Lung Nodule Classification in CT Images Using 3D DenseNet

Ge Zhang¹, Lan Lin*, Jingxuan Wang¹

¹ Department of Biomedical Engineering, College of Life Science and Bioengineering, Intelligent Physiological Measurement and Clinical Translation, Beijing Base for Scientific and Technological Cooperation, Beijing University of Technology, Chaoyang District, Beijing

*Corresponding author and e-mail: L Lin, lanlin@bjut.edu.cn

Abstract. Lung cancer is the main malignant tumour affecting the health of residents in China. Automatically discriminating benign and malignant pulmonary nodules can facilitate the early detection of lung cancer, which reduces lung cancer mortality. The rising quantity of public available lung CT datasets made it possible to use deep learning approaches for lung nodules malignancy classification. Unlike most of the previous models that focused on 2D convolutional neural nets (CNN), here we explore the use of the DenseNet architecture with 3D filters and pooling kernels. The performance of the proposed nodule classification was evaluated on publicly available LUNA16 dataset, a subset of lung image database consortium and image database resource initiative dataset (LIDC/IDRI). It achieved a 92.4% classification accuracy. The proposed method provides an independent module with encouraging prediction accuracy that can be easily incorporated with a lung cancer computer-aided diagnosis system.

1. Introduction

As one of the deadliest tumors, lung cancer causes the deaths of roughly daily 422 people around the world [1]. Early recognition of potentially cancerous pulmonary nodules is an important strategy in reducing the mortality rate for patients [2]. Compared with other cancers, the biological characteristics of lung cancer are very complex with no obvious symptoms in the early stage. A lot of patients diagnosed with lung cancer are already at an advanced stage, which is associated with a high mortality rate. Low-dose high resolution lung CT screening provides a common, noninvasive imaging modality for early diagnosis of cancerous lung tumors. However, precisely analyzing the pathological organization of pulmonary nodules from CT scans is not straightforward. Currently, radiologists are assessing CT scans to distinguish whether potential lung nodules are benign or malignant, which inevitably entails a high workload for radiologists. Inconsistent judgments are also made due to different experiences of those radiologists.

Naturally, a fast and accurate computer-aid diagnosis (CAD) system of automated classification of pulmonary nodules over CT scans is an effective means for timely medical intervention. Generally, nodule classification consists of two steps: nodule feature extraction and classification model. For the past two decades, extensive research has been done on building CAD systems, which help radiologists to identify malignant nodules in CT images. To classify pulmonary nodules, traditional classification models adopted different machine learning algorithm, i.e. support vector machine (SVM) [3], artificial neural network [4], K-nearest classifier [5], random forests [6], and ensemble classifiers [7], with image features such as texture, shape, form, density, small waves, histogram of oriented gradients. Scholars pay more attention to the design of handcrafted features before deep learning. Deep learning algorithm
attempts to learn top-notch features automatically [8]. By simulating the human brain with a layered model, it frees users from troublesome handcrafted feature extraction, and provides a strong classification framework with self-learning and feature extraction capability. Motivated by its superior performance in the tasks of medical image classification [9-11], deep learning, especially the well-known convolutional neural networks (CNN) have been used in lung nodule classification [12]. Abraham et al. [13] used three 2D CNN networks (AlexNet, VGG16 and SilNet) to perform lung nodule classification. Shen et al. [14] adopted a multiscale 2D CNN approach for recognition of potentially cancerous nodules. However, CT scans are 3D images, most existing CNN-based approaches use a 2D model, which can’t capture the spatial information between slices. Further improvements are needed to take full advantage of the 3D spatial information.

In this paper, we examine the problem of predicting the malignancy of lung nodules in CT scans using a 3D model. DenseNet [15], which encourages feature reuse and alleviates the vanishing gradient problem has been extended to 3D. The result shows the proposed model is helpful in classifying nodules in CTs. The remainder of this paper is structured like this: Section 2 introduces the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) dataset and explains the 3D DenseNet model. Section 3 contains implementation and corresponding results. And in the end, we give a brief conclusion in Section 4.

2. Method

2.1. LIDC/IDRI Data Set

Initiated and collected by seven academic centers and eight medical imaging companies, LIDC/IDRI [16] is one of the largest lung nodule datasets. It provides big data support for developing, training, and evaluating CAD methods for lung cancer detection and diagnosis. At present, the LIDC/IDRI database includes 244,527 spiral CT images of the chest from 1,018 sets of 1,010 different patients. Each of which includes DICOM images with marked-up annotations of pulmonary nodules from multiple radiologists. Marked lesions belong to one of three categories (nodule larger than 3 mm, nodule less than 3 mm, and non-nodule larger than 3 mm). More specifically, the database contains a total of 7,371 nodules by at least one radiologist, of which 2,669 lesions marked nodule > 3mm. CT scans have matrix size of 512×512 pixels, axial plane resolution ranges from 0.46×0.46 mm to 0.98×0.98mm, and slice thickness varies from 0.5 to 5mm. Nodule diameters in this dataset range from 2.03 mm to 38.12 mm. A preprocessing software [17] was used in this study to preprocess of lung CT images and visualize annotations.

2.2. 3D DenseNet

DenseNet is a kind of CNN with densely feature connection, which ensures maximum information flow between layers in the network. The architecture of DenseNet is different from the conventional CNN that alternatively stack convolutional and pooling layers; it further extended the method proposed by ResNet by connecting each layer in a dense block with their front layers. The network structures are progressively hierarchical. The layers are densely connected together in a dense block. This neural network structure improves information and gradients flow over the network, which can strengthen the feature propagation.

Figure 1 is the flowchart of lung nodule classification module. 3D DenseNet model consists of a convolutional layer, five fully connected dense blocks for connecting the features densely, four transition blocks for compressing channels, a pooling layer and a softmax layer for predicting benign or malignant of nodules. In each dense block, every dense layer receives features learned in all preceding layers through shortcut connection. When there was a direct connection between the lower layer and the higher layer, the gradient would be easier to propagate back from high layers to the low layers. Therefore, the dense block can effectively alleviate the problem caused by the vanishing gradient. Each dense block is a module, which in turn is composed of three convolution blocks. In order to ensure easily superimposing of the feature maps, the sizes of feature maps need to be
consistent in a dense block. Between two dense blocks, a transition layer is set to control the model complexity. Table 1 lists the corresponding network parameters in the 3D DenseNet model. The growth rate of dense blocks is set to \( k = 32 \), and compression rate of transition blocks are set at 0.5.

**Figure 1 Flowchart of nodule classification module**

| Layers            | Parameters | Output          |
|-------------------|------------|-----------------|
| Convolution       | 3x3x3 conv | 48x48x48x64     |
| Dense Block(1)    | \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \& 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 3 \) | 48x48x48x160 |
| Transition Layer(1)| \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \end{bmatrix} \times 3 \) | 24x24x24x80   |
| Dense Block(2)    | \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \& 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 3 \) | 24x24x24x176  |
| Transition Layer(2)| \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \end{bmatrix} \times 3 \) | 12x12x12x88   |
| Dense Block(3)    | \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \& 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 3 \) | 12x12x12x184  |
| Transition Layer(3)| \( \begin{bmatrix} 2 \times 2 \times 2 \text{ pool} \end{bmatrix} \times 3 \) | 6x6x6x96      |
| Dense Block(4)    | \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \& 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 3 \) | 6x6x6x188     |
| Transition Layer(4)| \( \begin{bmatrix} 2 \times 2 \times 2 \text{ pool} \end{bmatrix} \times 3 \) | 3x3x3x94      |
| Dense Block(5)    | \( \begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \& 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 3 \) | 3x3x3x190     |
| Classification Block | Softmax   | (, 190)       |
|                   |            | (, 2)         |
3. Experimental Result
This dataset used in this study comes from the Grand Challenge on Lung Nodule Analysis 2016 (LUNA16), a subset of the LIDC-IDRI. A total of 888 CT scans in LUNA16 dataset were used in this study. The nodule annotations were based on agreement of at least three out of four radiologists. Therefore, 590 individuals have one or more nodules, and 298 have none. Only some of those nodules were annotated by at least three out of four radiologists. Thus, we used at least three out of four radiologists’ annotation as a gold standard. Nodules less than 3mm in diameter considered independent of the lung cancer screening protocols, they are excluded from the algorithm evaluation. Malignancies of nodules in LIDC are not confirmed with biopsy. For each nodule, its degree of malignancy is visually assessed by a radiologist using a five-grade marking system (Table 2), suggesting an increasing degree of malignancy suspiciousness. The ground truth for training and testing based on LIDC data was derived with different settings, such as the median malignancy scores, average of malignancy scores, and the maximum malignancy scores. In this study, the maximum malignancy scores were used. Nodules having a malignancy score less than three and greater than three were labeled as benign and malignant, respectively. In our experimental setting, nodules with malignancy scores of three were not involved in training and evaluations since they have a vague malignancy status. This procedure resulted in using 793 nodules (437 benign and 356 malignant) for testing and training. A 48×48×48 image volume around the nodule mass center was extracted for each nodule. Trilinear interpolation was applied to normalize the CT volumes to achieve isotropic voxel dimensions. According to nodule contour delineation of the four radiologists, the nodules were segmented from their surroundings. See Figure 2 for an illustration of this procedure. After extraction of nodule patches, we applied those 3D arrays as input to the CNN. We trained the model in 200 epochs, using 80% of data (336 benign and 287 malignant) for training and 20% (101 benign and 69 malignant) for testing. Initial learning rate considered for this work was 0.001.

Figure 2 Lung nodule extraction procedure
Deep learning-based methods thus far are data hungry and need thousands or millions of images to make a plausible classification. Although we have big data on lung CT, most of these scans are normal and do not contain nodules. To avoid a problem like over-fitting, we opted to use 3D augmentation techniques to enlarge training dataset. The data were augmented through operations as random flips, shifts, Gaussian noise, and rotations that are not prone to shape deformation. The original training dataset was augmented to 36 times. The image has center-level 300 HU and window width of 1800 HU. Subsequently, the pixel values are normalized to have zero mean and unit variance. All experiments were conducted on a workstation equipped with a NVIDIA GeForce RTX 1070 graphics card, and 2.60GHz Intel Xeon processors with 16GB RAM memory.

Three performance metrics are measured using three metrics; accuracy, sensitivity and specificity,

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \\
\text{Specificity} = \frac{TN}{TN + FP} \times 100\%
\]

where TP, TN, FP, and FN represent the number of samples that are correctly predicted by the model as the positive class, the number of samples are correctly predicted by the model as the negative class, the number of samples in negative class are predicted by the model as the positive class, and the number of samples in positive class that are predicted by the model as the negative class. We get an accuracy of 92.4%, sensitivity of 87.0%, and specificity of 96.0%. Augmented dataset has a great impact on the accuracy of the results. The model accuracy is 92.4% and 88.2% for training set with and without the augmented data. This proves the importance of data augmentation. The result showed that contour of correctly identified malignant nodules normally have large size and irregular shape, and contour for correctly identified benign ones have regular shape. There are thirteen examples of the nodules incorrectly classified. These nodules demonstrate a high heterogeneity in the appearances.

4. Conclusions
In this study, we exploit the advantages of densely concatenated convolution to propose a 3D DenseNet architecture which tackles the challenging problem of computerized classification of lung nodules. The DenseNet is constructed to improve the information and gradients propagation throughout the network. The model is trained with LUNA16 dataset. In general, the model proposed has achieved a good classification performance on the malignancy suspiciousness of a lung nodule.

While the results are encouraging, more work remains to be done. First, one limitation of current study is lack of data. Although LIDC dataset is one of the largest datasets available, but identified nodules are still fairly small. Therefore, we can consider using a larger data set in our future work. Second, our approach is based on nodule contour delineation from radiologists, we plan to incorporate a scheme that can automatically detect the nodule candidates and segment them. Third, the accuracy of malignacies grade in LUNA16 in is unclear, because those grades are not confirmed with biopsy. A new dataset with biopsy is urgently needed. Fourth, there are great differences in shapes, sizes, internal characteristics of nodule, ensemble learning of multiple 3D CNN can be utilized to increase the overall performance.

| Malignancy Level | 1  | 2  | 3  | 4  | 5  | Total number |
|------------------|----|----|----|----|----|--------------|
| Number           | 152| 285| 393| 191| 165| 1186         |
| Percentage       | 12.8%| 24%| 33.13%| 16.1%| 13.9%| 100%         |

Table 2. Grade of benign and malignant lung nodules
Acknowledgement
The authors gratefully acknowledge the support from National Natural Science Foundation of China (81971683), Natural Science Foundation of Beijing Municipality (L182010) and the Scientific Research General Project of Beijing Municipal Education Committee (KM20181005033).

References
[1] R.L. Siegel, K.D. Miller, A. Jemal. Cancer statistics. CA: a cancer journal for clinicians. 2019, 69(1): 7-34.
[2] M.S. Brown, J.G. Goldin, R.D. Suh, M.F. McNitt-Gray, J.W. Sayre, D.R. Aberle. Lung Micronodules: Automated Method for Detection at Thin-section CT-Initial Experience. Radiology. 2003, 226(1): 256-262.
[3] Jing Zhang, Bin Li, Lianfang Tian. Lung nodule classification combining rule-based and SVM. 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA). 2010: 1033-1036.
[4] K. Suzuki, Feng Li, S. Song, K. Doi. Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. IEEE Transactions on Medical Imaging. 2005, 24(9): 1138-1150.
[5] Ying Li, Qian Yu. Lung Nodule Classification Using Supervised Manifold Learning Based on All-Class. 2012 International Conference on Computer Science and Service System. 2012: 2269-2272.
[6] S.S. Baboo, E. lyyapparaj. A Classification and analysis of pulmonary nodules in CT images using random forest. 2018 2nd International Conference on Inventive Systems and Control (ICISC). 2018: 1226-1232.
[7] A. Tartar, A. Akan, N. Kilic. A novel approach to malignant- benign classification of pulmonary nodules by using ensemble learning classifiers. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014: 4651-4654.
[8] Miao Tian, Lan Lin, Baiwen Zhang, Shuicai Wu. Study on the Application of Deep Learning in Neuroimaging. China Medical Devices. 2016, 31(12): 4-9.
[9] Yuchao Wu, Lan Lin, Jingxuan Wang, Shuicai Wu. Application of semantic segmentation based on convolutional neural network in medical images. Journal of Biomedical Engineering. 2020, 37(3): 533-540.
[10] Lan Lin, Ge Zhang, Shuicai Wu. Application of deep learning algorithm in classification of benign and malignant nodules on thyroid ultrasound images. Chinese Medical Equipment Journal. 2020, 41(6): 93-98.
[11] Lan Lin, Jingxuan Wang, Zhenrong Fu, Xuetao Wu, Xuetao Wu, Shuicai Wu. A review on brain age prediciton in brain ageing. Journal of Biomedical Engineering. 2019, 36(3): 493-498.
[12] Jingxuan Wang, Lan Lin, Sixuan Zhao, Xuetao Wu, Shuicai Wu. Research progress on computed tomography image detection and classification of pulmonary nodule based on deep learning. Journal of Biomedical Engineering. 2019, 36(4): 670-676.
[13] G.K. Abraham, P. Bhaskaran, V.S. Jayanthi. Lung Nodule Classification in CT Images Convolutional Neural Network. 2019 9th International Conference on Advances in Computing and Communication (ICACC). 2019: 199-203.
[14] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, Jie Tian. Multi-scale Convolutional Neural Networks for Lung Nodule Classification. Information Processing in Medical Imaging. 2015: 588-599.
[15] Gao Huang, Zhuang Liu, L. van der Maaten, K.O. Weinberger. Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 4700-4708.
[16] Lan Lin, Yuchao Wu, Shuang Song, Shuicai Wu. Lung Image Database Consortium and Image Database Resource Initiative: reviewing progresses toward improved computer-aided diagnosis of pulmonary nodule. Chinese Medical Equipment Journal. 2018, 39(10): 95-99.
[17] Jingxuan Wang, Baiwen Zhang, Lan Lin, Xuetao Wu, Shuicai Wu. A Software for the Lung Image Database Consortium and Image Database Resource Initiative. ICBBS 2019: Proceedings of 2019 8th International Conference on Bioinformatics and Biomedical Science. 2019: 59-63.