are biometrics, medical image processing, and so forth.

Weidong Jin was born in Jining city, in 1973. He received the M.S. degree in Political Economics from Yunnan University, Kunming city, Yunnan province, in 2004 and the Ph.D. degree in Political Economics from Nanjing University, Nanjing city, Jiangsu province, in 2007.

From 2011 to 2012, he had a part-time research at Institute of Educational Economics, Peking University. Since 2012, he has been a professor at the School of economics, Shandong University of Finance and Economics. He is the author of one book, more than 30 articles which were published in Statistical Research, Research on Education Tsinghua University, Economic Review, etc. His research interests include educational economics, human capital theory, and economic methods.

He won the third prize of Shandong Soft Science Outstanding Achievement Award in 2011 and the second prize in 2014, the third prize of Shandong College Outstanding Achievement Award (Humanities and Social Science) in 2012 and 2015.

Xinjian Chen received the Ph.D. degree from the Center for Biometrics and Security Research, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2006. After completing the graduation, he joined Microsoft Research Asia, where he was involved in research on handwriting recognition. From 2008 to 2012, he conducted the Post-Doctoral Research at several prestigious groups, such as the Medical Image Processing Group, University of Pennsylvania, the Department of Radiology and Image Sciences, National Institutes of Health, and the Department of Electrical and Computer Engineering, University of Iowa. In 2012, he joined the School of Electrical and Information Engineering, Soochow University, as a Full Professor. He has authored over 70 high quality international journal/conference papers. His research interests include medical image processing and analysis, pattern recognition, machine learning, and their applications. He is a Distinguished Professor with Soochow University, and serves as the Director of the University Level LaboratoryMedical Image Processing, Analysis and Visualization Laboratory

Yilong Yin received the Ph.D. degree from Jilin University, Changchun, China, in 2000.

He is the Director of the Machine Learning and Applications Group and a Professor with Shandong University, Jinan, China. From 2000 to 2002, he was a Post-Doctoral Fellow with the Department of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include machine learning, data mining, and computational medicine.