Applied Intelligence (2021) 51:4082–4093
https://doi.org/10.1007/s10489-020-02051-1

A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis

Xukun Li[1] · Yukun Zhou[1] · Peng Du[1] · Guanjing Lang[2] · Min Xu[2] · Wei Wu[2]

Accepted: 30 October 2020 / Published online: 26 November 2020

Abstract

The purpose of this study was to establish and validate a new deep learning system that generates quantitative computed tomography (CT) reports for the diagnosis of pulmonary tuberculosis (PTB) in clinic. 501 CT imaging datasets were collected from 223 patients with active PTB, while another 501 datasets, which served as negative samples, were collected from a healthy population. All the PTB datasets were labeled and classified manually by professional radiologists. Then, four state-of-the-art 3D convolution neural network (CNN) models were trained and evaluated in the inspection of PTB CT images. The best model was selected to annotate the spatial location of lesions and classify them into miliary, infiltrative, caseous, tuberculoma, and cavitary types. The Noisy-Or Bayesian function was used to generate an overall infection probability of this case. The results showed that the recall and precision rates of detection, from the perspective of a single lesion region of PTB, were 85.9% and 89.2%, respectively. The overall recall and precision rates of detection, from the perspective of one PTB case, were 98.7% and 93.7%, respectively. Moreover, the precision rate of type classification of the PTB lesion was 90.9%. Finally, a quantitative diagnostic report of PTB was generated including infection possibility, locations of the lesion, as well as the types. This new method might serve as an effective reference for decision making by clinical doctors.

Keywords Deep learning · Computed tomography · Convolution neural network · Tuberculosis

1 Introduction

Pulmonary tuberculosis (PTB) is one of the leading respiratory infectious diseases worldwide [1]. India, Indonesia, and China have the highest PTB burden [2, 3]. Also, in China, the PTB is the second-highest infectious disease after viral hepatitis [4]. Therefore, correct detection and diagnosis of PTB are crucial importance.

With the rapid development of big data and artificial intelligence (AI), deep learning method [5] has been gradually applied to computer-aided diagnosis (CAD),deep learning technologies, such as the convolutional neural network (CNN) with its strong ability of nonlinear modeling, have also been applied extensively in medical image processing. Relevant studies have been conducted on the diagnosis of pulmonary nodules, the classification of benign and malignant tumors worldwide. [6–12]. This approach has been used to improve the diagnosis of pulmonary nodules and lung cancer worldwide [13–16]. So far, several intelligent diagnostic systems have been available for lung nodule detection, such as the Dr. Watson system from IBM. At the same time, some well-known academic institutions and organizations have also launched competitions to encourage lung nodule detection on computed tomography (CT) images. Of these, the most famous was the Lung Nodule Analysis 2016 (LUNA16) [17] and the Data Science Bowl 2017 (DSB2017), which were held by the notable data science website Kaggle. These open-sourced datasets contain a series of detection and segment algorithms. However, so far, only a few studies have explored the detection and classification of PTB infection using this approach. Currently, very few open-sourced CT image datasets of PTB are available. Moreover, wider distribution and different characteristics of PTB lesion regions compared with those of lung nodules have further complicated the diagnosis. Despite the differences in morphological features between PTB lesion and pulmonary nodule, some of the open-

✉ Wei Wu
1198042@zju.edu.cn

[1] Artificial Intelligence Lab, Hangzhou AiSmartVision Co., Ltd., 259 Wensan Road, Hangzhou Zhejiang 310012, People's Republic of China

[2] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, School of Medicine, Zhejiang University, 79 QingChun Road, Hangzhou Zhejiang 310003, People's Republic of China

sourced intelligent detection methods for pulmonary nodules, such as data preprocessing and image segmentation, have a considerable reference value for PTB detection.

The traditional CT image processing methods are mainly used in the preprocessing stage, and the image segmentation and classification process is based on deep learning algorithms [18–20]. Huang et al. [21] proposed a noisy U-Net (NU-Net) that can enhance the neural network's sensitivity to small nodules by adding noise to the hidden layers in the training set, they found that the algorithm's sensitivity to small nodules with diameters of 3–5 mm (97.1%) was greater than the U-Net value (90.5%). Ciompi et al. [22] constructed a labeling system to automatically classify the morphological characteristics of pulmonary nodules into solid, sub-solid, calcified and non-solid lesions. Moreover, Cao et al. [23] proposed a two-stage convolutional neural network (TSCNN) for lung nodule detection. The first stage is based on the improved U-Net segmentation network with a high recall rate; after which, three 3D-CNN classification networks are designed to reduce false positive. Their experiments showed that the proposed TSCNN architecture could be used to obtain competitive detection performance. In order to improve the classification accuracy of lung nodules, Zuo et al. [24] proposed a method based on a three-dimensional (3D) convolution neural network (ConvNet) model, which is trained by transferring knowledge from a multi-resolution two-dimensional (2D) ConvNet model. The experimental results showed a competitive average score in the false-positive reduction, with the sensitivities of 0.619 and 0.642 at 0.125 and 0.25 FPs per scan. Additionally, Julian de Wit et al. [25], who won the second place in the DSB2017 competition, constructed a pulmonary nodule detector through a 3D CNN to predict the possibility of cancer. Besides, Pezeshk et al. [26] designed a two-stage computer-aided detection system for automatic detection of pulmonary nodules and discovered that the nodule detection system could achieve 91% sensitivity at two false positives per scan. Liao et al. [27] used a modified 3D u-net to detect all suspicious lesions (pulmonary nodules) and evaluate pulmonary malignancy. Huang et al. [28] studied the benign-malignant nodules classification, they merged a Deep Transfer Convolutional Neural Network (DTCNN) and Extreme Learning Machine (ELM) to achieved the classification with superb AUC (94.9%). Xu et al. [29] constructed a multi-resolution CT screening image dataset (DeepLN Dataset). They proposed an effective framework (called DeepLN) to detect lung nodules and demonstrated the proposed method can address the issue well.

In this study, four fine-tuned 3D CNN models were evaluated. The best model was then used to detect and classify the PTB lesion regions based on CT image datasets. Our main contributions can be summarized as follows: the spatial location of each lesion, the confidence (infection probability) of every single infection, presence of calcifications, the classification of lesion type, overall infection probability, and effective volume of the left and right lungs were digitally analyzed according to the output of the AI network model. These reports generated a quantitative evaluation of a single infection region and the whole PTB case, thus greatly assisting clinical doctors in making more accurate diagnostic decisions.

2 Method

2.1 Process

Figure 1 shows the whole process of PTB diagnostic report generation. First, the CT images were preprocessed to extract effective lung regions. Second, the 3D CNN model was used to detect and classify the lesion region. Then, the overall infection probability was calculated using the Noisy-Or Bayesian function. Finally, a quantitative diagnostic report, together with the corresponding labeled CT images were exported for reference.

2.2 Dataset introduction

Five types of active PTB lesions were defined according to the Expert Consensus of the Chinese Society of Radiology [30] (Fig. 2): miliary, infiltrative, caseous, tuberculoma, and cavitary.

This study used 501 CT imaging datasets from 223 patients, diagnosed with active PTB at the inpatient department of tuberculosis of the Affiliated Hospital of Zhejiang University from 2016 to 2019. Moreover, 501 CT image datasets from a healthy population collected in the same hospital were added as negative samples. Therefore, 1002 CT imaging datasets, all in Digital Imaging and Communications in Medicine (DICOM) format, were analyzed.

The ethics committee of the First Affiliated Hospital, College of Medicine, Zhejiang University approved this study, and all research was performed in accordance with relevant guidelines/regulations. All participants and/or their legal guardians signed the informed consent form before the study.

Two professional radiologists manually annotated the lesions of PTB. Together, 2884 (117 miliary, 2255 infiltrative, 135 caseous, 91 tuberculoma, and 286 cavitary) regions were labeled as PTB lesions.

2.3 Dataset preprocessing

To facilitate the detection of PTB lesions, the CT images were resampled to keep the voxel of CT image to $1 \times 1 \times 1$ mm^3 measured in the real space, following the rule of nearest-neighbor interpolation. A CT image contains not only the lung but also other tissues, and some of them may have similar shape and structure as the infections of tuberculosis, such as

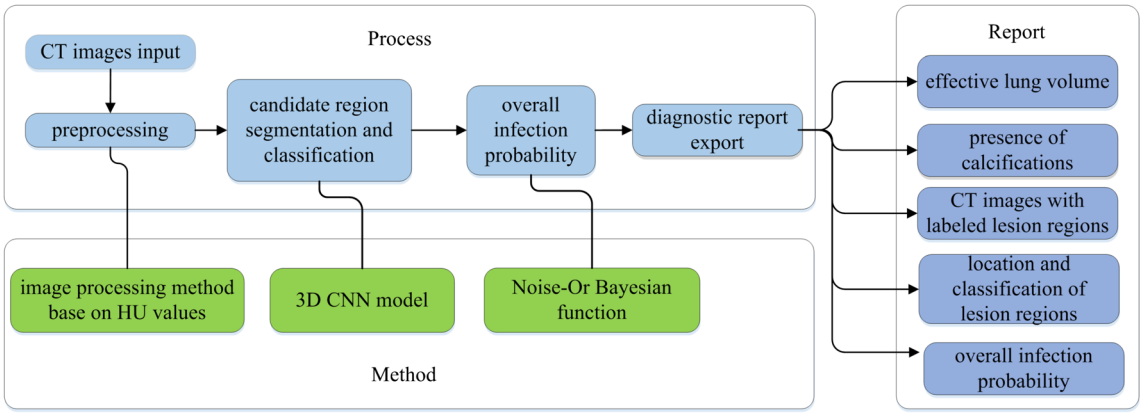

Fig. 1 Process flow chart

the stomach. Furthermore, it was straightforward to reduce irrelevant regions to speed up the convergence rate during the training process and to improve the accuracy of results. Hence, the resampled CT sets were preprocessed to generate masks of the effective lung so as to eliminate the unrelated regions before the training of the deep learning model, as follows:

(1) As the digital grayscale image had the pixel value ranging [0, 255], the resampled CT raw data were converted from the Hounsfield Unit (HU) to the interval of the aforementioned values accordingly. The HU data matrix was clipped within [−1200, 600] (any value beyond this was set to −1200 or 600 accordingly) and then linearly normalized to [0, 255] to fit into the digital image format as shown in Fig. 3a.

(2) A fixed threshold (−200) was used to binarize the resampled CT images, and bones and soft tissues, such as blood vessels and muscles with substantial HU values, were filtered out (Fig. 3b).

(3) All connected components smaller than 0.3 cm^2 and with eccentricity larger than 0.99 were removed to eliminate some high-luminance radial imaging noise. The components (usually clothes and accessories) with the distance to the center of the CT image more than 6.2 cm were also removed. Furthermore, the components with a volume between 450 and 7500 cm^3were kept, as shown in Fig. 3c. The range in the present study was expanded compared

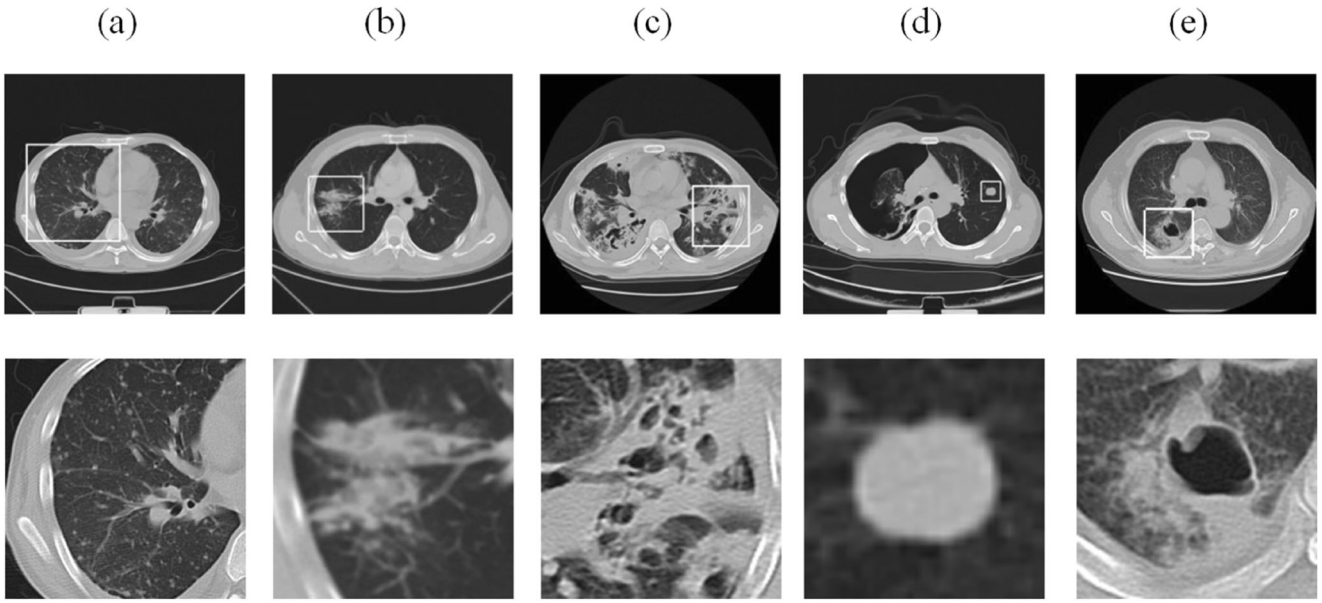

Fig. 2 Five PTB types. **(a)** Miliary; **(b)** infiltrative; **(c)** caseous; **(d)** tuberculoma; **(e)** cavitary. Top: the whole slices. Bottom: the zoomed-in regions

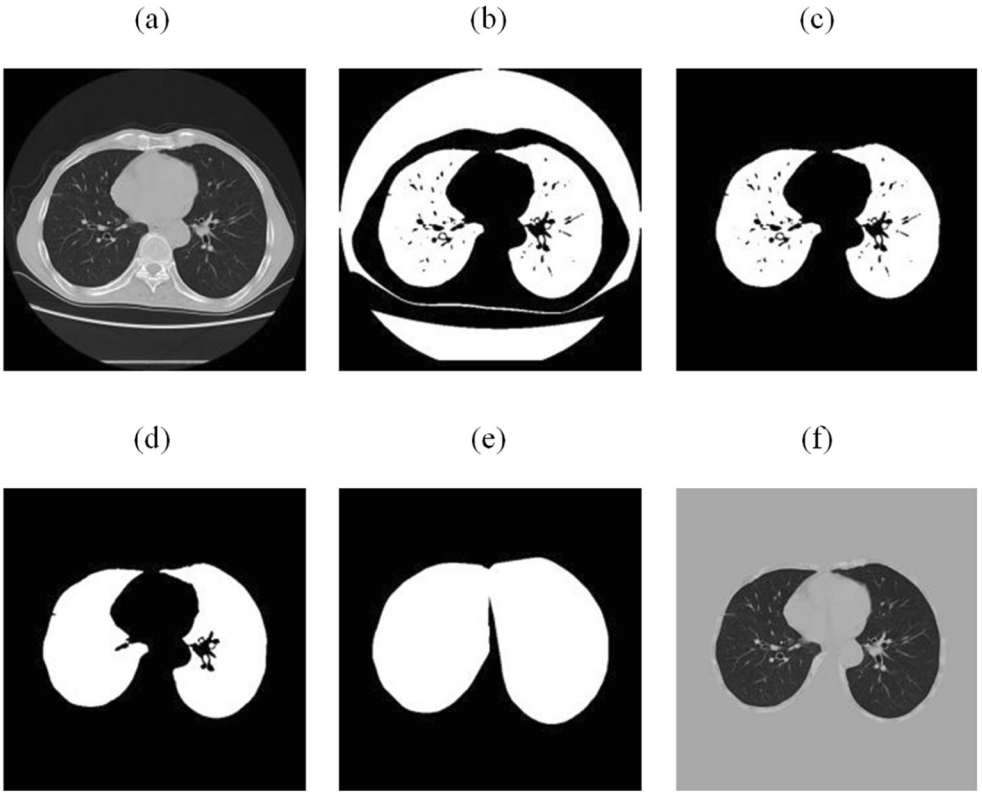

Fig. 3 Image data preprocessing. **(a)** Normalized resampled CT image; **(b)** binarized images with HU equal to −200; **(c)** removal of unrelated regions; **(d)** erosion and dilation; **(e)** convex hull operation to create the mask; **(f)** normalized CT image multiplied by the mask to generate a valid pulmonary region

with those reported by lung nodule detection studies [27], which ranged from 680 to 7500 cm^3. The nodule detection study usually focuses on small regions, while lesions could be more massive for PTB cases.

(4) The extracted mask in step 3 was eroded into two sectors and then dilated to the original size to remove small black holes (Fig. 3d).

(5) Convex hull operation was performed on the effective region, which was extracted from the previous step, to include lesion regions attached to the outer wall of the lung (Fig. 3e).

(6) The matrix data of images in step 1 were multiplied by the masks exported from step 5 to obtain the final effective pulmonary region for further processing. The space out of the mask was filled with 170, which was equivalent to 0 if converted back to HU value (Fig. 3f).

2.4 PTB data process and augment

To reduce the influence of the uneven distribution of different PTB lesion types in the present dataset, types with fewer specimens were expanded to ensure that the sampling frequency of different types is consistent during training. The lesion of miliary, caseous, and tuberculoma cases was duplicated 10 times; cavitary cases were duplicated 5 times during the training to balance the specimen number with the infiltrative type, which was the dominant type among all. At the same time, generic data augment mechanisms, random clipping and left-right flipping were performed on specimens to increase the number of training samples and prevent data overfitting [31].

2.5 Deep learning model for detection and classification

2.5.1 Network structure

DenseVoxNet [32], 3D U-Net [33, 34], and V-Net [35] networks are among the most widely used networks in the domain of medical image segmentation. These networks usually include two network paths: contracting and expanding. The images are firstly fed into the contracting path to finish the down-sampling process, and capture the context information. Then, the up-sampling process is completed in the symmetrical expanding path to obtain precise localization information of the targets.

The segmentation models had to be replaced since they were originally designed to generate the pixel-level mask of designated objects. As the border between a healthy region and the focus of infection usually is usually blurred and indistinct, it is extremely hard to label pixel-level masks for lesion regions of pulmonary tuberculosis. Furthermore, the variable

three-dimensional structure of the lesion regions aggravates this work. Therefore, the segmentation operation used in this study was more like the anchor operation in the object detection algorithm. A Region Proposal Network (RPN) [36] was used to detect the Region of Interesting (ROI) with 3D bounding boxes, instead of pixel-level segmented masks.

Four 3D CNN network models, with different feature extraction structures followed by the same RPN output layer, were designed and evaluated. The first model (referred to as DENSEVOXNET-RPN) used DenseBlock as the feature extraction part, and there have no concatenation operations for feature maps with the same dimensions. The second model (referred to as 3DUNET-RPN) used a 3D U-Net backbone as the feature extraction part. The third model (referred to as VNET-RPN) used the V-Net backbone as the feature extraction part.

In this paper, we used a modified V-Net backbone with inception-resnet [37] block added (fourth model), as shown in Fig. 4. V-Net is a 3D version of U-Net with resnet blocks [38], which has been demonstrated as a effective method on the feature extraction part of lung CT image, in the domain of pulmonary nodule detections. However, compared with nodules the lesion region of PTB is more widely distributed, and its shape and size are more diverse. Therefore, we use inception-resnet to extract features by using convolution kernel with different sizes. The inception-resnet block included convolution operations with different kernel sizes ($1 \times 1 \times 1$, $3 \times 3 \times 3$) to enable the model extracting more detailed features from various receptive fields and to greatly reduce network parameters. Furthermore, the concatenate operation within the inception block was used to amalgamate multi-scale features to enhance the training of feature extraction procedure.

Then, a RPN was used to detect the Region of Interesting (ROI) with 3D bounding boxes, instead of pixel-level segmented masks. Therefore, we name the model VNET-IR-RPN.

The original RPN algorithm [36] is a 2D network and we add an extra z dimension by transforming the description of anchor from 2D version $\{x, y, width, height\}$ to 3D version $\{x, y, z, d\}$, where d referred to the side length of this cube.

2.5.2 Definition of the loss function

The total loss L_{total} function included confidence loss L_{conf}, location regression loss L_{reg}, and classification loss L_{class}. The first two losses were the nodule-detection loss function, while the classification loss was added for PTB study. The confidence loss L_{conf} is a cross-entropy loss that measured whether this proposal was a valid target:

$$L_{conf} = -(p\log(\hat{p}) + (1-p)\log(1-\hat{p})) \quad (1)$$

where p is the ground truth and $\hat{p}$ is the predicted value.

The ground truth bounding box of a PTB lesion is denoted by (G_x, G_y, G_z, G_d) and the bounding box of an anchor is denoted by (A_x, A_y, A_z, A_d), where the first three elements stand for the center point and the last element, for the side length.

The regression labels of the bounding box included the regression of the center point (d_x, d_y, d_z) and the side length d_d

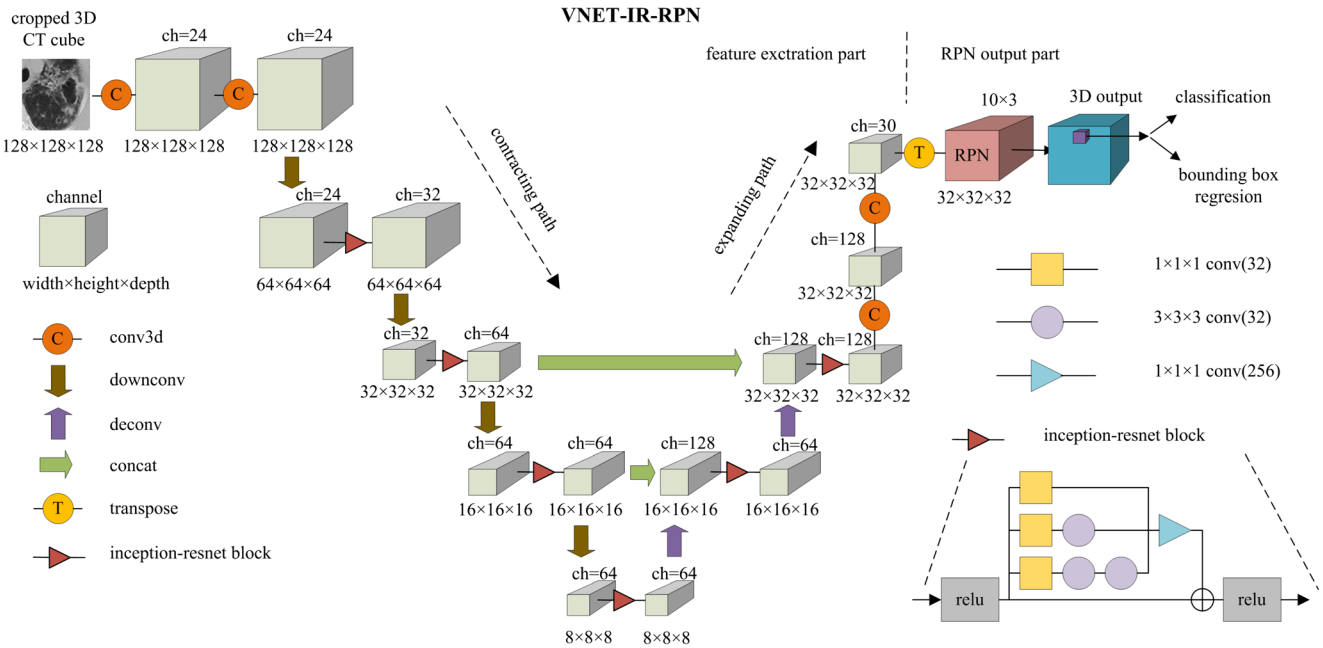

Fig. 4 VNET-IR-RPN network structure

$$\begin{aligned} d_x &= (G_x - A_x)/A_d \\ d_y &= (G_y - A_y)/A_d \\ d_z &= (G_z - A_z)/A_d \\ d_d &= \log(G_d/A_d) \end{aligned} \tag{2}$$

The corresponding predictions were ($\hat{d}_x, \hat{d}_y, \hat{d}_z, \hat{d}_d$). Then, the total regression location loss was defined by

$$L_{reg} = \sum_{k\in\{x,y,z,d\}} S(d_k, \hat{d}_k) \tag{3}$$

where the function S was defined as

$$S(d, \hat{d}) = \begin{cases} |d - \hat{d}|, if\, |d - \hat{d}| > 1 \\ (d - \hat{d})^2, else. \end{cases} \tag{4}$$

L_{class} is the cross entropy loss of five-classification dimension,

$$L_{\text{class}} = -\sum_i y_i \log \hat{y}_i \tag{5}$$

where y_i is the ground truth label and $\hat{y}_i$ is the predicted label.

Intersection over Union (IoU), which is equal to the overlapping area of the bounding boxes of two objects divided by their united area, is an evaluation metric used to measure the accuracy of an object detector on a particular target. It defines the tags of each anchor box in the present study. The anchor box with IoU larger than 0.5 was treated as a positive sample ($p = 1$), while that with IoU < 0.02 was regarded as a negative sample ($p = 0$). Others were neglected during training and validation.

Then, the loss function (L_{total}) is defined as follows:

$$L_{\text{total}} = L_{\text{conf}} + p(L_{\text{reg}} + \lambda L_{\text{class}}) \tag{6}$$

p equals to 1 and 0 when the box is a positive sample and a negative sample respectively; and λ is set to 0.5 according to the setting of Yolo [39], which is a well-tuned deep learning algorithm for 2D object segmentation and identification.

2.5.3 Training process

Patch-based input for training The 3D CT image patches were cropped from the lung images and then individually fed into the network. The Clipped 3D CT image patches were cropped from the lung scans to save GPU memories and then fed into the network individually. The size of the patch was $128 \times 128 \times 128 \times 1$ (height × length × width × channel). All four models' outputs of the last convolution network were reshaped to $32 \times 32 \times 32 \times 3 \times 10$ in the transpose layer; the last two dimensions corresponded to 3 anchors in the RPN network and the 10 regression dimension, respectively. Three different scales of anchors with the side length of 10, 40, and 80 mm were used. The 10 regression dimension was $\{p_i, x_i, y_i, z_i, d_i, t_0, t_1, t_2, t_3, t_4\}$ where p_i is the confidence; x_i, y_i and z_i denote for the center of the candidate; d_i is the side length of the region, and t_{0-4} is the possibility of 5 PTB types individually.

Transfer learning To accelerate the convergence rate of the PTB analysis models, transfer learning was utilized in the study by first training models for the task of lung nodule detection using two open-sourced pulmonary CT datasets LUNA16 and DSB2017, which contained 888 and 2101 lung CT nodule analysis cases, respectively.

Proposed network DENSEVOXNET-RPN, 3DUNET-RPN, VNET-RPN, and VNET-IR-RPN were firstly trained with the above two lung nodule dataset. The outputs of the detection models included the coordinates of the center point, the side length, and the confidence of the region. There were no classification vectors at this stage. Loss function L_{total} was used, while λ was set to 0 to exclude the classification factors as well. The models were relatively easy to converge as the morphological characteristic of nodule compared with tuberculosis. When the nodule detection models well converged, their parameters were used to initialize the network for PTB study.

PTB training In the next PTB training stage, only the output layer and the loss function were modified to include the lesion classification task, while the rest of the network structure remained unchanged.

At the beginning of the training, the PTB analysis network was initialized with the parameters from the pre-trained nodule detection model (as they had the exact same network structures) except for the output layer, which was randomly initialized with the normal distribution.

Performance evaluation A non-maximum suppression algorithm [40] was first performed on detected PTB lesion regions to remove repeated candidate bounding boxes. If the central coordinate of the remaining box was within the radius of the human annotated lesion region, the result was marked as true positive (TP); otherwise, it was false positive (FP). False-negative (FN) indicated that no predicted bounding box corresponded to a human annotated region to measure the number of issues missed by the model. Accordingly, *Recall*, *Precision,* and the more balanced *F*1_*score* were used to measure the performance of the deep learning model:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Quantitative diagnostic report The exported CT images were converted back to the original size for easier review. The final quantitative diagnostic report, based on the detection and classification of PTB information, included the overall infection probability, effective volume of the left and right lungs, classification of lesion type, the spatial location of the infection, and presence of calcifications. The original CT images with corresponding annotated lesion regions were also exported.

3 Overall infection probability of the left and right lungs

According to the confidence level of each single detected lesion, the overall infection level (P) of the left and right lungs was calculated using the probability formula of the Noisy-Or Bayesian function [41] as follows:

$$P = 1 - \prod_i (1 - P_i) \tag{10}$$

where P_i represents the infection possibility of the ith lesion in this single lung.

4 Effective volume of the left and right lungs

The effective volume of lungs has consulting value for doctors in medical diagnosis [42, 43]. The effective volume of a single lung was calculated by extracting the effective region in the original CT images according to the value of HU (threshold equals to −200). By removing the blood vessels, soft tissues, and lesion regions, the volume V_i of a single CT image was calculated as $V_i = S_i \times h$, where S_i is the effective lung region of the ith piece, and h is the physical thickness between two adjacent slices. Then, the total volume (in real physical size) of the effective lung was measured as follows:

$$V_{\text{total}} = \sum_{i=0}^{k} S_i \times h \tag{11}$$

5 The spatial location of the infection

As a 3D system, the number of CT image slices was used instead of coordinate z. The parameters x, y, and d (in pixel size) represented the center point and the side length of this single lesion region. The origin of coordinates of x and y was at the lower-left corner of each CT image.

6 Classification of lesion type

The detected PTB lesion regions were classified into (1) miliary, (2) infiltrative, (3) caseous, (4) tuberculoma, and (5) cavitarytypes.

7 Recognition of the presence of calcification

According to clinical experience, a HU value of more than 120 of the nodule with an effective region of at least 3 pixels indicated the presence of calcification.

8 CT image annotations

The location of lesions was annotated as the bounding box on the CT slices corresponding to the output of the deep learning model, together with their infection probabilities, types, and presence of calcification. Only the image slice with the center of the lesion was labeled to avoid confusion.

Table 1 M1-M4 referred to DENSEVOXNET-RPN, 3DUNET-RPN, VNET-RPN, and VNET-IR-RPN The detection and classification accuracy values were based on the training set. All values were presented with mean standard deviation within the brackets. The *detection accuracy* was the recall rate. The *classification accuracy* was calculated only when it was a true positive region

The average value in five folders		M1	M2	M3	M4
loss		0.201(±0.015)	0.263(±0.027)	0.151(±0.019)	0.134(±0.022)
detection precision	train	0.897(±0.016)	0.892(±0.012)	0.926(±0.023)	0.936(±0.021)
	validation	0.796(±0.041)	0.859(±0.036)	0.882(±0.028)	0.901(±0.023)
classification accuracy	train	0.921(±0.023)	0.918(±0.018)	0.942(±0.013)	0.948(±0.012)
	validation	0.895(±0.028)	0.902(±0.020)	0.923(±0.016)	0.934(±0.014)

9 Results

9.1 Evaluation platform

An Intel i7-8700k CPU together with NVIDIA GPU GeForce GTX 1080ti was used as the testing server.

9.2 K-fold cross-validation

In order to evaluate the generalizability of the designed models, we validated the results with 5-folder cross-validation. In total, 852 CT samples were selected for cross-validation, and the remaining 150 were used as the final test set. In each fold, 752 train sets and 100 validation sets were used. The initial learning rate was set to 0.01, and Stochastic Gradient Descent (SGD) was used as the optimizer. In the experiment, 5-folder cross-validation was used for each model. In order to evaluate the performance of each model, the average loss value of each model (if the loss was not significantly decreased) and the detection and classification accuracy of the test model in the training and verification sets were compared, as shown in Table 1.

9.3 Model performance on the test dataset

The performance of all four 3D CNN models was evaluated on the test set, which consisted of 150 cases, including 75 cases from the PTB group and 75 normal cases from a healthy group, with 412 valid PTB lesion regions. The best model in 5-fold cross-validation was selected. First, the detection accuracy was validated. The classification accuracy was calculated only when a true positive region was observed.

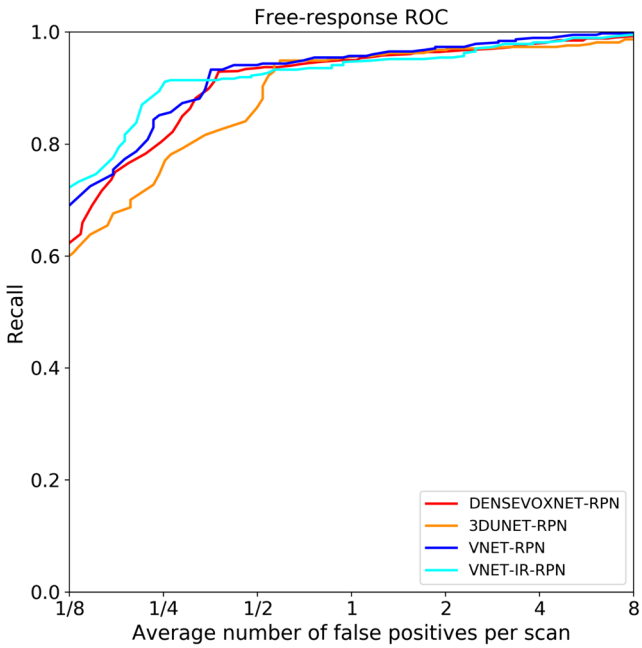

Fig. 5 The FROC curve of different models

Table 2 Results with and without preprocessing

Lesion region	TP	FP	FN	Recall	Precision	f1-score
without preprocess	340	83	72	0.825	0.804	0.814
with preprocessed	354	43	58	0.859	0.892	0.875

The Free-Response Operation Characteristic (FROC) analysis was utilized to evaluate the performance of different models on the test dataset, as shown in Fig. 6. To facilitate directly quantitative comparisons among models, FROC system score was calculated; FROC was the average of the recall at seven predefined false positives per scan (1/8; 1/4; 1/2; 1; 2; 4; and 8).

The corresponding FROC system scores for DENSEVOXNET-RPN 3DUNET-RPN, VNET-RPN, and VNET-IR-RPN were 0.893, 0.875, 0.901, and 0.917, respectively (Fig. 5). The VNET-IR-RPN had the best performance. This result also highlighted the effectiveness and efficiency of inception-resnet blocks in this 3D inspection architecture. Therefore, the VNET-IR-RPN model was used for the rest of this study.

To achieve the maximum value of *F*1_*score*, the threshold was set (classified as lesion region if the predicted probability is higher than the threshold) to 0.38. In the test dataset, 397 candidate regions were detected by the VNET-IR-RPN model, including 354 TP and 43 FP. Moreover, 58 regions were observed as FN. The corresponding *Recall*, *Precision*, and *F*1_*score* were 85.9%, 89.2%, and 0.875, respectively.

Furthermore, it was worthwhile to compare the performance of VNET-IR-RPN on original CT images (Fig. 3a) and preprocessed CT images (only valid lung regions shown in Fig. 3f). The same test benchmark was used; 5.2% bounding boxes were located out of the region of the lung, especially close to the stomach. There were 423 candidate regions detected by the VNET-IR-RPN model, including 340 TP and 83 FP. Moreover, 72 regions were observed as FN. The corresponding *Recall*, *Precision*, and *F*1_*score* were

Table 3 The *Recall, Precision,* and *F*1_*score* from the perspective of whole CT cases

	Recall	Precision	f1-score
Whole PTB case	0.987	0.937	0.961

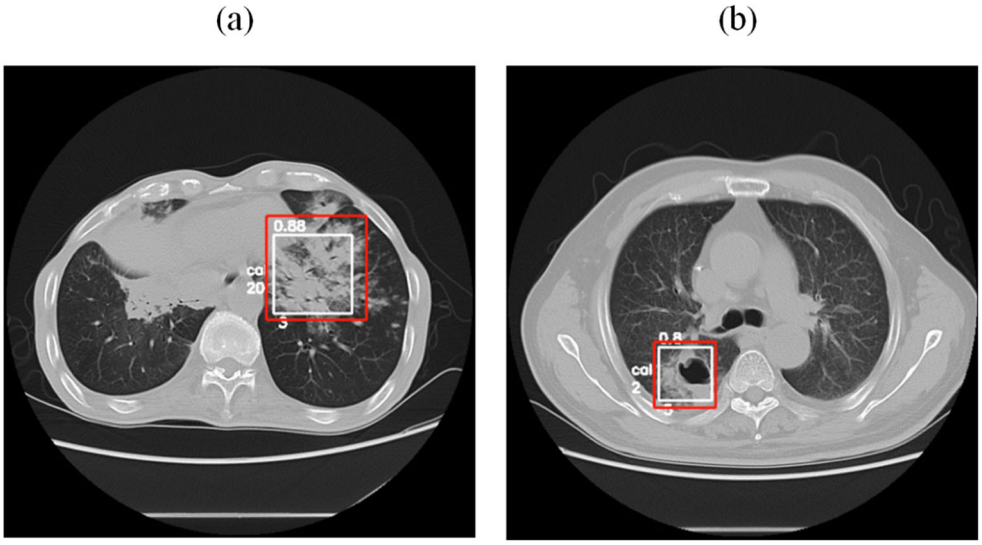

Fig. 6 Detection and classification results of PTB lesion regions. Red line represents the manual labels and the white line is the output of a deep learning network. **(a)** The digit 0.88 on the top of the bounding box denoted the probability of infection in this lesion region; the digit 3 at the bottom of the box for type 3 (caseous) PTB; the "cal" and digit 20 on the left of the box for the presence of calcifications with 20 pixels. **(b)** The digit 0.8 for the probability of infection in this lesion region; the digit 5 at the bottom for type 5 (cavitary) PTB; and the cal and digit 2 on the left for the presence of cal with 2 pixels. (Usually, at least 3 pixels indicated calcification)

82.5%, 80.4%, and 0.814, respectively. There was a 5.9% retrograde for *F*1_*score* if the input CT images were not preprocessed, as shown in Table 2.

From the perspective of a whole PTB case, 79 cases (74 cases from PTB group and 5 cases from a healthy group) were detected to have at least one lesion, while 71 (1 case from PTB group and 70 cases from a healthy group) cases were left as no findings. The *Recall, Precision,* and *F*1_*score* of infected cases as shown in Table 3.

In addition to the spatial label, the 354 detected TP samples were classified using the VNET-IR-RPN model at the same time with detection. The results showed that 322 regions were correctly cataloged, and the classification precision rate was 90.9%. The results of sample detection and classification case are shown in Fig. 6.

9.4 Example of the diagnostic report

An example of an exported diagnosis report, consisting of a summarized description report and a series of images with labeled lesions accordingly, is shown in Table 4 and Fig. 7.

10 Discussion

In this study, we have addressed the thorough research to the pulmonary tuberculosis(PTB), four state-of-the-art 3D deep learning models were adopted to analyze CT images of the lungs. A VNET-IR-RPN backbone with inception-resnet blocks achieved an excellent performance, both with reference to the accuracy of detection and classification. The exported

Table 4 Example of a diagnostic report. The number of CT image slice was used instead of *z*. The parameters *x*, *y*, and *d* (in pixel size) represented the center point and the side length of the lesion region. Cal., Presence of calcification; IP, infection probability

Quantitative diagnostic report of PTB
Name: xxx, Date of Birth: xxx, Gender: xxx, Study Date: xxx

Left lung:
Overall IP: 98.8%
Effective volume: 974.16 (cm^3)
24^{th} slice, x = 367, y = 377, d = 35, type: 2 (infiltrative), IP = 75.0%, Cal.: no
26^{th} slice, x = 400, y = 314, d = 37, type: 2 (infiltrative), IP = 80.0%, Cal.: no
34^{th} slice, x = 359, y = 383, d = 48, type: 2 (infiltrative), IP = 65.0%, Cal.: no
45^{th} slice, x = 323, y = 370, d = 59, type: 5 (cavitary), IP = 75.0%, Cal.: yes

Right lung:
Overall IP: 98.3%
Effective volume: 1352.57 (cm^3)
26^{th} slice, x = 164, y = 196, d = 32, type: 2 (infiltrative), IP = 71.0%, Cal.: no
39^{th} slice, x = 163, y = 315, d = 31, type: 2 (infiltrative), IP = 71.0%, Cal.: no
39^{th} slice, x = 179, y = 244, d = 26, type: 2 (infiltrative), IP = 65.0%, Cal.: no
44^{th} slice, x = 147, y = 226, d = 26, type: 2 (infiltrative), IP = 78.0%, Cal.: no
49^{th} slice, x = 177, y = 325, d = 37, type: 2 (infiltrative), IP = 78.0%, Cal.: no
50^{th} slice, x = 202, y = 239, d = 34, type: 2 (infiltrative), IP = 73.0%, Cal.: no

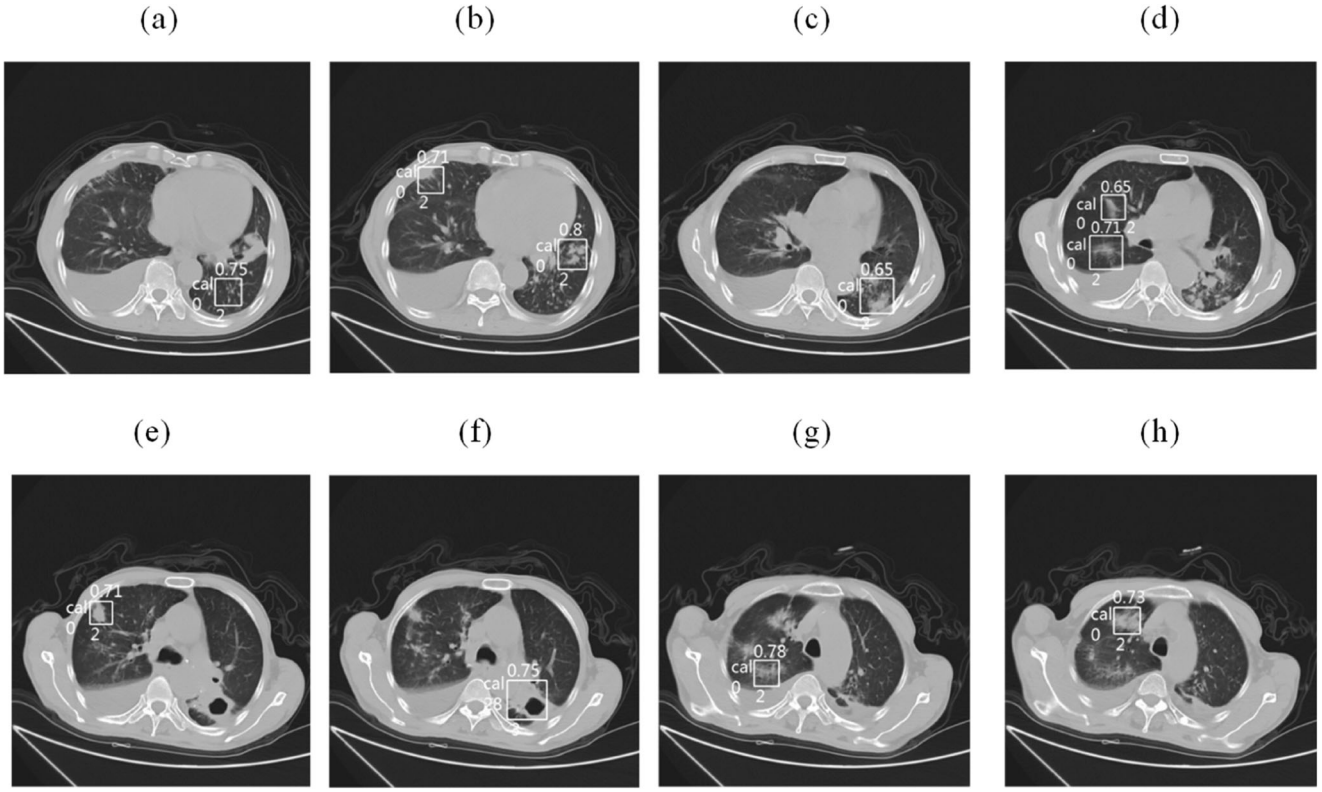

Fig. 7 Example of a diagnostic result corresponding to the same case as in Table 2 **(a)** 24th slice; **(b)** 26th slice; **(c)** 34th slice; **(d)** 39th slice; **(e)** 44th slice; **(f)** 45th slice; **(g)** 49th slice; and **(h)** 50th slice

quantitative report, with overall infection probability, calcification information, effective lung volume, lesions with spatial coordination, and corresponding labeled images, may serve as an effective reference for doctors to make clinical decisions.

11 Conclusions

The pulmonary tuberculosis lesion area in the lung morphology is various, it is difficult to quantify the analysis. In this paper, after studying a large number of related papers and methods, the detection of pulmonary tuberculosis has been effectively studied, and a quantitative diagnosis report has been generated. It makes up for the blank of using deep learning to explore tuberculosis.

This study has a few limitations. First, besides PTB, signs, including lesions in the pleural cavity, pericardium, bones, and lymph nodes, may be used for evaluating intrathoracic TB. In this study, we were mainly focused on the five typical types of pulmonary lesions while ignoring the other signs. Second, due to the lack of relevant training samples, other pulmonary lesions, such as infectious diseases (bacteria, fungi, viruses, parasites, and so on) and non-infectious diseases (tumors and vasculitis and so on) could not be correctly identified and could be misjudged as a certain type of PTB. In our next study, we plan to include samples of pulmonary lesions other than TB. Third, the CT samples in this study were collected from inpatient PTB cases with a relatively massive region of lesions. The current model might be less sensitive to trivial PTB lesions. Moreover, the proposed model might misjudge some of the lesions due to the false-positive rate; therefore, doctors still needed to review the full CT scan to confirm the result.

Future investigations can be improved from the following aspects. First, when extracting effective lung regions, this study applied a fixed threshold method to extract masks. For improvement, more effective detection methods can be used in data preprocessing due to the wide distribution and different types of PTB lesion regions. For example, a better pulmonary mask can be achieved by extracting the lung contour by a deep learning regression method. Second, during the full complete TB treatment cycle for one patient, clinical doctors were more concerned about the changes in PTB lesions. Hence, patients needed to be scanned several times. The comparisons should be made before, during, and after the therapy to assess the treatment effect. An artificial intelligence system can be used in the future to analyze all CT cases of one patient along the time sequence with a quantitative comparison of the whole PTB treatment.

Acknowledgments This study was supported by the China National Science and Technology Major Project Fund (20182X10101-001).

This study was supported by the National Key Research and Development Program of China (2019YFC0840600 & 2019YFC0840609).

Author's contributions Wei Wu and Xukun Li initiated the paper and provided guidance on the concept. Xukun Li and Yukun Zhou developed the network architecture and data/modeling infrastructure, training, testing setup, and statistical analysis. Wei Wu, Xukun Li and Peng Du wrote the manuscript. Wei Wu and Min Xu provided clinical expertise and guidance on the study design. Guanjing Lang and Min Xu created the datasets, interpreted the data, and defined the clinical labels. Wei Wu supervised the project.

Data availability All data included in this study are available from the corresponding author upon reasonable request.

Compliance with ethical standards

Competing interests All authors declare no competing interests.

References

1. Bibbins-Domingo K, Grossman DC, Curry SJ et al (2016) Screening for latent Tuberculosis infection in adults. JAMA 316(9):962–969
2. Ankrah AO, Glaudemans AWJM, Maes A, van de Wiele C, Dierckx RAJO, Vorster M, Sathekge MM (2018) Tuberculosis[J]. Semin Nucl Med 48(2):108–130
3. GBD Tuberculosis Collaborators (2018) The global burden of tuberculosis: results from the global burden of disease study 2015. Lancet Infect Dis 18(3):261–284
4. Li T, Shewade HD, Soe KT, Rainey JJ, Zhang H, du X, Wang L (2019) Under-reporting of diagnosed tuberculosis to the national surveillance system in China: an inventory study in nine counties in 2015. BMJ Open 9(1):e021529
5. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
6. Firmino M, Angelo G, Morais H, Dantas MR, Valentim R (2016) Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. Biomed Eng Online 15(1):2
7. Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 14(3):390–406
8. Muyoyeta M, Kasese NC, Milimo D et al (2017) Digital CXR with computer aided diagnosis versus symptom screen to define presumptive tuberculosis among household contacts and impact on tuberculosis diagnosis. BMC Infect Dis 17(1):301
9. Hinton B, Ma L, Mahmoudzadeh AP et al (2019) Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers: a case-case study. Cancer imaging 19(1):41
10. Yanagawa M, Niioka H, Hata A et al (2019) Application of deep learning (3-dimensional convolutional neural network) for the prediction of pathological invasiveness in lung adenocarcinoma: a preliminary study. Medicine 98(25):e16119
11. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, du N, Fan W, Xie X (2019) AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys 46(2):576–589
12. Wang S, Summers RM (2012) Machine learning and radiology. Med Image Anal 16(5):933–951
13. Javaid M, Javid M, Rehman MZU, Shah SIA (2016) A novel approach to CAD system for the detection of lung nodules in CT images. Comput Methods Prog Biomed 135:125–139
14. Kobayashi H, Ohkubo M, Narita A et al (2016) A method for evaluating the performance of computer-aided detection of pulmonary nodules in lung cancer CT screening: detection limit for nodule size and density. Br J Radiol 90(1070):20160313
15. Gupta A, Saar T, Martens O, Moullec YL (2018) Automatic detection of multi-size pulmonary nodules in CT images: large-scale validation of the false-positive reduction step. Med Phys 45(3): 1135–1149
16. Li X, Guo F, Zhou Z, Zhang F, Wang Q, Peng Z, Su D, Fan Y, Wang Y (2019) Performance of deep-learning-based artificial intelligence on detection of pulmonary nodules in chest CT. Chinese journal of lung cancer 22(6):336–340
17. Setio AAA, Traverso A, de Bel T, et al (2016) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge, arXiv preprint arXiv:1612.08012
18. Tu X, Xie M, Gao J, Ma Z, Chen D, Wang Q, Finlayson SG, Ou Y, Cheng JZ (2017) Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in CT images with convolutional neural network. Sci Rep 7(1):8533
19. XieY ZJ, Xia Y (2019) Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT. Med Image Anal 57:237–248
20. Sim Y, Chung MJ, Kotter E et al (2019) Deep convolutional neural network–based software improves radiologist detection of malignant lung nodules on chest radiographs. Radiology 294(1):182465
21. Huang W, Hu L (2019) Using a Noisy U-net for detecting lung nodule candidates. IEEE access, PP(99):1-1
22. Ciompi F, Chung K, Van Riel SJ et al (2017) Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 7:46479
23. Cao H, Liu H, Song E, et al (2020) A two-stage convolutional neural networks for lung nodule detection. IEEE journal of biomedical and health informatics PP(99):1-1
24. Zuo W, Zhou F, He Y, et al (2019) Automatic classification of lung nodule candidates based on a novel 3D convolution network and knowledge transferred from a 2D network[J]. Med Phys, 46(12)
25. de Wit J, Hammack D (2017) Kaggle national data science bowl 2017 2nd place code. https://github.com/juliandewit/kaggle_ndsb2017
26. Pezeshk A, Hamidian S, Petrick N, et al (2018) 3-D Convolutional Neural Networks for Automatic Detection of Pulmonary Nodules in Chest CT. IEEE Journal of Biomedical and Health Informatics PP(99):1–1
27. Liao F, Liang M, Li Z, et al (2017) Evaluate the malignancy of pulmonary nodules using the 3D deep leaky Noisy-or network. arXiv preprint arXiv:1711.08324

28. Huang X, Lei Q, Xie T, Zhang Y, Hu Z, Zhou Q (2020) Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on CT images. Knowl-Based Syst 204:106230
29. Xu X, Wang C, Guo J et al (2020) DeepLN: a framework for automatic lung nodule detection using multi-resolution CT screening images. Knowledge-Based Syst 189:105128
30. Committee on Infectious Disease Radiology Chinese Society of Radiology (2018) Expert consensus on imaging diagnosis of hierarchical diagnosis and treatment for tuberculosis. Electronic Journal of Emerging Infectious Diseases, 3(2):118–127 (Chinese)
31. Srivastava N, Hinton G, Krizhevsky A et al (2014) Dropout: a simple way to prevent neural networks from Overfitting. J Mach Learn Res 15(1):1929–1958
32. Yu L, Cheng J Z, Dou Q, et al (2017) Automatic 3D cardiovascular MR segmentation with densely-connected volumetric ConvNets, MICCAI 2017
33. Çiçek, Özgün, Abdulkadir A, Lienkamp SS, et al (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv preprint arXiv: 1606.06650
34. Chen J, Yang L, Zhang Y, et al (2016) Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. arXiv preprint arXiv:1609.01006
35. Milletari F, Navab N, Ahmadi SA (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. arXiv preprint arXiv:1606.04797
36. Ren S, He K, Girshick R, et al (2016) Faster R-CNN: towards real-time object detection with region proposal networks. arXiv preprint arXiv :1506.01497
37. Szegedy C, Ioffe S, Vanhoucke V, et al (2016) Inception-v4, inception-ResNet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261
38. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society
39. Redmon J, Divvala S, Girshick R et al (2016) You only look once: unified. Real-Time Object Detection IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
40. Neubeck A, Gool LJV (2006) Efficient Non-Maximum Suppression. 18th International Conference on Pattern Recognition (ICPR 2006), 20–24 August 2006, Hong Kong, China. IEEE Computer Society
41. Oniśko A, Druzdzel MJ, Wasyluk H (2001) Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. Int J Approx Reason 27(2):165–182
42. Washko GR, Fan VS, Ramsey SD, Mohsenifar Z, Martinez F, Make BJ, Sciurba FC, Criner GJ, Minai O, Decamp MM, Reilly JJ, National Emphysema Treatment Trial Research Group (2008) The effect of lung volume reduction surgery on chronic obstructive pulmonary disease exacerbations. Am J Respir Crit Care Med 177(2):164–169
43. Klooster K, Ten Hacken NHT, Franz I et al (2014) Lung volume reduction coil treatment in chronic obstructive pulmonary disease patients with homogeneous emphysema: a prospective feasibility trial[J]. Respiration, 2014 88(2):116–125

Xukun Li receieved his master degree in Control Engineering from Zhejiang University of Technology. He has worked in the area of deep learning, such as face recognition, object detection and classification, as well as the analysis of CT images and pathological images, such as detection and segmentation of tuberculosis, liver cancer and cervical cancer, etc.