

“© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Using Multi-level Convolutional Neural Network for Classification of Lung Nodules on CT images

Juan Lyu, Sai Ho Ling, *Senior Member, IEEE*

Abstract— Lung cancer is one of the four major cancers in the world. Accurate diagnosing of lung cancer in the early stage plays an important role to increase the survival rate. Computed Tomography (CT) is an effective method to help the doctor to detect the lung cancer. In this paper, we developed a multi-level convolutional neural network (ML-CNN) to investigate the problem of lung nodule malignancy classification. ML-CNN consists of three CNNs for extracting multi-scale features in lung nodule CT images. Furthermore, we flatten the output of the last pooling layer into a one-dimensional vector for every level and then concatenate them. This strategy can help to improve the performance of our model. The ML-CNN is applied to ternary classification of lung nodules (benign, indeterminate and malignant lung nodules). The experimental results show that our ML-CNN achieves 84.81% accuracy without any additional hand-craft preprocessing algorithm. It is also indicated that our model achieves the best result in ternary classification.

I. INTRODUCTION

According to the American Cancer Society, colon and rectum, breast, prostate, and lung and bronchus are the four major cancers [1]. Lung cancer is one of the major cancer and the estimated number of death for lung and bronchus cancers is 155,870 in 2017 (accounting for about 26% of deaths of four major cancers). Moreover, the five-year survival rate of lung cancer is only 18%, which is one of the lowest cancers for years 2006 to 2012. As we all know, the early diagnosis can improve the chance of cure and five-year survival rate, however, the clinical symptoms of lung cancer usually present on the advanced stage.

Generally, Computed Tomography (CT) is a main and most effective method to diagnose the lung nodule in the early stage. While the interpretations of CT images are mainly carried out by physicians, however, they are limited by the human's subjectivity, large variations across interpreters, and fatigue [2]. Therefore, use of Computer Aided Diagnosis (CAD) systems is an effective method to assist the expert's workflow for the early diagnosis and interventions.

Inspired by the tremendous success of deep learning in computer vision [3-5], more and more researchers study how to diagnose and classify lung cancers with deep learning models, especially convolutional neural networks (CNNs) [6-8]. It is proved that CNN has shown a far superior performance to hand-craft methods [4], such as sparse

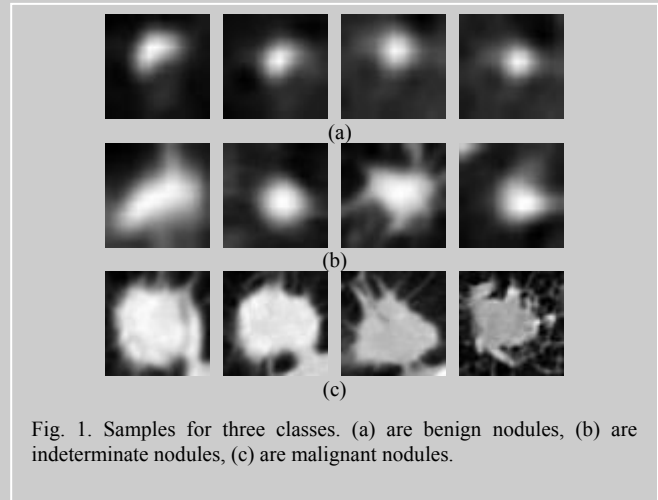


Fig. 1. Samples for three classes. (a) are benign nodules, (b) are indeterminate nodules, (c) are malignant nodules.

coding [9] and scale-invariant feature transform plus Fisher Vectors [10] (SIFT+FVs). In [11], they tested the feasibility of deep structure algorithms in lung cancer diagnosis by comparing the performance of three classical deep learning algorithms including convolutional neural network (CNN), deep belief networks (DBNs) and stacked denoising autoencoder (SDAE), with traditional CAD. The binary classification accuracies of those deep learning models are 79.76%, 81.19% and 79.29%, respectively. However, they just used three basic deep learning structures to train the input data with few parameters, while the input data include 114,728 nodules, so it has lower representation ability for input data.

In 2015, Shen et al. [12] proposed a multi-scale CNN (MCNN) consisting of three CNNs to train three scales of input and achieved a higher binary classification accuracy of 86.84%. Its resulting output is the concatenation of the outputs of three fully-connected layers, which was directly fed to the final classifier. However, the fully-connected layer can largely increase the training parameter and the complexity of the model is significantly increased. Thus, the ability of extracting effective features is limited.

In 2017, they extended their research to solve the lung nodules classification which includes uncertain nodules and presented a multi-crop CNN (MC-CNN) [13] with a new strategy based multi-crop pooling strategy. This method crops the convolutional feature maps twice by center crop with a quarter size in each pooling layer, then applies max-pooling different times. However, it can lead to parts of useful information loss so that there was an impact on extracting effective features, especially for the larger nodules. Finally, the binary classification accuracy is 87.14%.

In classification of lung nodules, ternary classification is

Juan Lyu is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001, China (e-mail: Juanlyu91@gmail.com).

Sai Ho Ling is with the School of Biomedical Engineering, University of Technology Sydney, Ultimo, NSW, Australia (e-mail: Steve.Ling@uts.edu.au).

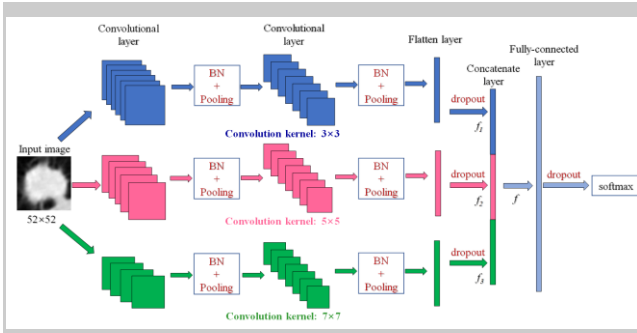


Fig. 2. The architecture of proposed ML-CNN for lung nodule malignancy classification. Each color represents one level of CNN, where are processing by different convolutional kernels. The number of squares in the first convolutional layer of each level is different and corresponds to the number of feature maps produced. Meanwhile, same numbers of feature maps are produced in the second convolutional layer of each level.

more comprehensive and realistic than binary classification, because it includes all kinds of nodules. Fig. 1 shows the samples of lung nodules for each class. In [13], ternary classification is also used and the accuracy is only 62.46%.

Due to the variety of nodules sizes and morphologies, multi-scale features of input are attractive. To ensure useful information is not missed and reduce the computational complexity, we propose a multi-level convolutional neural network (ML-CNN) to enhance the performance of lung nodule malignancy classification in thoracic CT images. We classify the nodules into three classes: benign, indeterminate and malignant. Our approach is evaluated on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database [14].

The ML-CNN is based on three main ideas. First, to tackle the variations of nodule size and morphology, we design three levels of CNNs where each level has the same structure and same numbers of feature maps in the second convolutional layers, but they have different convolutional kernels. This idea is aimed to extract multi-scale features of input effectively. Second, unlike [12], we flattened all feature maps of every level and concatenated them to a one-dimensional vector, then connected with a fully-connected layer. This strategy can help to improve the classification accuracy of our model. Third, we add dropout technique between the flatten and concatenate layer to reduce the influence of overfitting. Experimentally, when the dropout parameter (keep) is 0.4, the model can achieve a better performance. This strategy guaranteed all scales of features can contribute to the classification result. By designing the model in this way, we can use the extracted multi-scale features to enhance the classification result.

II. METHODS

A. Convolutional Neural Networks

CNNs are feed-forward neural networks inspired biologically by the human visual system [15], which can extract multi-level hierarchies of features of input [16]. CNNs are typically involving four types of layers: convolutional, activation, pooling and fully-connected layers [17].

In convolutional layer, like human visual system, the

connections between neurons and input are sparse, different neurons respond to different local areas of the input [6]. In addition, the neurons of each map share the same weights, this mechanism exponentially reduces the numbers of parameters of model and improves its efficiency. The convolutional function between the input and output of convolutional layer is defined by:

$$y_l = \sum_i x_l * w + b_l, \quad (1)$$

where x_l represents the input of the l -th layer, it could be the input image or feature map, w denotes the parameters of filter and b_l denotes the bias scalar, y_l denotes the output of the l -th layer, $*$ is the convolution operation.

Followed by the convolutional operation, a non-linear activation layer is formed to perform a non-linear transformation from input to output space. In this paper, we adopt the rectified linear unit (ReLU) [18] as the activation function and it can be expressed as:

$$z = \max(0, y), \quad (2)$$

where y denotes the input of activation function as well as the convolution output y_l in (1), z denotes the output of activation function.

Following the convolutional layer, there is a max-pooling layer, which applies the designed operator to merge nearby feature values into the maximum one. It is formulated as:

$$u_{(i,j)} = \max \{ z_{(m,n)} : i \cdot s \leq m < i \cdot s + p, j \cdot s \leq n < j \cdot s + p \}, \quad (3)$$

where p denotes the pooling size, s denotes the stride of pooling. $u_{(i,j)}$ is the neuron at position (i, j) in the output map, $z_{(m,n)}$ presents the neuron at position (m, n) in the input map. The reason we choose the max-pooling is due to its translation invariability, which is especially efficient to deal with not well-aligned nodule images.

Moreover, one fully-connected layer is put in place followed by a softmax classifier. The training of CNNs is performed similarly to that of other artificial neural networks (ANNs), by minimizing a loss function using stochastic gradient descent [16] (SGD) based methods and back propagation [19] of the error.

B. Multi-level Convolutional Neural Network

As mentioned above, lung nodules have different sizes and various morphologies, in order to extract effective features of each lung nodule, the multi-scale convolution strategy [20] is used. Due to this reason, we propose a new CNN structure named multi-level CNN (ML-CNN) and shown in Fig. 2. We can see in this Figure that the structure of ML-CNN includes three parallel levels of CNNs, while unlike MCNN, fully-connected are not included layer in these CNNs. The motivation is that if we use multi-scale convolution strategy in a single level CNN, for example, a CNN contains 3 convolutional layers with 3 kinds of convolutional kernel sizes, each convolutional layer only can extract one scale features. Consequently, we design a

multi-level CNN to extract multi-scale features in every convolutional layer.

For ML-CNN, there are two convolution layers followed by batch normalization (BN) [21] and pooling layers. BN is used after the convolution operation and before the activation operation. It is used to reduce the internal covariate shift. The problem is formally known as covariate shift when the distribution of network activations changes between training and production stages.

In ML-CNN, we have 3 levels and they have same structures and same number of feature maps in the last convolution step. However, their convolutional kernels are different.

We designed 3 kinds of convolutional kernels to improve the ability of extracting features, they are 3×3 , 5×5 and 7×7 , respectively. Meanwhile, each level is using the same kernel size. After the second pooling layer, 60 feature maps are produced in every level. In our approach, the strategy of flatten first, concatenated second is introduced. The output of the concatenate layer f is defined as:

$$f = [f_1 \quad f_2 \quad f_3], \quad (4)$$

where f_1 , f_2 and f_3 are the outputs of dropout in every layer.

The reason for this is that we want every scale of features to contribute to the final classification result. As a result of experiment, when the parameter of dropout (keep) is equal to 0.4, which means 40% parameters are kept, a best result can be achieved. Following the concatenate layer, a fully-connected layer and a softmax classifier are connected in turn.

III. EXPERIMENTS AND RESULTS

A. Data

In this research, the lung nodule images are obtained from the Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) database, which is a publicly available reference for the medical imaging research community [22, 23]. There are 1,018 cases originated from a total of 1,010 patients and each consists of images from a thoracic CT scan, as well as the annotations provided by 4 radiologists. The images comply with the Digital Imaging and Communications in Medicine (DICOM) standard, and have an in-plane pixel size of 512×512 per 2D scan. In this dataset, it contains three categories of nodules, nodule $\geq 3\text{mm}$, nodule $< 3\text{mm}$ (small nodule) and non-nodule $\geq 3\text{mm}$.

B. Data Setup

According to the annotation documents, only nodules larger than 3mm have malignancy annotations, so we excluded 142 cases with only small nodule or non-nodule markings as noted in the nodule collection report [24]. Besides, there are 11 cases we deleted as some information are missed. Finally, the proposed methodology was applied to 865 CT scans. As annotations in the database, there are five different ratings of the malignancy levels ranging from 1 to 5, level 1 and 2 are considered as benign nodule, level 3 as indeterminate nodule and level 4 and 5 as malignant

nodule. As every nodule is annotated by more than one

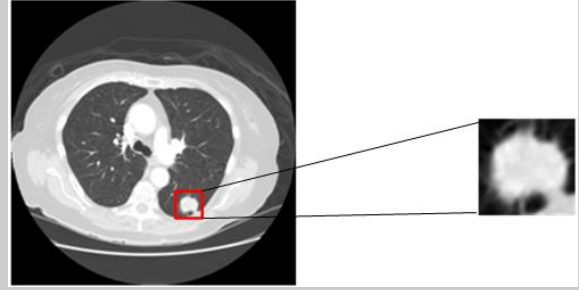


Fig. 3. Nodule cropped samples. We cropped the nodule according to the red frame from original image, which is 3 pixels out of the maximum contour coordinate values

radiologist, the diagnosis is attributed to the class with the highest frequency. If two classes both have the highest frequency, we consider this nodule as indeterminate nodule.

Finally, we obtain 5,376 benign, 11,000 indeterminate and 22,499 malignant nodules from the database. In order to balance data size for each class, we rotated all benign nodules by 90 degrees [13, 25] counterclockwise and added the rotated images to the set, yielding a total of 10,752 benign nodules. At the same time, a half of malignant nodules are deleted randomly, so the final number of malignant nodules is 11,249. The rate of three classes is close to 1:1:1. Finally, the dataset contains 33,001 nodules totally.

For the training process, it is expensive to use original images directly. Because of difference of nodules sizes, we cropped them according to the contour annotations of 4 radiologists, shown as Fig. 3, then they are resized by 52×52 .

C. Experiment Setup

In our experiment, the ML-CNN is applied to classify the ternary classification of lung nodules, the ratio of the training and testing data is 85:15. In the training process, we initialized the learning rate with 0.001, momentum with 0.9 and weight decay parameters with 5×10^{-4} , respectively. We completed the training process in 500 epochs using a batch size of 200. The loss function we used is cross entropy [26]. The configuration of three levels CNNs are shown in Table I. Moreover, the number of neurons in fully-connected layer is 300, and the dropout parameter followed by fully-connected layer we set is 0.5.

We use accuracy as the assessment criterion and defined as:

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (5)$$

where TP is true positive, FN is false negative, TN is true negative and FP is false positive.

D. Results and Discussion

It is important to note that our comparison just uses the single channel 2-dimension images cropped from original CT images to classify the nodules automatically, which does not include the approaches with complex hand-craft preprocessing algorithm, like [25] and [27]. Besides, [25] used the different data with our work, even based on the same database.

TABLE I. ML-CNN CONFIGURATION

Levels	Filter size in convolutional layers	The number of feature maps	Filter size in pooling layers	Dropout (keep)
Level 1	3*3	40 60	2*2 7*7	0.4
Level 2	5*5	28 60	2*2 7*7	0.4
Level 3	7*7	24 60	2*2 7*7	0.4

We validate the effectiveness of our network with the state-of-the-art method, MC-CNN [13], based on the accuracy under the above conditions. The MC-CNN includes 3 convolutional layers followed by 3 pooling layers. When the convolutional layers contained 64 feature maps and the multi-crop pooling layer was in the position of the first pooling layer, they obtained the best result. The size of convolutional kernel is 3×3. The comparison testing result is tabulated in Table II. The result shows that the accuracy of our method ML-CNN is more competitive for ternary classification than MC-CNN, the accuracy of ML-CNN is 84.81% while the accuracy MC-CNN is only 62.46%. The performance of our method is significantly improved 35.78%.

It is indicated that the enhancement of performance attributes to the proposed multi-scale convolution method, which helps the network to extract more effective features. Also, the strategy of flatten first, concatenated second which guarantees all three scales of feature maps can contribute to the final classification result.

TABLE II. THE TESTING RESULT OF TERNARY CLASSIFICATION

Methods	Accuracy (%)
MC-CNN [13]	62.46
ML-CNN (Our work)	84.81

IV. CONCLUSION

In this paper, a deep learning computational architecture, multi-level convolutional neural networks is proposed to classify multi-classes malignancy of lung nodules. We conducted ternary classification including benign, indeterminate and malignant categories. The evaluation results showed that the proposed multi-level convolutional neural network can help to improve the performance significantly in ternary lung nodules classification. In the future, we will study lung nodule detection problem and then crop the nodules from original CT images automatically using deep learning models.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2017," (in eng), *CA Cancer J Clin*, vol. 67, no. 1, pp. 7-30, 2017.

[2] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153-1159, 2016.

[3] Ross Girshick, Jeff Donahue, Trevor Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet

classification with deep convolutional neural networks," *Communications of the Acm*, vol. 60, no. 2, pp. 1106-1114, 2012.

[5] N. Wang, S. Li, A. Gupta, and D.-Y. Yeung, "Transferring Rich Feature Hierarchies for Robust Visual Tracking," 2015, [online] Available: <http://arxiv.org/abs/1501.04587v2>.

[6] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207-1216, 2016.

[7] R. Golan, C. Jacob, and J. Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks," *International Joint Conference on Neural Networks (IJCNN)*, pp. 243-250, 2016.

[8] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images," *Journal of Healthcare Engineering*, vol. 2017, Art. no. 8314740, pp. 1-7, 2017.

[9] A. C. Berg, J. Deng, and L. Fei-Fei, *Large scale visual recognition challenge 2010*, 2010, [online] Available: <http://image-net.org/challenges/LSVRC/2010/index>.

[10] J. Sanchez and F. Perronnin, "High-dimensional signature compression for large-scale image classification," in *CVPR 2011*, pp. 1665-1672, 2011.

[11] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," in *SPIE Medical Imaging*, 2016.

[12] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale Convolutional Neural Networks for Lung Nodule Classification," in *Information Processing in Medical Imaging*, Cham, pp. 588-599: Springer International Publishing, 2015.

[13] W. Shen *et al.*, "Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, pp. 663-673, 2017.

[14] S. G. Armato III *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," (in eng), *Medical physics*, vol. 38, no. 2, pp. 915-931, 2011.

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, no. 1, pp. 106-154, 1962.

[16] J. V. Dahl *et al.*, "Convolutional networks and applications in vision," in *IEEE International Symposium on Circuits and Systems*, pp. 253-256, 2013.

[17] K. Kang and X. Wang, "Fully Convolutional Neural Networks for Crowd Segmentation," *Computer Science*, vol. 49, no. 1, pp. 25-30, 2014.

[18] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Computer Vision and Pattern Recognition*, pp. 2892-2900, 2015.

[19] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *International 1989 Joint Conference on Neural Networks*, vol.1, pp. 593-605, 1989.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," pp. 1-9, 2014.

[21] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," pp. 448-456, 2015.

[22] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013.

[23] S. G. Armato III *et al.*, *The Cancer Imaging Archive*, 2015, [online] Available: <http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX>.

[24] A. P. Reeves and A. M. Biancardi, "The Lung Image Database Consortium (LIDC) Nodule Size Report," 2011, [online] Available: <http://www.via.cornell.edu/lidc/>.

[25] K. Liu and G. Kang, "Multiview convolutional neural networks for lung nodule classification," *International Journal of Imaging Systems and Technology*, vol. 27, no. 1, pp. 12-22, 2017.

[26] L.-Y. Deng, "The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning," *Technometrics*, vol. 48, no. 1, pp. 147-148, 2006.

[27] G. L. F. da Silva, O. P. da Silva Neto, A. C. Silva, A. C. de Paiva, and M. Gattass, "Lung nodules diagnosis based on evolutionary convolutional neural network," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 19039-19055, 2017.