PAPER • OPEN ACCESS

Computer-Aided Detection of Mediastinal Lymph Nodes using Simple Architectural Convolutional Neural Network

To cite this article: E Kurniawan *et al* 2020 *J. Phys.: Conf. Ser.* **1505** 012018

View the article online for updates and enhancements.

Computer-Aided Detection of Mediastinal Lymph Nodes using Simple Architectural Convolutional Neural Network

E Kurniawan[1], P Prajitno[1,*], D S Soejoko[1]

[1] Department of Physics, Faculty of Mathematics and Natural Science, Universitas Indonesia, Depok, 16424, Indonesia.

* Email: prawito@sci.ui.ac.id

Abstract. Lung cancer is the most common and the deadliest cancer in the world. Lung cancer staging usually was done by radiologist by detecting mediastinal lymph node (LN) enlargement. Mediastinal LN is difficult to be detected visually due to its low contrast to the surrounding tissues, various size and shape, and sparse location. Therefore, computer-aided detection (CADe) system has been developed as a tool for radiologist to detect mediastinal LN automatically. The state of the art mediastinal LN CADe system use complex architectural convolutional neural network (CNN). However, more simple architecture of the CNN is needed to reduce the computational complexity of the CADe system, especially if the system was intended to be used in a regular computer. Therefore, in this experiment we used simple architectural 2D CNN which is converted to fully convolutional network (FCN) to detect mediastinal LN candidate in a stack of CT images. Then, the mediastinal LN candidates were classified using 3D CNN to reduce the false positive (FP). The best performance of this CADe system was 65% of sensitivity at 5 FP/ patient.

Keywords: mediastinal lymph nodes, simple architectural convolutional neural network, CT image, computer-aided detection, deep learning

1. Introduction

Lung cancer is the most common and deadliest type of cancer in the world. Based on data from the Global Cancer Statistics (GLOBOCAN) 2018, lung cancer occupies 11.6% of total cancer cases and 18.4% of total cancer deaths [1].

Determination of the stage of lung cancer is done by a radiologist by examining the presence or absence of enlargement of the mediastinal lymph nodes based on CT images of the chest area of lung cancer patients [2]. Up to now, radiologist does the work manually. But lately, a system that helps radiologist to do that activities is has been developed, namely computer-aided detection (CADe) systems [3].

The best method currently in the research of mediastinal LN CADe systems is the deep learning-based detection method conducted by Shin et al. [4]. They used deep convolutional neural networks (CNN) to detect mediastinal LN. Complex CNN architectures, such as AlexNet and GoogLeNet, were used to detect mediastinal LN on CT images of the thoracic region. Their CADe system is able to detect mediastinal LN with a sensitivity of 85% / 3 false positive (FP) per patient. However, the CNN architecture which they used was complex enough so that a high-performance computer was needed to operate the system.

A simpler CNN architecture in a mediastinal LN CADe system is needed to reduce the computational complexity so that the CADe system can be used well even on low-performance computer. Therefore, in this research, a mediastinal LN CADe system was made using CNN with a simple architecture. A simple CNN architecture with the right amount of data and the proper training process has been proven to provide good CADe system performance like that of Pezeshk et al. [5]. The complete methodology of this work is described in Section 2. We show the results of our approach in Section 3 and the conclusions are stated in Section 4.

2. Method

This study uses online dataset sources (www.cancerimagingarchive.net) [6] which contains 388 mediastinal LN in CT images of 90 patients. Each image has a dimension of 512×512 pixels with a voxel size of $0.85 \times 0.85 \times 1.00$ mm^3. We only perform normalization of HU units to the images before the images is used in the deep learning system. The data is used both in the CNN training process and the CADe system evaluation process. Evaluation of CADe system performance was done by referring to the annotations that have been included in the online dataset.

The tools consist of hardware and software. The hardware was an ordinary desktop computer with an Intel Core i7 processor, 4 GB RAM (DDR4), 1 TB HDD, and Windows OS 10. The software were Python 3.6 using Spyder as IDE, Keras and Tensorflow (for deep learning purposes), MITK Work Bench, and Image J.

The mediastinal LN detection method in this study basically adopts the method which has been used by Pezeshk et al. [5] to detect pulmonary nodules. But, in this study, we developed the method in 2D way at the screening stage in order to reduce the computational complexity. The method [5] was considered suitable for this experiment because mediastinal LN are also in the thoracic region as well as pulmonary nodules.

The mediastinal LN detection method was divided into two stages, namely stage 1: screening and stage 2: false positive reduction. The screening stage aims to detect mediastinal LN candidates quickly. Ideally, the screening stage should be very sensitive so it is hoped that all mediastinal LNs will be detected as candidates for mediastinal LN. However, the more sensitive the screening stage, the greater the number of false positives. Therefore, the second stage is needed, namely the false positive reduction stage.

2.1. Stage 1: screening

The purpose of this screening stage is to quickly determine the mediastinal LN candidates which need further analysis. At this stage, 2D CNN has been trained and converted to 2D fully convolutional networks (FCN). FCN is a CNN which the fully connected layers is converted into convolution layers so that FCN can receive inputs that are larger than the training sample size. Meanwhile, CNN can only accept inputs that are the same size as the training sample. Thus, the screening process using FCN is much faster than using CNN directly. The conversion illustration is presented in Figure 1.

2D CNN for screening purposes is trained using 2D samples extracted from CT images. The sample is divided into two class, namely positive samples and negative samples. Positive samples are taken from 2D region of interest (ROI) centred on each annotated mediastinal LN. Negative samples are taken from 2D ROI in five groups of regions (left-sided chest, right-sided chest, back, lung, and mediastinum) with the sample coordinates randomly determined in each region. ROI size of positive and negative samples are made the same, which is 24×24 pixels. The sample size is considered to be the optimum size after observing the size of annotated mediastinal LN.

CNN requires a lot of training data to achieve good performance. Meanwhile, the amount of training data in the form of positive samples is limited. Therefore, we did the data augmentation process to increase the number of positive samples. Each positive sample is subjected to 13 types of transformation, namely flip up down, flip left right, 90° rotation, 180° rotation, shear with a factor of ± 0.2 and a combination of these transformations. Thus, from the data augmentation process, the total number of positive samples were 5044 samples. Data augmentation is not carried out on the negative samples because the negative samples can be taken as many as desired. The total number of negative samples used in this study were 15357 samples.
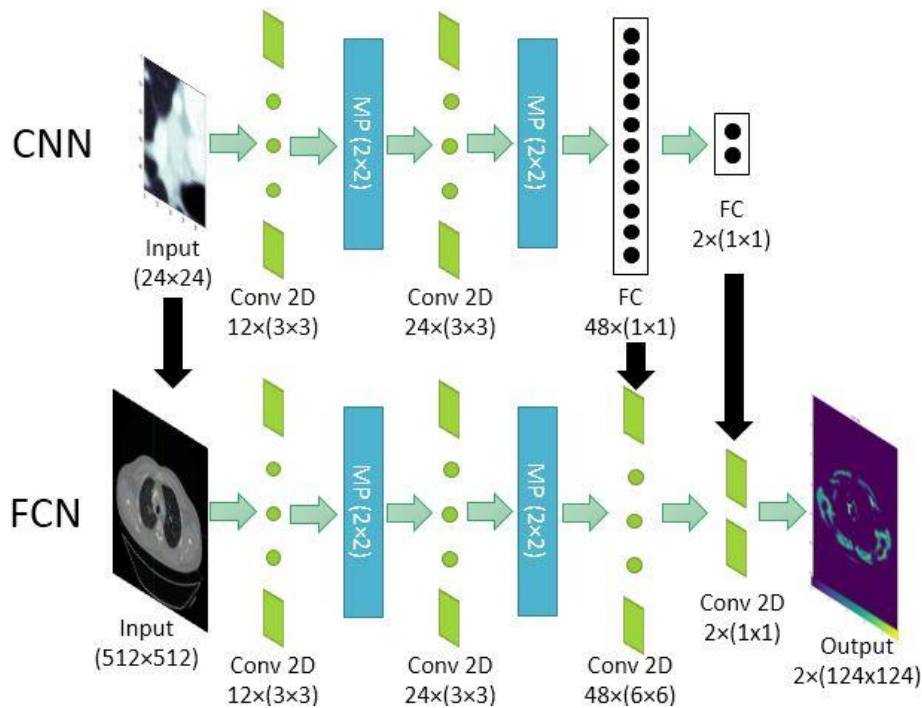
Figure 1. The 2D CNN architecture and the conversion of 2D CNN into 2D FCN

Table 1. Details of the 2D CNN architecture

Layers	Kernel Size	Stride	Output Size	Number of Feature
Input	-	-	24×24	-
Convolution 1	3×3	1	24×24	12
Max Pool 1	2×2	2	12×12	12
Convolution 2	3×3	1	12×12	24
Max Pool 2	2×2	2	6×6	24
Fully Connected 1	-	1	1×1	48
Fully Connected 2	-	1	1×1	2

The 2D CNN architecture which was used in this study consists of two layers of convolution and two layers of max-pooling followed by two fully connected layers. The illustration of the 2D CNN architecture is presented in Figure 1. At the convolution layer, zero padding is applied to the input so the output of the convolution layer will be the same as the input. This is done to maintain edge properties. Details of the 2D CNN architecture is presented in Table 1.

The weight value in 2D kernel is initialized using He normal initialization. The activation function at each layer (except the last fully connected) is the Rectified Linear Unit (ReLU). The type of pooling layer in this study is max pooling. The 2D max pooling (MP) operation size is 2×2. The stride at the MP layer is 2 so that there is no overlapping in the MP operation. Based on this arrangement, the size of the output data from the MP layer will be half of the size of the input data.

The training process is carried out as many as 120 epochs with a batch size of 32. Weights that provide the smallest validation loss in the training process is stored and the fully connected weights is reshaped to be suitable for 2D FCN. The loss function used in this training process is binary cross entropy which is generally used for the binary classifier training process. A dropout rate of 0.2 is added for each convolution layer and 0.5 for the fully connected layer to minimize overfitting.

After the 2D CNN has been trained, the 2D CNN is converted to the 2D FCN so that it can be used to scan a full size of CT slice (512×512 pixels), slice by slice in a stack of CT images. The conversion is done by changing the weight form of the fully connected layer on 2D CNN which

initially takes the form of a vector into a 2D array that is the same size as the output size of the layer before the fully connected layer. 2D array containing the weight acts as the kernel so that the FC layer is now a convolution layer. The weights used are only changed in shape (not in value) so that the resulting 2D FCN architecture does not require retraining and will produce the same output as the original 2D CNN architecture.

The results of the screening stage are score map, which is the distribution of mediastinal LN probabilities in a slice of CT images. Based on the score map, the position of mediastinal LN candidates can be estimated which are points that have a high probabilities.

2.2. Stage 2: false positive reduction

Mediastinal LN candidates resulting from the screening stage were then analysed using 3D CNN to determine whether the candidate was indeed believed to be a mediastinal LN or not. This is done to reduce the number of false positives (FP) resulting from the screening stage.

Mediastinal LN candidates that have been produced from the screening stage are used as input for 3D CNN. In this study 3D CNN was made with the architecture illustrated in Figure 2. The scheme was adopted from Pezeshk et al. [5] with a slight change in sample size and kernel. The details of the 3D CNN architecture is shown in Table 2.

The data used to train the 3D CNN is the same as the data used to train the 2D CNN at the screening stage. The only difference is the data size is made larger in the X and Y direction and has a size in the direction of Z, which is 28×28×12 voxels. By using this set up, the 3D CNN at this stage can load larger mediastinal LNs so that it is expected to be more accurate in doing the classification.

3. Results and Discussion

The screening stage is carried out using 3D CNN which has been converted to FCN. The results of the screening stage is a score map showing the distribution of the probabilities of finding a mediastinal LN in a slice of CT images. An example of a score map that results from the screening stage is shown in Figure 3 (b). Based on the location of points on the score map that have a probability of > 95%, it can be determined which points are candidates for mediastinal LN. The point is used as the focal point for taking 3D array (28×28×12 voxel) as a mediastinal LN candidate. Figure 4 (a) – 4 (c) show some examples of mediastinal LN candidates obtained from the screening stage.

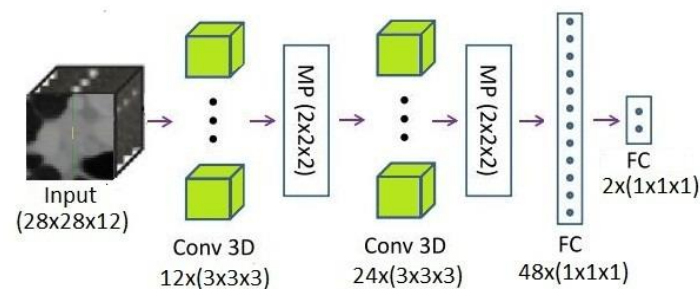**Figure 2.** The 3D CNN architecture for FP reduction stage

Table 2. Details of the 3D CNN architecture

Layers	Kernel Size	Stride	Output Size	Number of Feature
Input	-	-	$28 \times 28 \times 12$	-
Convolution 1	$3 \times 3 \times 3$	1	$28 \times 28 \times 12$	12
Max Pool 1	$2 \times 2 \times 2$	2	$14 \times 14 \times 6$	12
Convolution 2	$3 \times 3 \times 3$	1	$14 \times 14 \times 6$	24
Max Pool 2	$2 \times 2 \times 2$	2	$6 \times 6 \times 3$	24
Fully Connected 1	-	1	$1 \times 1 \times 1$	48
Fully Connected 2	-	1	$1 \times 1 \times 1$	2

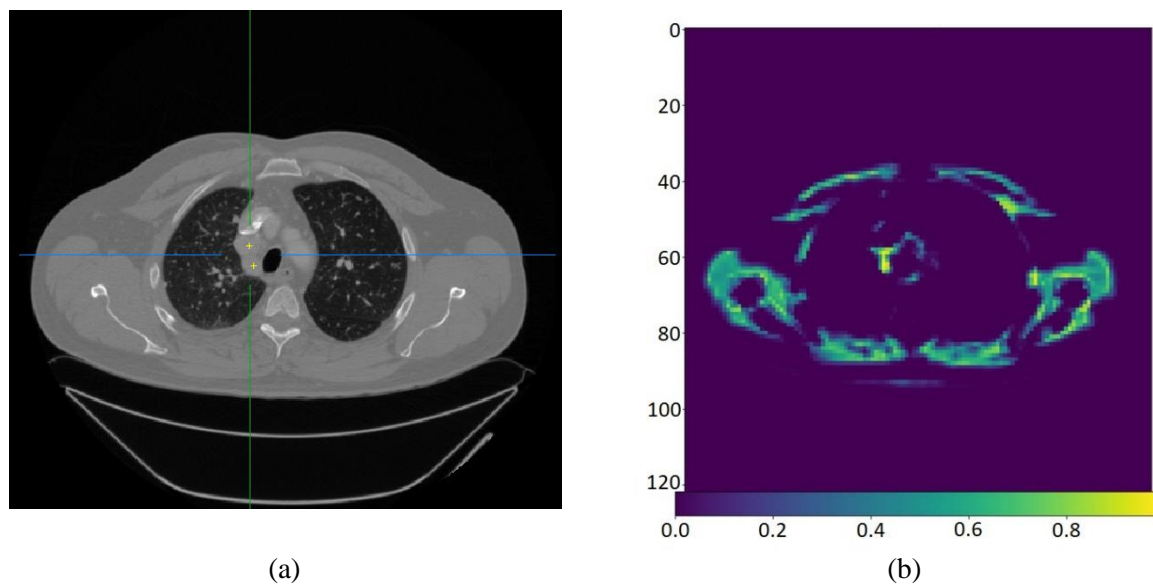(a)   (b)

Figure 3. Examples of an original CT image (a) and a score map (b) generated from (a)

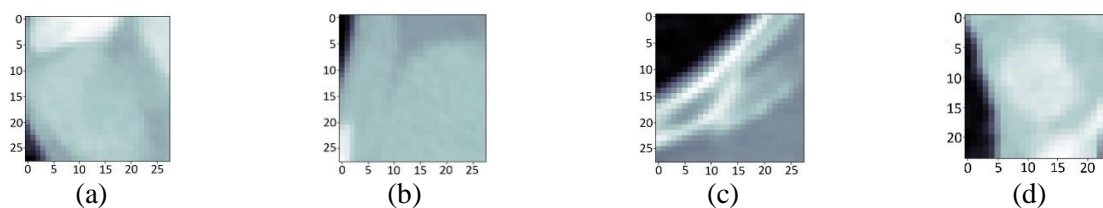(a)   (b)   (c)   (d)

Figure 4. Examples of mediastinal LN candidates (a) – (c), and a true mediastinal LN (d)

Based on Figure 4, it appears that Figure 4 (a) does indeed look like a true mediastinal LN, which is shown in Figure 4 (d), so it is proper for Figure 4 (a) to be a mediastinal LN candidate. However, Figure 4 (b) and (c) are clearly seen as muscle and bone that should not be candidates for mediastinal LN. This can occur due to several factors as follows.

1. There are points on the score map that have a fairly high probability when in fact they are not mediastinal LN. If considered further, it is known that these points are soft tissue (muscle) which have characteristics similar to mediastinal LN. Therefore, 3D FCN can incorrectly predict the soft tissue as a mediastinal LN because the characteristics of the two are similar.

2. Retrieval of mediastinal LN candidate coordinates cannot actually be done precisely because the size of the resulting score map is different from the size of the original CT image due to the convolution and max pooling processes. Mapping the location of a point on the score map cannot be determined with certainty because the process of max pooling selects a maximum point that is unknown to the user when reducing dimensions. Therefore, the mediastinal LN candidates obtained can be in the form of bones like Figure 4 (c) because there may be soft tissue similar to the mediastinal LN next to bones and an error in mapping the score map coordinates to the original image occurred.

Mistakes in taking mediastinal LN candidates might cause false positives (FP) in the final result of the CADe system. Therefore, the false positive reduction stage is needed. Mediastinal LN candidates resulting from the screening stage were further analysed using the 3D CNN described previously.

Table 3 shows the final performance of our CADe systems after FP reduction stage compared to the performance of the other CADe systems. Our system performance is still below the others due to the lack of training data and the lack of size of the 3D CNN input. Pezeshk et al. (2018) [5] uses far more training data and larger input sizes for the 3D CNN in the FP reduction stage. This cannot be done yet in this study, considering that this action will prolong computing time whereas we used a regular computer without GPU.

Table 3. Comparison of several mediastinal LN CADe system performances

Mediastinal LN CADe system	Sensitivity (%)	FP/patient
Shin et al., 2016 [4]	85	3
Roth et al., 2016 [7]	77	3
Roth et al., 2014 [6]	70	3
Jiamin liu et al., 2016 [8]	60	3
This work	65	5

The top three in Table 3 use deep learning to detect mediastinal LN. This shows the advantages of deep learning over other machine learning mechanisms. Jiamin Liu et al. (2016) [8] used a non-deep learning mechanism, namely random forest, as a method of detecting mediastinal LN candidates and continued to use support vector machines (SVM) as final classifiers. Nevertheless, the performance of the Jiamin Liu et al. (2016) [8] system is still better than our CADe even though we used deep learning. This indicates that there are still parameters that are not yet optimum in our system. However, our system is capable of scanning a stack of CT images at 4 slices/s using a regular computer with 65% sensitivity at 5 FP/patient. This shows that the simple architectural CNN-based mediastinal LN CADe system can be used on a regular computer with performance that is still possible to be improved.

4. Conclusion
The conclusion of this study is that a mediastinal LN CADe system based on simple architectural CNN has been developed with 65% sensitivity at 5 FP/patient and 4 slices/s scanning speed. For future research, we suggest performing hard negative mining to improve the CADe system performance.

5. Acknowledgements

6. Reference
[1]   Bray F, Ferlay J and Soerjomataram I 2018 Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries *Ca Cancer J Clin* **0** 1-31

[2]   Feuerstein M, Deguchi D, Kitasaka T and Iwano S 2009 Automatic Mediastinal Lymph Node Detection in Chest CT *Proc. SPIE* **7260** 1-11

[3]   Rafael I, Valente S, César P, Cavalcanti E, Marques J, Hugo V, Albuquerque C D, Manuel J and Tavares R S 2015 Automatic 3D pulmonary nodule detection in CT images : A survey *Computer Methods and Programs in Biomedicine* **124** 1-17

[4]   Shin H, Roth H R, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D and Summers R M. 2016 Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures , Dataset Characteristics and Transfer Learning *IEEE Trans. Med. Imaging* **35** 1285-1298

[5]   Pezeshk A, Hamidian S, Petrick N and Sahiner B 2018 3D convolutional neural networks for automatic detection of pulmonary nodules in chest CT *IEEE J. Biomed. Heal. Informatics*

[6]   Roth H R, Lu L, Seff A, Cherry K M, Hoffman J, Wang S, Liu J, Turkbey E, Summers R M 2014 A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations *MICCAI 2014 Part I LNCS* **8673** 520-527

[7]   Roth H R, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L and Summers R M 2016 Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation *IEEE Trans. Med. Imaging* **35** 1170-1181

[8]   Liu J, Hoffman J, Zhao J, Yao J, Lu L, Kim L, Turkbey E B and Summers R M. 2016 Mediastinal lymph node detection and station mapping on chest CT using spatial priors and random forest *Med. Phys.* **43** 4362–4374