# Learning deep spatial lung features by 3D convolutional neural network for early cancer detection

Taolin Jin[1], Hui Cui[1], Shan Zeng[2], Xiuying Wang[1*], *Member, IEEE*

[1]Biomedical and Multimedia Information Technologies Research Group, The University of Sydney, Australia
[2]College of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China
*Corresponding author: xiu.wang@sydney.edu.au

***Abstract*- Accurate early lung cancer detection is essential towards precision oncology and would effectively improve the patients' survival rate. In this work, we explore the lung cancer early detection capacity by learning from deep spatial lung features. A 3D CNN network architecture is constructed with segmented CT lung volumes as training and testing samples. The new model extracts and projects 3D features to the following hidden layers, which preserves the temporal relations between neighboring CT slices. The well-built 3D CNN model consists of 11 layers which generates 12,544 neurons and 16 million parameters classifying whether the patient is diagnosed as cancer or not. ReLU nonlinearity and Sigmoid function are used as activation and classification methods. The model achieves a prediction accuracy of 87.5% where only the biomedical images themselves are used as the input dataset. The model's lowest error rate reaches 12.5% that improves the traditional AlexNet architecture by 2.8%.**

***Keywords—Lung cancer prediction, 3D CNN, deep features***

## I. INTRODUCTION

Lung cancer remains the 5th most commonly diagnosed cancer in the Australia within recent years, and is regarded as the leading cause of cancer deaths worldwide [1]. According to statistical report by Australian Government, the 5 year lung cancer survival rate remains 16% where most of the cancer patients are detected at late stage [1]. The best chances for cure on non-small cell lung cancer happens on the surgery of early detected cancer of stages I to II [2], where the 5-year survival possibility can be raised up to over 70% [3].

Over the decade, computer aided diagnosis has attracted innumerable research interest and has been investigated by both industry and research centers. With the development of computational techniques, computerized biomedical image segmentation and machine learning based cancer detection have been widely investigated towards precision oncology.

Machine learning algorithms, especially deep architected learning algorithms, have attracted intensive research interest recently. Convolutional neural network (CNN) [4] is one of the most popular deep learning architectures and has been widely investigated in feature extraction for classification and segmentation tasks. The CNN model, or the so-called AlexNet, extracts the features from the top to bottom with a 2-dimensional (2D) filter, and projects the features in a flat plane. For the 3D medical imaging volumes, however, the 2D feature extraction has its drawbacks. For example, the valuable spatial information or the correlations between slices may be neglected during the learning and training process.

Although 3D CNN model has been investigated on AlexNet [4], it was built for natural images with RGB channels. This model projects current feature map onto 2D maps via a 3D filter. The restricted 3D convolution layer limits the input layer, the following layers, and the filter windows in 3 dimensions. A novel 3D CNN model [5] is proposed for action recognition by extracting both spatial and temporal dimensional features. The 3D convolutions, however, are only used in some of the convolution stages. Such a construction connects the continuous video frames into a cube to capture motion information. Other CNN based studies for lung lesion diagnosis are mostly conducted with the purpose of nodule classification. For instance, a 2D CNN model is learned from multiple views of lung CT scans to classify whether the lung nodule is benign or malignancy. [6] The CNN model is also used for lung nodule detection and segmentation with weakly labeled data, where the 3D CNN model is trained for nodule detection with 3D convolutional filters, and another unsupervised learning model allows the system to grow out a 3D lung nodule region that is ready for segmentation. [7]

In this work, we aim to investigate the performance of deep spatial lung features in early cancer detection. In order to preserve the spatial information, we construct a full 3D CNN model. The major contributions are improving the traditional 2D CNN to a full 3D architecture for lung cancer detection, which include (a) using segmented lung volume as an independent instance of data input, (b) 3D deep feature extraction, and (c) a classification function expertizing on binary classes. The paper

978-1-5386-2839-3/17/$31.00 ©2017 IEEE

is organized as: in Section II we introduce the lung volume segmentation and 3D CNN construction procedures; Section III gives the experimental results and discusses the contributions and limitations of the research.

## II. METHODS

We propose a cancer predication method using deep 3D CT lung volume features extracted from a 3D CNN model. The two major steps are segmenting the entire lung volume from thorax CT volume and constructing a 3D CNN model for deep spatial feature extraction, training and testing.

### *A. Dataset*

The Kaggle Data Science Bowl 2017 dataset [8] is obtained from the US National Cancer Institute which includes 1,397 instances of high-risk patients' lung low-dose CT scans saved in DICOM standard. For each patient's scan, it contains a series of axial 2D slices. The CT slices are reconstructed in size of 512 x 512 pixels while the number of slices in each chest cavity scan varies from 128 to 220 depending on the machine and patient. Each instance (patient scan) is labeled as 0 or 1, where 0 means the patient is not diagnosed as cancer, and 1 represents being diagnosed as cancer. Within the dataset, 1,035 instances are annotated as 'non-cancer', and the rest 362 are 'cancer'.

### *B. Lung volume segmentation*

The initial step in our framework is lung segmentation and data resizing by the following procedures as shown in Figure 1.

To segment a lung volume, the CT image is firstly converted to a binary mask and the next step is to choose a threshold value for lung separation (as shown in Figure 1 a)). As lung region has lower Hounsfield Unit (HU) values than the other surrounding soft tissues in CT images where the value of air is -1000 and that for lung is about -500 [9], -400 is chosen as an initial safe threshold as suggested by [10]. The following step 3 aims to removing the 'blobs', e.g. patient bed, connected to the border of the image as shown in Figure 1 b) and steps 4-6 are erosion and dilation operations to remove the blood vessels. In order to prevent excluding any potential region of interests such as the objects attached to the chest wall, a closure operation is implemented with a disk size of radius 10. Figure 1(g) is the closure operation result where the lung region is enlarged and all the inside holes are filled in full as shown in Fig. 1(h). The final step is to apply the extracted mask on the original CT volume and accomplish the whole lung volume segmentation (as shown in Fig. 2).

As the segmented lung volumes are too large and are of inconsistent sizes to be feed into a deep learning model, we perform a further step to resize and unify the segmented volumes. Considering the two principles that the resizing step should not lose the majority of the important information contained in the original CT scan and it should efficiently release the server's stress on memory, a new volume of (128, 128, 20) is used for resizing. The lung volumes whose slice number are not 20 are resampled. After these pre-processing steps, a patient's lung volume and its cancer/non-cancer label are associated in pair ready for training.

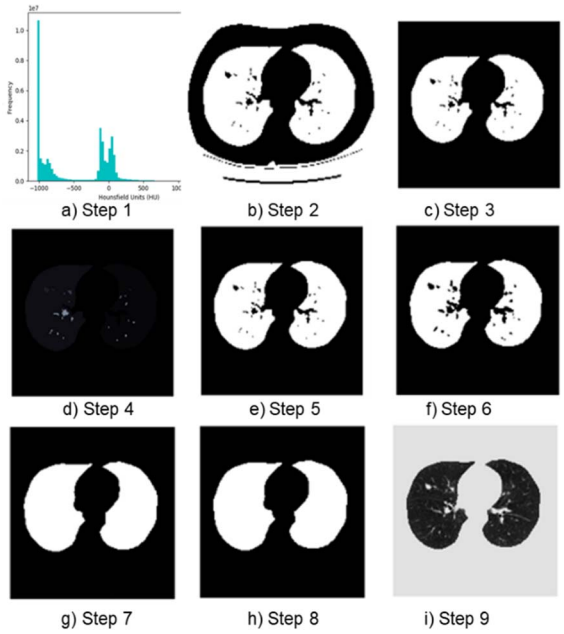


Figure 1. The proposed lung volume segmentation procedure. a) initial binary mask generation by modeling HU values; b) initial mask; c) patient bed and d)-f) blood vessels removal by erosion and dilation operations; g) closure operation and h) lung region enlargement to remove tissues attached to chest wall; i) the extracted lung region by applying the final mask.

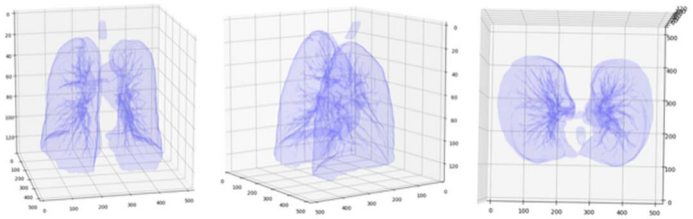

Figure 2. 3D visualization of segmented lung volume.

### *C. 3D deep feature extraction architecture construction*

A convolutional neural network, AlexNet [4], is used as the basis of our learning model. In order to train and extract 3D spatial deep features from the segmented 3D lung volumes, the original AlexNet designed for 2D data classification is improved with respect to the 3D projection process and a Sigmoid

regression instead of SoftMax as the classification in our project is a binary task.

The new architecture is given in Figure 3 with all the parameter settings. It describes how to compute the feature map size, and how to set up the initial neuron number for a fully connected layer. The weight is represented as filter size that means the window size when extracting from the previous layer's output. The bias is described as filter number in the table, with the description on how many features to compute at the current layer. The stride is the step length when switching the filter window over the image chunks or feature maps.

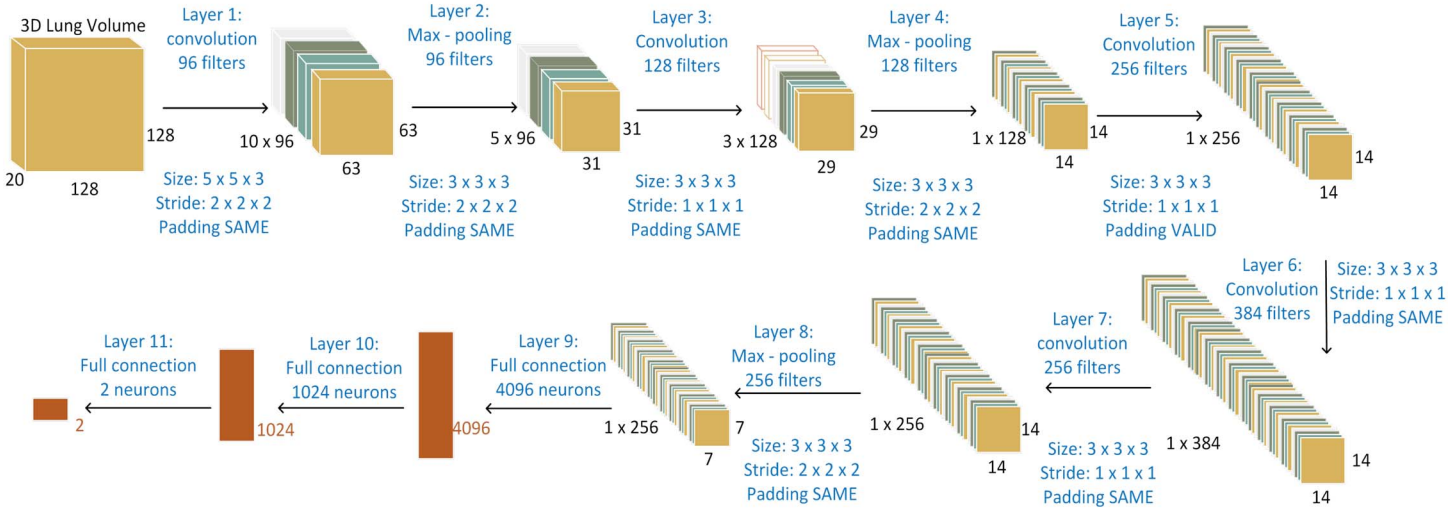


Figure 3. The 3D deep learning architecture in our cancer detection model.

The overall structure is constructed with 11 layers where there are 5 convolutions, 3 max pooling and 3 full connections layers. In the convolutional layer, the feature filter is a cubic block with the specific size for each layer and projects a 3D block to the next layer of feature map as shown in Figure 4. For some convolutional layers, the subsampling step is necessary, with the respect of keeping the most important features. The feature map's depth on a hidden layer is decided according to the filter number. In the first convolutional layer in our model, for instance, 96 3D feature maps are generated because of the 96 filter channels.

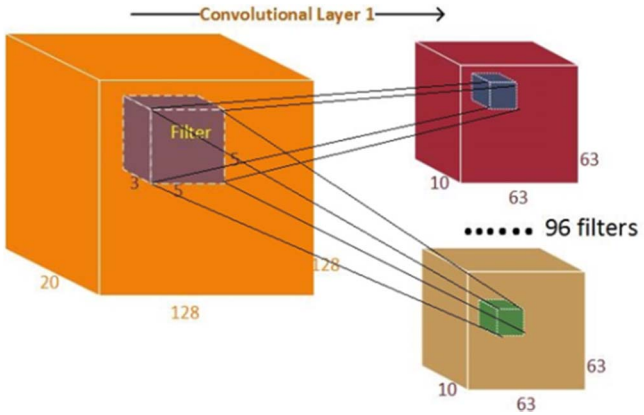


Figure. 4. Illustration of 3D feature projection process from a convolutional layer to the next layer

To prevent the potential risk on gradient disappearance in the converging process, Rectifier (ReLU) activation function is used for each of the convolutional and fully connected layers in our model due to its computation efficiency [4].

As each input volume is of size (128,128,20), it will generate 12,544 neurons ($7 \times 7 \times 1 \times 256$) before moving to the fully connected layers. In order to reduce the overfitting problems, dropout regularization [11] is performed to where the dropout rate is set as 0.5.

## III. EXPERIMENTAL EVALUATION

### *A. Running environment and evaluation*

The server used for implementation is with Intel Core i7-7700K 4-Cores CPU processor, NVIDIA GeForce GTX 1080 8G GPU, 32GB Dual Channel DDR4 RAM, 256GB SSD (Boot) and 2TB SATA (storage) hard disk. Ubuntu operating system is installed. This work was implemented by Python 3.5 and programmed based on Tensor Flow 1.0.

The prediction accuracy is calculated as

$$Accuracy = (TP+TN) / (TP+TN+FP+FN) \quad (1)$$

where TP (true positive) and TN (true negative) represent the positive and negative sample number of right classification; FP (false positive) and FN (false negative) are the negative and positive sample number of false classification.

*B. Feature and neuron sizes*

The parameters regarding the filter, filter size stride and padding of the 3D CNN are listed in Figure 3. Suppose an image (feature map) to be fed into a layer is of size $N_1 \times N_2 \times N_3$, the filter is of size $F_1 \times F_2 \times F_3$, the stride is of size $S_1 \times S_2 \times S_3$, the output size represented by $O_1 \times O_2 \times O_3$ is calculated by:

$$O_i = \frac{N_i - F_i}{S_i} + 1 \qquad (2)$$

Accordingly, the feature and neuron sizes after each layer is calculated and given in Tables 1 and 2.

**Table 1. Output feature sizes on each layer in the 3D CNN model (constructed as Figure 3)**

| After Layer# | Feature Size |
|---|---|
| 0 (Input) | 128 x 128 x20 |
| 1 | 63 x 63 x 10 x 96 |
| 2 | 31 x 31 x 5 x 96 |
| 3 | 29 x 29 x 3 x 128 |
| 4 | 14 x 14 x 1 x 128 |
| 5 | 14 x 14 x 1 x 256 |
| 6 | 14 x 14 x 1 x 384 |
| 7 | 14 x 14 x 1 x 256 |
| 8 | 7 x 7 x 1 x 256 |

**Table 2. Neurons in full connection layers**

| Layer# | Input neuron | neuron |
|---|---|---|
| 9 | 7 x 7 x 1 x 256 x predict instances # | 4096 |
| 10 | 4096 | 1024 |
| 11 | 1024 | 2 |

*C. Results and Discussion*

The finalized network took 84 hours to train the deep learning model with the segmented 3D lung volumes. The 3D CNN model is trained with 1,300 instances, and tested with 40 instances randomly selected each time.

The accuracy evaluation results are given in Table 3. The best accuracy was achieved when the input volume size as defined as 128×128×20. Our model's lowest error rate reaches 12.5% which improves the traditional AlexNet architecture by 2.8% [4]. The lowest accuracy of 75.0% was obtained without the pre-segmentation of lung volume.

To further investigate the impact of epoch and input lung volume size on the prediction results, the model is also trained with different epoch numbers and volume sizes. As shown by the experiment 2 and 5 in Table 3, when there are more epochs being considered in training, the higher accuracy will be obtained. The input volume sizes also have an impact on the prediction results. As shown by Figure 5, the best result in our experiments was achieved when the volume size was resampled to (128, 128, 20). It is understandable smaller resample sizes will result in the loss of lung volume information leading to decreased accuracy. Without the segmentation of lung volume, an accuracy of 75.0% was obtained even though the epoch was 1000 as compared by Experiments in Table 3.

It was also found that with the increasing number of convolutional layers constructed in the neural network, it would continuously generate more complex feature maps and improve the accuracy.

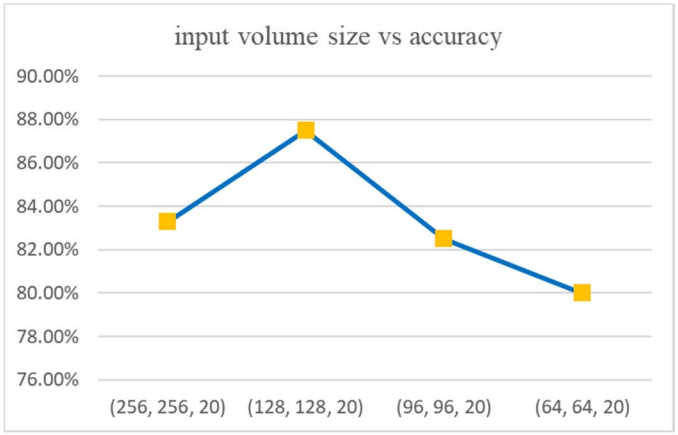


Figure 5. Impact of the sizes of input resampled segmented lung volume on the prediction accuracy.

**Table 3. Accuracy with different epochs and different input volume sizes**

| Experiment# | Input Volume | Train | Test | Epoch | Accuracy |
|---|---|---|---|---|---|
| 1 | (256, 256, 20) | 1,300 | 30 | 500 | 83.3% |
| 2 | **(128, 128, 20)** | **1,300** | **40** | **500** | **87.5%** |
| 3 | (96, 96, 20) | 1,300 | 40 | 500 | 82.5% |
| 4 | (64, 64, 20) | 1,300 | 40 | 500 | 80.0% |
| 5 | (128, 128, 20) | 1,300 | 40 | 50 | 77.5% |
| 6 | without lung volume segmentation | 1,300 | 30 | 1000 | 75.0% |

As our cancer detection model only considered lung volumes and the corresponding deep spatial features instead of segmented lung nodules as in other models, the trained model has no information about the nodules. A further investigation was performed to inspect the relations of the nodules and given cancer/non-cancer labels.

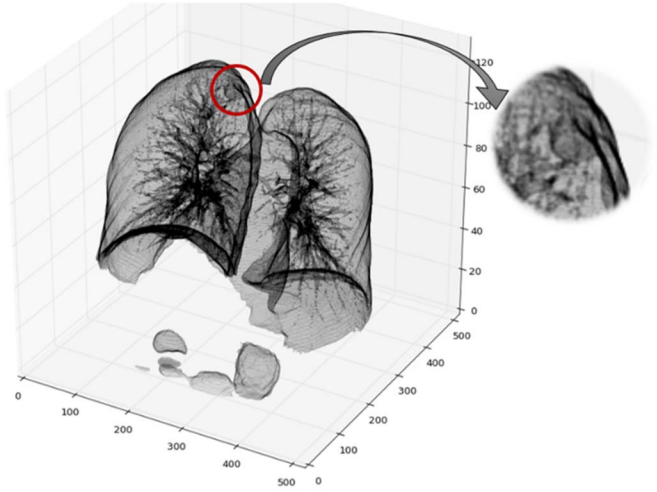

Figure 6. One case labled as non-cancer but with a benign nodule. (Patient ID 56e0f421b7faddcae027d7feb6dae8cd)

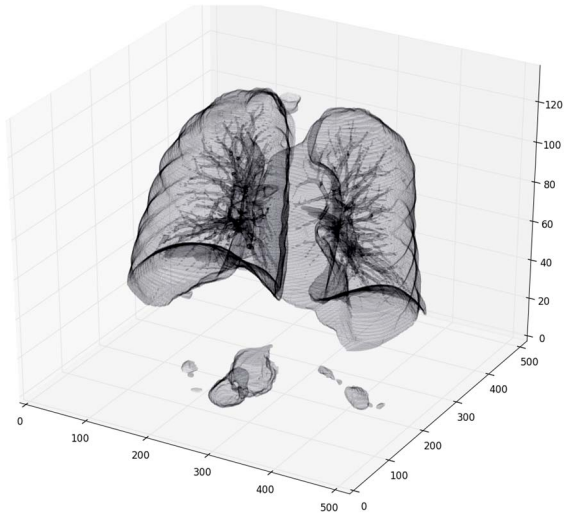

Figure 7. One case labeled as "cancer" but without any detected nodules. (Patient ID 00edff4f51a893d80dae2d42a7f45ad1)

As shown in Figure 6, the case is labeled as "non-cancer" but with a nodule. The nodule has smooth surface and regular spherical shape which matches the properties of a benign nodule. Although the benign nodule is not the fatal cause on leading the cancer, it brings the large disturbances on learning the classification model. Comparing with the entire lung, a nodule is a relatively tiny feature in the learning process which is hard for our model to distinguish whether such a tiny nodule is shaped with benign nodule's smooth contour or the malignant nodule's lobular appearance.

The other case in Figure 7 is labeled as "cancer" but no nodules are detected. As this is not a singular case through the dataset, 4 out of 30 randomly selected 'cancerous' lung volumes are detected in the similar situations. Therefore, it directs our future work including constructing a cancer prediction model incorporating both deep whole lung volume features and nodule features.

## IV. CONCLUSION

An entire full 3D convolutional neural network architecture is constructed with the segmented lung volumes as input from low-dose CT scans for early cancer detection. The highest accuracy achieved by the system is 87.5% by training the deep learning model with 1,300 instances which improved the traditional 2D AlexNet architecture.

## REFERENCES


[1] A. Government. (2017). *Lung cancer statistics*. Available: https://lung-cancer.canceraustralia.gov.au/statistics

[2] C. M. H. N. F. Kanarek, L. Mathieu, H. L. Tsai, C. M. Rudin, J. G. Herman and M. V. Brock, "Survival After Community Diagnosis of Early-stage Non-small Cell Lung Cancer," *Am J Med,* vol. 127, no. 5, pp. 443-449, 2015.

[3] A. H. S. a. S. Islam, "Early Lung Cancer Detection, Mucosal, and Alveolar Imaging," *Interventional pulmonology,* vol. 22, no. 3, pp. 271-280, 2016.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," presented at the Neural Information Processing Systems (NIPS), pp. 1097-1105, 2012.

[5] W. X. Shuiwang Ji, Ming Yang, Kai Yu, "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 35, no. 1, pp. 221-231, 2013.

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

[7] R. A. J. J. T. T. B. H. Kim, "Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data," presented at the Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 2016.

[8] Kaggle. (2017). *Data Science Bowl 2017*. Available: https://www.kaggle.com/c/data-science-bowl-2017

[9] S. Khalid, A. C. Shaukat, A. Jameel, and I. Fareed, "Segmentation of Lung Nodules in CT Scan Data: A Review," in *Medical Imaging: Concepts, Methodologies, Tools, and Applications*: IGI Global, 2017, pp. 601-613.

[10] A. Teramoto and H. Fujita, "Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter," *International journal of computer assisted radiology and surgery,* pp. 1-13, 2013.

[11] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929-1958, 2014.