

# An Assisted Diagnosis System for Detection of Early Pulmonary Nodule in Computed Tomography Images

Ji-kui Liu<sup>1</sup> · Hong-yang Jiang<sup>2</sup> · Meng-di Gao<sup>2</sup> · Chen-guang He<sup>3</sup> · Yu Wang<sup>2</sup> · Pu Wang<sup>1</sup> · He Ma<sup>2</sup> · Ye li<sup>1</sup>

Received: 20 May 2016 / Accepted: 7 December 2016 / Published online: 28 December 2016  
© Springer Science+Business Media New York 2016

**Abstract** Lung cancer is still the most concerned disease around the world. Lung nodule generates in the pulmonary parenchyma which indicates the latent risk of lung cancer. Computer-aided pulmonary nodules detection system is necessary, which can reduce diagnosis time and decrease mortality of patients. In this study, we have proposed a new computer aided diagnosis (CAD) system for detection of early pulmonary nodule, which can help radiologists quickly locate suspected nodules and make judgments. This system consists of four main sections: pulmonary parenchyma segmentation, nodule candidate detection, features extraction (total 22 features) and nodule classification. The publicly available data set created by the Lung Image Database Consortium (LIDC) is used for training and testing. This study selects 6400 slices from 80 CT scans containing totally 978 nodules, which is labeled by four radiologists. Through a fast segmentation method proposed in this paper, pulmonary nodules including 888 true nodules and 11,379 false positive nodules are

segmented. By means of an ensemble classifier, Random Forest (RF), this study acquires 93.2, 92.4, 94.8, 97.6% of accuracy, sensitivity, specificity, area under the curve (AUC), respectively. Compared with support vector machine (SVM) classifier, RF can reduce more false positive nodules and acquire larger AUC. With the help of this CAD system, radiologist can be provided with a great reference for pulmonary nodule diagnosis timely.

**Keywords** Computer aided diagnosis (CAD) · Pulmonary nodule detection · Ensemble classifier · LIDC

## Introduction

According to investigation in [1], 12.9% cancer cases around the world concentrated on lung cancer by 2012, which remained the most common cancer in the world. Moreover, the mortality of lung cancer reached up to 19.4% because of its high case fatality. As some studies show, tobacco smoking is the main indicted for lung cancer [2, 3], and the lung cancer risk for a smoker is over 15 times higher than a non-smoker [4]. Moreover, another lethal factor of lung cancer is air pollution. In recent years, most first-tier cities and some medium-sized cities of China suffered from fog and haze, which contains excessive amounts of carcinogenic particles [5]. Thus serious lung health hazard has gradually appeared with high-speed urban development. In this sense, people should enhance their health management. However, most important of all, to avoid missing the best treatment opportunities an effective method of lung nodule early detection is urgently needed.

With the development of medical imaging techniques, X-ray, ultrasonography, positron emission tomography (PET) and computerized tomography (CT) are applied to lung nodule detection. Through many researchers' validation, CT

---

This article is part of the Topical Collection on *Patient Facing Systems*

Ji-kui Liu and Hong-yang Jiang contributed equally to this work.

---

✉ He Ma  
mahe@bmie.neu.edu.cn

✉ Ye li  
ye.li@siat.ac.cn

<sup>1</sup> Key Laboratory for Health Informatics of the Chinese Academy of Sciences (HICAS), Shenzhen Institutes of Advanced Technology, Shenzhen 518055, Guangdong, China

<sup>2</sup> Sino-Dutch Biomedical and Information Engineering School, Hunnan Campus, Northeastern University, Shenyang 110169, Liaoning, China

<sup>3</sup> Software School, North China University of Water Resources and Electric Power, Zhengzhou 450045, Henan, China

medical imaging technology is superior to others on the sensitivity and accuracy [6, 7]. Lung nodules are diminutive masses, whose sizes are approximately from 5 mm to 40 mm in the human lung [8], and the shapes of them are usually spherical. However, sometimes these nodules can be distorted due to the influence of surrounding anatomical structures [9]. Besides, some false positive cases such as ribs, blood vessels, and airway walls, which are mistakenly considered as lung nodules, make the task of the lung nodule detection challenging [10]. In general, compared with computer aided detection (CAD) system, the manual lung nodule detection method is error-prone and time-consuming.

This paper proposed an efficient assisted diagnosis system for detection of early pulmonary nodule in computed tomography images. We first used a novel two-dimensional OTSU's curve thresholding method to segment the pulmonary parenchyma region. Then using spatial fuzzy C-means (SFCM) method the suspected pulmonary nodules were extracted. Finally the random forest (RF) classifier was used to make classification, which obtained a preferable result. Moreover, we also discussed the information security issues when the CAD system was applied to remote diagnosis.

## Related works

An increasing number of radiologists have accepted the help of computer aided system and recognized its significance, especially in the pulmonary nodule detection, segmentation and diagnosis [11–13]. The main target of the CAD system mentioned here was to precisely locate the pulmonary nodule. Multifarious CAD methods designed for detecting lung nodule have been proposed by many researchers until now [14]. An automated detection method based on shape and texture features proposed by Taşçı, E. et al was used to detect juxtal-pleural nodule [15], which acquired more than 90% accuracy of detection. Penedo M G et al used a two-level artificial neural network (ANN) architecture to detect lung nodule on digitized chest radiographs, which provided results of 89–96% sensitivity and 5–7 FP's/image [16]. Diego M. Peña et al took advantage of minimal characteristics extraction technique to detect nodules, whose sizes are larger than 4 mm [17], and the sensitivity, specificity and accuracy are 94.23, 84.75, 89.19%, respectively. To reduce false negative cases, Akram Sheeraz et al used hybrid features to design a pulmonary nodule detection system based on support vector machine (SVM) [18], which can obtain the sensitivity of 95.31%. Focusing on small lung nodules of sizes between 2 mm and 10 mm, Alex Martins Santos et al developed an automated detection system using Gaussian mixture models. Besides, Tsallis's and Shannon's entropy calculated as a kind of texture descriptor was combined with SVM

classifier for suspected regions classification. The results demonstrated that small nodules were detected with a sensitivity of 90.6%, a specificity of 85% and an accuracy of 88.4% [19]. As the tremendously growth of data, Atsushi Teramoto et al developed a fast lung nodule detection method using a cylindrical shape filter to reduce computation cost. The correct detection rate is 80% and the detection speed is 4–36 times faster than other existing methods [20]. In addition, Gonçalves L et al used Hessian-based strategies for the lung nodule segmentation in chest CT scans. Moreover, they combined both the shape index (SI) and curvedness (CV) approach and the central adaptive medialness approach [21]. Jiang H Y et al designed a risk assessment system about lung cancer with the help of CAD and acquired anticipated results [22].

## Materials

This study used Lung Image Database Consortium (LIDC) data set, which has been collected from five different sites in America [23]. A total of 80 patients' CT scans were obtained from LIDC for analysis and the number of CT slices each patient varied between 130 and 270. The standard distance between two slices was 2.5 mm and the thickness of each slice is between 1.3 and 3.0 mm. Among these CT slices, some images without pulmonary parenchyma can be easily detected and were regarded as invalid images. Thus the valid slices can be reduced to approximately between 20 and 100 per patient's CT scan. This data set contained various lung nodules which may appear in any part of lung, thus based on nodules' positions in the lung, four types of lung nodules are detected including well-circumscribed, vascularized, juxtal-pleural and pleural-tail, which were named in [24]. Each scan of CT slices was reviewed and diagnosed by four radiologists individually and the diagnostic records were stored in xml files as a golden standard.

## Methods

To precisely screen out the suspected nodules from numerous lung-CT slices and make predictions about them, this study presented a novel solution to pulmonary nodule detection. Considering speed of calculation and accuracy of detection, this paper designed an efficient CAD system that performed four fundamental tasks: pulmonary parenchyma segmentation, pulmonary nodule candidate detection, feature extraction and pulmonary nodule classification. Figure 1 displays the whole flow chart of this method.

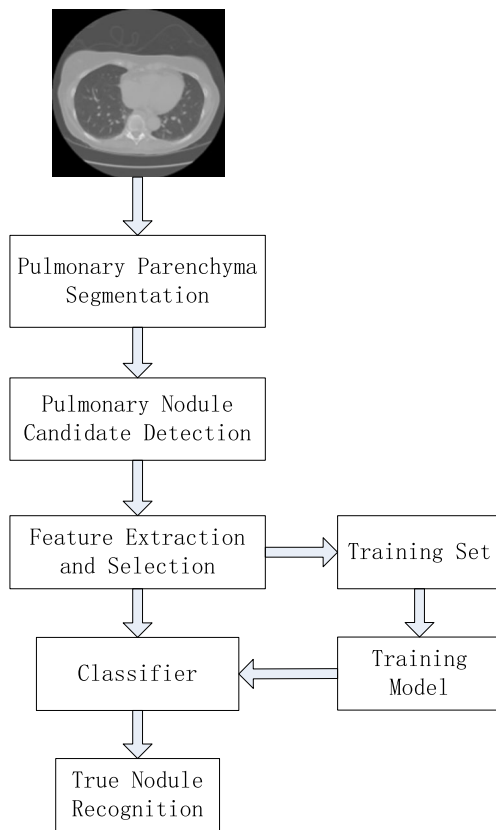


Fig. 1 Flow chart of proposed method

**Pulmonary parenchyma segmentation**

Most of the CAD systems need to preprocess raw CT slices to remove noises generated by the equipment or environment. However, according to the characteristics of lung CT images, this study abandon this part to improve the processing speed. As the two pulmonary lobes locate approximately the center of thoraxes, which can be clearly seen in the CT images, it is convenient to detect and depict lung region through differences of gray value. Generally, the contrast ratio of pulmonary lobes and their external is quite large thus this study utilized gray value statistics method to segment pulmonary parenchyma.

This paper employed a novel two-dimensional OTSU’s curve thresholding segmentation method based on gray and non-local spatial gray feature, which was first proposed by Zhao Feng et al [25]. Firstly, we obtained a two-dimensional threshold point  $I(a, b)$  in the given gray scale image. Here  $a$  was gray value of threshold point  $I$  which was calculated by traditional OTSU method and  $b$  was non-local mean gray value of threshold  $I$  which was calculated by equation (1).

$$b = \sum_{p \in W^r} w_p i_p, \tag{1}$$

$$0 \leq w_p \leq 1, \tag{2}$$

$$\sum_{p \in W^r} w_p = 1, \tag{3}$$

where  $W$  represented a calculation window whose size was  $r \times r$  centered on point  $I$ . Besides,  $w_p$  and  $i_p$  was the weight and gray value of point  $p$  in the  $r \times r$  window, respectively. The value of  $w_p$  can be set depending on the similarity between point  $I$  and point  $p$  and this study used Euclidean distance to compute the similarity, which is shown as equation (4).

$$w_p = \frac{1}{\|v(N_I) - v(N_p)\|_2}, \tag{4}$$

where  $v(N_I)$  and  $v(N_p)$  are vectors whose elements were composed of gray values around the center point  $I$  and  $p$  with the size of  $k \times k$ , which can be called similarity window.

However, we did not just use this single point  $I(a, b)$  to distinguish between the target and background. Secondly, a curve passing the point  $I(a, b)$  was made across the two-dimensional histogram and this curve can be a straight line  $S(a, b)$  perpendicular to the diagonal, which is shown in Fig. 2.

Finally, we used line  $S(a, b)$  to classify the target and background with the discriminant (5).

$$z(i, j) = \begin{cases} 0 & \text{if } x + y \leq a + b \\ 255 & \text{if } x + y > a + b \end{cases} \tag{5}$$

With the help of two-dimensional OTSU’s curve thresholding segmentation method, the noises such as salt and pepper noise can be well eliminated. Moreover, through a series of logical operations, the pulmonary parenchyma was easily acquired, which is shown in Fig. 3.

**Pulmonary nodule candidate detection**

To correctly detect the pulmonary nodule, this paper used an improved fuzzy C-means (FCM) algorithm, which was proposed in our previous work [26]. Clustering algorithm was an unsupervised statistical method which can classify or cluster the data element to a certain class or cluster. The aim of clustering was to increase the difference between each class so that the data elements in the same class could be as similar as possible. Generally, one element belongs to only one cluster by classical C-means algorithm. However, with the help of

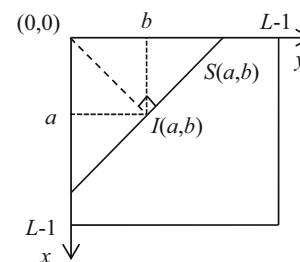
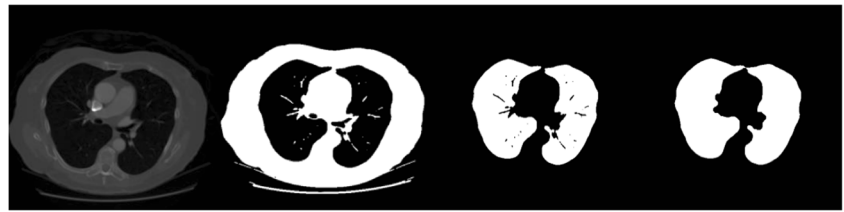


Fig. 2 Two dimensional histogram of threshold line

**Fig. 3** Segmentation process by two-dimensional OTSU's curve thresholding method



fuzzy C-means method, the same element may be sorted into several different clusters which made it widely used in the classification of multi-class complex images.

However, the basic FCM algorithm was sensitive to noise and incapable of displaying the images' space information so that this study adopted a novel FCM method, which is named as spatial fuzzy C-means (SFCM) [26]. The special function  $h_{ij}$  of SFCM is defined as (6), which can enhance the performance of the original algorithm.

$$h_{ij} = \sum_{k \in NB(x_i)} u_{ik} \tag{6}$$

where  $NB(x_i)$  is a square window whose center point is  $x_i$  and  $u_{ik}$  is a matrix whose elements represent degree of affiliation between pixel point  $x_i$  and cluster  $j$ . Besides, the improved fuzzy affiliation with spatial information is expressed as (7).

$$\mu_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{j=1}^c u_{ij}^p h_{ij}^q} \tag{7}$$

In equation (7), the parameter  $p$  and  $q$  are represented for the weight of center pixel and spatial information respectively thus the fuzzy affiliation can be connected to both the center pixel and pixels around it. Actually, two steps need to be implemented for SFCM method. First, we calculated the affiliation matrix and cluster centers based on classical FCM method. Second, spatial information was integrated to obtain a new fuzzy affiliation matrix  $\mu_{ij}$ .

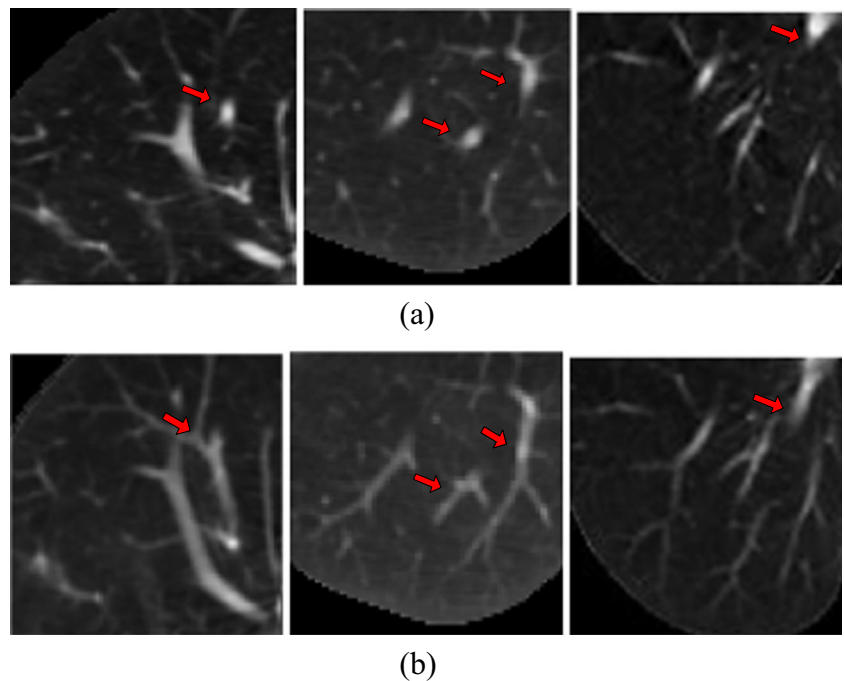
To avoid the curse of dimensionality caused by high-dimensional space, this study took advantage of a Gaussian kernel function  $k(x_i, y_j)$ , which is shown as (8).

$$k(x_i, y_j) = e^{-\frac{\|x_i - y_j\|^2}{\sigma^2}} \tag{8}$$

As a consequence, traditional inner product was replaced by (8) and thus we can get the new objective function, which is shown as (9).

$$J_m = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|\phi(x_i) - \phi(x_j)\|^2 \cdot \|\phi(x_i) - \phi(x_j)\|^2 = k(x_i, x_i) - 2k(x_i, v_j) + k(v_j, v_j) \tag{9}$$

**Fig. 4** Performance of blood vessels enhancement: (a) Original images; (b) Enhanced blood vessels images



In addition, the fuzzy affiliation matrix and cluster centers were also updated, which are expressed as (10), (11), respectively.

$$u_{ij} = \frac{\left(1-k(x_i, y_j)\right)^{\frac{-1}{(m-1)}}}{\sum_{j=1}^c \left(1-k(x_i, y_j)\right)^{\frac{-1}{(m-1)}}} \quad (10)$$

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot k(x_i, y_j) \cdot x_i}{\sum_{i=1}^n u_{ij}^m \cdot k(x_i, y_j)} \quad (11)$$

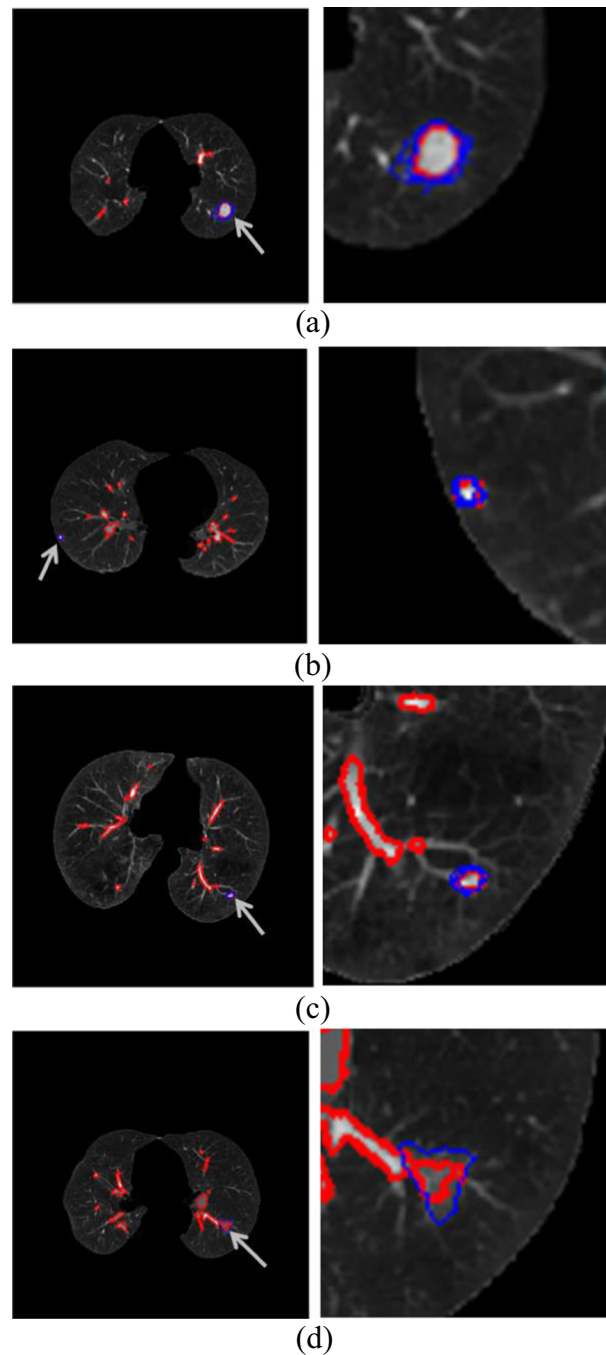
Fully utilizing the novel SFCM method as well as morphology method, this study can segment suspected pulmonary nodules including true nodules and blood vessels in the pulmonary parenchyma. However, the shape of blood vessels was usually circular or oval which was quite similar to that of the pulmonary nodules. It was therefore quite difficult for the computer to recognize true pulmonary nodules. In this study, we proposed that vascular imaging can be enhanced by utilizing differences among adjacent slices of nodules and blood vessels. Figure 4 shows that one single slice combined the vascular information of the previous one with the rear one which can be utilized to reconstruct blood vessels. Thereafter, using the above improved method we can segment the suspected pulmonary nodules which are shown in Fig. 5. The regions of red markers represent the segmented suspected pulmonary nodules and the regions labeled by blue circles are recognized as true nodules which were labeled by experienced experts.

**Feature extraction and classifiers**

In this section, we designed a feature set to incorporate texture and shape information for pulmonary nodule feature description. It was widely considered by the radiologists that features, especially texture and shape features between true nodules and false positives are different and diacritical [27, 28]. In this study, we extracted 22 texture and shape features of all the nodule candidates, which were used for classification.

According to texture features depicted in [29], we extracted seven invariant moment features, which can keep good invariance under translation, rotation, scale and mirror. Four images gray statistical characteristics including mean gray value, maximum gray value, minimum gray value and gray standard deviations from the suspected nodules region were calculated. Furthermore, the average gray scale of neighbor region has been utilized to encode the texture features and the gray scale contrast value between the nodule area and the surrounding area were also used for acquiring the texture properties. A total of thirteen texture features were obtained in this study.

It was also desirable to capture the shape characteristics of the lung nodules images in conjunction with the textures, so as



**Fig. 5** Segmentation results of various forms of nodules, The regions of red markers represent the suspected pulmonary nodules and the regions labeled by blue circles are recognized as true nodules which were labeled by experienced experts. Four types of nodules are: (a) Isolated nodule, (b) Juxta-pleural nodule, (c) Vascularized nodule and (d) Ground-glass nodule

to increase the identification precision. In this work, we extracted the length of nodule (the long axis of the smallest fitting ellipse), the width of nodule (the shortest axis of the smallest fitting ellipse), the nodule’s eccentricity, the area of nodule which was expressed by the number of pixels, the ratio of pixel compared with the minimum convex side surrounded

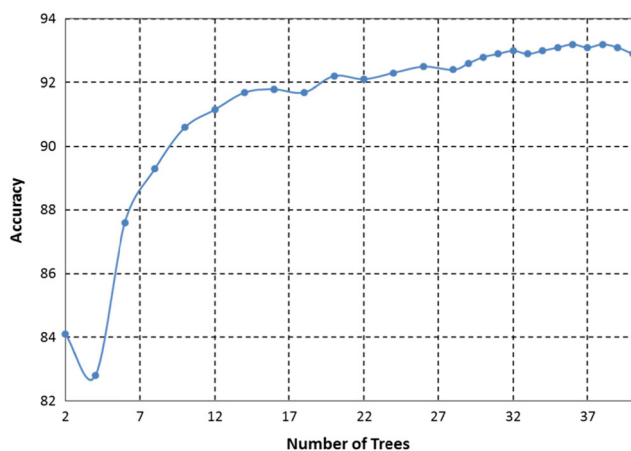
the area, the ratio of pixel compared with the minimum surrounding rectangle, the circular degree and the position of nodules. Nine shape features in total were used in this article.

As many researchers verified, ensemble classifiers emerged as a fast and robust classification method and performed better than single classifiers [30]. The generalization capability of ensemble classifiers benefited from various members of ensemble, which can be designed depending on the special application. Because of difference of the classifiers, the ensemble of them can weaken each classifier and average their votes. Then we can select the optimum hypothesis result. Beyond that, the ensemble classifiers can be executed in a parallel manner, which consumed less time for large data training. According to the significance of classifiers, different weights also can be allocated to these classifiers in some circumstances.

In this study, random forest (RF) approach was used to make classification, which consisted of multiple decision trees. By means of sequential forward selection, some valuable features formed the nodes of the tree and thus each decision tree can independently make judgments based on its features after data training stage. Using a total of 22 features, this study attempted different numbers of trees to experiment for comparison, whose results are shown in Fig. 6. The experiment demonstrated that RF with more than 20 trees can acquire an accuracy of over 92%. Meanwhile we can get an optimal accuracy when we chose 36 trees.

## Experiments and results

The proposed system was implemented with MATLAB 2014 based on a Windows environment. The development and test platform was a PC equipped with Quad-core 2.5GHz processor and 2GB of RAM memory. In our experiments, we used the 6400 slices of 80 CT scans from the data set LIDC, in which four radiologists in total marked 978 true nodules. All



**Fig. 6** Accuracy of RF with different numbers of trees

the nodule candidates were divided into two categories (positive and negative samples) to train the nodule recognition model. Then we evaluated the performance of the pulmonary nodule candidate detection method by the recall ratio, the sensitivity value and other indicators.

## Performance evaluation of the pulmonary nodule candidate detection method

The novel pulmonary nodules segmentation method proposed by this study was carried out, and a total of 12,267 pulmonary nodule candidates were segmented by this algorithm, which contained 888 true pulmonary nodules and 11,379 false pulmonary nodules. This study adopted recall ratio to reflect the performance of the pulmonary nodules segmentation method and the definition of recall ratio is shown as (12). The recall ratios of four radiologists and this system are shown in Table 1, respectively. We can see that the recall ratio of this system is 90.8%, better than Radiologist 1(83.02%) and approximate to Radiologist 2 (91.1%). The result showed that the proposed pulmonary nodules segmentation method had a higher sensitivity and it was nearly equal to the results given by the experienced experts.

$$Recall = \frac{\text{number of detected true nodules}}{\text{number of true nodules in the database}} \times 100 \quad (12)$$

## Performance evaluation of the true nodule identification method

To reduce the false positive as far as possible, 22 discriminating features were extracted including 13 texture features and 9 shape features and then designed an effective machine learning method to classify true pulmonary nodules from the pulmonary nodule candidates. In order to increase the ratio of discrimination of the shape features, this work proposed the first blood vessels imaging remodeling and enhancing, as shown in the Fig. 4.

**Table 1** True nodules detection accuracy of each radiologist and proposed system

	No. of detected true nodules	No. of true nodules	Recall (%)
Radiologist 1	812	978	83.02
Radiologist 2	891	978	91.1
Radiologist 3	911	978	93.15
Radiologist 4	929	978	94.99
Proposed System	888	978	90.8

The red arrow indicated that the isolated suspected nodule was remodeled into blood vessel. It can reduce the false positive rate effectively. In this study, three classifiers including RF, supervised vector machine (SVM) and back-propagation (BP) network were used for comparison. Here 36 trees were used in the RF, the polynomial kernel was used in the SVM and the BP network consisted of three layers structure. The performance indicators were calculated for the random 100 runs of the ten-fold cross-validation, which is shown in Table 2.

Table 2 illustrated that RF and SVM classifiers performed better than BP network in all the indicators, the values of which were all more than 91%. Besides, RF preferably acquired these indicators with over 92%. Compared with the SVM, the accuracy of the RF was slightly less than the SVM, but the sensitivity and AUC of RF was 94.8 and 97.6%, respectively, which was superior to those of SVM, 91.3 and 94.7%. The sensitivity reflected detection rate of positive sample and the higher the value, the lesser omission ratio of positive sample, which was a worthy indicator in medical statistics.

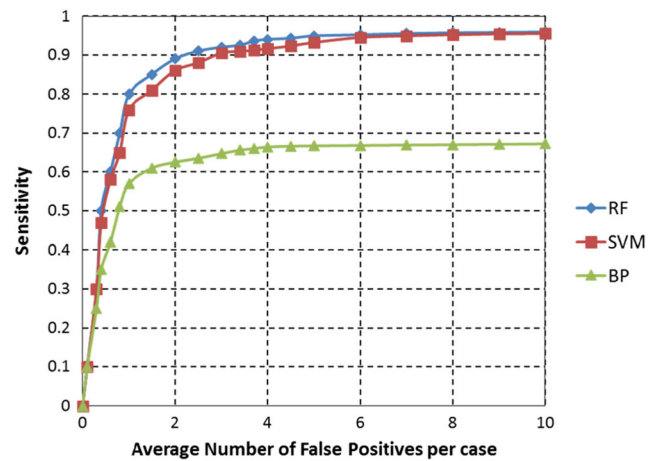
Figure 7 showed the receiver operating characteristic (ROC) curves using three classifiers, RF, SVM and BP, for the nodule recognition under 22 basic features. The results demonstrated that RF classifier was able to control the number of false positive nodules within 10 per case, which was a little better than SVM in this study.

### Discussion

For a CAD system, it can acquire a smaller false positive rate while raising the detection rate of smaller nodules. In Fig. 4, in order to reduce false positive rate of pulmonary nodule candidates, we first enhanced and reconstructed the vessel shape by utilizing vessel information of adjacent slices before the pulmonary nodule segmentation. This step was crucial to increase discriminability of shape feature. In Table 3, the performance of latest pulmonary nodule detection schemes are listed to compare with our work. The universal problem of these latest researches was that their data sets were small volumes thus influencing the reliability of their results. Moreover, the sizes of nodules in their data sets were in specified range, so it was difficult to popularize them to different cases. This work

**Table 2** Analysis result of RF, SVM and BP classifiers

	RF	SVM	BP
Accuracy (%)	93.2	94.1	73
Specificity (%)	92.4	95.6	81.8
Sensitivity (%)	94.8	91.3	65.7
AUC (%)	97.6	94.7	79



**Fig. 7** ROC curves of three classifiers: RF, SVM and BP

contained both great and small nodules recognition. The sensitivity rates acquired by this work were between the latest work. As vast nodules were included in this data set, the false positive (FP) nodules per scans were over the average level of other work. From the Table 3, we can see that our method found more number of nodules than the other approaches. That was because smaller size of nodule was used in this paper. However, it causes increase of the false positive rate. Therefore, FP/Scan of this work was slightly higher than some of schemes. But it was meaningful work to increase detection rate of smaller nodules. In addition, FP/Scan of this work can become smaller if we divided the nodules into different size ranges, which is next goal for improving our results.

As the ability of the remote terminals, back servers, and telecommunications are more powerful than before, the telemedicine is more and more acceptable and popular. The approach proposed in the study can also be used in remote diagnosis because of the easy configuration and implementation of the algorithm on mobile or other terminals with well-managed software and hardware, e.g. with the higher resolution camera or novel cloud

**Table 3** Performance comparison between this proposed work with other schemes within 5 years

Schemes (within 5 years)	Data set	Sensitivity (%)	FP/Scan
Saien S et al.[31]	42 scans/124 nodules	95.9	38.3
Santos A M et al.[19]	28 scans	90.6	1.17
Lu L et al.[32]	98 scans/223 nodules	85.2	3.13
W. Guo et al.[33]	86 scans/111 nodules	85	2.6
Choi W J et al.[34]	43 scans/151 nodules	95.28	2.27
Netto S M B et al.[35]	29 scans/48 nodules	85.93	0.138
Choi W J et al.[36]	84 scans/148 nodules	97.5	6.76
Proposed work	80 scans/978 nodules	94.8	4.5

processing units. A typical scenario as follows: the patient can take a photo copy of CT or MRI image with his/her smart phone and then upload it onto the Internet, the analysis result can be fed back after being processed in cloud server executing diagnosis algorithm [37]. Hence, the security is another issue that we should concern about when this remote diagnosis system is utilized and prevailed, e.g. user authentication and privacy protection. Wang et al [38] investigated the question of several anonymous two-factor authentication schemes, and indicated the flaws of them to give a practical suggestion, especially with an impressive list of security requirements, e.g., resistance to smart card loss attack. Subsequently, Wang et al [39] proposed an enhanced scheme with provable security defined by a formal model, which showed the feature of user un-traceability, was much proper for smart phone applications. Additionally, in our previous work [40, 41], a novel tri-factor mutual authentication scheme was also proposed with the assistant of biometric method in health care applications, as well as a user access control policies were adopted in a practical HIS (hospital information system). Thus, as a part of the health care monitoring platform, the lung nodule detection system can be implemented properly and achieve the security criterion.

## Conclusion

In this study, an efficient and stable pulmonary nodule detection system was proposed, which mainly focused on a large amount of data with a small number of texture and shape features. A novel pulmonary nodule segmentation method was proposed to detect all the suspected nodules. Using random forest classifier we can reduce many false positives and obtain a preferable accuracy, sensitivity and specificity. As the rapid growth of medical image data, radiologists themselves were difficult to finish all the work of detecting pulmonary nodules within a short time. This study can help radiologists screen out amount of nodule candidates and provide reference information to them. Moreover, it is convinced that this study can be popularized and applied in the remote medical field. However, at present the CAD systems are in the development stage, which still need promoting its accuracy. Meanwhile early warning for the generation of pulmonary should be paid more attention in the future.

**Acknowledgements** This work was supported in part by National 863 project of China (SS2015AA020109), National Natural Science Foundation of China (No. 61502472 and No. 31300816) and STS funding from Chinese Academy of Sciences (JCYJ20140417113430655 and JCYJ20140417113430619).

## References

1. Ferlay, J., Soerjomataram, I., Dikshit, R., et al., Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136(5):E359–E386, 2015.
2. Della, P. C., Cigarette smoking: the most widespread and insidious weapon of mass destruction. *Arch. Med. Vet.* 1(1):5, 2015.
3. Pesch, B., Kendzia, B., Gustavsson, P., et al., Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. *Int. J. Cancer* 131(5):1210–1219, 2012.
4. World Health Organization International, *Tobacco smoke and involuntary smoking*. Lyon, France, 2002.
5. Chen, X., Shao, S., Tian, Z., et al., Impacts of air pollution and its spatial spillover effect on public health based on China's big data sample. *J. Clean. Prod.* 2016. doi:10.1016/j.jclepro.2016.02.119.
6. Diederich, S., Lentschig, M., Overbeck, T., et al., Detection of pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs and single-image reporting. *Eur. Radiol.* 11(8):1345–1350, 2001.
7. Valente, I. R. S., Cortez, P. C., Neto, E. C., et al., Automatic 3D pulmonary nodule detection in CT images: a survey. *Comput. Meth. Programs Biomed.* 2015. doi:10.1016/j.cmpb.2015.10.006.
8. Swensen, S. J., Viggiano, R. W., Midthun, D. E., et al., Lung nodule enhancement at CT: multicenter study 1. *Radiology* 214(1):73–80, 2000.
9. Xu, D. M., van der Zaag-Loonen, H. J., Oudkerk, M., et al., Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. *Radiology* 250(1):264–272, 2009.
10. Li, Q., Sone, S., and Doi, K., Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Med. Phys.* 30(8):2040–2051, 2003.
11. Ozekes, S., and Osman, O., Computerized lung nodule detection using 3D feature extraction and learning based algorithms. *J. Med. Syst.* 34(2):185–194, 2010.
12. Noor, N. M., Than, J. C. M., Rijal, O. M., et al., Automatic lung segmentation using control feedback system: morphology and texture paradigm. *J. Med. Syst.* 39(3):1–18, 2015.
13. Ma, Q., Ji, B., Jia, B., et al., Differential diagnosis of solitary pulmonary nodules using  $^{99m}\text{Tc}$ -3P4-RGD2 scintigraphy. *Eur. J. Nucl. Med. Mol. Imaging* 38(12):2145–2152, 2011.
14. Firmino, M., Morais, A. H., and Mendoca, R. M., Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects. *Biomed. Eng. Online* 13:1–16, 2014.
15. Taşçı, E., and Uğur, A., Shape and texture based novel features for automated juxtapleural nodule detection in lung CTs. *J. Med. Syst.* 39(5):1–13, 2015.
16. Peña, D. M., Luo, S., and Abdelgader, A., Auto diagnostics of lung nodules using minimal characteristics extraction technique. *Diagnostics* 6(1):13, 2016.
17. Penedo, M. G., Carreira, M. J., Mosquera, A., et al., Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. *IEEE Trans. Med. Imaging* 17(6):872–880, 1998.
18. Akram, S., Javed, M. Y., Akram, M. U., et al., Pulmonary nodules detection and classification using hybrid features from computerized tomographic images. *J. Med. Imag. Health Inform.* 6(1):252–259, 2016.
19. Santos, A. M., de Carvalho Filho, A. O., Silva, A. C., et al., Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. *Eng. Appl. Artif. Intell.* 36:27–39, 2014.
20. Teramoto, A., and Fujita, H., Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter. *Int. J. Comput. Assist. Radiol. Surg.* 8(2):193–205, 2013.



21. Gonçalves, L., Novo, J., and Campilho, A., Hessian based approaches for 3D lung nodule segmentation. *Expert Syst. Appl.* 61: 1–15, 2016.
22. Jiang, H. Y., Ma, H., Qian, W., et al., Risk analysis for pathological changes in pulmonary parenchyma based on lung computed tomography images. *J. Comput. Assist. Tomogr.* 40(3):357–363, 2016.
23. Messay, T., Hardie, R. C., and Rogers, S. K., A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Image Anal.* 14(3):390–406, 2010.
24. Zhang, F., Song, Y., Cai, W., et al., Lung nodule classification with multilevel patch-based context analysis. *IEEE Trans. Biomed. Eng.* 61(4):1155–1166, 2014.
25. Feng, Z., Jiulun, F. A. N., Xiaoying, P. A. N., et al., Two-dimensional Otsu's curve thresholding segmentation method based on gray and non-local spatial gray feature. *Appl. Res. Comput.* 5:104, 2012.
26. Chunying, P., Jikui, L., and Lixi, H., White blood cells image classification based on improving the connection of FCM and LFP. *J. Imag. Graph.* 18(5):545–551, 2013.
27. Wang, H., Guo, X. H., Jia, Z. W., et al., Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image. *Eur. J. Radiol.* 74(1):124–129, 2010.
28. Ye, X., Lin, X., Dehmeshki, J., et al., Shape-based computer-aided detection of lung nodules in thoracic CT images. *IEEE Trans. Biomed. Eng.* 56(7):1810–1820, 2009.
29. Gonzales, R. C., Woods, R. E., and Eddins, S. L., *Digital image processing using MATLAB*. Pearson Prentice Hall, New Jersey, 2004.
30. Lee, S. L. A., Kouzani, A. Z., and Hu, E. J., Random forest based lung nodule classification aided by clustering. *Comput. Med. Imaging Graph.* 34(7):535–542, 2010.
31. Saien, S., Pilevar, A. H., and Moghaddam, H. A., Refinement of lung nodule candidates based on local geometric shape analysis and Laplacian of Gaussian kernels. *Comput. Biol. Med.* 54:188–198, 2014.
32. Lu, L., Tan, Y., Schwartz, L. H., et al., Hybrid detection of lung nodules on CT scan images. *Med. Phys.* 42(9):5042–5054, 2015.
33. Guo, W., and Li, Q., High performance lung nodule detection schemes in CT using local and global information. *Med. Phys.* 39(8):5157–5168, 2012.
34. Choi, W. J., and Choi, T. S., Automated pulmonary nodule detection system in computed tomography images: a hierarchical block classification approach. *Entropy* 15(2):507–523, 2013.
35. Netto, S. M. B., Silva, A. C., Nunes, R. A., et al., Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Comput. Biol. Med.* 42(11):1110–1121, 2012.
36. Choi, W. J., and Choi, T. S., Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput. Meth. Programs Biomed.* 113(1):37–54, 2014.
37. He, C., Fan, X., and Li, Y., Toward ubiquitous healthcare services with a novel efficient cloud platform. *IEEE Trans. Biomed. Eng.* 60(1):230–234, 2013.
38. Wang, D., He, D., Wang, P., et al., Anonymous two-factor authentication in distributed systems: certain goals are beyond attainment. *IEEE Trans. Dependable Secur. Comput.* 12(4):428–442, 2015.
39. Wang, D., Wang, N., Wang, P., et al., Preserving privacy for free: efficient and provably secure two-factor authentication scheme with user anonymity. *Inf. Sci.* 321:162–178, 2015.
40. He, C. G., Bao, S. D., and Li, Y., A novel tri-factor mutual authentication with biometrics for wireless body sensor networks in healthcare applications. *Int. J. Smart Sens. Intell. Syst.* 6(3):910–931, 2013.
41. He, C. G., Cao, C. Z., Bao S. D. An enhanced role-based access control mechanism for hospital information systems. *Int. Conf. Comput. Intell.* 1001-1005, 2011