Expert knowledge-infused deep learning for automatic lung nodule detection

Jiaxing Tan[a], **Yumei Huo**[b], **Zhengrong Liang**[c,*], and **Lihong Li**[d]

[a]Department of Computer Science, City University of New York, the Graduate Center, NY, USA

[b]Department of Computer Science, City University of New York at CSI, NY, USA

[c]Department of Radiology, State University of New York at Stony Brook, NY, USA

[d]Department of Engineering Science and Physics, City University of New York at CSI, NY, USA

Abstract

BACKGROUND: Computer aided detection (CADe) of pulmonary nodules from computed tomography (CT) is crucial for early diagnosis of lung cancer. Self-learned features obtained by training datasets via deep learning have facilitated CADe of the nodules. However, the complexity of CT lung images renders a challenge of extracting effective features by self-learning only. This condition is exacerbated for limited size of datasets. On the other hand, the engineered features have been widely studied.

OBJECTIVE: We proposed a novel nodule CADe which aims to relieve the challenge by the use of available engineered features to prevent convolution neural networks (CNN) from overfitting under dataset limitation and reduce the running-time complexity of self-learning.

METHODS: The CADe methodology infuses adequately the engineered features, particularly texture features, into the deep learning process.

RESULTS: The methodology was validated on 208 patients with at least one juxta-pleural nodule from the public LIDC-IDRI database. Results demonstrated that the methodology achieves a sensitivity of 88% with 1.9 false positives per scan and a sensitivity of 94.01% with 4.01 false positives per scan.

CONCLUSIONS: The methodology shows high performance compared with the state-of-the-art results, in terms of accuracy and efficiency, from both existing CNN-based approaches and engineered feature-based classifications.

Keywords

Computer aided detection (CADe); computed tomography (CT) imaging; pulmonary nodules; deep learning; image features analysis

*Corresponding author: Zhengrong Liang, Department of Radiology, Electrical and Computer Engineering, and Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA. Jerome.Liang@stonybrookmedicine.edu.

1. Introduction

Lung cancer is the leading cause of cancer-related deaths in the United States. In 2018, there were approximately 234,030 new cases reported and the estimated deaths were 154,050 [1]. And its overall five-year survival rate is merely 18%. Computed tomography (CT) imaging is the most popular imaging techniques for detecting and diagnosing the lung cancer, which typically consists of the task of detecting the pulmonary nodules on CT scans as the first stage and then the task of diagnosing the malignancy of the detected nodules. Pulmonary nodule detection on CT scans is a challenging task given the high dimensionality of the data (512 X 512 X 140 voxels per CT scan), the low signal to noise ratio of the dynamic sequence and the variable size and shape of pulmonary nodules. Therefore, a computer-aided detection (CADe) system that automatically localizes the pulmonary nodules in CT images would be a useful tool to facilitate radiologists' diagnosis. Meanwhile, the high dimensionality of CT images requires computationally efficient methods for pulmonary nodule detection to be developed for practical and feasible purposes.

Based on their locations, lung nodules can be classified into isolated nodules, which are within the lung area, and juxta-pleural nodules, which are typically attached to the lung wall. Since juxta-pleural nodules have similar intensities as the lung wall, they are even more difficult to detect. As a result, the traditional methods such as region growing and active contour model, usually fail in the classification of juxta-pleural nodules.

In the past decades, a lot of research has been done in this field. Initially in 1970s nodule detection was performed by a series of low-level image processing (e.g. adaptive filtration and histogram equalization [2]) and geometry modeling (circularity, size [3]). Then starting from the 1990s, supervised techniques with engineered (hand-crafted) features were commonly applied for nodule detection. Specific methods ranged from grey level thresholding method [4], surface normal overlap method [5], combined volumetric shape index and dot maps followed by adaptive thresholding and modified expectation-maximization method [6], to support vector machine (SVM) based methods [7] etc. All these methods are usually implemented through a series of procedures with manually empirical parameter adjustments and thus are computationally complex and potentially sub-optimal, resulting in false positives (FP) and missed detections.

Nevertheless, from these researches, a lot of features have been found very effective in pulmonary nodules detection and diagnosis. For example, Ye et al. [6] demonstrated that the combination of the shape index (local shape information) and ldquodotrdquo features (local intensity dispersion information) provides a good structure descriptor for the initial nodule candidate (INC) generation. Their proposed feature-based classifier obtained an average detection rate of 90.2%, with approximately 8.2 FPs per scan. Han et al. [7] extracted features by four categories, including 10 geometric or shape features, 16 intensity features, 15 gradient features, and 8 Hessian eigenvalue based features. Their feature-based classifier obtained an overall sensitivity of 82.7% at a specificity of 4 FPs/scan, and 89.2% sensitivity at 4.14 FPs/scan for the classification of juxta-pleural INCs only.

Recently, due to its power for automatically feature learning, deep learning based techniques started to be applied to computer aided detection in medical imaging including the pulmonary nodules detection on CT scans ([8, 9]). Deep learning based methods, following the end-to-end design, are totally automatic with fewer steps and more data invariant. However, due to the complexity of medical imaging (e.g. anatomical structure details and low image contrast among soft tissues), it is very hard to obtain effective features from raw data by self-learning and achieve comparable or even better performance than the traditional hand-crafted feature based techniques in terms of the metrics of the receiver operating characteristic (ROC), the Area Under the ROC curve (AUC) and the False Positive Rate. Moreover, limited available annotated training sets make their application in medical image analysis even more challenging.

Two related solutions are proposed so far. Shin et al. [10] utilize the well- trained benchmark weights from computer vision to solve medical imaging problem via transfer learning. However, models trained in computer vision have constraints on the input size and requires the input to have three channels. The method to convert single channel input into three channels is hard to design and will result in information loss. Wang et al. [11] proposed a hybrid model for nodule malignancy risk prediction by adding shape and texture feature as one part of the model inputs, which demonstrates the possibility of combining traditional features to improve deep learning based model.

In this research, we introduce a novel system for pulmonary nodule detection from CT scan images inspired by the simplicity of deep learning and the efficiency of features obtained from the previous study of the traditional hand- engineered features based methods. Our main goal is to achieve a reduction of run time complexity and prevent CNN from overfitting without losing detection accuracy. To this endeavor, we design the expert knowledge-infused deep learning- based method, in which two-stage fusion (early fusion and late fusion) is proposed for integrating different types of expert knowledge features. We evaluated our methodology on 208 patients with at least one juxta-pleural nodule (which is more challenging to be detected) from the largest public database founded by the Lung Image Database Consortium and Image Database Resource initiative (LIDC-IDRI). Results showed that our methodology achieved comparable and even better accuracy compared to the state of the art ([7, 9]), while attaining reduced running time resulted from the CNN neural network with less overfitting.

1.1. Related works for lung nodule detection

In view of characteristics of feature design, CADe systems for lung nodule detection can be classified into hand-engineered features based systems or self-learned features based systems. Hand-engineered features based systems are built on the domain knowledge about lung nodule characteristics. Tan et al. [12] developed a two-stage CADe system. The detection stage is to generate nodule candidates based on nodule and vessel enhancement and locate the centers of the nodule clusters. The classification stage is based on the proposed extract invariant features to differentiate between real and false nodules. A feature selection with genetic algorithms and artificial neural networks in classification stage was used for a better combination of all designed features. Choi et al. [13] focused on the shape

characteristics of the lung nodules. After nodule candidates are detected through multi-scale dot enhancement filtering, three-dimensional shape-based features are extracted from these candidates, and then used for classification by means of support vector machine. Pena et al. [14] took advantage of the three-dimensional (3D) shape characteristics of nodule candidates and proposed a 3D blob algorithm associated with a connectivity algorithm for nodule shaped candidate selection. In their method, eight minimal representative features are extracted for candidates discrimination with SVM classifier. And their proposed CADe system considers both internal and juxta-pleural nodules detection. In [15], after lung parenchyma is segmented by region growing and the opening process based lung segmentation, a stable 3D Mass-Spring Model associated with a spline curves reconstruction is used for suspected nodular lesions segmentation and extraction. Gray value, contour and shape features extracted from nodule candidates are used with neural network for a further false positives reduction. Sivakumar et al. [16] designed an efficient CADe system by utilizing fuzzy based clustering for nodule segmentation. Besides, SVM with three different kernel functions is used for good nodule classification results. To increase the accuracy of lung nodules identification, Camarlinghi et al. [17] proposed to ensemble the outputs of three different CADe systems developed by the Italian MAGIC-5 collaboration.

A self-learned feature based system, in contrast, does not have pre-defined features but can automatically find useful features from the input data. Inspired by the feature learning power of deep learning methods [18], several successful attempts have been made. Shin et al. [10] evaluated the performance of different deep CNNs with the help of transfer learning to take advantage of the pre-trained CNN model from computer vision to perform CT scan analysis. Two CADe tasks were experimented on: thoraco-abdominal lymph node detection and interstitial lung disease classification. Anirudh et al. [8] have applied 3D CNN on a weakly labeled lung nodule dataset. Given only center point and estimated region size as a label, unsupervised segmentation is used to grow out a 3D region for patch preparation. In [19], we designed an end-to-end voting based CNN method for lung nodule detection and obtained AUC of 0.82.

Due to the complexity of medical imaging with limited data size, it is essential for us to design a deep learning based method to be able to effectively learn useful features and achieve efficient self-learning from those features. Setio et al. [20] designed a model that fuses features learned from differently oriented planes extracted from each candidate for final decision making. Teramoto et al. [21] proposed to combine CT image and positron emission tomography image for a better decision making. Jiang et al. [9] proposed to provide the candidate mask as a second input, which can denote the location of the candidate in the input patch. And the candidate is centered with center pooling applied to enhance feature learned inside the candidate. Recently, in the area of computer vision, to overcome the challenges of the limited existing datasets and the difficulty for network to learn various types of robust features simultaneously, some researchers have proposed methods to incorporate the hand-crafted features in combination with CNN based models ([22–27]).

Inspired by these designs, we propose to infuse domain knowledge to help the feature learning and design an efficient deep learning based CADe for lung nodule detection. From the research of the traditional hand-engineered features based CADe systems, a lot of

domain knowledge have been shown to be effective for lung nodule detection. For example, Ye et al. [6] proposed to combine local shape information and local intensity dispersion information to improve initial nodule candidates generation; Tac et al. [28] applied shape and texture based features for the juxta-pleural nodule detection and showed the effectiveness of 33 features; Santos et al. [29] considered Tsallis's and Shannon's entropy measurements as texture features combined with SVM for designing a CADe for small nodules (diameters between 2 and 10 mm); and in [7], a contrast of inside/outside candidate region is shown to be beneficial in the task of candidate classification.

In this paper, we consider to infuse various domain knowledge and examine the performance of our expert knowledge infused deep learning based pulmonary nodule detection model. Starting from our basic deep learning based model, we explore ways of incorporating expert knowledge and compare their performance. Experimental results showed that our proposed knowledge infused deep learning based method outperforms the basic model by a large scale with more than 10% in AUC, and moreover achieves comparable and even better accuracy compared to the state of the art ([7, 9]) while attaining reduced running time resulted from the CNN with less overfitting and limited dataset issues.

2. Methods

As Shown in Fig. 1, our pulmonary nodule detection method consists of two steps: INC detection and INC classification. In the INC detection step (Fig. 1 (a)), we segment lung area from the thoracic CT scan and then extract the INCs within the lung area with a sensitivity of 100% while maintaining a low percentage of FPs. INCs are then centered and resized, and used as the input of the INC classification step. In the INC classification step (Fig. 1 (b)), we design a CNN based classification model infused with various expert knowledge which includes the contrast information of the INCs and the outer environments (which are denoted as inside and outside INCs knowledge respectively), and the important shape and texture features obtained from the hand engineered methods. In the following, we first describe the data preparation for our nodule detection method, and then give the details of INC detection step and INC classification step respectively.

2.1. Data preparation

The original RAW CT data is acquired from the largest public database founded by the LIDC-IDRI database [30]. This public database contains radiologists visual assessment of the risk of malignancy on the pulmonary nodules. Radiologists label the position of the nodule for each CT slice.

Each CT slice has a size of 512×512. We then transform the RAW data into a matrix file which is the CNN friendly format. Most of the juxta-pleural nodules in our datasets, as shown in Fig. 2, are relatively small with a radius less than 15 mm, which increases the difficulty of conducting nodule detection. Compared with non juxta-pleural nodules, juxta-pleural nodules are usually semispherical or speculate in shape, making them more difficult to detect. In addition, because they have similar intensities with the pleural wall, correcting the lung borders to include these nodules is another challenge task. This study focuses on

evaluate our CADe system on 208 patients with irregularly-shaped juxta-pleural nodules to demonstrate the robustness of our system.

2.2.   INC detection

2.2.1.   Lung segmentation—In order to ensure nodule detection being performed within the lung area, the segmentation of lungs from the body volume is desired. First of all, we need to separate the body structure from its surrounding CT background, where a threshold of –500 HU was imposed on the entire scan. Based on the obtained binary image, the chest body volume was then extracted by removing the outside component that connects to image margins. Our next task is to segment lungs via a vector quantization (VQ) algorithm from the chest body volume. In this study, we choose the first-order 3D neighbors for constructing local intensity vectors of 7 elements or dimensions. The principal component analysis [31] is utilized to generate a series of feature vectors via the Karhunen-Loeve transformation [32]. The first few principal components that sum up, at least, 95% percent of the total variance are chosen for optimizing the dimensions of the feature vectors. Then a self-adaptive online VQ algorithm is applied to these feature vectors for classification, where the maximum number of classes for classification is preset based on the anatomical knowledge. For the chest body volume, we observe a clear separation of two major classes (i.e., the air, tissue and other dense body tissue). Since the lung parenchyma and the air in other organs have similar image intensities, they were classified into the low-intensity class. The initial lung mask corresponds to the largest and the second largest (if the left and right lungs are disconnected due to pathologic abnormalities) connected components in the low- intensity class. Then a flood-fill operation [33] is utilized to fill the holes inside the extracted lung mask. Furthermore, morphological closing [34] is operated on the lung boundaries to include the juxta-pleural nodules (i.e., nodules grow near or originated from the parenchyma wall) into consideration.

2.2.2.   INCs extraction via VQ—After the extraction of the lung fields by the initial two-class VQ, a second- stage VQ aims to simultaneously detect and segment INCs within the lung area. Detection of INCs is a very important and challenging task in the computer-aided detection of lung nodules which require our algorithm to accurately characterize all nodules with a sensitivity of 100% while maintaining a low percentage of FPs. Thus, in this stage, we aimed at developing an accurate VQ scheme to detect all INCs. Statistically, we observed the image intensity distribution can be decomposed into four normal-tailed Gaussian mixtures. Based on physicians input, we interpret the four classes as the low-intensity parenchyma, the high-intensity parenchyma, blood vessels, and INCs. As a result, we applied a four- class VQ algorithms for extracting INCs in this stage. Because the intensity of lung nodule class is relatively higher than the other three classes, the class with the highest average intensity was extracted as the INCs.

2.2.3.   INC resizing—After the INCs are extracted, we centered the INCs in each extracted patch. CNN based model could only take fixed input size, while the detected INCs have a large range in diameters from 3 mm to 20 mm approximately. One simple way is to take a patch with the smallest size which can contain all nodules of various sizes. However, such INC patch extraction method will make small nodules too subtle in the patch for

detection and thus provides too little information for candidate classification. This issue worsens as the diameter of nodules in our dataset has a large variance.

Considering this situation, we propose a proportional patch extraction method. As shown in Fig. 3 and Equation (1), we cut the patch based on a predefined parameter *r* which is the ratio of nodule diameter *d* to the environment *b*. Then we resize the patches to the same size with image resizing method, for which various methods have been proposed and used. In this study, we applied the bilinear resize method.

$$r = \frac{d}{b} \quad (1)$$

2.3. Deep learning based INC classification

CNN has shown a great success in computer vision on the ImageNet challenge. Since AlexNet [35] was proposed, the performance improvement has been achieved for almost every year with deeper and deeper structures supported by high- performance computing facilities. CNN could learn features from data itself, which turns out to be efficient and automatic. Some progress has been made in medical imaging using CNN to solve classification problems, including lung nodule detection ([8, 10, 18, 36]). However, due to the limited medical image dataset, it is very hard to design and train the CNN models to achieve the same performance as in computer vision. On the other side, with the decades of study on the traditional hand engineered methods, some useful features have been found to be effective for lung cancer detection and diagnosis. In this paper, we aim at designing a deep learning based pulmonary nodule detection system with expert knowledge infused. This model will leverage the limited medical image dataset as well as achieve the best possible detection results by applying the expert knowledge to the deep learning model. In the following, we first introduce our basic deep learning model for pulmonary nodule detection, and then we will present ways that expert knowledge including environment (inside and outside) of the INC knowledge, shape-based knowledge, and texture based knowledge, are infused in our deep learning based INC classification model.

2.3.1. Basic model design—Generally speaking, CNN is a deep neural network, inspired by the biology study of human cortex and constructed by four types of layers: input layer, multiple convolution layers with each followed by a subsampling layer, and several fully- connected layers in the end.

A convolution layer has K different kernels, and each has the shape $m \times n$ and performs convolution operation (denoted as $\ast$) on each of J sub-images of the input image u. A non-linear activation function g will be applied to the convolution result with a learnable bias $\hat{b}$ added. The whole procedure is shown in Equation (2), where (Wk denotes the learnable weight matrix of kernel k ($k \in K$).

$$f = g\left(\left(W_k * u_{1:J}\right) + \hat{b}\right). \quad (2)$$

Correspondingly, the output of the convolution layer is K feature maps, which are generated by a convolution operation with one kernel applied on the whole image. Rectified linear unit (ReLU) function, as shown in Equation (3), are commonly used as activation functions after the convolution layers [9] as it can help the CNN structure reduce the training error rate faster than other activation functions [35]. Given a value t, the ReLU function is defined in Equation (3).

$$\mathrm{Re}\, LU(t) = \max(t,\, 0). \quad (3)$$

However, there might be potential problems caused by ReLU, as it sets 0 for all negative values. To alleviate this, we employ LeakRelu as the activation function which was first proposed in [37]. As shown in Equation (4), LeakyReLU sets a small non-zero gradient NegativeSlope, denoted as α, which is user pre-defined, to negative values.

$$leakyReLU\,(t,\, a) = max(t,\, 0) + a \times min(t,\, 0). \quad (4)$$

As for the pooling layer, we take the commonly used max-pooling operation. Our basic model is based on a commonly used CNN structure proposed for nodule detection in [36]. The design is illustrated in Fig. 4. The details are shown in Table 1. It contains two convolution layers with kernel size 5×5 and 3×3 respectively, each with LeakyReLU as the activation function and followed by a Max-pooling layer. The last column shows how many different kernels are used for a convolution layer or how many neurons are used in a fully connected layer. The first convolution layer has 32 kernels and the second one has 64 kernels. The first fully connected layer has 500 neurons. The output layer has two nodes representing the two classes to be classified: nodule and non-nodule. We are not specifying the input layer in the table since all ConvNets take input size of 22×22. We employ the Binary Cross-Entropy (BCE) loss for training as this is a binary classification problem in nature. Given a label y and model prediction probability $\hat{y}$, the cross-entropy loss is calculated as:

$$BCE(y,\, \hat{y} = -(ylog(\hat{y}) + (1-y)log\,(1-\hat{y})) \quad (5)$$

2.3.2. Early fusion of the knowledge from inside and outside of INCs—As addressed in [7], a contrast of inside and outside INC knowledge is beneficial in the task of candidate classification. To help CNN learn better features, after we obtain the INC patches from the raw image as shown in Section 2.2, we extract the INC knowledge and infuse it into our deep learning detection model. There are two possible ways to extract the INC

knowledge. The first way is to obtain the INC mask indicating the position of INC such that the values within the region of INC are set to 1 and the values outside are set to 0. In this way, we can take the original CT patch with INC mask as the input. The second way is to generate two separate images from the original INC patches, one is the inside INC image which contains only the INC with all the pixel values outside set to 0 and the other is the outside INC image which contains only the environment with all the pixel values within INC set to 0. Figure 5 shows the inside INC images and the outside INC images separated from the INC patches for nodule INCs and non-nodule INCs. In this way, we can take both the inside INC image and the outside INC image as inputs.

For both types of INC knowledge, we will use them as inputs for the first convolution layer. Since this fusion is performed in the early stage of the CNN model, we denote it as early fusion. Moreover, there are two different designs to use this INC knowledge as inputs: The first way is to take two images, which are the original image and the INC mask image for the first type or the inside INC image and the outside INC image for the second type, as dual input paths. That is, each as the separate input passes through an individual convolution path and the outputs from these two paths will be combined afterwards as illustrated in Fig. 6(a). The second way is to take both images as single input path, i.e., combine two images as one input with two channels that pass through one convolution path as illustrated in Fig. 6(b).

2.3.3. Early fusion of the histogram of oriented gradients (HOG) feature—

Shape characteristics of the lung nodules have been shown effective in the task of INC classification by Ye et al. [6], Ta ci et al. [28], and Choi et al. [13] etc. In this research, we will extract the shape features from the obtained INC images and infuse them into our CNN based detection model.

We utilize the HOG feature, which is designed based on the idea that the distribution of local intensity gradients in an image could characterize the appearance and shape of an object [38]. HOG feature generates a normalized histogram of gradient orientations in localized portions of an image. Similar to edge map, as it operates on local cells, it is invariant to geometric and photometric transformations.

The first-order derivatives of HOG features are computed in the following three steps:

Gradient Calculation: For each pixel in original image I, we generate two matrices, Ix for gradients in x-axis and Iy for gradients in y-axis, which are calculated as:

$$I_x(x,\ y)\ =\ I(x+1,\ y) - I(x-1,\ y). \quad (6)$$

$$I_y(x,\ y)\ =\ I(x,\ y+1) - I(x,\ y-1). \quad (7)$$

Then the gradient magnitude m could be calculated as

$$m(x, \ y) \ = \ \sqrt{I_x(x, \ y)^2 + I_y(x, \ y)^2}, \quad (8)$$

And orientation θ is calculated as:

$$\theta(x, \ y) \ = \ \arctan\left(\frac{l_y(x, \ y)}{I_x(x, \ y)}\right), \quad (9)$$

Orientation Bining The second step is to generate cell histograms. The whole image is divided into cells and then HOG is computed for each cell.

Block Normalization The last step is to perform block normalization. At this step, the cells are grouped into blocks and the HOG features are normalized locally. Then the whole HOG feature of the input image is generated by combining all the blocks.

HOG columns in Fig. 7 show the HOG images extracted from the INC patches for nodule INCs and non-nodule INCs. Since the HOG feature, as matrix, is CNN input friendly, we use the produced HOG feature as an input for the first convolution layer. That is, we perform the early fusion of the HOG knowledge.

2.3.4. Early fusion of the local binary patterns (LBP) feature—Besides the shape characteristics, we also infuse the texture features in our detection model. We utilize LBP feature which is a powerful texture descriptor and computes a local representation of texture [39]. It has shown a performance enhancement when jointly used with HOG both in Computer Vision [40] and Medical Imaging [11].

We adopt the most commonly used 8-neighbor setting of LBP. For each pixel c in the grayscale image, based on its eight neighbors, the LBP feature of c is then calculated as below:

$$LBP_{N, R}(c) \ = \ \sum_{i \ = \ 0}^{N-1} \delta(l_i - l_c)2^i, \ \delta(x) \ = \ \begin{cases} 1, & x \ge 0 \\ 0, & Otherwise \end{cases} \quad (10)$$

where N is number of neighbors and is set to 8, Ic is the gray value of the central pixel c in the local neighborhood, and Ii ($0 \ i \ N-1$) is the gray values of N equally spaced pixels (neighbors) on a circle centered at c with the radius of R. Therefore, the signs of the pixel value differences between the neighbors and the center c are interpreted as an N-bit binary number, resulting in $2N$ (e.g., $28 = 256$) distinct values for the binary pattern. Due to the fact that the nodule images could not contain all 256 patterns, the selection of effective patterns is adapted.

LBP columns in Fig. 7 show the LBP images extracted from the INC patches for nodule INCs and non-nodule INCs. Similar to HOG feature, LBP feature is CNN input friendly, so

we perform the early fusion of LBP knowledge feeding the LBP feature as the input to the first convolution layer.

2.3.5. Late fusion of Haralick features—Different from the LBP features that describe the local texture features, another group of text features, named Haralick features [41], describe the global texture and have also been shown to be effective in the task of nodule detection [42, 43]. In this research, we also infuse Haralick features in our detection model.

Haralick features are statistical texture features extracted from 2D intensity or gray-level image, which are calculated from a gray level co-occurrence matrix (GLCM) to capture the gray-level correlations among resolution cells or image pixels in a 2D image slice. Along each direction through the image, 14 texture measures are calculated from the GLCM. The 14 texture measures are listed in [41]. These measurements are utilized to describe the overall texture of the image using measures such as entropy and sum of variance.

For 2D Haralick features, four directions, as shown in Fig. 8, are defined on the image plane (0, 45, 90, 135 degrees) which are sufficient to span over the image slice. Assuming a similarity among the four directions, the mean and range for each of the 14 sets of four directions are calculated resulting in 28 features. Different from HOG and LBP features which are extracted as a matrix, Haralick features are extracted as a vector of 28 feature values and thus cannot be used directly as an input of the convolution layers. As a result, we design to integrate Haralick features into the extracted features obtained from the last max-pooling layer and are fed as an input into the fully-connected layers to give the classification results. Since this fusion is performed in the late stage of the CNN model, we denote it as late fusion.

2.3.6. Extension of feature fusion model—As addressed above, we propose our Feature Fusion Model by combining traditional features with deep learning based detection model. The whole design of our model is illustrated in Fig. 1 (b). For this model, we apply BCE loss as our training loss. One can extend this model by infusing more features. If the extracted features are the same size as the original image such as HOG and LBP features, the early fusion of these features can be applied; Otherwise, if the extracted features are not the same size and are very hard to be converted to the same size of the original image, as Haralick feature, the late fusion of these features can be employed.

3. Experiment

3.1. Experiment setting

As we described in Section 2.2, we extract INCs with a size of 24×24 from the original CT images, center the INCs, and resize them using the bilinear interpolation method. To augment the dataset, we apply rotation in four directions, which results in 5604 INCs containing nodules. To train this binary classifier, in each iteration, we randomly sample the same number of negative samples to build a balanced dataset such that in each iteration we trained our model with a dynamically sampled balanced dataset. Then for each INC patch, we separate the inside INC image and outside INC image to obtain the inside and outside

INC knowledge as addressed in Section 2.3.2, and extract the HOG feature, LBP feature, and Haralick features to obtain the shape, local texture, and global texture knowledge as addressed in Section 2.3.3, Section 2.3.4, Section 2.3.5, respectively. All these knowledge is infused into our deep learning based pulmonary nodule detection model in the way of either early fusion or late fusion. As for the parameter settings, for the HOG feature, we set the cell size to be 8×8 pixels and each block contains 2×2 cells, and for LBP feature, we take 8 neighbors from the circle centered at each center pixel with the radius R set to 2.

In all the experiments, we conduct 10-fold cross validation. That is, we randomly evenly split data into 10 folds. Each time we train our CNN model with 9 folds and test with the rest one. During the training, the data has been further split into 95% for training and 5% for validation to avoid overfitting. We repeat this procedure 10 times.

In the experiment, we implement our model with PyTorch [44]. We use Adam [45] optimizer with a batch size of 20. The learning rate is set to 10–5, momentum is set to 0.9, and weight decay is set to 0.0005. The network is initialized with a Gaussian Distribution. All the programs were executed on the server which had a 64-bit operating system, 64GB RAM and 20-core Intel Core i7–6950X processor with the main frequency of 1.9GHz. The GPU we worked on is Nvidia GTX 1080.

3.2. Experiment result

3.2.1. Comparison of the effect of INC resizing—First, we experiment on the effect of INC resize operation. This experiment is performed on our basic deep learning INC classification model which is described in Section 2.3.1. We use two datasets both containing the INC patches of the size 24×24 and having the INCs centered. The only difference is that the first dataset contains the INC patches directly cut from the CT image without any resize operation performed while the second one with the resize operation performed following the proportional patch extraction described in Section 2.2.3. The experimental results are shown in Fig. 9(a). Comparing basic model's performance on two datasets with the obtained AUC of 0.89 and 0.86 respectively, we observed great performance improvement by applying our INC resize method based on the proportional patch extraction.

3.2.2. Comparison of different ways of infusing INC knowledge—As addressed in Section 2.3.2, we can extract two types of INC knowledge from INC patches which are the original image and the INC mask image for the first type and the inside INC image and the outside INC image for the second type. For each type of INC knowledge, we can use two images as the input in two ways namely, (1) use two images as dual input paths as illustrated by Fig. 6(a) and (2) use two images as single input path as illustrated by Fig. 6(b). Therefore, we come up with four different designs to infuse INC knowledge in our deep learning detection model.

- MaskNet: The original image and the INC mask image each as the separate input passes through an individual convolution path and the outputs from these two paths will be combined afterward.

- MaskChannel: The original image and the INC mask image are combined as one input with two channels and pass through one convolution path.

- SepNet: The inside INC image and the outside INC image each as the separate input passes through an individual convolution path and the outputs from these two paths will be combined afterward.

- DualChannel: The inside INC image and the outside INC image are combined as one input with two channels and pass through one single path.

We conducted an experiment to compare these four designs based on our deep learning detection model. As shown in Fig. 9(b), the SepNet with AUC of 0.931 outperforms all other designs. And on the other hand, the combination of the inside INC image and the outside INC image yields better performance than taking mask as reference for the original image. One possible explanation is that dividing original image into two separate images will be less confusing than using a reference mask attached to the original image patch. The CNN model needs to find the correlation between the reference mask and the original image. As a result, we decide to select the SepNet as a design criteria to infuse the INC knowledge as well as other expert knowledge in our deep learning INC classification model.

3.2.3. Step by step experiment result comparison—In this experiment, we examine the effect of our deep learning pulmonary nodule detection model with the fusion of expert knowledge. Our experiments are thus performed step by step: first on the basic model, then the basic model with INC knowledge infused, and last the basic model with INC knowledge and HOG, LBP and Haralick features infused. Formally, we define them as the following three structures:

- Basic: Our basic model design.

- Dual: The basic model with INC knowledge infused as the SepNet design.

- Dual + Features: Based on the structure Dual, the HOG, LBP and Haralick features are further infused addressed in Section 2.3.3, Section 2.3.4, Section 2.3.5

The experimental results are shown in Fig. 9(c). We can see each of our proposed improvement steps could result in an improvement of model performance.

Moreover, a Wilcoxon signed-rank test [46] was performed for comparing the prediction results on the dataset in each step, where all of the P-values are much less than 0.05, indicating that the proposed method could significantly improve model performance.

Our basic model achieves 0.89 in AUC and the Dual model increased performance by around 4%. Then with combined traditional features, the model achieves 1% more in AUC score, which is 0.943 as our final model performance. A comparison of free-response ROC curve (FROC) of these three models is illustrated in Fig. 10. The training loss per 300 epochs is shown in Fig. 11. Note that the loss of the final training is not the lowest by observation. In our experiments, the training weights are stored every 20 epochs. Through the experiments, we found that the weights stored at the final epoch achieve the best performance.

3.2.4. Experiment on different proportional extraction ratios—We now conduct experiments on the effect of different proportional extraction ratio used for INC resize on our final deep learning based expert knowledge infused INC classification model (Dual + Features). Fig. 9(d) shows the performance comparison when the ratio is set to 20%, 40%, 60% and 80% respectively. The best performance is achieved with the ratio of 60%. A similar performance is achieved with the ratio of 40%. However, when the ratio of INCs is too small or too large, the performance will exacerbate. This could be caused by the fact that either the separated INC or the environment is too small to provide enough information in the comparison.

3.2.5. Comparison of our model with other related work—Finally, we compare the performance of our model with other state of art work as shown in Table 2. Our work focuses on the juxta-pleural dataset which has 208 patients. The method listed in the table, either rely on the traditional features such as HOG, LBP, Haralick, etc. or rely on deep learning method only. Compared to H. Han et al. [7], which acquires the best performance in the literature for juxta-pleural nodules with a sensitivity of 89.2% and lower FP/scan of 4.14, our method achieves a sensitivity of 88% with the FP/scan of 1.89, and the sensitivity of 94.01% with the FP/scan of 4.01.

4. Discussion

In medical imaging tasks, the capacity of CNN is limited by the insufficient data. This problem becomes more severe for CT lung image data, which is complex in terms of anatomical structure, noise, and low image contrast among soft tissues. To cope with this issue, data augmentation and additional features (HOG, LBP, etc.) have been used to help CNN effectively learn a better feature representation. We attribute the success of our method to the ability to "augment" the data in a novel way, rather than improving the depth (capacity) of the network. Current deep learning method purely relies on the model itself to find a meaningful pattern from data. However, when only limited data is available, the power of model's feature learning is limited. Our research shows that domain knowledge could serve as a good way to help the deep learning model effectively learn from the limited data. Domain knowledge provides more insights and guides the model to learn better. This design not only greatly reduces model complexity and parameter size but also demonstrates the similar or even better performance than the current state-of-the-art. Furthermore, our method is simple and can be easily generalized to different settings/problems. In addition, our proportional patch extraction method is adaptive to INC sizes with high variance, thus improving the detection performance. Such design could be applied to other tasks.

In this paper, we infuse shape and texture knowledge including LBP, HOG and Haralick features into the deep learning model. There exist other image-based features such as curvature, entropy, etc. that have been considered in the traditional hand-engineered methods. For lung nodule detection, the effective features of nodules could also include the clinical-based features, such as age, sex, race, smoking status, etc. It would be our future work to infuse these features into our CADe framework. It is also worthwhile to investigate the most effective set of features that could compensate CNNs self-learned features or supervise CNN to achieve better performance.

5. Conclusion

We presented a novel knowledge-infused deep learning-based system for automated detection of juxta-pleural pulmonary nodules in chest CT scans. This system is designed to relieve the dataset limitation challenge for the general deep learning-based methods in the medical imaging field and significantly reduces the complications of the traditional procedures for pulmonary nodules detection while retaining and even outperforming the state-of-art accuracy. With the proposed two-stage fusion method (early fusion and late fusion), this system has shown good scalability and adaption capability in that one can easily combine more useful expert knowledge in the CNN-based model and apply it to other medical imaging problems.

It is interesting to investigate whether there are other effective methods to infuse expert knowledge as well as to augment the data. It is also interesting to investigate the difference between hand-craft features and CNN self-learned features and whether we can achieve better performance by combining both.

Acknowledgments

References

[1]. A.C. Society, Cancer Facts and Figures 2018, Atlanta: American Cancer society (2018).

[2]. Sherrier RH, Chiles C, Wilkinson WE, et al., Effects of image processing on nodule detection rates in digitized chest radiographs: ROC study of observer performance, Radiology 166(2) (1988), 447–450. [PubMed: 3336719]

[3]. Giger ML, MacMahon H, et al., Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields, Medical Physics 15(2) (1988), 158–166. [PubMed: 3386584]

[4]. Armato SG, Giger ML, Moran CJ, et al., Computerized detection of pulmonary nodules on CT scans, Radiographics 19(5) (1999), 1303–1311. [PubMed: 10489181]

[5]. Paik DS, Beaulieu CF, Rubin GD, et al., Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT, IEEE Transactions on Medical Imaging 23(6) (2004), 661–675. [PubMed: 15191141]

[6]. Ye X, Lin X, Dehmeshki J, et al., Shape-based computer-aided detection of lung nodules in thoracic CT images, IEEE Transactions on Biomedical Engineering 56(7) (2009), 1810–1820. [PubMed: 19527950]

[7]. Han H, Li L, Han F, et al., Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme, IEEE Journal of Biomedical and Health Informatics 19(2) (2015), 648–659. [PubMed: 25486657]

[8]. Anirudh R, Thiagarajan JJ, Bremer T and Kim H, Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data, in: Medical Imaging 2016: Computer-Aided Diagnosis, 9785 (2016), 978532.

[9]. Jiang H, Ma H, Qian W, et al., An automatic detection system of lung nodule based on multi-group patch-based deep learning network, IEEE Journal of Biomedical and Health Informatics (2017).

[10]. Shin HC, Roth HR, Gao M, et al., Deep Convolutional Neural Networks for Computer-Aided Detection: CNN architectures, dataset characteristics and transfer learning, IEEE Transactions on Medical Imaging 35(5) (2016), 1285–1298. [PubMed: 26886976]

[11]. Wang H, Zhao T, Li LC, et al., A hybrid CNN feature model for pulmonary nodule malignancy risk differentiation, Journal of X-ray Science and Technology 26(2) (2018), 171–187. [PubMed: 29036877]

[12]. Tan M, Deklerck R, Jansen B, et al., A novel computer-aided lung nodule detection system for CT images, Medical Physics 38(10) (2011), 5630–5645. [PubMed: 21992380]

[13]. Choi WJ and Choi TS, Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor, Computer Methods and Programs in Biomedicine 113(1) (2014), 37–54. [PubMed: 24148147]

[14]. Penã DM, Luo S and Abdelgader A, Auto diagnostics of lung nodules using minimal characteristics extraction technique, Diagnostics 6(1) (2016), 13.

[15]. Cascio D, Magro R, Fauci F, et al., Automatic detection of lung nodules in CT datasets based on stable 3d mass–spring models, Computers in Biology and Medicine 42(11) (2012), 1098–1109. [PubMed: 23020972]

[16]. Sivakumar S and Chandrasekar C, Lung nodule detection using fuzzy clustering and sup- port vector machines, International Journal of Engineering and Technology 5(1) (2013), 179–185.

[17]. Camarlinghi N, Gori I, Retico A, et al., Combination of computer-aided detection algorithms for automatic lung nodule identification, International Journal of Computer Assisted Radiology and Surgery 7(3) (2012), 455–464. [PubMed: 21739112]

[18]. LeCun Y, Bengio Y and Hinton G, Deep learning, Nature 521(7553) (2015), 436. [PubMed: 26017442]

[19]. Tan J, Huo Y, Liang Z, et al., A Fast Automatic Juxta-pleural Lung Nodule Detection Framework Using Convolutional Neural Networks and Vote Algorithm, Patch-Based Techniques in Medical Imaging (2018).

[20]. Setio A, Ciompi F, Litjens G, et al., Pulmonary Nodule Detection in CT images: False Positive Reduction Using Multiview Convolutional Networks, IEEE Transactions on Medical Imaging 35(5) (2016), 1160–1169. [PubMed: 26955024]

[21]. Teramoto A, Fujita H, Yamamuro O and Tamaki T, Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique, Medical Physics 43(6Part1) (2016), 2821–2827. [PubMed: 27277030]

[22]. Lan Z, Lin M, Li X, Hauptmann AG and Raj B, Beyond gaussian pyramid: Multi-skip feature stacking for action recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), 204–212.

[23]. Simonyan K and Zisserman A, Two-stream Convolutional Networks for action recognition in videos, A dvances in Neural Information Processing Systems (2014), 568–576.

[24]. Wang L, Qiao Y and Tang X, Action recognition with trajectory-pooled deep-convolutional descriptors, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), 4305–4314.

[25]. Yue-Hei Ng J, Hausknecht M, Vijayanarasimhan S, et al., Beyond short snippets: Deep networks for video classification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), 4694–4702, IEEE.

[26]. Zha S, Luisier F, Andrews W, et al., Exploiting image-trained CNN architectures for unconstrained video classification, arXiv Preprint arXiv:150304144 (2015).

[27]. Park E, Han X, Berg TL and Berg AC, Combining multiple sources of knowledge in deep CNNs for action recognition, Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on (2016), 1–8, IEEE.

[28]. Tas E and Uǵur A, Shape and texture based novel features for automated juxtapleural nodule detection in lung CTs, Journal of Medical Systems 39(5) (2015), 46. [PubMed: 25732079]

[29]. Santos AM, de Carvalho Filho AO, Silva AC, et al., Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM, Engineering Applications of Artificial Intelligence 36 (2014), 27–39.

[30]. Armato SG, McLennan G, Bidaut L, et al., The lung image database consortium (LIDC) and image database resource initiative (IDRI):A completed reference database of lung nodules on CT scans, Medical Physics 38(2) (2011), 915–931. [PubMed: 21452728]

[31]. Wold S, Esbensen K and Geladi P, Principal component analysis, Chemometrics and Intelligent Laboratory Systems 2(1–3) (1987), 37–52.

[32]. Dony R et al., Karhunen-loeve transform, The Transform and Data Ompression Handbook 1 (2001), 1–34.

[33]. Glassner AS, Graphics Gems (2013).

[34]. Serra J, Image analysis and mathematical morphology (1983).

[35]. Krizhevsky A, Sutskever I and Hinton GE, Imagenet classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems (2012), 1097–1105.

[36]. Tajbakhsh N and Suzuki K, Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs, Pattern Recognition 63 (2017), 476–486.

[37]. Maas AL, Hannun AY and Ng AY, Rectifier nonlinearities improve neural network acoustic models, Proc icml 30 (2013), 3.

[38]. Dalal N and Triggs B, Histograms of Oriented Gradients for Human Detection 1 (2005), 886–893, IEEE.

[39]. Ojala T, Pietikainen M and Maenpaa T, Multiresolution gray-scale and rotation in- variant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7) (2002), 971–987.

[40]. Wang X, Han TX and Yan S, An HOG-LBP human detector with partial occlusion handling (2009), 32–39, IEEE.

[41]. Haralick RM, Shanmugam K et al., Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics (1973), 610–621.

[42]. Han F, Wang H, Zhang G, et al., Texture feature analysis for computer-aided diagnosis on pulmonary nodules, Journal of Digital Imaging 28(1) (2015), 99–115. [PubMed: 25117512]

[43]. Hu Y, Liang Z, Song B, et al., Texture feature extraction and analysis for polyp differentiation via computed tomography colonography, IEEE Transactions on Medical Imaging 35(6) (2016), 1522–1531. [PubMed: 26800530]

[44]. Paszke A, Gross S, Chintala S, et al., Automatic differentiation in PyTorch (2017).

[45]. Kingma DP and Ba J, Adam: A method for stochastic optimization, arXiv preprint arXiv: 14126980 (2014).

[46]. Wilcoxon F, Individual comparisons by ranking methods, Biometrics Bulletin 1(6) (1945), 80–83.
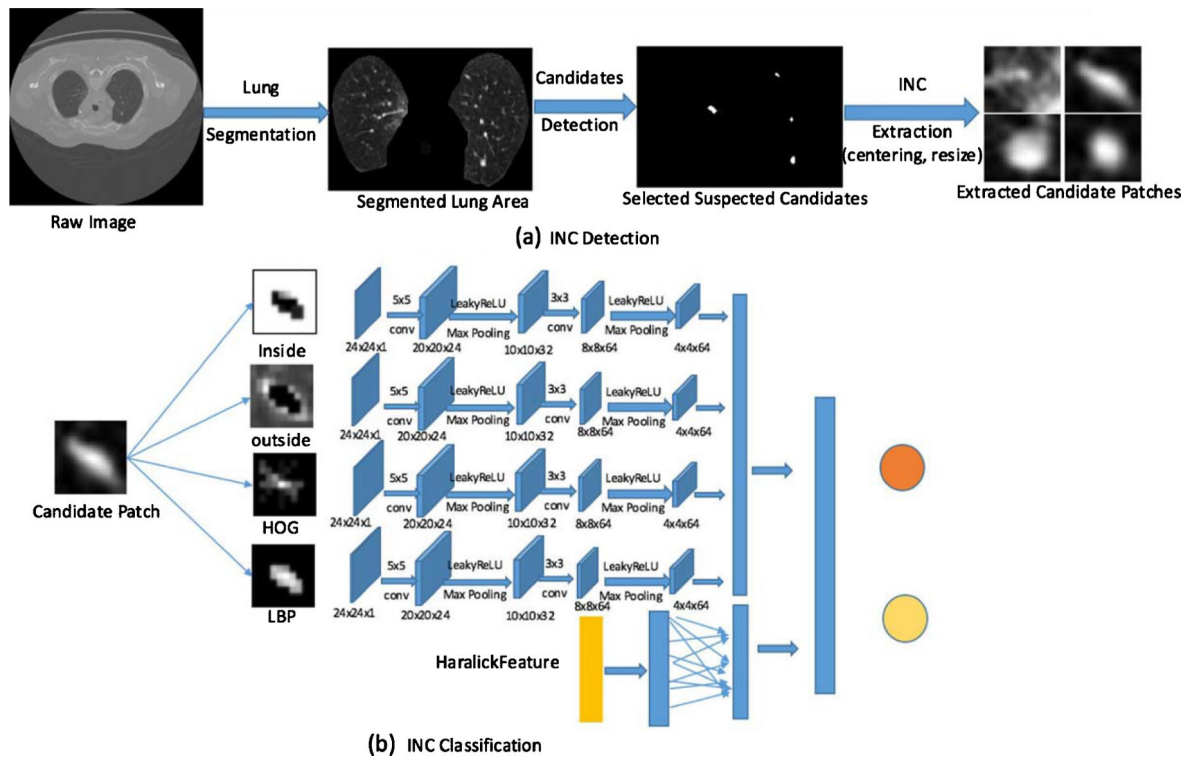
Fig. 1.
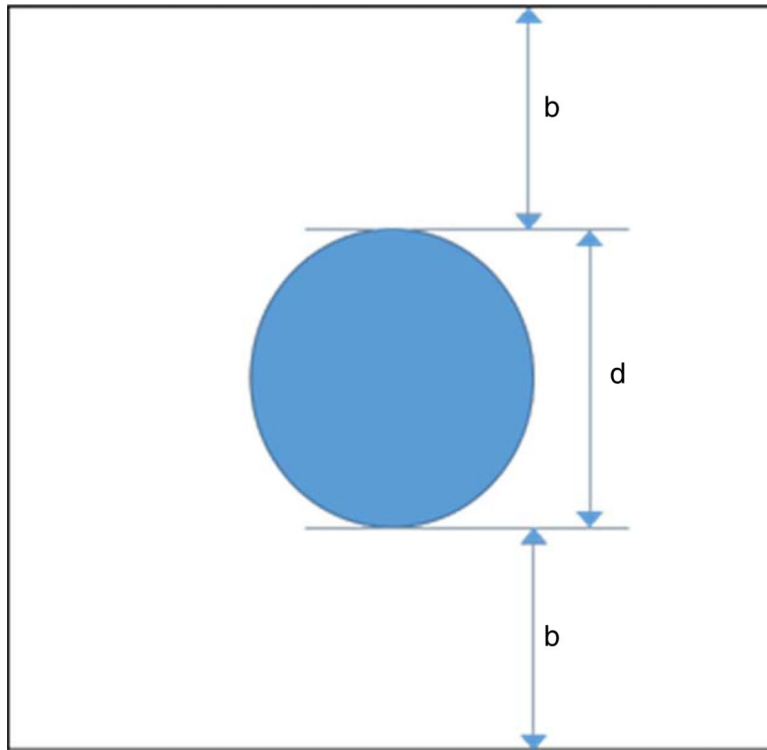The outline of our proposed method, which contains (a) INC detection and (b) INC classification.
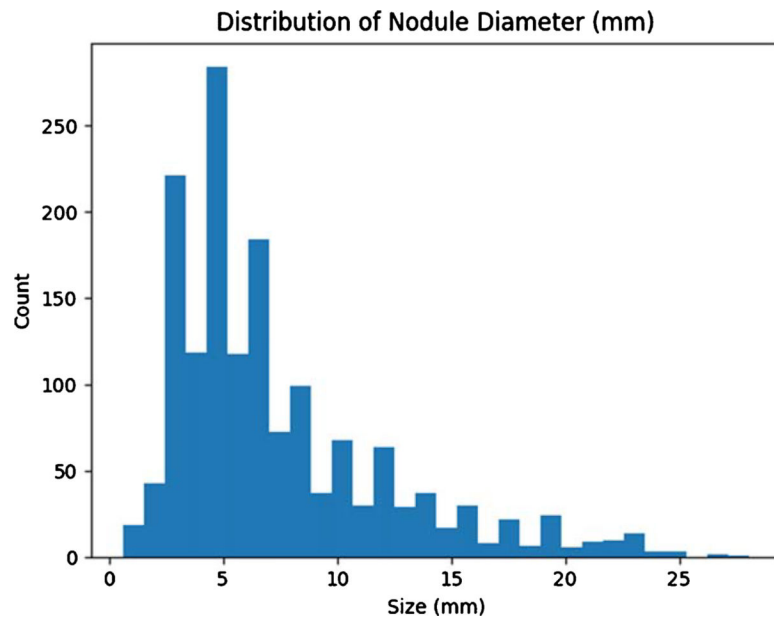
Fig. 2.
Diameter distribution of the juxta-pleural nodules.

Fig. 3.
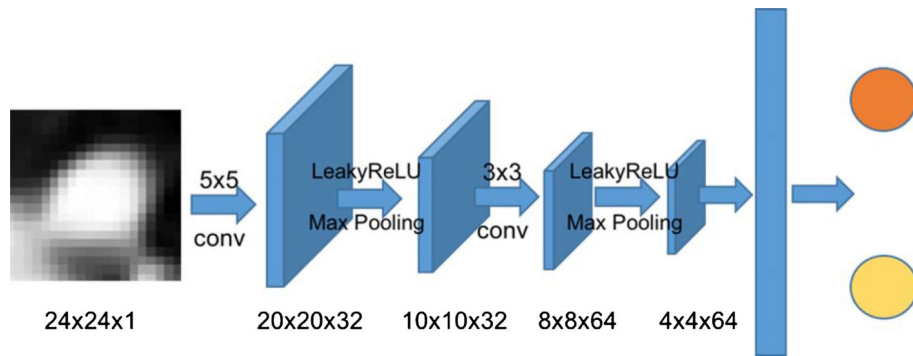Design of our patch cut, the nodule is centered. The environment is taken by a ratio considering the diameter of the nodule.

Fig. 4.
Our Basic CNN model Design, which contains two convolution layers with kernel size 5 × 5 and 3 × 3 respectively, each with LeakyReLU as activation function and followed by a Max-pooling layer. The output layer has two nodes representing the two classes: nodule and non-nodule.
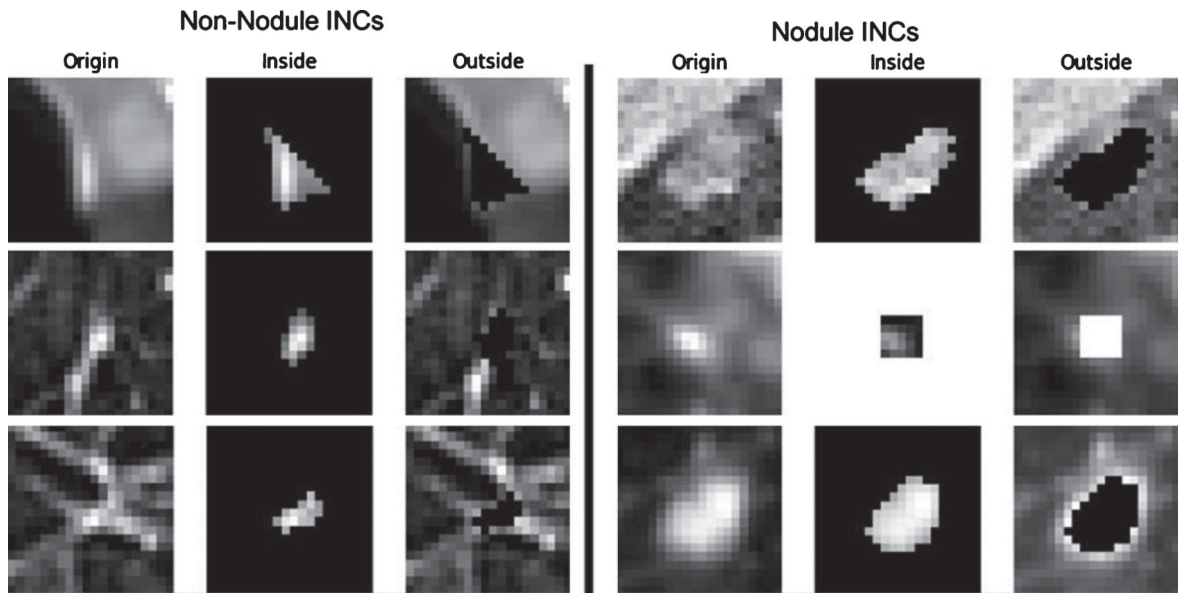
Fig. 5.
Illustration of separation of the inside and outside for both non-nodule INCs and nodule INCs.
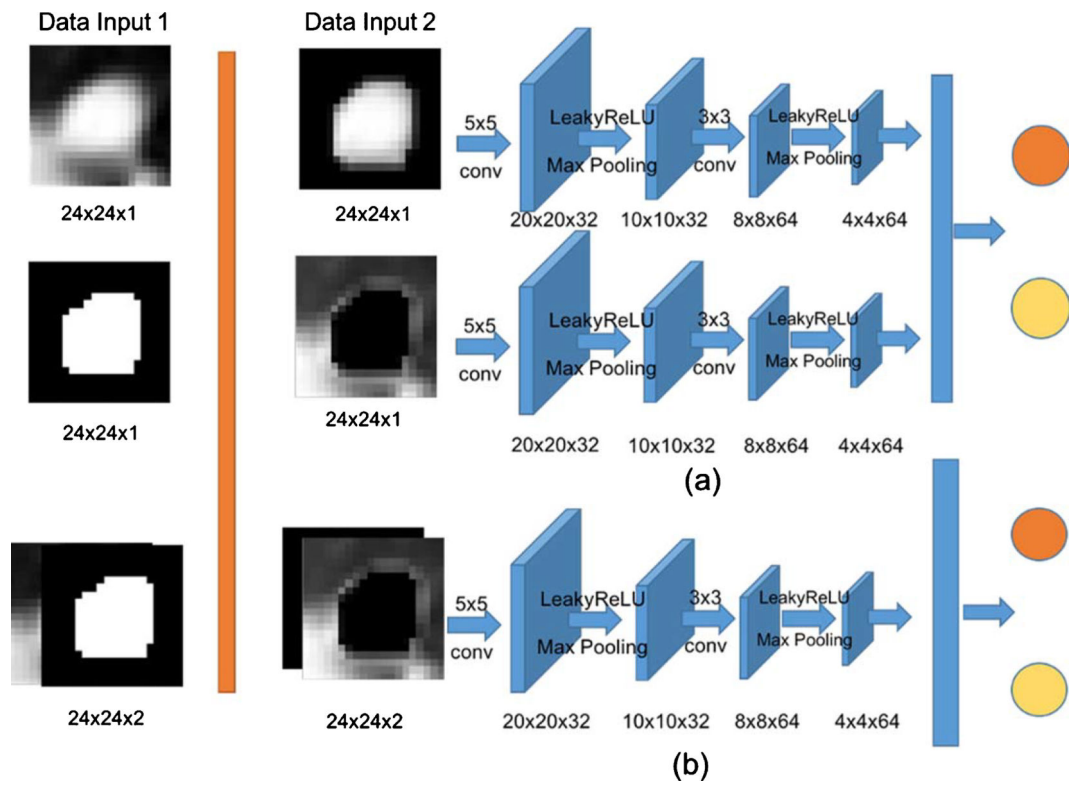
Fig. 6.
Two network designs with inside/outside INC knowledge infused. (a) The late fusion solution, where two separate inputs fused after passed through individual convolution path. (b) The early fusion solution, where two inputs first fused into a single input with two channels and then passed into the same convolution path.
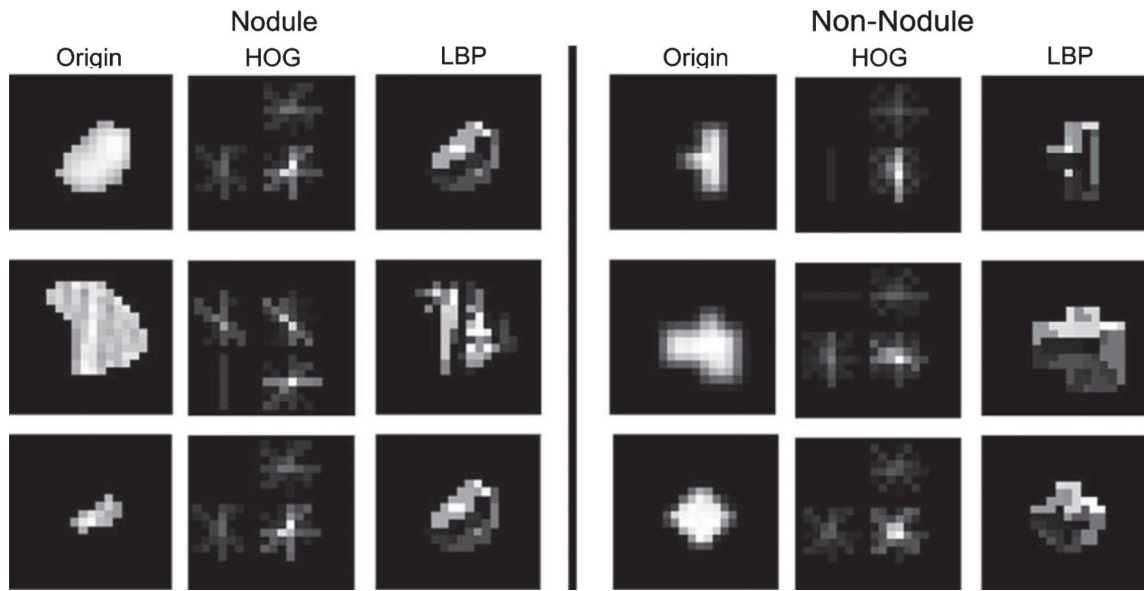
Fig. 7.
Illustration of HOG and LBP features for both non-nodule INCs and nodule INCs.
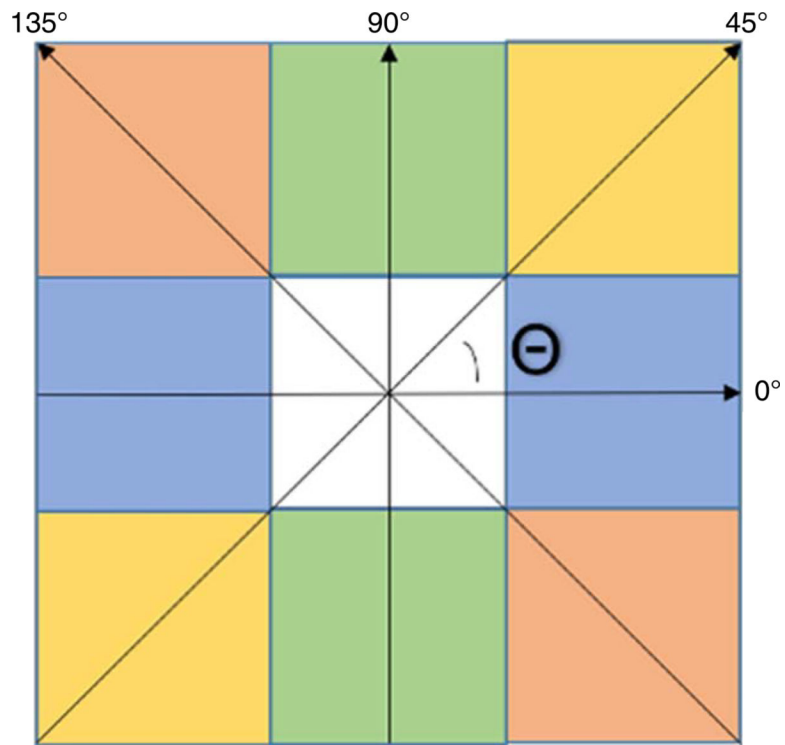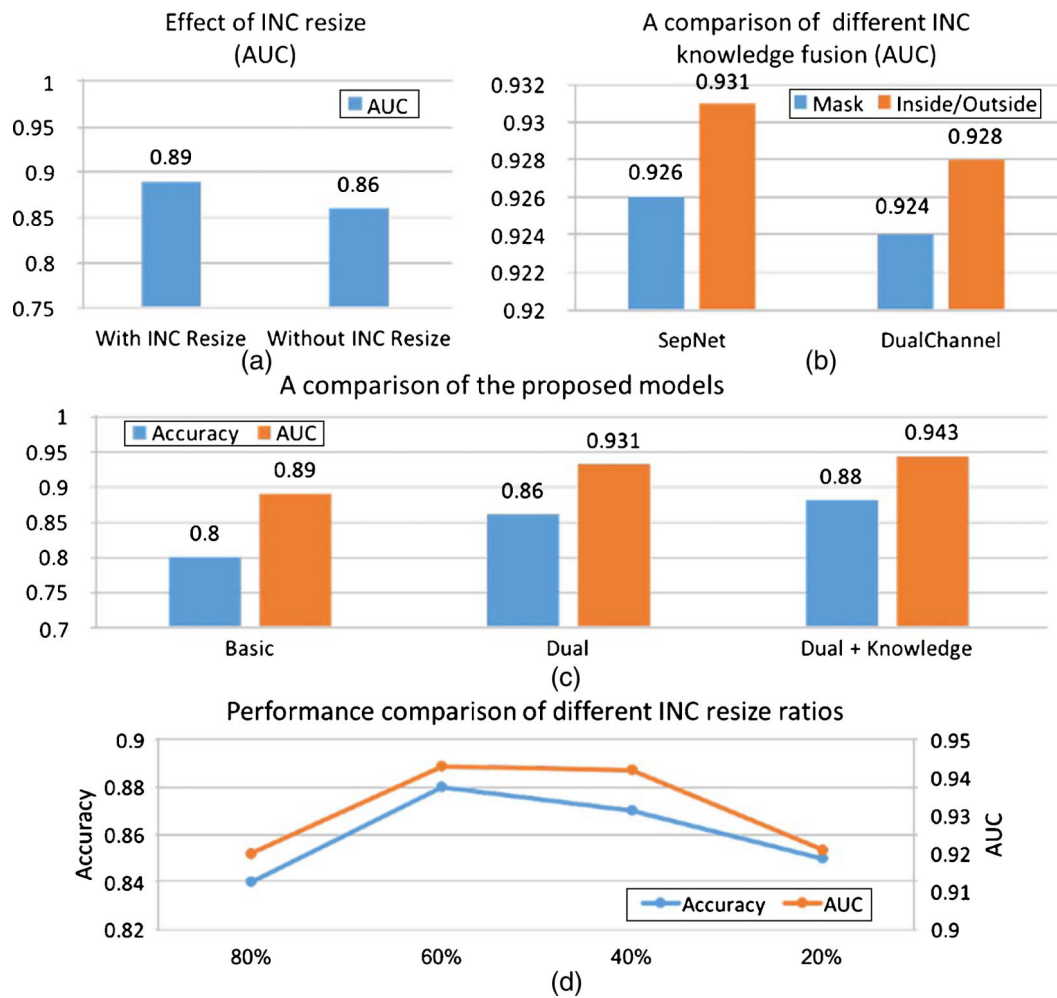
Fig. 8.
Illustration of the 2D Haralick method for extraction of texture features with image pixel size unit of d = 1 and four directions in an image slice.

Fig. 9.

Illustration of experiment results. (a) AUC measurement with INC resize vs without INC resize; (b) Four ways of INC knowledge fusion; (c) Performance comparison of three model structures; (d) Performance comparison with different proportional extraction ratios.
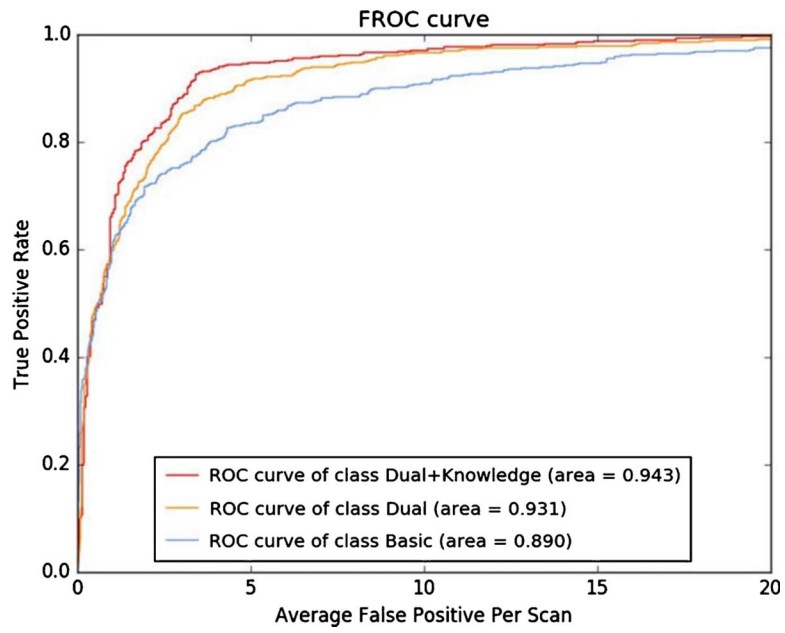
Fig. 10.
FROC curve for our final model.
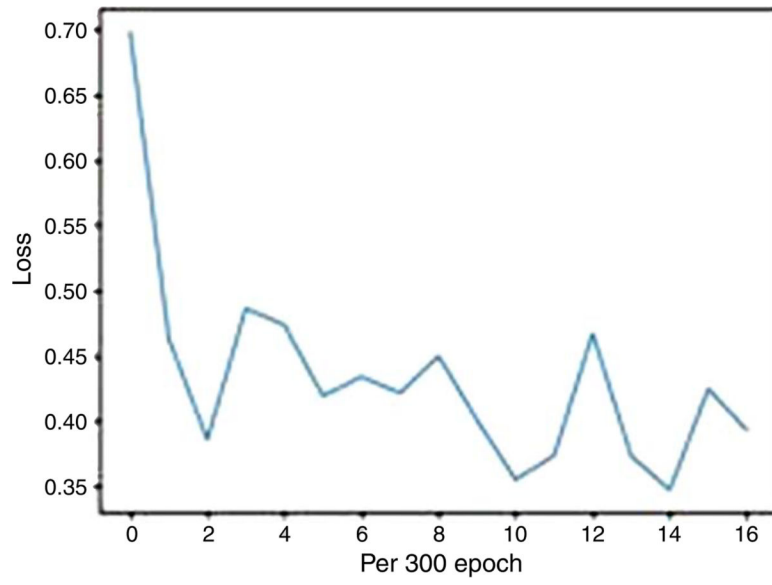
Fig.11.
Training loss of our final model.

Table 1

Our basic model design

Layer	Type	Kernel	No.
1	Convolution	5×5	32
2	Max-pooling	2×2	-
3	Convolution	3×3	64
4	Max-pooling	2×2	-
5	Fully connected	-	500
6	Fully connected	-	2

Table 2

Performance comparison of our model with the state of the art work

Researches	Sensitivity	FP/ scan	Number of Cases
M. Firmino et al. [47]	94.4%	7.04	420
M. Tan et al. [12]	87.5%	4	360
W. J. Choi et al. [13]	97.5%	6.67	84
D. M. Pena et al. [14]	94.23%	0.2	45
D. Cascio et al. [15]	88%	2.5	84
H. Han et al. [7]	89.2%	4.14	205
Teramoto A et al. [21]	90.1%	4.9	84
Setio et al. [20]	85.4%/90.1%	1/4	888
A. Teramoto et al. [48]	80%	4.2	84
Netto et al. [49]	85.93%	0.138	29
Akbarizadeh G et al. [50]	85.5%	1.99	22
Dou et al. [51]	84.8%/90.7%	1/4	888
Jiang et al. [9]	80.06%/85.06%/90.13%/94%	4.7/6.9/11.3/15.1	1006
Our work	80.10%/84.80%/90.14%/94.01%	1.89/2.65/3.175/4.01	208