A CADe system for nodule detection in thoracic CT images based on artificial neural network

Xinglong LIU[1], Fei HOU[2*], Hong QIN[3] & Aimin HAO[1]

[1]*State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China;*
[2]*School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore;*
[3]*Department of Computer Science, Stony Brook University, Stony Brook 11790, USA*

Abstract   Lung cancer has been the leading cause of cancer-related deaths in 2015 in United States. Early detection of lung nodules will undoubtedly increase the five-year survival rate for lung cancer according to prior studies. In this paper, we propose a novel rating method based on geometrical and statistical features to extract initial nodule candidates and an artificial neural network approach to the detection of lung nodules. The novel method is solely based on 3D distribution of neighboring voxels instead of user-specified features. During initial candidates detection, we combine organized region properties calculated from connected component analysis with corresponding voxel value distributions from statistical analysis to reduce false positives while retaining true nodules. Then we devise multiple artificial neural networks (ANNs) trained from massive voxel neighbor sampling of different types of nodules and organize the outputs using a 3D scoring method to identify final nodules. The experiments on 107 CT cases with 252 nodules in LIDC-IDRI data sets have shown that our new method achieves sensitivity of 89.4% while reducing the false positives to 2.0 per case. Our comprehensive experiments have demonstrated our system would be of great assistance for diagnosis of lung nodules in clinical treatments.

Keywords   artificial neutral network, Lung nodule, computed tomography, computer aided detection, nodule detection

1   Introduction

According to the most recent statistics from the American Cancer Society [1], lung cancer has been the leading cause of cancer-related deaths in 2015 in United States. A total of 1658370 new cancer cases and 589430 cancer deaths were projected to occur in the United States in 2015, while the overall five-year survival rate for lung cancer is merely 16%. The survival rate would increase to 52% if the nodules could be localized, and decrease to 4% if nodules are metastasized. Therefore, it is of great importance to detect lung cancer at early stages.

---

* Corresponding author (email: houfei@ntu.edu.sg)

For early cancer detection, computed tomography (CT) has been proven a more sensitive imaging modality because of its ability of providing reliable image textures for detection of small nodules. Nonetheless, this might also give rise to more ambiguities and massive data, which makes it hard for radiologists to distinguish nodules and healthy structures. In recent years, computer aided detection (CADe) systems have become promising tools to make more precise diagnosis while reducing the overall workload of radiologists [2].

It is commonly agreed that detection of pulmonary nodules could provide great assistance for better pulmonary cancer diagnosis. However, pulmonary nodule detection tends to be a nontrivial task, which is not only because the appearance of nodules varies in a wide range, but also because the low contrast of nodule densities persists in comparison with adjacent vessels and lung tissues. Despite earlier difficulties, pulmonary nodule detection systems using CT imaging modality have been making great progresses over past decades [3,4]. Generally speaking, available CADe systems consist of three steps: image preprocessing, initial nodule candidate extraction, and false positives (FPs) reduction.

1.1 Previous work

In the first stage, the system aims at reducing data size and localizing search space to lungs, where a segmentation of lung from the entire chest volume is usually required. Because of large image contrast between lung region and the body tissues, image intensity-based thresholding or edge preserving smooth is effective and most commonly applied [5]. Ukil et al. [5] and van Rikxoort et al. [6] further enhanced this method as a pre-processing step for lung lobe segmentation. Gao et al. [7] proposed another threshold-based approach consisting of four steps which take anatomical information into consideration. Although simple and fast, it is extremely hard to determine the threshold since it is greatly related with the image acquisition types, protocols, intensities, especially when there are severe lung pathologies. Meanwhile, Itai et al. [8] extracted the lung region with a 2D parametric deformable model using the lung borders as an external force. Silveira et al. [9] adopted a 2D geometrical level set active contour initialized at the chest boundary. The main drawbacks of deformable modeling approaches are their excessive sensitivity to initialization and the inability of traditional external forces to capture natural inhomogeneity in lung regions.

After locating the search space, initial nodule candidates detection is the next essential step. Conformal nodule filtering [10] and un-sharp masking [11] can enhance nodules while suppressing other structures by the way of single or multiple gray-level thresholding [12]. Also, 3D cylindrical and spherical filters are applied to detect small lung nodules. Circular or semi-circular nodule candidates can be detected by template matching. However, these methods are not robust against nodules' various appearances. In [13], they proposed morphological operator to detect lung nodules. The drawbacks of these methods are difficulties in detection of pulmonary wall nodules. Meanwhile, there are also other pattern-recognition techniques such as clustering [14], linear discriminate functions, connected component analysis [15], gray scale distance transform [14] and patient specific priori model [16], etc.

After getting the initial nodule candidates, a rule-based filtering operation [17,18] is usually employed to reduce obvious FPs at low cost, so that their influence on more time-consuming processes can be eliminated. To further reduce FPs, a number of supervised reduction approaches have been developed towards the characteristics of nodules, such as linear discriminant analysis (LDA) [13], artificial neural network (ANN) [19] and support vector machine (SVM) [20, 21]. Approaches based on hierarchical vector quantization (VQ) [22] are also proposed to overcome drawbacks of global thresholding, providing comparable detection rates at lower computational cost.

Most recently, Shen et al. [23] proposed a multi-crop convolutional neural network to automatically extract nodule's salient information for classification of malignancy suspiciousness. They directly took raw nodule patches as inputs and applied an end-to-end machine learning architecture. Cheng et al. [24] carried out a comprehensive study on CAD systems for nodule malignancy detection based on deep learning technologies. Ginneken et al. [25] proposed a method based on features trained from natural images for nodule detection in CT scans, which shows great potentials of CNN features in medical detec-
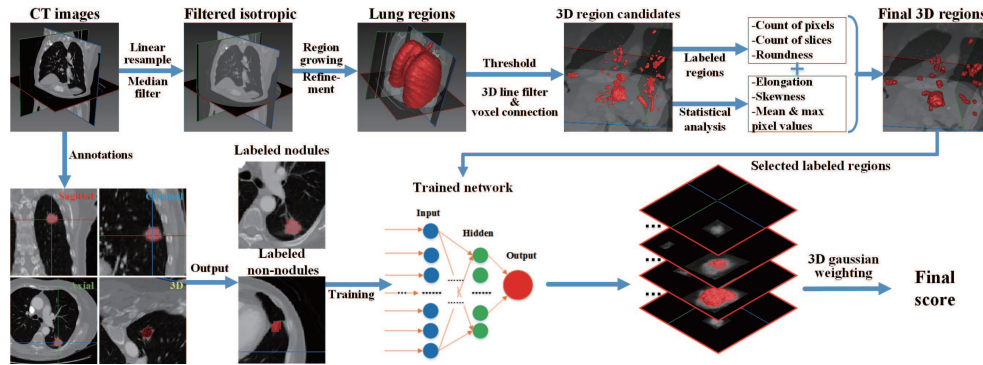
Figure 1 (Color online) Pipeline of the proposed method. The top row constitutes the initial candidate extraction. The bottom row constitutes the training and forecasting.

tion tasks. Lu et al. [26] proposed a hybrid method unifying morphological operation, dot-enhancement based on Hessian matrix, fuzzy connectedness segmentation, local density maximum algorithm, geodesic distance map, and regression tree classification. Colin et al. [27] investigated the performance of the state-of-the-art CAD systems for pulmonary nodule detections. Their experiments have shown that the best commercial CAD system could achieve the best performance with sensitivity of 82% at an average of 3.1 false positive detections per scan. Chen et al. [28] proposed a method to bridge the gap between computational and semantic features. They applied the multi-task regression scheme with Haar-like features to better describe nodules. Setio et al. [29] proposed a multi-view approach to extract nodule features. They applied different classification schemes towards comprehensive comparison among different methods. They also proposed an automatic method for large solid nodule detection based on morphological processing, thresholding, and feature descriptions. Their CAD system has manifested great performance for large nodule detection.

1.2 Our approach

Inspired by the aforementioned work, this paper presents an artificial neural network based approach to the extraction of lung nodules from chest CTs. Different from classical methods, we focus on the inner structures of nodule voxels and apply ANN to generalize these characteristics. We are working in 3D space consisting of only voxels instead of processing slice by slice in CT volume. The major contributions of our work include:

• A framework which can be easily integrated into existing CADe systems and rapidly accommodate and process new data streams with few human interactions.

• A novel voting method based on geometrical and statistical features to better extract initial candidate regions while suppressing ambiguous structures.

• A nodule detection approach with multiple trained ANNs based on 3D massive sampling of candidate voxels instead of user-specified features with a goal to reduce various false positives.

• Reaching a comparable sensitivity as high as 89.4% while decreasing a FPs/case rate as low as 2.0/case.

The pipeline of our method is documented in Figure 1. Firstly, we re-sample input CT data using a linear interpolation method to make data isotropic. Secondly, we apply median filter to smooth the image and adopt a 3D region growing method to extract lung regions from the smoothed images. After that, to recover excluded nodules attached to lung walls, we will conduct refinement using the "rolling ball" [30] method. With lung region extracted, we threshold the reduced data and use a Hessian based line filter to remove vessels and airways. Next, we organize voxels from CT data into regions using connected voxel analysis and compute the voting score based on geometrical and statistical features for each region. Following that, selected regions are sent to trained ANN and outputs from each ANN are organized using the scoring method. The final scores indicate probability for each region of being nodules. Finally, We use logical AND to operate on thresholded final scores from all multiple ANNs to calculate final true
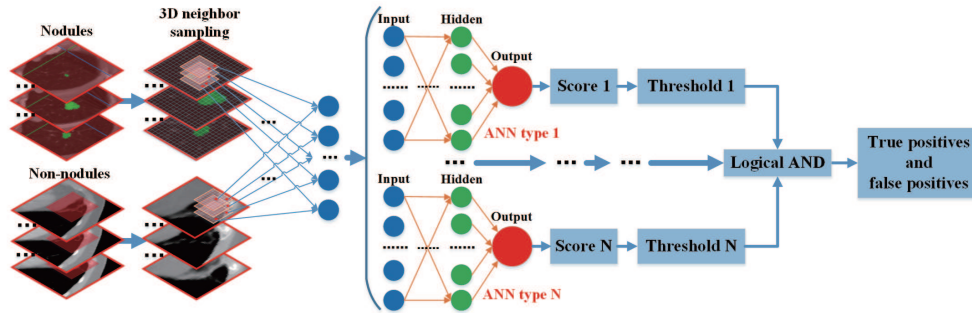
Figure 2 (Color online) Architecture of the proposed network. Each network consists of multiple ANNs and all outputs from these ANNs are organized using a 3D scoring method. Finally, a logical AND operator is used to organize all thresholded data and to distinguish true positive and false positive nodules.

positive (TP) and false positive (FP) nodules. The architecture of our multiple ANNs is illustrated in Figure 2.

This paper is organized as follows. Section 2 describes the data source we have been training and validating our ANN detailedly. Then, the architecture, creation, and training of the proposed ANN are described in Section 3. After that, Section 4 introduces our potential lung nodule extraction method. Following that, validation on 107 CT cases and comparisons with other CADe systems are presented in Section 5, and statistics table for sensitivity rate and FPs/case, as well as figures showing typical detected, missed and mistaken FP nodules are also shown in this section. Finally, a brief discussion and conclusion in Section 6 is presented.

2 Materials

Data source selection is of high importance for creating and testing a CADe system. Both our training and testing procedures are based on the lung image database consortium and image database resource initiative (LIDC-IDRI) [31] which consist of totally 1012 chest CT image cases and 1356 nodules. Each case in the database includes images from a clinical thoracic CT scan and an associated eXtensible Markup Language (XML) file that records the locations of the image pixels on the nodule boundary in each image and nine characteristics of the nodule malignancy given by up to four experienced thoracic radiologists [32].

To test and validate our methods, we choose 107 chest CTs with various imaging qualities of different patients from the LIDC-IDRI image data base. The CT scanning parameters include reconstruction intervals of 0.6–3.0 mm [0.6 mm (n=1, 1% of the database), 1.25 mm (27, 25.2%), 2 mm (6, 5.6%), 2.5 mm (60, 56%), 3 mm (13, 12.2%)], and tube currents of 120 mA with 200–400 kVp. Each slice is sampled with 512×512 pixels, with in-plane spacing as 0.4–1.0 mm [0.4–0.6 mm (7, 6.8%), 0.6–0.8 mm (76, 73.8%), 0.8–1.0 mm (25, 23.4%)]. Nodule information for all the data sets is stored in an XML format file along with image data from LIDC-IDRI. Among all the 611 nodules in 107 data sets, we are focusing on 252 (41%, average 2.4/case) nodules whose diameters are larger than 3 mm [3–15 mm (180, 71.4%), 15–30 mm (46, 18.3%), >30 mm (26, 10.3%)].

3 Network creation and training

3.1 Network architecture and creation

Different from methods based on human-specified features, ANN based, supervised nonlinear image processing techniques are perfect alternatives to retrieve and distinguish the distribution of nodules from other different structures.

Since different lung nodules may vary in sizes, locations, and pixel distributions, one single ANN can

not describe all the features of all types of nodules. To reduce various type of FPs, we extend one ANN to a group of ANNs and each ANN in the group corresponds to one specific type of nodule, e.g., ANN No.1 is trained to distinguish nodules attached to lung walls, ANN No.2 is trained to distinguish nodules with GGO features, and so on. The network architecture is illustrated in Figure 2.

Each ANN is designed as a typical multi-layer, fully connected network and consists of one input layer, one hidden layer, and one linear output layer. The neutral number for input layer is related to maximum size of the nodule while the number of hidden units is selected by use of a method for designing the structure of an ANN [33]. The ANN is working directly on voxel data from the chest CT images and should output the corresponding 'likeness' to nodules. The input to the network consists of voxel values P which is defined as $P = \{I_{V_r}(x-i, y-j, z-k) | -r < i, j, k < r\}$, where $I(x, y, z)$ is the voxel value located at space index (x, y, z), V_r is the subvolume which centers at (x, y, z) from the input data, r represents the sample radius, and P is eventually arranged as a single row vector.

3.2 Network training

The arranged vector consists of neighboring values of voxel $I(x, y, z)$. However, pixel size with a CT section is generally different from the reconstruction interval between sections. Besides, the reconstruction interval is often different at different institutions and under different imaging protocols. All of these factors make it meaningless to define 'neighbors' for voxels. To reduce such variations in different CT data, we rectify and normalize the data as pre-processing. We apply linear interpolation techniques to convert the original data into isotropic volume data. After that, the voxel values of the isotropic volumes are linearly scaled from -1000 Hounsfield Unit (HU) to 1000 HU while values $\leqslant -1000$ HU correspond to 0.0 and values $\geqslant 1000$ HU correspond to 1.0.

$$T(x, y, z) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{x^2 + y^2 + z^2}{2\sigma^2}\right), & \text{nodule,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The output of the ANN is a continuous value indicating the likelihood for the corresponding voxel of being nodules. During training, to enhance lung nodules and suppress other structures such as vessels, we set each teaching volume from ground truth using a 3D Gaussian distribution with standard deviation σ. The distribution represents the likelihood in the region of being a nodule.

The peak of it should be at the center of the nodule while the longer distance from the center, the lower likelihood the teaching output of the voxel gets. For non-nodule teaching volumes, the teaching output should be always zero.

To train the ANN, we denote the error to be minimized by

$$E = \frac{1}{C} \sum_{V_i \in V} \sum_{(x,y,z) \in V_i} \left\{ T_i(x, y, z) - O_i(x, y, z) \right\}^2, \tag{2}$$

where C is the total number of training voxels, $T_i(x, y, z)$ and $O_i(x, y, z)$ are respectively the teaching and output values for the ith training volume at coordinate (x, y, z). V denotes all the teaching sub-volumes in teaching set. V_i is the ith volume of V. The expert ANN is trained by a linear-output back-propagation (BP) algorithm. In our implementation, we apply the RPROP [34] algorithm instead of the original BP method. After training, the expert ANN is expected to output highest value when the voxel is located at the center of a nodule and a lower value while the distance from the voxel to the center increases.

4 Potential lung nodule extraction

Thoracic CT provides tremendous information to assist doctors with better diagnosis and treatment. However, the tremendous information also causes ambiguities and significant time expenses during processing for computer aided detection (CADe) systems. It is essential to reduce data complexity and extract the region of interests (ROIs) for both better performance and better efficiency.
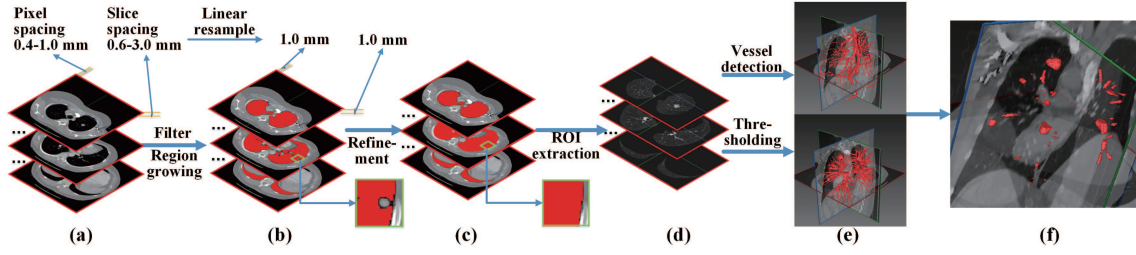
Figure 3 (Color online) Pipeline of the potential nodule extraction procedure. (a) Original data. Pixel and slice spacing varies in different data sets. (b) Re-sampled data. The data set is isotropic and pixel spacing as well as slice spacing are interpolated to 1 mm. (c) Extracted lung regions. Juxta-wall nodules are recovered. (d) ROI filtered lung regions. (e) Top: filtered vessels; bottom: thresholded candidates. (f) Final potential nodule candidates.

The pipeline of our potential lung nodule extraction procedure is described in Figure 3. In our implementation, we re-sample the CT data set using a linear interpolation method to make it isotropic. Then, we apply a region growing approach together with the "rolling ball" refinement procedure [30] to extract the correct lung region. The refinement procedure ensures nodules attached to lung walls could be extracted correctly. After that, we threshold the original data from −1000 HU to 1000 HU in the region of the extracted lung mask. Within these extracted regions, we compute the vesselness for each voxel and subtract vessels from the thresholded lung region. Finally, we combine the statistical and geometrical features of these candidate regions and use the voting score to decide the final lung nodule candidates.

4.1 Lung extraction

It is an essential requirement for CADe systems that input data set should be isotropic. However, most clinical CT images are totally anisotropic with different inner pixel spacings and slice spacings. Based on the characteristics of radioactive decay along distance, it is reasonable to use a linear interpolation method to re-sample data set to make it isotropic. Re-sampled results are described in Figure 3(a) and (b).

Meanwhile, to reduce data size and ambiguities, we apply a region growing approach to extract the lung regions from CT data. First, the upper threshold α_1 and lower threshold α_2 are calculated based on voxel values of two user placed seeds within left and right pulmonary lobes respectively. Then, the two seeds are treated as the initial voxels of lung voxel group. We recursively search the connected 22 neighbors of each voxel in the group and put voxels with values between α_1 and α_2 into the group until no more voxels are included. Due to intensity features, nodules attached to lung walls can not be extracted using the region growing process. Therefore, we use the "rolling ball" [30, 35] approach to refine region boundaries. With the refined lung region as mask, we extract the region of interest (ROI) from original data for further processing. The original data, lung region, and ROI are described in Figure 3(a)–(c).

4.2 Vessel regions removal by 3D line filter

Most false positives (FPs) are caused by vessel regions. Therefore, we use 3D line filter [36] to remove vessels from extracted regions. The 3D line filter is based on eigenvalue analysis of Hessian matrix. Then, the output of the filter is defined as

$$
\lambda = \begin{cases} |\lambda_3| \left(\dfrac{\lambda_2}{\lambda_3}\right)^{\gamma_{23}} \left(1 + \dfrac{\lambda_1}{|\lambda_2|}\right)^{\gamma_{12}}, & \lambda_3 < \lambda_2 < \lambda_1 \leqslant 0, \\ |\lambda_3| \left(\dfrac{\lambda_2}{\lambda_3}\right)^{\gamma_{23}} \left(1 + \alpha \dfrac{\lambda_1}{|\lambda_2|}\right)^{\gamma_{12}}, & \lambda_3 < \lambda_2 < 0 < \lambda_1 \text{ and } \lambda_1 < \dfrac{|\lambda_2|}{\alpha}, \\ 0, & \text{otherwise}, \end{cases}
\tag{3}
$$

where $\lambda_1, \lambda_2, \lambda_3$ $(\lambda_1 > \lambda_2 > \lambda_3)$ represent the eigenvalues of H, γ_{12} $(\gamma_{12} \geqslant 0), \gamma_{23}$ $(\gamma_{23} \geqslant 0)$ and α are control parameters.
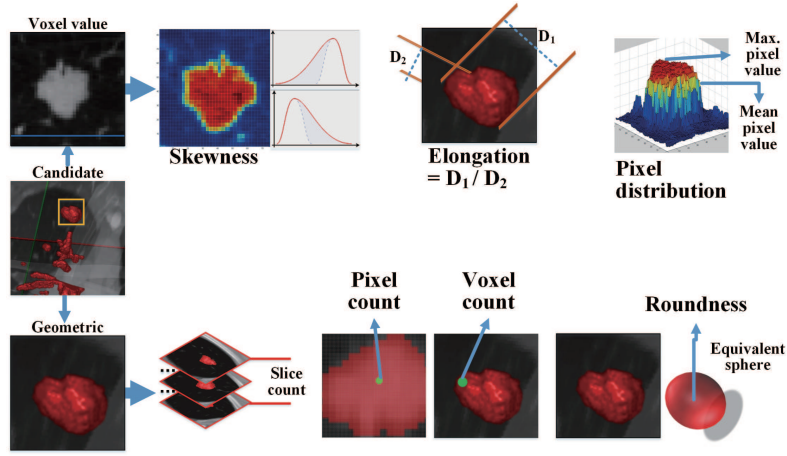
Figure 4 (Color online) Features used for voting in potential nodule extraction. Both geometrical and statistical features are used.

In our implementation, we adopt a multi-scale approach based on the size σ_H used to compute H to detect vessels of different sizes and remove these vessel regions by threshold processing. On the other hand, it has commonly been recognized [32] that HU for almost all lung nodules are ranging from -1000 HU to 1000 HU. Therefore, we threshold voxel values with extracted ROI in the last step as mask and subtract the detected vessels. Finally, we obtain regions for voting process. Thresholded regions and vessels are described in Figure 3(e).

4.3 Nodule candidates extraction

Based on thresholded and subtracted regions from the last step, we will first conduct morphological opening process with a very small radius to remove tiny regions. Then all valid voxels are classified into different groups through connected component analysis. The image with classified voxels is denoted by mask I_{m}. Based on the 3D mask image I_{m} and the original image I, we obtain 3D voxel value image $I_{\mathrm{v}} = I \cdot I_{\mathrm{m}}$, where $\cdot$ means voxel-by-voxel multiplication. Based on I_{m} and I_{v}, we have calculated four geometrical and three statistical features of candidate regions to vote for final nodule candidates. These features are described in Figure 4.

The voting scores for slice count, pixel count, and voxel count are defined as

$$f_{i,k} = 1 + \exp\left(\frac{N_{i,k} - N_{k,\min}}{N_{k,\max} - N_{k,\min}}\right), \tag{4}$$

where $k = 1, 2, 3$, i denotes the i-th 3D region, $N_{i,k}$ represents corresponding count of slices ($k = 1$), pixels ($k = 2$) and voxels ($k = 3$) within this region, $N_{k,\min}$ and $N_{k,\max}$ are minimum count and maximum count assigned by the input. These three scores are designed to give a penalty to candidates with too tiny or too large size. Meanwhile, for candidate regions whose slice count, pixel count, and voxel count are between $N_{k,\min}$ and $N_{k,\max}$, these scores will exponentially increase, because these three parameters are relatively stable during calculation for each region, larger values are more preferred.

The voting score for roundness is defines as

$$f_{i,4} = \frac{A_h(r)}{A_{\mathrm{m}}}, \tag{5}$$

where A_{m} is the measured area of the region, which is equal to the count of voxels on region surface, r and A_h are the radius and the area of the equivalent sphere respectively, and $A_h(r) = 4\pi r^3$. This score is used to calculate the comparability between a region and a sphere. This is based on the assumption that nodules, especially solid isolated nodules, are distributed like sphere in 3D space.

The voting score for skewness and elongation is defined as

$$
f_{i,k} = \begin{cases} 1 + \dfrac{(D_{k,\max} - D_{i,k})}{D_{\max}}, & D_{i,k} \leqslant D_{k,\max}, \\ \dfrac{1}{\exp\left(D_{i,k} - D_{k,\max}\right)}, & D_{i,k} > D_{k,\max}, \end{cases} \tag{6}
$$

where $D_{i,k}$ is the skewness ($k = 5$) and elongation ($k = 6$) of the voxel value distribution, $D_{k,\max}$ is the maximum preferred value assigned by the input. These two scores are used to measure the voxel value distributions. For a "perfect" lung nodule, the ratio of its long and short axis should be near 1, and the distribution should satisfy a standard Gaussian distribution with skewness around 0. We use skewness and elongation to measure the offset from the candidate region to a "perfect" nodule. These two parameters would change rapidly if skewness or elongation of the nodule voxel distribution varies. Therefore, our strategy is to allow these features linearly increasing while their values are smaller than given thresholds and exponentially decreasing otherwise. In this way, we make sure that statistical values will achieve higher feature values while they are within specified threshold and will get increasing penalty while they are larger than given thresholds.

The voting score for maximum voxel value and mean voxel value of the test region is designed to enhance tiny nodules with large dynamic range. Although almost all valid nodule voxel values are ranging from -1000 HU to 1000 HU, there indeed exist some nodule voxels whose values are larger than 1000 HU. This score will be extra high when the maximum and mean voxel value are larger than a given value within the test region.

Based on these 7 features, the final voting score for the i-th region is defined as

$$
F_i = \sum_{k=1}^{7} w_k f_{i,k}, \quad \text{where} \sum_{k=1}^{7} w_k = 1. \tag{7}
$$

For candidate regions from each data set, we select 40 regions at most by sorting F. For each voxel in a selected region, we iterate its neighbors with a radius of 3 and arrange all the CT values from each neighbor in a column vector, giving rise to a feature vector with 343 dimensions (i.e., $7\times7\times7$). This vector represents the key distribution features of CT values for each voxel and will be used both in training and testing stages.

4.4 3D nodule candidates scoring

To combine all the outputs from the group of trained ANN outputs, we propose a 3D scoring method. For the n-th ANN at the i-th test region in the volume, we define the score as

$$
S_{n,i} = \sum_{(x,y,z)\in V_i} f_G(x, y, z; \sigma) \times O_i(x, y, z), \tag{8}
$$

where $O_i(x, y, z)$ is the output value for the i-th region volume. V_i denotes the volume of the i-th region. f_G is the normalized 3D Gaussian weighting function with standard deviation σ defined as

$$
f_G(x, y, z; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{x^2 + y^2 + z^2}{2\sigma^2} \right\}, \tag{9}
$$

where its center corresponds to the center of V_i, and σ is linearly changing along with the region size V_i. The 3D Gaussian weighting function is used because the expected output of the trained ANNs would be similar to a 3D Gaussian distribution used in the teaching volume. This scoring method can combine the output of the test volumes in the data set. High score values indicate volumes may be near or contain nodules, while a lower value indicates a non-nodule. By thresholding $S_{n,i}$, regions scored by the n-th ANN are selected. By organizing all of these selected regions using an AND operator, detected regions from all ANNs are collected.

5 Experiments and results

The proposed CADe system is validated on a subset of LIDC-IDRI, which are available online at http://cancerimagingarchive.net. In our system, we have chosen 154 cases from LIDC-IDRI, within which 47 cases with 175 nodules are used to train ANNs, while the other 107 cases with 252 nodules are used to validate the performance. According to the nodule type, we classify the training nodules into four types: isolated, juxta-pleural, juxta-vessel, and ground-glass opacity (GGO), which respectively lead to four trained ANN models. It should be noted that although training cases are limited, the key point is that real training data are organized by each voxel into vectors, leading to huge amount of final training data. In our implementation, each type ANN is trained through over fifty thousand pieces of voxel vector data. Each ANN model outputs the score using the method proposed in Subsection 4.4 and the final results are determined using an AND operator over these scores from four ANN trained models.

In practical clinical treatments, isolated nodules can be more easily identified by radiologists. Meanwhile, the other three types of nodules sometimes may be missed because they are attached to tissues and do not have accurate contours or boundaries. Therefore, during the procedure of selecting test cases, we intentionally reduce cases with isolated nodules, while including more cases with juxta-pleural and juxta-vessel nodules to show the performance of our CADe system for all four types of nodules. Among all the 252 nodules from the 107 testing cases, there are 112 (44.4%) isolated, 78 (30.9%) juxta-pleural, 52 (20.7%) juxta-vessel, and 10 (4.0%) GGO nodules, respectively. It should be noticed that, even though we only have 10 GGO cases because cases with GGO nodules in LIDC-IDRI are very limited, our system does have the ability to extract all the GGO nodules correctly.

5.1 Parameter selection

Commonly used parameters during processing for all images are described as follows: (1) In pre-processing, image and slice spacings for all images are re-sampled to 1 mm, while radius is 3 for median filtering. (2) In region growing, although seeds are placed by hands, $-800 \leqslant \alpha_2 \leqslant \alpha_1 \leqslant -600$ is very common in all cases. (3) In training (Eq. (1)) and 3D scoring (Eq. (9)), σ is changing along with region size but the standard value is 4. (4) In vessel detection and Eq. (3), σ_H is changing from 1.0 to 4.0 with spatial steps of 0.3, $\gamma_{12} = \alpha = 0.5$, $\gamma_{23} = 100$. (5) In nodule candidates voting, $N_{1,\min} = 3$, $N_{1,\max} = 15$, $N_{2,\min} = 4$ (2 pixels $\times$ 2), $N_{2,\max} = 900$ (30 $\times$ 30), $N_{3,\min} = 8$ (2 voxels $\times$ 2 $\times$ 2), $N_{3,\max} = 8000$, $D_{5,\max} = 1.0$, $D_{6,\max} = 3.0$. The weighting parameters w_1 to w_7 are assigned to [0.15, 0.15, 0.1, 0.2, 0.15, 0.15, 0.1], respectively.

5.2 Performance

It may be noted that, nodule detection is not a so easy task, even different radiologists may have different diagnosis on the same region. However, the LIDC-IDRI database provides a venue for us to conduct quantitative comparison between methods. The LIDC-IDRI data sets are labeled and non-blindly double checked by four radiologists. The least count of radiologists identifying one region as nodule is called agreement level (i.e., agreement level 1 means the corresponding nodule is annotated by at least one of the four radiologists) and nodules at higher agreement level are subset of nodules at lower agreement level (i.e., nodules at agreement level 3 are a subset of nodules at agreement level 2).

In total, we have 252 nodules comprising 252 (100%) nodules at agreement level 1, 214 (84.9%) nodules at agreement level 2, 187 (74.2%) nodules at agreement level 3, and 161 (63.9%) nodules at agreement level 4. To validate the performance of our proposed system, we have shown the average false positives per case and the nodule sensitivity at different agreement levels in Table 1 in details.

The table reveals that our CADe system achieves high sensitivity at agreement level 4 but much lower sensitivity at agreement level 2. Nonetheless, it should be noticed that, correct results at agreement levels 1 and 2 might not be truly indicative for the purpose of nodule detection since the LIDC annotations are non-blindly double checked by four radiologists and the majority ($\geqslant 2$) of the four radiologists did not indicate them as nodules.

Table 1 Performance of the proposed CADe system at different agreement levels on LIDC-IDRI data sets

Agreement level	Nodule count	True positives	False positives	Average (FPs/case)	Sensitivity (%)
2	214	171	227	2.12	79.9
3	187	163	224	2.09	87.2
4	161	144	216	2.01	89.4

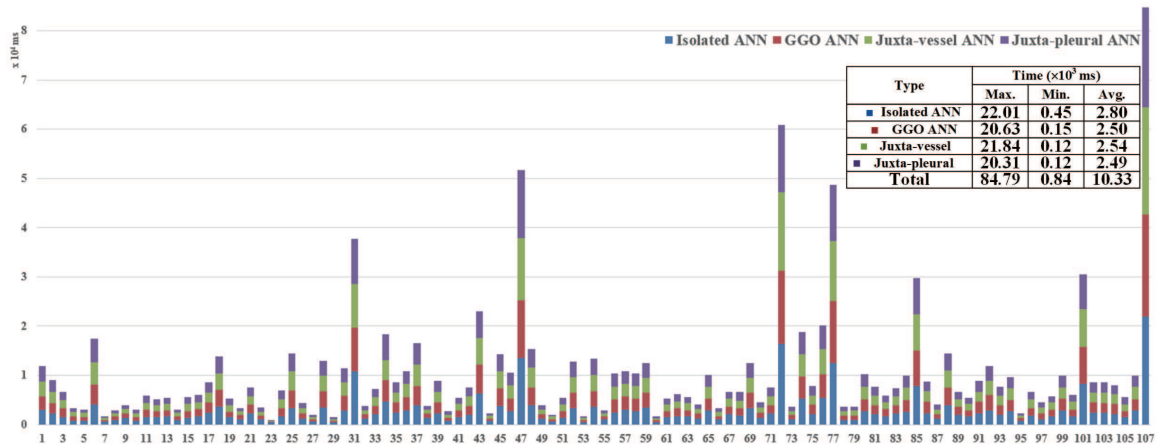**Figure 5** (Color online) Testing time statistics for all 107 datasets. Maximum, minimum and average time for every nodule at each trained ANN are shown with different colors. The horizonal axis represents indices of the CT cases. The vertical axis represents the corresponding time performance.
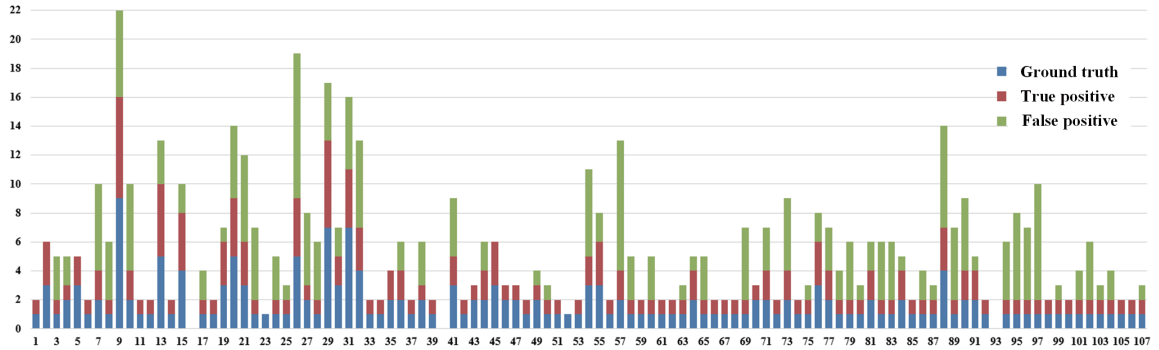
Figure 6 (Color online) Statistics for all 107 data sets. The horizonal axes represents the CT case number. The vertical axes represents the corresponding count. We use three colors to label corresponding ground truth, true positive, and false positive nodule counts. It should be noticed that all statistics are at agreement level 3.

Examining time performance, we realize that, it would take almost one day to train each ANN for specified nodule type on an Intel(R) Core(TM)-i7 processor without GPU acceleration. Specifically, we have shown the testing time statistics of separate ANN for each case in Figure 5. Besides, we have also shown the ground truth, true positive and false positive nodule count distribution at agreement level 3 for the entire 107 data sets used in our validation in Figure 6.

We have documented different types of detected nodules at different sizes and conditions in Figures 7–10. In each figure, original axial (top row) and corresponding ANN output (bottom row) regions are shown. Meanwhile, some false positive nodules as well as some typical missed true positivenot removed by our system are described in Figures 11 and 12, respectively. Detailed descriptions are documented in captions for each figure.

5.3 Comparison

To put the performance of our system into perspectives, we have compared the proposed system with reported operating aspects from previously published CADe systems. It is a difficult task to make

Table 2 Performance comparison of the proposed CADe system with other methods

CADe Systems	Casecount	Nodule count	Applied nodule size (mm)	Average (FPs/case)	Sensitivity (%)
Yuan et al. (2006) [37]	N/A	337	$\geqslant 4$	3.0	83.09
Opfer et al. (2007) [38]	N/A	59	$\geqslant 4$	4.0	91.0
Gori et al. (2007) [39]	N/A	45	$\geqslant 5$	3.8	74.7
Murphy et al. (2009) [40]	813	1525	N/A	4.2	80.0
Golosio et al. (2009) [41]	84	148	3–30	4.0	79.0
Ye et al. (2009) [17]	54	122	$\leqslant 20$	8.2	90.2
Messay et al. (2010) [42]	84	143	3–30	3.0	80.4
Tan et al. (2011) [43]	125	259	$\geqslant 3$	4.0	87.5
Han et al. (2014) [22]	205	490	$\geqslant 3$	4.0	82.7
The proposed system	**107**	**252**	$\geqslant \mathbf{3}$	**2.0**	**89.4**

definitive comparison between these CADe systems due to variability in data sets. Nevertheless, we believe that it is still important to make an attempt towards a relative comparison while the average false positives per CT case and the nodule detection sensitivity could be comparable between different systems. To make validation more convincible, we have compared with systems operating on both LIDC-IDRI and other data sets. The detailed statistics are described in Table 2.

• Yuan et al. [37] analyzed the commercially available CAD system called ImageChecker CT LN-1000 system by R2 Technology on a database of 150 patients with 1.25 mm axial slice images. The sensitivity of the ImageChecker is 83.09% at 3 FPs/case. They use 337 nodules classified by consensus review.

• Opfer and Wiemker [38] proposed a validation study of a CAD system from LIDC and discussed the influence of underlying ground truth during detection. They use 59 nodules from the LIDC database and results are at agreement level 4.

• Gori et al. [39] proposed a voxel-based neural network approach to the detection of lung nodules in the framework of the MAGIC-5 Italian project. They use 45 internal nodules, excluded sub-pleural nodules annotated by experienced radiologists.

• Murphy et al. [40] proposed a method using local image features and k-nearest-neighbor classification in the framework of the Nelson Trial lung cancer screening program. The results are obtained from 813 scans with 1525 nodules annotated by two radiologists.

• Golosio et al. [41] reported the performance of a novel multi-threshold method on 23 scans from the Italung-CT database and 83 scans from LIDC databases. Their results are counted with 148 nodules at agreement level 4.

• Ye et al. [17] proposed a method using volumetric shape index map, antigeometric diffusion and modified expectation-maximization. It is applied on 122 nodules from 108 thoracic CT scans using a wide range of X-ray tube currents. It was reporting a sensitivity of 90.2%, at a FP rate of 8.2 FPs/case.

• Messay et al. [42] presented a CADe system based on thresholding, morphological operations, 3D nodule features and linear and quadratic discriminant classifiers. Their validation was on 143 nodules at agreement level 1 from 84 CT scans of LIDC and achieves a sensitivity of 80.4% and 3 FPs/case.

• Tan et al. [43] proposed a fixed-topology ANN classifier based on 45 geometric, position, and intensity features. They validated the system on 259 nodules from 125 LIDC databases and achieved sensitivity 87.5% at 4.0 FPs/case.

• Han et al. [22] proposed a hierarchical vector quantization scheme. They validated the system on 323 nodules from 205 LIDC-IDRI database and achieved sensitivity 82.7% at 4.0 FPs/case.

From Table 2, it can be observed that the performance of our method compares well with other methods in terms of sensitivity and FP rates. When comparing with Opfer and Weimker [38] and Ye et al. [17], our method offers a much lower FPs/case while retaining very close sensitivity. Besides these two, our sensitivity is higher than other methods while the FPs/case is reduced to about 2.0. It should be noticed that methods validated on LIDC database are more comparable with our method since we use the same data source. We have annotated those in description of Table 2 and labeled the agreement
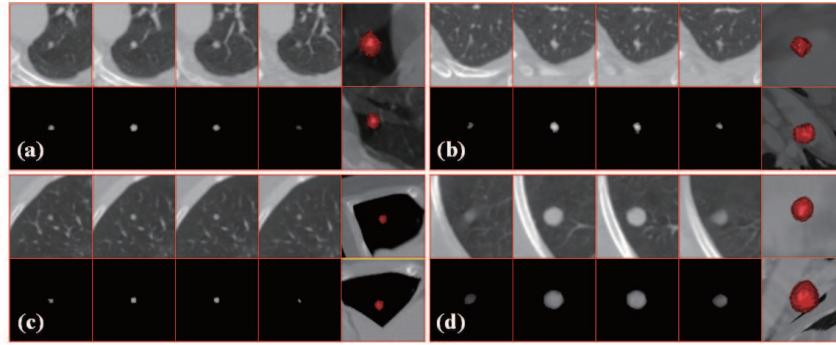
Figure 7   (Color online) Four isolated nodules of different sizes with 3D surfaces on the right. (a) and (c) Nodule diameter is 3–5 mm; (b) nodule diameter is over 5 mm; (d) Nodule diameter is over 30 mm. 3D surfaces are enlarged for a better view.
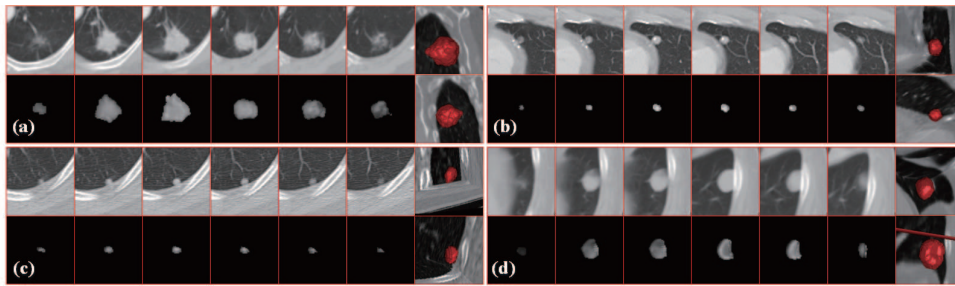

Figure 8   (Color online) Juxta-pleural nodules of different sizes and different conditions with 3D surfaces on the right. (a) and (b) show cases where nodules are slightly attached onto the lung wall; (c) and (d) show cases where nodules are totally attached. Diameters for (a) and (d) are over 30 mm. Diameter for (b) is 5–15 mm. Diameter for (c) is 3–5 mm. 3D surfaces are enlarged for a better view.
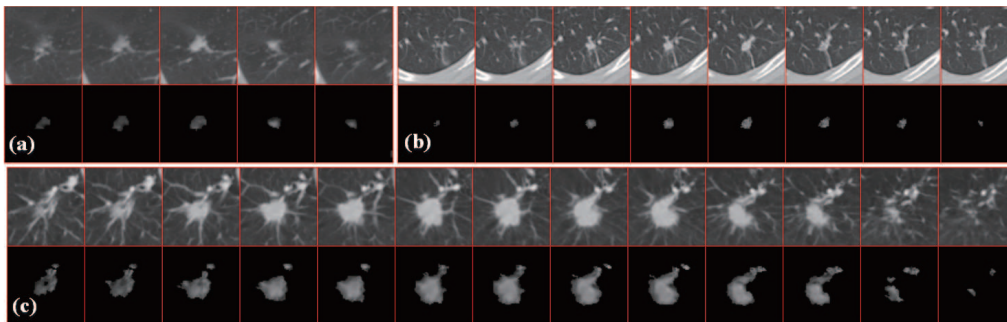

Figure 9   (Color online) Juxta-vessel nodules of different sizes and different conditions. Diameters for (a) and (b) are 5–15 mm. Diameter for (c) is over 30 mm. Only a few vessels are around in (a). (b) is surrounded by tiny vessels. (c) is surrounded by more severe and obvious vessels.

level (indicating the number of regions identified as nodule by the four radiologists) for tests using LIDC database.

5.4   Discussion and limitation

We have shown some typical cases of four type nodules in Figures 7–10, respectively. Descriptions are documented in captions for each figure. It should be noticed that our CADe system can detect all of these four types of nodules whose diameters vary from 3 mm to 30 mm or over 30 mm. However, there indeed exist some cases in which we had missed some nodules. In Figure 12(a), this nodule is missed since its location is changing slice by slice and its distribution is more like an oblique cylinder than a sphere, so our system classified its region as vessels. In Figure 12(b), the nodule is surround by a massive count of
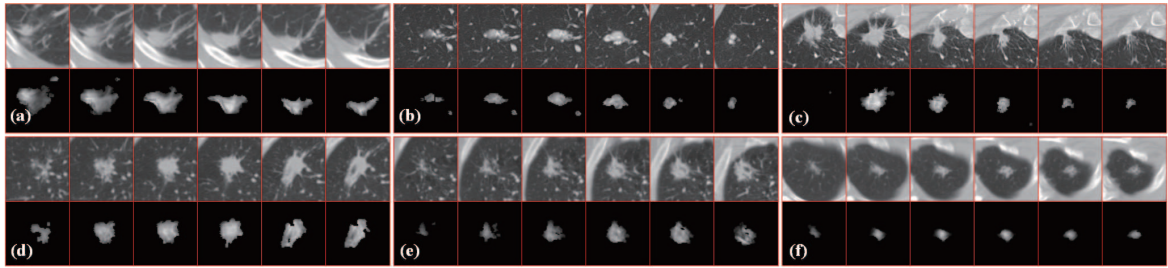
Figure 10 (Color online) Ground-glass optical (GGO) nodules of different sizes and different conditions. Appearances of GGO nodules may vary significantly. (a), (c), and (e) Nodules are slightly attached to lung walls or vessels; (b), (d), and (f) nodules are isolated.
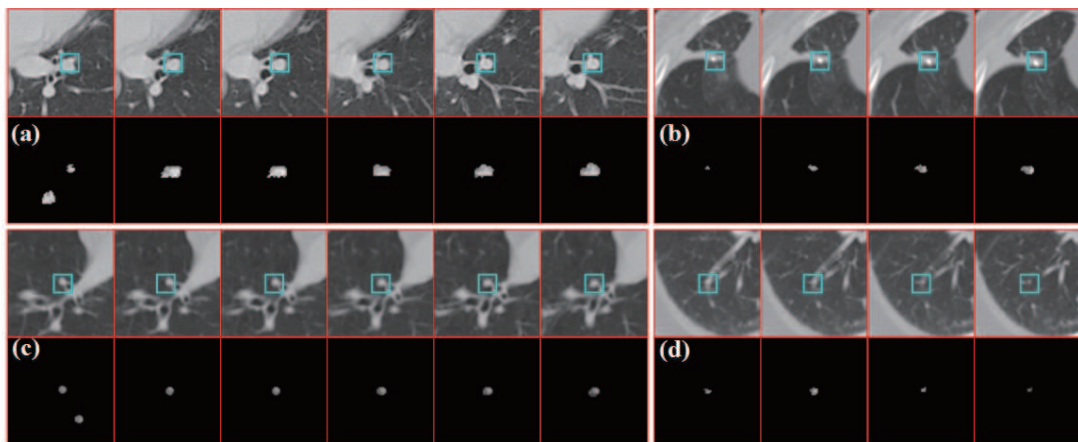

Figure 11 (Color online) Some typical types of false positives not removed by our system. (a) and (c) Nodules are not removed since their appearances and structures are very similar with juxta-vessel nodules. Their locations are stable and not changing along slices. (b) Nodule is not removed since it has a very high intensity inside the vessel which is a direct evidence of being nodules. (d) Nodule is not removed because it is inside a vessel and its size is too small to afford enough structure information.
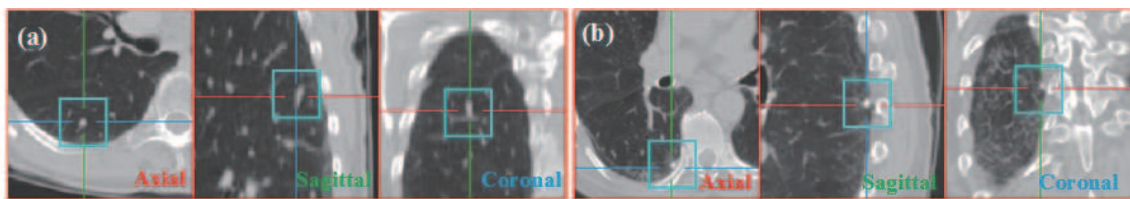

Figure 12 (Color online) Some typical cases of missed nodules. The nodules are located at the center of the box and the cross hair. (a) Missed nodule is very similar to vessels since its location is changing slice by slice and its distribution is more like an oblique cylinder than a sphere, but it is annotated as nodule by radiologists; (b) very tiny nodule is attached onto lung walls and surrounded by ambiguous voxels, making it hard to find either the correct outline or the distribution. Therefore, our system also refused to classify it as a nodule.

tissue voxels, making it hard to find either the correct outline or the distribution. Therefore, our system also refused to classify it as a nodule.

Although our method can handle nodules over 3 mm well, the sensitivity rate would fall and the FPs would rise if the candidate nodule remains too tiny (i.e., 3–5 mm) since the structure and neighbor information are not adequate due to its small sizes. This is perhaps more severe if there are many tiny nodules in the same case. As described in cases 26, 88 and 97 in Figure 6, more false positives are included to extract all the tiny lung nodules. In the near future, we would introduce more size-sensitive features in order to enhance correct detection for tiny nodules.

6 Conclusion

In this paper, we have presented a complete CADe system for lung nodule detection. For initial candidate nodules detection, we have organized voxels into regions and introduced 3D geometrical as well as statistical features to constitute a voting method. After that, we have devised multiple ANNs with massive sampling from the candidate regions for each type of nodules, and a scoring method to combine all the outputs from all the ANNs. The experiments on 252 nodules from 107 CT cases in LIDC-IDRI database have shown that our methods can achieve high sensitivity of 89.4% for all isolated, juxta-pleural, juxta-vessel, and GGO nodules at a FPs/case rate as low as 2.0. It is our expectation that our novel system would be of great assistance during diagnosis of lung nodules for radiologists in CT scans.

Conflict of interest The authors declare that they have no conflict of interest.

Supporting information The supporting information is available online at info.scichina.com and link. springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Siegel R L, Miller K D, Jemal A. Cancer statistics, 2015. CA-Cancer J Clin, 2015, 65: 5–29
2 El-Baz A, Suri J S. Lung Imaging and Computer Aided Diagnosis. Boca Raton: CRC Press, 2011
3 van Ginneken B, Armato S G, de Hoop B, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Med Image Anal, 2010, 14: 707–722
4 El-Baz A, Beache G M, Gimel'farb G, et al. Computer-aided diagnosis systems for lung cancer: challenges and methodologies. Int J Biomed Imag, 2013, 2013: 942353
5 Ukil S, Reinhardt J M. Anatomy-guided lung lobe segmentation in X-ray CT images. IEEE Trans Med Imag, 2009, 28: 202–214
6 van Rikxoort E M, de Hoop B, van de Vorst S, et al. Automatic segmentation of pulmonary segments from volumetric chest CT scans. IEEE Trans Med Imag, 2009, 28: 621–630
7 Gao Q, Wang S J, Zhao D, et al. Accurate lung segmentation for X-ray CT images. In: Proceedings of the 3rd International Conference on Natural Computation (ICNC 2007), Haikou, 2007. 2: 275–279
8 Itai Y, Kim H, Ishikawa S, et al. Automatic segmentation of lung areas based on SNAKES and extraction of abnormal areas. In: Proceedings of the 17th IEEE International Conference on Tools With Artificial Intelligence (ICTAI'05), Tokyo, 2005. 5: 381
9 Silveira M, Nascimento J, Marques J. Automatic segmentation of the lungs using robust level sets. In: Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 2007. 4414–4417
10 Shih-Chung B L, Freedman M T, Lin J S, et al. Automatic lung nodule detection using profile matching and back-propagation neural network techniques. J Digit Imag, 1993, 6: 48–54
11 Mao F, Qian W, Gaviria J, et al. Fragmentary window filtering for multiscale lung nodule detection: preliminary study. Acad Radiol, 1998, 5: 306–311
12 Ko J P, Betke M. Chest CT: automated nodule detection and assessment of change over time — preliminary experience 1. Radiology, 2001, 218: 267–273
13 Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance 1. Radiology, 2004, 230: 347–352
14 Yamada N, Kubo M, Kawata Y, et al. ROI extraction of chest CT images using adaptive opening filter. In: Proceedings of SPIE Medical Imaging 2003, San Diego, 2003. 869–876
15 Saita S, Oda T, Kubo M, et al. Nodule detection algorithm based on multislice CT images for lung cancer screening. In: Proceedings of SPIE Medical Imaging 2004, San Diego, 2004. 1083–1090
16 Brown M S, McNitt-Gray M F, Goldin J G, et al. Patient-specific models for lung nodule detection and surveillance in CT images. IEEE Trans Med Imag, 2001, 20: 1242–1250
17 Ye X, Lin X, Dehmeshki J, et al. Shape-based computer-aided detection of lung nodules in thoracic CT images. IEEE Trans Bio-Med Eng, 2009, 56: 1810–1820
18 Riccardi A, Petkov T S, Ferri G, et al. Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and Zernike MIP classification. Med Phys, 2011, 38: 1962–1971

19 Gurcan M N, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys, 2002, 29: 2552–2558

20 Fetita C I, Prêteux F, Beigelman-Aubry C, et al. 3D automated lung nodule segmentation in HRCT. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Montreal, 2003. 626–634

21 Tanino M, Takizawa H, Yamamoto S, et al. A detection method of ground glass opacities in chest X-ray CT images using automatic clustering techniques. In: Proceedings of SPIE Medical Imaging 2003, San Diego, 2003. 1728–1737

22 Han H, Li L, Han F, et al. Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme. IEEE J Biomed Health, 2015, 19: 648–659

23 Shen W, Zhou M, Yang F, et al. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn, 2016, 61: 663–673

24 Cheng J Z, Ni D, Chou Y H, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep, 2016, 6: 24454

25 Ginneken B V, Setio A A A, Jacobs C, et al. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: Proceedings of IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, 2015. 286–289

26 Lu L, Tan Y, Schwartz L H, et al. Hybrid detection of lung nodules on CT scan images. Med Phys, 2015, 42: 5042–5054

27 Colin J, van Rikxoort E M, Keelin M, et al. Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. Eur Radiol, 2016, 26: 1–9

28 Chen S, Ni D, Qin J, et al. Bridging computational features toward multiple semantic features with multi-task regression: a study of CT pulmonary nodules. In: Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2016. 53–60

29 Setio A A, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imag, 2016, 35: 1160–1169

30 Armato III S G, Giger M L, Doi K, et al. Computerized lung nodule detection: comparison of performance for low-dose and standard-dose helical CT scans. In: Proceedings of SPIE Medical Imaging 2001, San Diego, 2001. 1449–1454

31 Armato III S G, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys, 2011, 38: 915–931

32 Wiemker R, Bergtholdt M, Dharaiya E, et al. Agreement of CAD features with expert observer ratings for characterization of pulmonary nodules in CT using the LIDC-IDRI database. In: Proceedings of Medical Imaging 2009, Florida, 2009. 72600H

33 Suzuki K. Determining the receptive field of a neural filter. J Neural Eng, 2004, 1: 228

34 Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proceedings of the IEEE International Conference On Neural Networks, San Francisco, 1993. 586–591

35 Armato III S G, Giger M L, MacMahon H. Automated detection of lung nodules in CT scans: preliminary results. Med Phys, 2001, 28: 1552–1561

36 Sato Y, Nakajima S, Shiraga N, et al. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Med Image Anal, 1998, 2: 143–168

37 Yuan R, Vos P M, Cooperberg P L. Computer-aided detection in screening CT for pulmonary nodules. Am J Roentgenol, 2006, 186: 1280–1287

38 Opfer R, Wiemker R. Performance analysis for computer-aided lung nodule detection on LIDC data. In: Proceedings of SPIE Medical Imaging 2007, San Diego, 2007. 65151C

39 Gori I, Fantacci M E, Martinez A P, et al. An automated system for lung nodule detection in low-dose computed tomography. In: Proceedings of SPIE Medical Imaging 2007, San Diego, 2007. 65143R

40 Murphy K, van Ginneken B, Schilham A M R, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. Med Image Anal, 2009, 13: 757–770

41 Golosio B, Masala G L, Piccioli A, et al. A novel multithreshold method for nodule detection in lung CT. Med Phys, 2009, 36: 3607–3618

42 Messay T, Hardie R C, Rogers S K. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal, 2010, 14: 390–406

43 Tan M, Deklerck R, Jansen B, et al. A novel computer-aided lung nodule detection system for CT images. Med Phys, 2011, 38: 5630–5645