

PAPER

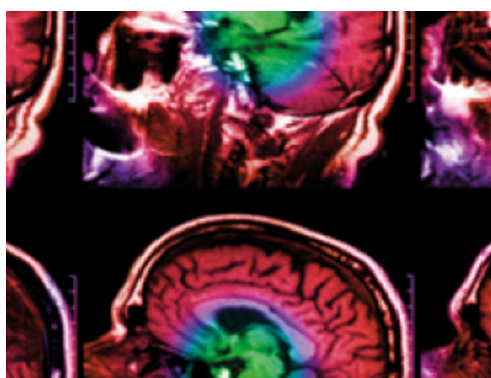
Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification

To cite this article: Yaojun Dai *et al* 2018 *Phys. Med. Biol.* **63** 245004

View the [article online](#) for updates and enhancements.

Recent citations

- [Deep Learning and Lung Cancer: AI to Extract Information Hidden in Routine CT Scans](#)
Kitt Shaffer
- [Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations](#)
Guobin Zhang *et al*
- [A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification](#)
Ying Ren *et al*



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.



PAPER

Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification

Yaojun Dai¹, Shiju Yan^{1,3}, Bin Zheng² and Chengli Song¹¹ School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai, People's Republic of China² School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, United States of America³ Author to whom any correspondence should be addressed.E-mail: 162672020@st.usst.edu.cn**Keywords:** deep learning, attribute, benign and malignant classification, pulmonary nodule

Abstract

Existing deep-learning-based pulmonary nodule classification models usually use images and benign-malignant labels as inputs for training. Image attributes of the nodules, as human-nameable high-level semantic labels, are rarely used to build a convolutional neural network (CNN). In this paper, a new method is proposed to combine the advantages of two classifications, which are pulmonary nodule benign-malignant classification and pulmonary nodule image attributes classification, into a deep learning network to improve the accuracy of pulmonary nodule classification. For this purpose, a unique 3D CNN is built to learn image attribute and benign-malignant classification simultaneously. A novel loss function is designed to balance the influence of two different kinds of classifications. The CNN is trained by a publicly available lung image database consortium (LIDC) dataset and is tested by a cross-validation method to predict the risk of a pulmonary nodule being malignant. This proposed method generates the accuracy of 91.47%, which is better than many existing models. Experimental findings show that if the CNN is built properly, the nodule attributes classification and benign-malignant classification can benefit from each other. By using nodule attribute learning as a control factor in a deep learning scheme, the accuracy of pulmonary nodule classification can be significantly improved by using a deep learning scheme.

1. Introduction

Lung cancer accounts for the highest number of mortalities among all cancers in the world. According to global cancer statistics in the year of 2000, 1.2 million people (12.3% of the total cancer patients) are diagnosed with lung cancer every year. Meanwhile, among those patients, there are 1.1 million deaths (17.8% of the total cancer mortality) caused by lung cancer each year (Parkin 2001). However, if lung cancer can be diagnosed and treated at the early stage, the patient's five-year survival rate will rise from 14% to 49% (Strauss *et al* 1997). Therefore, early diagnosis and timely treatment are critical to increase lung cancer patient's survival rate and decrease patient's mortality rate. According to Schneider's study (Schneider *et al* 2000), 90% lung cancer lesions can be found out at early stage. Usually, these lung cancer lesions are displayed as pulmonary nodules on medical images. Low-dose computed tomography (CT) is the most commonly used noninvasive method for pulmonary nodule assessment. Thus, it is crucial for doctors to classify correctly whether pulmonary nodule is benign or malignant base on CT images at diagnosis stage and further treatment stage.

Computer-aided diagnosis (CAD), which is an auxiliary scheme that can provide computational diagnostic references to doctors, has been proved to be an effective method in reducing both false positive rate and false negative rate of lung cancer diagnosis (Awai *et al* 2004, 2006, Jacobs *et al* 2015). Previous computer-aided pulmonary nodule benign-malignant classification studies usually focus on how quantitative features of pulmonary nodule image affect the accuracy of benign-malignant classification model. However, attribute as a human-nameable high-level semantic label, is rarely took into consideration in these studies. Meanwhile, attributes of a pulmonary

nodule are important for benign-malignant classification (McWilliams *et al* 2013, Patel *et al* 2013), diagnosis and further treatment (MacMahon *et al* 2005, Gould *et al* 2013, Naidich *et al* 2013). For instance, as illustrated in the diagnostic guidelines from several societies (MacMahon *et al* 2005, Travis *et al* 2011, Gould *et al* 2013, Naidich *et al* 2013), the spiculation, a kind of high-level semantic morphology feature of nodule, is important for lung cancer diagnosis. Other high-level semantic features like sphericity, calcification and lobulation, etc are recommended to be helpful references for diagnosis of lung cancer as well.

Existing pulmonary nodule benign-malignant classification CAD algorithms can be sorted in two main categories. One category classifies lung nodules with various deep neural networks based on pulmonary nodule images and benign-malignant labels. The other category uses traditional machine learning algorithm based on hand-crafted features which extracted from pulmonary nodule image. Researchers who studied both categories fail to pay enough attention to the importance of attributes in the pulmonary nodule benign-malignant classification.

The purpose of this paper is to develop a new convolutional neural network (named attribute-lung-nodule classification (ALNC)) which combines benign-malignant classification and attributes classification to train the network effectively. Instead of using singleton classification of benign or malignant, ALNC assorts attributes into several quantitative levels which can offer references to clinical usage. Benign-malignant pulmonary nodule classification focuses on general information of image, but attributes classification focus on specific details of pulmonary nodule image. CNN may fail to identify the differences between pulmonary nodule images if two images are similar in general. But ALNC can make more precise judgments when it observes specific details of the image. If the convolutional neural network is built properly, the attributes classification and benign-malignant classification have potential to benefit to both.

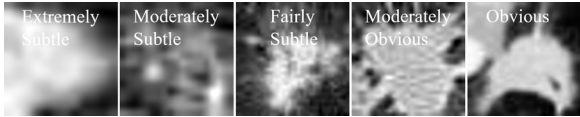
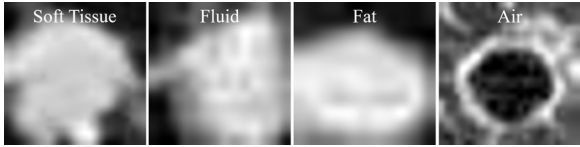
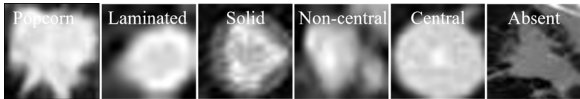
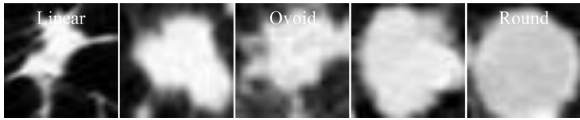
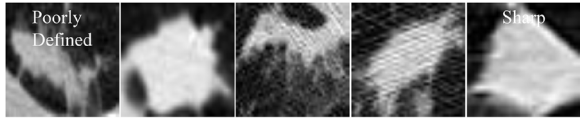
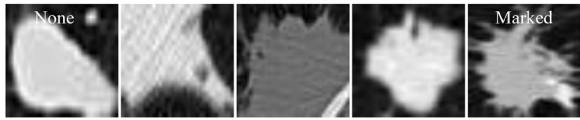


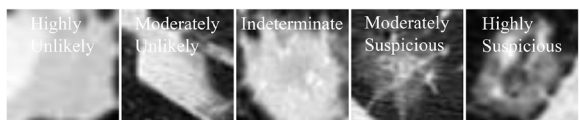
The proposed method is experimentally validated on the Lung Image Database Consortium (LIDC), which provides a public lung cancer screening thoracic CT dataset including nodule attribute level related diagnostic data. The LIDC database contains 1018 CT scans from 1010 patients at several medical centers in United States. Each thoracic CT comes with an associated XML file including annotations which were marked by four experienced radiologists with two rigorous blinded and unblinded image reading sessions. There is a total of 12 radiologists from different institutes participating in the annotation process. These XML files contain coordinates, malignancy and eight attributes quantitative scores of pulmonary nodules of which diameter larger than 3 mm. As shown in table 1, eight kinds of pulmonary nodule attributes and malignancy of nodules are annotated and rated by radiologists on several levels.

According to Naidich *et al* (2013) and Manos *et al* (2014), study of pulmonary nodule CAD should base on thin-slice thoracic CT scans. Therefore, CT scans with slice thickness larger than 3 mm are removed from this study. As a result, 888 CT scans are left in total. Among these CT scans, there are 1186 pulmonary nodules annotated by at least three radiologists. Because four radiologists may have different diagnoses on the same pulmonary nodule, the average number of malignancy scores of the same pulmonary nodule are considered as final rating score. Since the lack of biopsy and histopathology test results on these detected or annotated nodules, we divided these pulmonary nodules into two groups of high and low risk to be malignant based on the final rating scores of the radiologists. Specifically, the nodules with the rating score higher than three are placed into the high-risk group and the nodules with the rating scores lower than three are placed into the low risk group. After excluding nodules with final score equal to three which represents uncertainty, a list of 1011 pulmonary nodules are left for experiments in this study. For the same reason, levels of eight rated attributes of each pulmonary nodule are determined by the average scores (round down) from radiologists. The objective of this study is to classify these selected pulmonary nodules into two classes of being high and low risk of malignancy. To make it simple, we also defined high-risk class as malignant class and low risk class as benign class in the rest of the paper. It should be noticed that instead of being confirmed by pathohistological confirmation, the benign and malignant classification in this paper is based on subjective judgements from radiologists.

The majority of traditional computer-aided pulmonary nodule benign-malignant classification methods are based on quantitative image features extracted from pulmonary nodule CT images. For instance, Uchiyama *et al* (2003) extracted the Histogram feature from pulmonary nodule image, Farag *et al* (2011) extracted scale invariant feature transform (SIFT) feature, Sorensen *et al* (2010) extracted local binary patterns (LBP) feature, Song *et al* (2013) extracted Histogram of Oriented Gradients (HOG) feature. After extracting features, they classified pulmonary nodules based on these features by using traditional machine learning algorithms such as support vector machine (SVM) and random forest (RF). Nevertheless, the extracted features are mainly determined by the parameter setting. Thus, it is difficult for researchers to use extracted features to describe the variability of pulmonary nodules.

There is a growing trend shows that deep learning has made remarkable achievements in computer vision area. Convolutional neural network achieves great performance in understanding images because it can select and extract high level features from images automatically. Researchers in the field of computer aided diagnosis started to use deep neural network to classify pulmonary nodule. For instance, Sun *et al* (2016) used LeNet (LeCun *et al* 1990) to classify lung nodules. Hussein *et al* (2017) predicted malignancy of nodules based on

Table 1. Pulmonary nodule patterns with respect to the annotated levels in LIDC dataset.

| | | |
|--------------------|--|---|
| Subtlety |  | Example represents the degree of detection difficulty by human with five levels |
| Internal structure |  | Example represents different types of nodule constitution with four levels |
| Calcification |  | Example represents different calcification patterns of a nodule with six levels |
| Sphericity |  | Example represents the degree of roundness of a nodule in shape with five levels |
| Margin |  | Example represents the degree of how well-defined the nodule margin is with five levels |
| Lobulation |  | Example represents the degree of lobulation of a nodule in shape with five levels |
| Spiculation |  | Example represents the degree of spiculation of a nodule in shape with five levels |
| Texture |  | Example represents the solidity degree of a nodule with five levels |
| Malignancy |  | Example represents the likelihood of a nodule malignancy with five levels |

AlexNet (Krizhevsky *et al* 2012). Nibali *et al* (2017) used ResNet (He *et al* 2016) to study benign-malignant lung nodule classification problems. However, limitation of their studies is they only used the original architecture of these off-the-shelf networks and failed to design special network for pulmonary nodule benign-malignant classification. In addition, Song *et al* (2017) applied three types of deep neural networks (CNN, deep neural network (DNN) and stacked autoencoder (SAE)) for lung cancer classification. These networks are applied to the CT image classification task with some modification. Shen *et al* (2017) employed a novel multi-crop pooling strategy which crops different regions from convolutional feature maps and then applies max-pooling different times. These studies attempted to make some modifications to the network structure and basic operations of DNNs. Nevertheless, they did not take attribute information into consideration when build a neural network.

In this study, a new network architecture is put forward to study how radiologists-rated (or human-labeled) attributes affect performance of the convolutional neural network classification on benign and malignant pulmonary nodules. To the best of our knowledge, there are few related studies that have conducted thorough exploration of the combination of the rating attributes and convolutional neural network for pulmonary nodule classification. Chen *et al* (2017) exploited three multi-task learning (MTL) schemes to use features derived from deep learning models of stacked denosing autoencoder (SDAE) and CNN, as well as hand-crafted Haar-like and HOG features, for the assessment of pulmonary nodule attribute scores. Li *et al* (2017) trained a CNN regression model to predict the nodule malignancy and design a multitask learning mechanism to simultaneously share knowledge among different nodule characteristics to improve the final prediction result.

Chen *et al* (2017) focused on provide quantitative assessments over nine semantic features (including malignancy) for a pulmonary nodule depicted in the CT images. CNN is one of the methods that they applied extract features from pulmonary nodule images to. Thus, Chen *et al* trained nine 3-layer CNN models to extrated features for joint feature learning across different tasks. Then Chen *et al* (2017) employed linear regression, random forest to encode the relations of features across different tasks. Li *et al* (2017) interpreted the nodule malignancy prediction as a regression problem to predict continuous malignancy level. Li *et al* (2017) trained a 5-layer CNN regression model to learn discriminative deep features. Li *et al* (2017) updated Euclidean loss to impose inherent characteristics sharing while solving multiple correlated tasks simultaneously.

In contrast to these methods, our work focuses on improving accuracy of benign-malignant nodule classification by incorporating pulmonary nodule attributes into a deep 3D CNN. Firstly, these studies treated nine attributes (including malignancy) equally. They failed to emphasize the importance of malignancy. Therefore, in this study we design a novel loss function to highlight benign-malignant classification and balance the impact between attributes classification and benign-malignant classification. Secondly, both studies utilized 2D lung nodule images as input data, which do not contain a complete 3D information of a pulmonary nodule image. In this study, we design a 3D CNN for nodule characterization while combining 3D features to obtain improved characterization performance. Thirdly, the CNN they used is relatively shallow which cannot extract high-level features of pulmonary nodule images. In order to help CNN learn high-level information from pulmonary CT images, we need to build a 3D network deeper than existing ones. Thus, we chose 3D-ResNet-50 and 3D-DenseNet-40 (Huang *et al* 2017) as architectures of our base networks.

2. Methods

2.1. Architecture

In practice, radiologists make diagnosis by checking multiple slices of a pulmonary nodule and considering 3D information of the nodule. Most of previous methods, only use single or multi-view 2D images as input, which do not contain a complete 3D information of a lung nodule image. To eliminate the limitation, our proposed network use pulmonary nodules 3D CT cubes as input. The new network architecture is called attribute-lung-nodule classification (ALNC), which contains a basic 3D convolutional neural network, and $N + 1$ (in this paper, $N = 8$) full connected layers named as FC_1 to FC_N and FC_0 (FC_1 to FC_N represent N different full connected layers for corresponding attributes classification, and FC_0 represents full connected layer used for benign-malignant classification). Each full connected layer connects to a corresponding loss layer. N loss layers are used to calculate losses of attributes classification and the rest ones are used to calculate benign-malignant classification loss. ALNC can be applied to assort levels for N attributes of a nodule and can be used to classify benign-malignant category for each input pulmonary nodule image. 3D-DenseNet-40 and 3D-ResNet-50 are utilized separately as base convolutional neural networks in experiments.

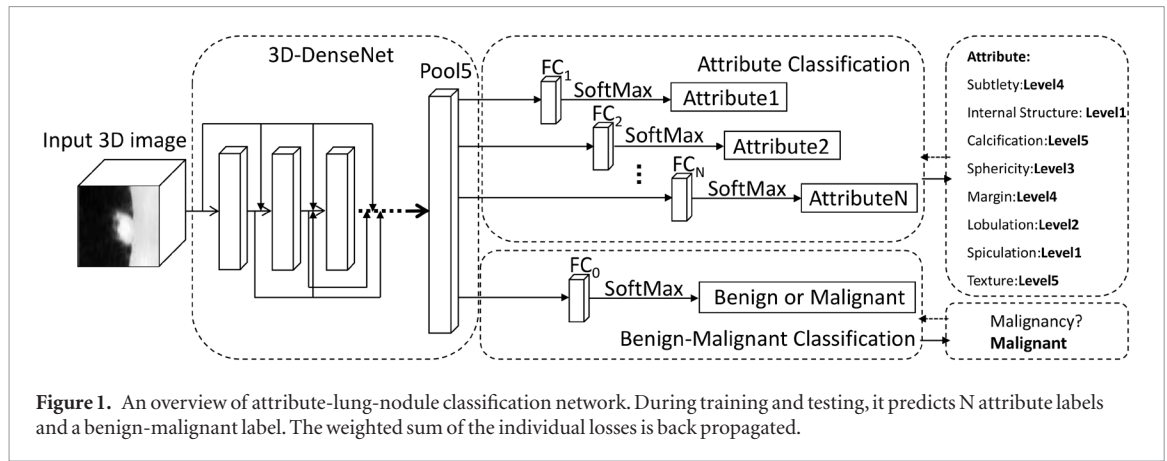
For 3D-DenseNet-40, we use standard DenseNet structure with 3D convolution and configurations ($L = 40, k = 12$). In comparison with other popular network architectures, such as AlexNet or VGGNet, ResNet with ‘pre-activation’ use fewer parameters while typically achieving better results. Hence, the setting for 3D-DenseNet-40 is kept the same as in the previous section. This means in a 3D dense block we will perform ‘batch normalization—ReLU—3D convolution—batch normalization—ReLU—3D convolution’ to extra features then concatenate with identify mapping. The 3D-DenseNet-40 used in our experiments has three 3D dense blocks that each has 13 layers. Before entering the first dense block, a 3D convolution with 16 output channels is performed on the input nodule images. For 3D convolutional layers with kernel size $3 \times 3 \times 3$, each side of the inputs is zero-padded by one pixel to keep the feature-map size fixed. We use $1 \times 1 \times 1$ convolution followed by 3D average pooling as transition layers between two contiguous dense blocks. As shown in the figure 1, at the end of the last dense block, a global average pooling (Pool5) is performed and then FC layers are connected to Pool5 layer to extract features for classification.

2.2. Loss function computation

Assume that we have n images, and each image contains N categories of attributes. Make $D_i = x_i, m_i, a_i$, which x_i represents the i th image, m_i represents benign or malignant of x_i , $a_i = \{a_i^1, \dots, a_i^N\}$ represents different levels of N attributes of x_i .

Take x for instance, the network firstly computes the descriptor v of Pool5 layer (we take 3D-DenseNet-40 as an example). Pool5 layer is connected to FC_0 layer which has output $y = [y_1, y_2]$. y_1 and y_2 are used to classify pulmonary nodule is benign or malignant, respectively. We choose SoftMax as classifier, so the prediction probability of malignancy of each category $k \in 1, 2$ is calculated as follows:

$$p(k|x) = \frac{e^{y_k}}{\sum_{i=1}^2 e^{y_i}}. \quad (1)$$



The loss function of the benign-malignant classification formulation is as follows:

$$L_{BM}(v, m) = - \sum_{k=1}^2 \log(p(k))q(k). \quad (2)$$

If the network prediction is correct, $q(k) = 1$, otherwise, $q(k) = 0$. In this case, minimizing the loss is equal to maximizing the possibility of being assigned to the ground-truth class.

Similarly, we also use N SoftMax classifiers for attributes classification. We assumed that there are l levels for a certain attribute, and the probability of assigning sample x to the attribute class $j \in 1, \dots, l$ can be calculated as:

$$p(j|x) = \frac{e^{y_j}}{\sum_{i=1}^l e^{y_i}}. \quad (3)$$

The loss function of this attributes classification can be written as follows:

$$L_{ATT}(v, m) = - \sum_{j=1}^l \log(p(j))q(j). \quad (4)$$

ALNC needs to learn multi-attributes classification and benign-malignant classification at the same time. In order to train ALNC networks properly, the final loss function is designed as:

$$L = \alpha L_{BM} + \frac{1}{N} \sum_{i=1}^N L_{ATT}. \quad (5)$$

Where L_{BM} and L_{ATT} represent loss of benign-malignant classification and attributes classification. α is a hyperparameter and determined by experiments to balance the influence of two kinds of losses. Thus, in ALNC, the evaluation criterion of the network training is determined by fusion of two loss functions which related to the benign-malignant classification and eight attributes classifications.

3. Experiments and results

3.1. Implementation details

Our implementation was realized in Python, using a custom version of the Caffe framework, which was enabled to perform volumetric convolutions via CuDNN v5. All the training and experiments were performed on a standard workstation equipped with 64 GB of memory, an Intel Xeon E5-1620 CPU working at 3.5 GHz and a NVidia GTX TITAN X with 12 GB of video memory. We let our model train for 55 epochs, or 198 K iterations. The batch size is set to 10. Learning rate is initialized to 0.0001 and changed to 0.00001 in last five epochs. The stochastic gradient descent (SGD) is implemented in each mini-batch to update the parameters.

3.2. Validation of α hyperparameter

The result of loss function indicates the error between network's prediction and ground truth. In this paper, ALNC's target is to minimize the result of loss function by modifying its parameters of network repeatedly according to gradients in the training phase. As shown in equation (5), both benign-malignant classification and attributes classification have impact on loss function. Thus, α is the key parameter to balance those effects. When using 3D-DenseNet-40 as basic network, 10 sets of cross validation experiments were conducted to choose the best parameter. As shown in figure 2(right), we performed cross validation over 1011 nodules. We randomly selected 211 lung nodule samples as test set while the other 800 nodules as training and validation set. Then we

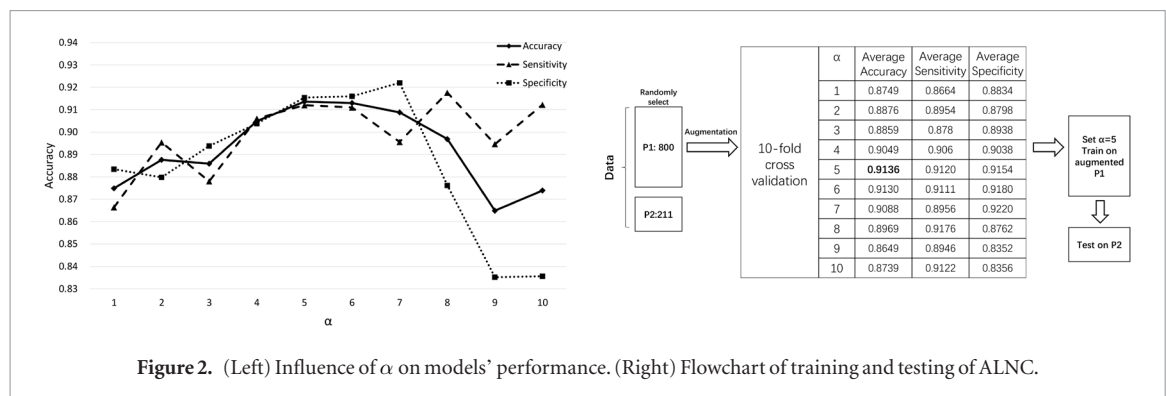


Figure 2. (Left) Influence of α on models' performance. (Right) Flowchart of training and testing of ALNC.

performed 10-fold cross validation over training and validation set (P1) to choose the best hyperparameter. It is difficult to train the parameters adequately because the train set is still relatively small if we train network from the start. We perform data augmentation to yield 49 extra samples corresponding to each sample in the training data. Data augmentation is performed by using salt and pepper noise, Gaussian noise, mean filtering, median filtering, Gaussian filtering and random rotation. We use the same amount of benign and malignant nodules to keep the network training process balanced and bias-free. As shown in figure 2(left), the accuracy raises at first and falls down after. When α equals to five, ALNC could obtain relatively high accuracy, sensitivity and specificity. Therefore, α is set as five when using 3D-DenseNet-40 as basic network.

3.3. Influence of attribute on accuracy of benign-malignant classification

In order to study the influence of attributes on the accuracy of ALNC benign-malignant classification, a total of eight experiments were conducted. An attribute is removed from training for each experiment. The results are shown in figure 3. Each column in the graph represents the accuracy of benign-malignant classification when certain attribute label is removed. The experiments show that every attribute can improve the benign-malignant classification of the model, removal of any attribute label will decrease the accuracy benign-malignant classification of the model ('3D-D', '3D-R' represent using 3D-DenseNet-40 or 3D-ResNet-50 as basic network).

3.4. Benign-malignant pulmonary nodule classification

Song *et al* (2017) designed three types of deep neural networks (CNN, deep neural network (DNN), and stacked autoencoder (SAE)) for lung cancer calcification. Shen *et al* (2017) cropped and pooled feature maps twice in different scales then performed concatenation operation instead of performing original max-pool operation directly. Shen *et al* used this method to study lung nodule benign-malignant classification problem. However, Shen *et al* (2017) and Song *et al* (2017) chose relatively simple networks that are less than 10 layers. Both of them failed to take attributes into consideration. It should be noticed that Song *et al* and Chen *et al* use different sets of included nodules. However, we all performed cross-validation to ensure our experiments results valid and logical. The major difference between above works and ALNC is that we build a deeper network and design a special loss function. Moreover, 3D ALNC performs a 3D convolution on nodule data which can extra high-level features than Shen *et al* (2017) and Song *et al* (2017). Results prove that 3D convolution is a better method to extra nodule features than multi-crop or simply 2D convolution. ALNC combines the advantages of attribute learning and deep learning. As shown in table 2, the accuracy of ALNC classification for benign and malignant pulmonary nodule classification is higher than their algorithms reported in the literatures. ALNC also retains the balance between sensitivity and specificity, which means that it is equally good at classifying both malignant and benign nodules. The ROC curves in figure 4 prove that ALNC demonstrates the best performance in terms of both sensitivity and specificity across a wide range of threshold values. As a conclusion, above experiments proved that incorporating attributes into deep learning network significantly improved the accuracy of benign-malignant pulmonary nodule classification.

3.5. Pulmonary nodule attributes classification

To improve the final prediction result, Li *et al* (2017) trained a CNN regression model to predict the nodule malignancy and designed a multitask learning mechanism to simultaneously share knowledge among different nodule characteristics. However, the network that Li *et al* (2017) designed has only five layers and is a two-dimensional CNN. Besides, Li *et al* (2017) did not try to balance attributes classification and benign-malignant classification. As shown in table 3, ALNC(3D-D) achieved good performance in attributes classification. The accuracies of eight attributes are all higher than 83%, which better than the attributes classification results reported in Li *et al* (2017). The result of experiment conducted showed that the architecture of ALNC can

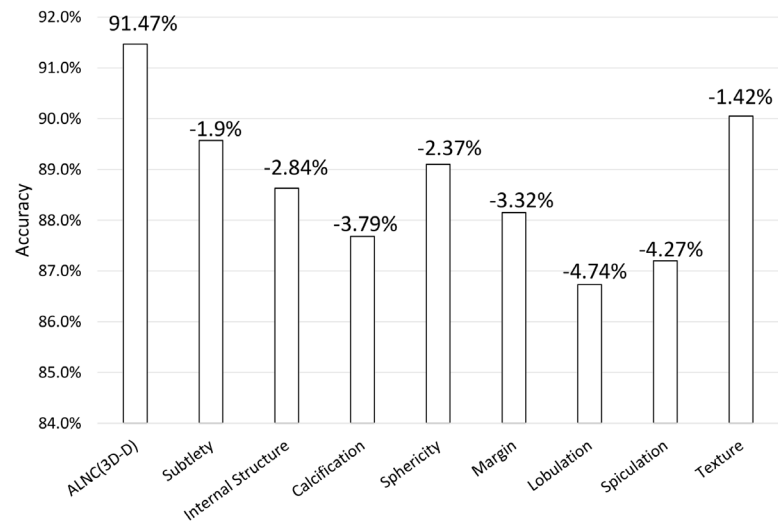


Figure 3. Benign-malignant classification accuracy on test set. We remove one attribute from the system at a time. Accuracy changes are indicated above the bars.

Table 2. Comparison between models' performance of different methods.

| Methods | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC |
|--------------------------|--------------|-----------------|-----------------|-------|
| Song <i>et al</i> (2017) | 84.15 | 83.96 | 84.32 | 0.916 |
| Shen <i>et al</i> (2017) | 87.14 | 77.00 | 93.00 | 0.930 |
| ALNC(3D-R) | 89.57 | 89.32 | 89.81 | 0.932 |
| ALNC(3D-D) | 91.47 | 91.26 | 91.67 | 0.969 |

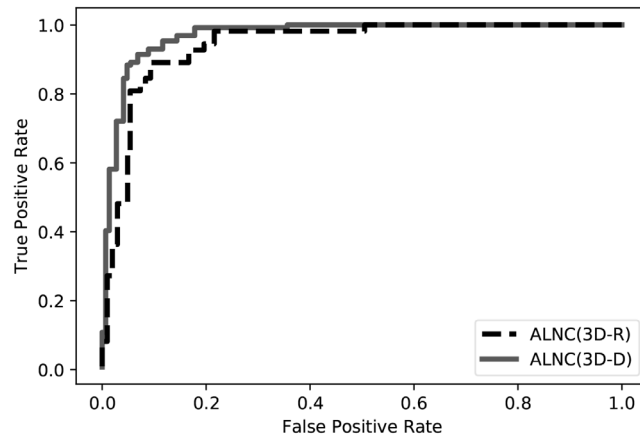
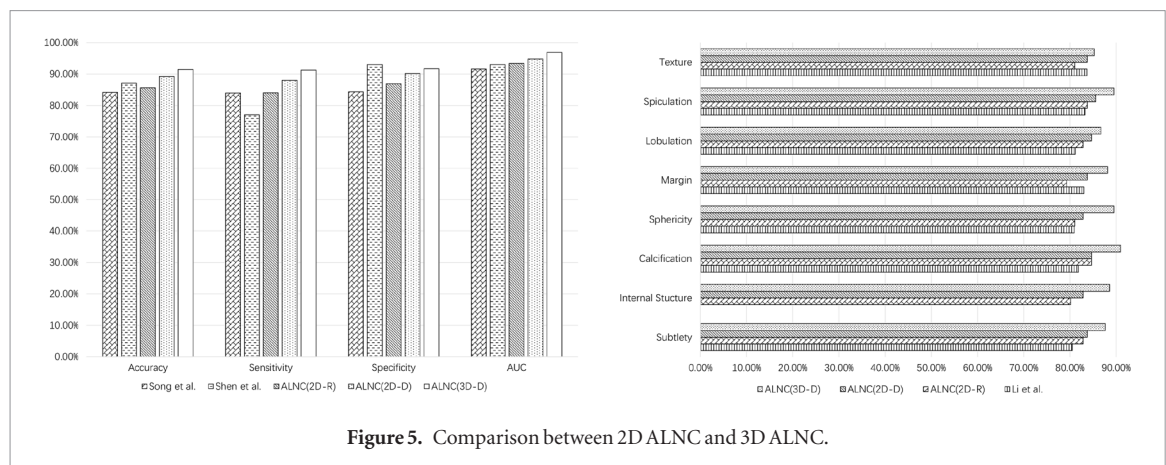


Figure 4. The receiver operating characteristic curve (ROC curve) of ALNC based on different basic models.

Table 3. Comparison between different methods about eight attributes classification.

| Methods | Li <i>et al</i> (2017) | ALNC(3D-R) | ALNC(3D-D) |
|------------------------|------------------------|------------|------------|
| Subtlety (%) | 80.5 | 86.73 | 87.68 |
| Internal structure (%) | — | 87.20 | 88.63 |
| Calcification (%) | 81.88 | 88.63 | 91.00 |
| Sphericity (%) | 80.96 | 83.89 | 89.57 |
| Margin (%) | 83.03 | 81.99 | 88.15 |
| Lobulaiton (%) | 81.19 | 85.78 | 86.73 |
| Spiculation (%) | 83.26 | 86.26 | 89.57 |
| Texture (%) | 83.72 | 84.36 | 85.31 |



combine two kinds of classifications more successfully than Li *et al* (2017), and can effectively improve the accuracy of attributes classification.

3.6. Comparison between 2D ALNC and 3D ALNC

By design, ALNC can improve accuracy of benign-malignant nodule classification by incorporating pulmonary nodule attributes. We conduct an experiment to investigate if 2D ALNC takes advantage of the attribute information. We designed a two-dimensional ALNC network structure to conduct experiments. To ensure a fair comparison between the two networks, we eliminate factors such as differences in dataset partitioning and data augmentation by using the exactly the same data and training configure. We choose 2D-DenseNet and 2D-ResNet as base network and optimize 2D ALNC by adjusting hyperparameters.

The input of convolutional neural network is two-dimensional image, but pulmonary nodule itself is three-dimensional. Therefore, in this paper we use the median intensity projection (MIP) (Hussein *et al* 2017) method, which performs the median intensity projection on the external cube of three-dimensional pulmonary nodule along three axes of x , y and z .

The three two-dimensional median intensity projection images are concatenated to a three-dimensional tensor. The three-dimensional tensor can be considered as a colorful two-dimensional image with three channels corresponding to Red, Blue, Green (RGB) channel, which can be put into a convolutional neural network to train the model. After performing median intensity projection, we get final training RGB samples of pulmonary nodules.

2D-ALNC takes the RGB images of lung nodules as input of network. And we use 2D-ResNet and 2D-DenseNet as basic networks. The same method as above are performed to select hyperparameter α for network training. Figure 5 shows two bar graphs for testing results. From the figure we can see that 3D-ALNC performs better than 2D-ALNC, while 2D-ALNC can still achieve better results than previous studies. The results of experiments prove that ALNC's structure has the ability to improve the performance of the model by incorporating attributes learning into a convolutional neural network. Interestingly, 2D-ALNC used pretrained models but 3D-ALNC did not, and 3D-ALNC still achieves better results than 2D-ALNC. We speculate that the reason is 3D input contains much more information than that in 2D input. The information can be used to extract features of lung nodules from more three-dimensional angles. Thus the model can 'learn' attributes classification and benign-malignant classification better.

4. Discussion

Although a number of computer-aided schemes for pulmonary nodule classification have been developed by using deep-learning-based algorithms and reported in the literatures, this study has a number of advantages and unique characteristics. Firstly, we proposed a unique ALNC classification model. According to the study results, ALNC demonstrated its capability of improving the accuracy of lung nodule benign-malignant classification no matter in 2D or 3D network. The decisive reason for ALNC to yield the superior results is that it effectively combines attributes classification and benign-malignant classification. The CT image of pulmonary nodule contains pixel correlation information of lesion, it can be also used to extract several kinds of lesion image features for lung nodule classification as do by radiologists in reading and interpreting pulmonary CT images among the clinical practice. In addition, the nodule attributes are high-level semantic representations of the nodules which annotated by experimental radiologists. It can provide important visual reference or information for radiologists to predict the possibility of the nodule being malignant. ALNC aims to take the full advantage

of nodule attributes, benign-malignant label and CT image data to raise the benign-malignant classification accuracy.

Secondly, unlike previous studies that select a region of interest (ROI) from one CT slice as input of the CNN based deep learning algorithm, we considered the variation between three-dimensional lung nodule existence and two-dimensional input of the CNN. In this paper, we utilized 3D networks to classify pulmonary nodules in a CT image into benign or malignant categories. Compared to those algorithms which uses 2D slices or approximate 3D image with multi-views, ALNC directly works on 3D volumes yields better results for the lung nodule classification problem when the slice thickness is consistent.

Thirdly, although previous studies (Chen *et al* 2017, Li *et al* 2017) have already tried to add attribute information into lung nodule benign-malignant classification, they did not solve the problem of maintaining the balance between benign-malignant classification and attributes classification in convolutional neural network. The results of this study support our hypotheses of why the ALNC outperforms the existing networks that was used in previous studies as follows:

1. Compared with traditional CNN like LeNet, AlexNet, VGGNet (Simonyan and Zisserman 2014), ALNC has deeper architecture, so it can obtain higher representational power for pulmonary nodule benign-malignant classification.
2. Previous works failed to design a proper loss function to explore the impact of attributes classification on benign-malignant classification. We establish a special loss function with a hyparparameter α to archive balance between attributes classification and benign-malignant.
3. Instead of using 2D or multi-view images as input of CNN, we feed ALNC with 3D images to get overall information of pulmonary nodules.

5. Conclusions

Nowadays, the deep-learning based computer aided classification of benign and malignant pulmonary nodules is mainly focus on the images and benign-malignant labels. As human-nameable high-level semantic label, nodule attributes can also help improve classification of pulmonary nodules to some extent. Nevertheless, these two sources of information have not been combined effectively in existing methods.

In order to solve this problem, we propose and test a new method to combine nodule attribute learning with the deep learning of nodule classification. In the aspect of data preprocessing, we chose to use pulmonary nodules 3D CT cubes as input samples of the network. In terms of network architecture, we propose a ALNC network structure, and design a loss function to balance the effect of two kinds of classifications of network training. The purpose of this study is not only to improve the classification accuracy of benign and malignant pulmonary nodules, but also to enhance the attributes classification ability. Experiments results show that the accuracy of attributes classification and accuracy of benign-malignant classification of the pulmonary nodules obtained from ALNC network are higher than those obtained from the existing networks.

Meantime, we should keep in mind that medical image is only a part of information of certain disease, even well-trained radiologists cannot confirm the diagnosis of some disease merely depend on medical images. It is not enough for researchers to study the correlation between images and benign-malignant labels. After doctors' annotation, attributes can contain large amount of valuable information about patients' diseases. By incorporating attribute information into deep learning successfully, ALNC has a promising prospect on helping doctors make diagnosis based on medical image dataset with a large number of high quality attribute labels. Theoretically, ALNC can be widely applied to improve the diagnosis accuracy of other diseases based on corresponding medial images and attributes datasets as well. However, the structure of ALNC and its loss function still need to be improved. In this study, we only take quantitative visual attributes into consideration, more valuable non-visual attribute labels like age, gender, genomics data, family history and histological information can also be incorporated into network training to further improve classification performance.

ORCID iDs

Yaojun Dai  <https://orcid.org/0000-0002-9782-2347>

References

- Awai K, Murao K, Ozawa A, Komi M, Hayakawa H, Hori S and Nishimura Y 2004 Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists detection performance *Radiology* **230** 347–52

- Awai K, Murao K, Ozawa A, Nakayama Y, Nakaura T, Liu D, Kawanaka K, Funama Y, Morishita S and Yamashita Y 2006 Pulmonary nodules: estimation of malignancy at thin-section helical CT: effect of computer-aided diagnosis on performance of radiologists *Radiology* **239** 276–84
- Chen S, Qin J, Ji X, Lei B, Wang T, Ni D and Cheng J Z 2017 Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images *IEEE Trans. Med. Imaging* **36** 802–14
- Farag A, Ali A, Graham J, Farag A, Elshazly S and Falk R 2011 Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest *IEEE Int. Symp. on Biomedical Imaging: from Nano to Macro* (IEEE) pp 169–72
- Gould M K, Donington J, Lynch W R, Mazzone P J, Midthun D E, Naidich D P and Wiener R S 2013 Evaluation of individuals with pulmonary nodules: when is it lung cancer?: diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines *CHEST J.* **143** e93S–120S
- He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- Huang G, Liu Z, van der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* vol 1 p 3
- Hussein S, Gillies R, Cao K, Song Q and Bagci U 2017 Tumornet: lung nodule characterization using multi-view convolutional neural network with gaussian process *IEEE Int. Symp. on Biomedical Imaging* pp 1007–10
- Jacobs C, van Rikxoort E M, Scholten E T, de Jong P A, Prokop M, Schaefer-Prokop C and van Ginneken B 2015 Solid, part-solid, or non-solid?: classification of pulmonary nodules in low-dose chest computed tomography by a computer-aided diagnosis system *Investigative Radiol.* **50** 168–73
- Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* pp 1097–105
- LeCun Y, Boser B E, Denker J S, Henderson D, Howard R E, Hubbard W E and Jackel L D 1990 Handwritten digit recognition with a back-propagation network *Advances in Neural Information Processing Systems* pp 396–404
- Li X, Kao Y, Shen W, Li X and Xie G 2017 Lung nodule malignancy prediction using multi-task convolutional neural network *SPIE Medical Imaging* vol 10134 (Bellingham, WA: SPIE) p 1013424
- MacMahon H, Austin J H, Gamsu G, Herold C J, Jett J R, Naidich D P, Patz E F Jr and Swensen S J 2005 Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner society *Radiology* **237** 395–400
- Manos D, Seely J M, Taylor J, Borgaonkar J, Roberts H C and Mayo J R 2014 The lung reporting and data system (LU-RADS): a proposal for computed tomography screening *Can. Assoc. Radiol. J.* **65** 121–34
- McWilliams A et al 2013 Probability of cancer in pulmonary nodules detected on first screening CT *New Engl. J. Med.* **369** 910–9
- Naidich D P et al 2013 Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner society *Radiology* **266** 304–17
- Nibali A, He Z and Wollersheim D 2017 Pulmonary nodule classification with deep residual networks *Int. J. Comput. Assist. Radiol. Surg.* **12** 1799–808
- Parkin D M 2001 Global cancer statistics in the year 2000 *Lancet Oncol.* **2** 533–43
- Patel V K, Naik S K, Naidich D P, Travis W D, Weingarten J A, Lazzaro R, Gutterman D D, Wentowski C, Grosu H B and Raoof S 2013 A practical algorithmic approach to the diagnosis and management of solitary pulmonary nodules: part 2: pretest probability and algorithm *CHEST J.* **143** 840–6
- Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions *Phys. Med. Biol.* **45** 459
- Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y and Tian J 2017 Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification *Pattern Recognit.* **61** 663–73
- Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *Proc. Int. Conf. on Learning Representations* (arXiv:1409.1556)
- Song Q, Zhao L, Luo X and Dou X 2017 Using deep learning for classification of lung nodules on computed tomography images *J. Healthc. Eng.* **2017** 8314740
- Song Y, Cai W, Zhou Y and Feng D D 2013 Feature-based image patch approximation for lung tissue classification *IEEE Trans. Med. Imaging* **32** 797–808
- Sorensen L, Shaker S B and De Bruijne M 2010 Quantitative analysis of pulmonary emphysema using local binary patterns *IEEE Trans. Med. Imaging* **29** 559–69
- Strauss G M, Gleason R E and Sugarbaker D J 1997 Screening for lung cancer: another look; a different view *Chest* **111** 754–68
- Sun W, Zheng B and Qian W 2016 Computer aided lung cancer diagnosis with deep learning algorithms *SPIE Medical Imaging* vol 9785 (Bellingham, WA: SPIE) p 97850Z
- Travis W D et al 2011 International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma *J. Thoracic Oncol.* **6** 244–85
- Uchiyama Y et al 2003 Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography *Med. Phys.* **30** 2440–54