La radiologia medica (2020) 125:374–383
https://doi.org/10.1007/s11547-019-01130-9

COMPUTED TOMOGRAPHY

Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations

Guobin Zhang[1] · Zhiyong Yang[1] · Li Gong[1] · Shan Jiang[1,2] · Lu Wang[1] · Hongyun Zhang[1]

Received: 27 August 2019 / Accepted: 27 December 2019 / Published online: 8 January 2020

Abstract
Lung cancer is pointed as a leading cause of cancer death worldwide. Early lung nodule diagnosis has great significance for treating lung cancer and increasing patient survival. In this paper, we present a novel method to classify the malignant from benign lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations (SE-ResNeXt). The state-of-the-art SE-ResNeXt module, which integrates the advantages of SENet for feature recalibration and ResNeXt for feature reuse, has great ability in boosting feature discriminability on imaging pattern recognition. The method is evaluated on the public available LUng Nodule Analysis 2016 (LUNA16) database with 1004 (450 malignant and 554 benign) nodules, achieving an area under the receiver operating characteristic curve (AUC) of 0. 9563 and accuracy of 91.67%. The promising results demonstrate that our method has strong robustness in the classification of nodules. The method has the potential to help radiologists better interpret diagnostic data and differentiate the benign from malignant lung nodules on CT images in clinical practice. To our best knowledge, the effectiveness of SE-ResNeXt on lung nodule classification has not been extensively explored.

Keywords Lung nodule · Classification · Squeeze-and-excitation · CT images · Deep learning

Introduction

Lung cancer has become the leading cause of cancer death around the world with the highest number of mortalities. According to the cancer statistics in the year of 2019, 142,670 lung cancer deaths (23.51% of the total cancer patient) are projected to occur in the United States [1]. The patient's 5-year survival rate will increase from 19 to 54% if lung cancer can be diagnosed at an early stage [1, 2]. Therefore, early lung cancer diagnosis is critical to increase survival effectively.

A lung nodule is a rounded opacity circumscribed parenchymal lesions with a diameter smaller than 3 cm. Lung nodules can be classified into malignant and benign nodules as shown in Fig. 1. Generally, lung nodules can be detected on various medical images such as magnetic resonance imaging (MRI), positron emission tomography (PET), and computed tomography (CT). Among them, low-dose CT has become the most sensitive and commonly used imaging modality for nodule detection and diagnosis. The major advantages of CT screening are high spatial resolution, non-invasiveness, and cost-effective. Nodules in CT scans are observed as regular or an irregular white shadow which makes them more likely to be detected. The national lung screening trial (NLST) also proved that using CT screening achieved a 20% reduction in the mortality of lung cancer [3].

Nodule diagnosis by radiologists is a complex and time-consuming task, because they need to analyze masses of nodule information (margin, size, texture, and shape, etc.) from numerous CT images to obtain an accurate interpretation. Besides, factors such as fatigue, distraction, and the limitation of professional experience will also lead to a misinterpretation of the available image data. Computer-aided diagnosis (CAD) systems, which consist of computer-aided detection (CADe) and computer-aided diagnosis (CADx)

✉ Shan Jiang
shanjmri@tju.edu.cn

[1] School of Mechanical Engineering, Tianjin University, Tianjin 300350, China

[2] Centre for Advanced Mechanisms and Robotics, Tianjin University, 135 Yaguan Road, Jinnan District, Tianjin 300350, China

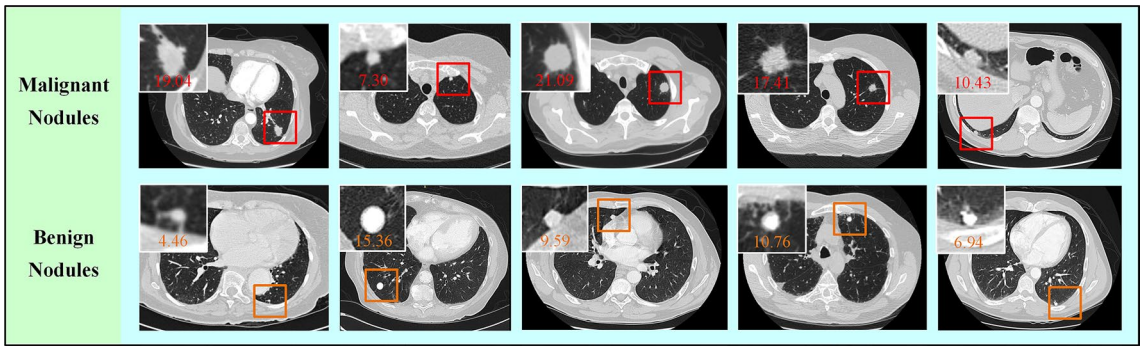

Fig. 1 CT images examples of benign and malignant nodules. The number below each nodule is the nodule size (mm)

systems, have been proven to be an efficient tool in both lung cancer detection and diagnosis [4, 5]. This paper mainly discusses CADx systems, which are aimed at characterizing and classifying the suspicious lesions with CT images.

A popular CADx system mainly consists of four major steps: data acquisition, nodule segmentation, feature extraction, and nodule classification. Among them, feature extraction plays a significant role in nodule diagnosis. Generally, the existing nodule classification algorithms can be divided into two main categories. One category classifies malignant and benign nodules based on handcrafted features which extracted by traditional feature descriptors, the other is based on deep features learned by various deep convolutional neural networks (DCNNs).

Handcrafted features usually include texture, shape (geometry), intensity, and morphology. They can reflect the malignancy and heterogeneity of nodules and offer useful information for nodule diagnosis [6]. de Carvalho et al. [7] extracted texture features using phylogenetic diversity to describe lung nodules and applied support vector machine (SVM) algorithm to classify nodules. Orozco et al. [8] used wavelet feature to represent nodules. Li et al. [9] calculated texture, shape, and intensity features based on the gray-level co-occurrence matrix (GLCM), local binary pattern (LBP), and Gabor filter methods to characterize lung nodules and applied them to an improved random forest (RF) algorithm to classify malignant and benign nodules. Similarly, Gong et al. [10] and Wu et al. [11] found that the combination of texture, shape, and intensity features had an excellent performance in characterizing nodules. Ferreira et al. [12] selected margin sharpness and GLCM-based texture features to describe the nodules. The work in [13] showed that the shape properties of the lung nodules achieved a promising result (93.19% accuracy) in classifying nodules. Akram et al. [14] selected intensity-based statistical features to describe nodules and applied SVM algorithm for nodule classification. Farahani et al. [15] reported on sixteen morphological and statistical features, and an ensemble of three classifiers consisting of multilayer perceptron (MLP), *k*-nearest neighbor (KNN), and SVM to classify benign or malignant lung nodules. Lakshmanaprabu et al. [16] proposed an automatic approach for lung cancer classification. The features they used were histogram features, texture feature, and wavelet features. All the features were fed into an optimal deep neural network (ODNN) for classification. Kaya et al. [17] used shape, size, and texture-based features to characterize nodules, they proven that ensemble classifiers with SVM and RF base classifiers achieved the best performance compared to LDA, KNN, and AdaBoost base classifiers.

Though handcrafted features are popular in characterizing lung nodules, there are still many limitations. For instance, nodule segmentation is a critical step in handcrafted feature-based classification. However, the accuracy of nodule segmentation is affected by many factors such as the irregular shapes of nodules and different image qualities. Additionally, handcrafted features could not fully characterize lung nodules due to its heterogeneity. Therefore, researches are increasingly geared toward utilizing DCNNs to improve classification task. The DCNN offers a uniform framework to learn high discriminative features from various raw image data automatically, which make them more competitive than handcrafted feature-based techniques. Additionally, it has proven to be useful models for tackling various classification tasks [18, 19]. Initially, Hua et al. [20] employed a 2D deep belief network (DBN) and a 2D convolutional neural network (CNN) for nodule classification in CT images. Two handcrafted feature computing methods, scale-invariant feature transform (SIFT) and fractal analysis were implemented for comparison. The results indicated that both CNN and DBN achieved better performance than SIFT and fractal analysis. Tajbakhsh et al. [21] compared the performance of 2D massive-training artificial neural networks (MTANNs) and 2D CNNs in nodule classification. With the development of deep learning techniques, more and more networks were optimized based on CNNs, such as 3D multi-view convolutional neural networks (MV-CNN) [22] and 3D multi-crop convolutional neural network (MC-CNN) [23]. Compared with the 2D networks, 3D CNNs can encode richer

spatial information from CT images to learn more distinguishable features. In general, deeper CNNs are thought to have increased representational power [24]. Therefore, some competitive networks were created for classification task. Nibali et al. [25] applied 2D deep residual networks (ResNet) [26] for lung nodule malignancy classification. Subsequently, dense convolutional network (DenseNet) [27] was created with the advantages of strengthening feature propagation. Liu et al. [28] presented a network named 3D dense convolutional binary-tree network (DenseBTNet), which was based on DenseNet, to tackle the task of nodule classification. Moreover, 3D dual path network (DPN) [29], which integrated the advantages of ResNet and DenseNet, was applied in the work [30] for lung nodule detection and classification. Besides, Qiang et al. [31] proposed a 2D dual-modal supervised deep autoencoder (DSDAE) framework to classify lung nodules. The frameworks used in [32, 33] were based on CNN. Mao et al. [34] used 2D unsupervised deep autoencoder (DAE) for nodule classification.

Several selected works fused handcrafted features and deep features to characterize lung nodules. Xie et al. [35] fused the shape, texture, and deep features at the decision level, employed an AdaBoosted back propagation neural network (BPNN) for nodule classification. Kaya et al. [36] found that combining size, shape, texture features, and deep features (learned by Alexnet) to represent nodules could lead to a promising performance. Paul et al. [37] fused deep features and classical radiomics features to predict malignant nodules. Thought the fused feature-based classification techniques achieved promising results, the limitations suffered in handcrafted feature-based methods were also occurred.

In this paper, a novel 3D DCNN-based framework is presented to study the performance of malignant and benign nodule classification. The presented framework is based on SE-ResNeXt, which integrates the advantages of ResNeXt for feature reuse and SENet for feature recalibration. The proposed method has been validated on the LUNA16 dataset and achieves promising performance in lung nodule classification. To our best knowledge, the effectiveness of SE-ResNeXt on lung nodule classification has not been extensively explored.

In the present work, "Methods" section introduce the proposed method in the classification of malignant and benign lung nodules. Then, experiments and results are presented in "Experiments and results" section to evaluate the proposed method. Finally, discussion and conclusion are given in "Discussion" and "Conclusion" sections.

Methods

The proposed method for the classification of benign and malignant lung nodules consists of three major stages: (1) image data acquisition from the LUNA16 database; (2) nodule extraction; and (3) training SE-ResNeXt for nodule classification. A short summary of the proposed method is presented in Fig. 2.

Data acquisition

The image database employed in this work for both training and testing is based on the LUNA16 database [38], which is refined from the publicly available Lung Image Database Consortium image collection (LIDC–IDRI) [39, 40]. As recommended by the American College of Radiology [41], thin-slice CT scans should be used for nodule detection and classification. Therefore, scans with a slice thickness greater than 3 mm, inconsistent slice spacing or missing slices were discarded. On top of that, nodules whose size is smaller than 3 mm are also removed because these nodules are usually not dangerous according to doctors' experience [42, 43]. After screening, we selected 888 CT scans with 1004 nodules for training and testing in this work. Each CT scan comes with associated XML files with annotations offered by four experienced radiologists. The annotations in the dataset were collected during a two-phase image annotation process accepted by at least 3 out of 4 radiologists. It almost includes all certain characteristics such as nodule locations, sizes, and malignancy levels (from 1 to 5). Level 1: highly unlikely for cancer; Level 2: moderately unlikely for cancer; Level 3: intermediate likelihood; Level 4: moderately suspicious for cancer; and Level 5: highly suspicious for cancer.

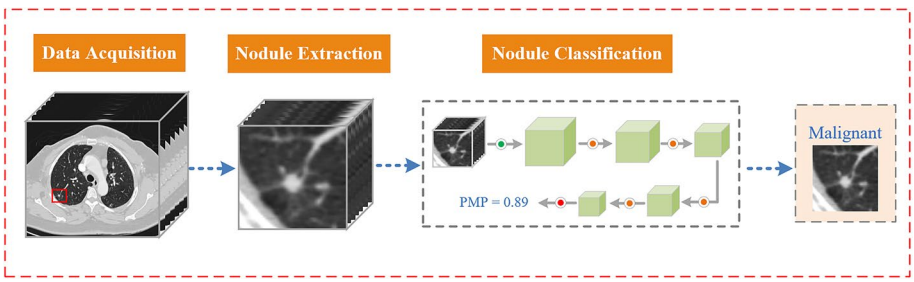

Fig. 2 Framework of the proposed nodule diagnosis method

Table 1 The distribution of nodule sizes in the LUNA16 database

Number of nodules	Nodule size (mm)				
	(0, 10]	(10, 20]	(20, 30]	(30, 32]	Total
Malignant	201	202	45	2	450
Benign	539	15	0	0	554

Nodule extraction

To train learning model efficiently, 3D patches containing lung nodules are cropped from raw CT images and fed into network. Each patch is centered on the nodule center. According to the annotations from professional radiologists, the two largest size of selected nodules are 32.27 mm and 30.61 mm, and the rest nodule sizes are between 3 mm and 30 mm. The distribution of selected nodule sizes in the LUNA16 database is described in Table 1. It is obvious that benign nodules usually have smaller sizes than malignant nodules. To better capture most of selected nodules, the size of input patch is set to $32 \times 32 \times 32$.

SE-ResNeXt

The SE-ResNeXt adopted in this work for nodule classification combines the ResNeXt and SENet. ResNeXt is a highly modularized network for image classification. Compared to ResNet, it introduces the concept of "cardinality," which is an essential factor in classification task. It is well known that the depth and width play a significant role in the part of representational power to a deep neural network. However, with the network going deeper or wider, the vanishing/exploding gradients or degradation problem will be exposed. In this case, the depth and width might start to give diminishing returns. Fortunately, the ResNeXt manages to avoid this problem and explores an efficient way to gain higher classification accuracy, which is increasing cardinality. It has been proven that increasing cardinality can achieve a better classification accuracy than gong wider or deeper [44]. Figure 3 provides a structure of a ResNeXt block with cardinality of 32.

A ResNeXt module can be defined as:

$$U^{\text{res}} = X + \sum_{i=1}^{C} \zeta_i(X) \tag{1}$$

where $\sum_{i=1}^{C} \zeta_i(X)$ represents the aggregated transformation, C is the size of the set of transformations, X is the input and U^{res} is the output. The operation $X + \sum_{i=1}^{C} \zeta_i(X)$ is performed by a shortcut connection and element-wise addition. The major advantage of shortcut connection adopted in ResNeXt is the feature reusage which can reduce the feature redundancy.

As shown in Fig. 4, SE module consists of five steps sequentially: a global average pooling (Avg Pool), a fully connected (FC) layer, a ReLU, a FC layer, and a sigmoid (Sig). There are two operations in a SE module: squeeze operation and excitation operation. The squeeze operation is achieved by global average pooling, and the excitation operation is made up of two FC layers, a ReLU and a sigmoid.

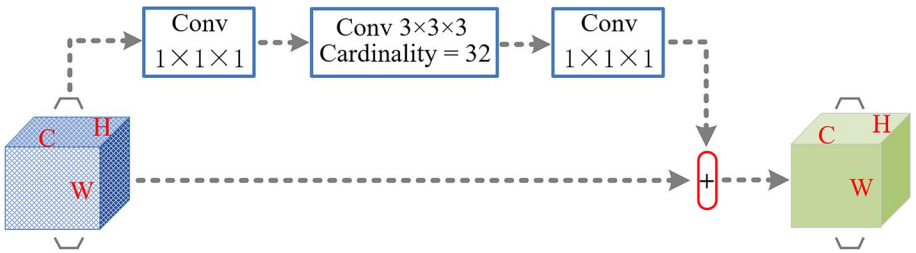

Fig. 3 Structure of a ResNeXt module

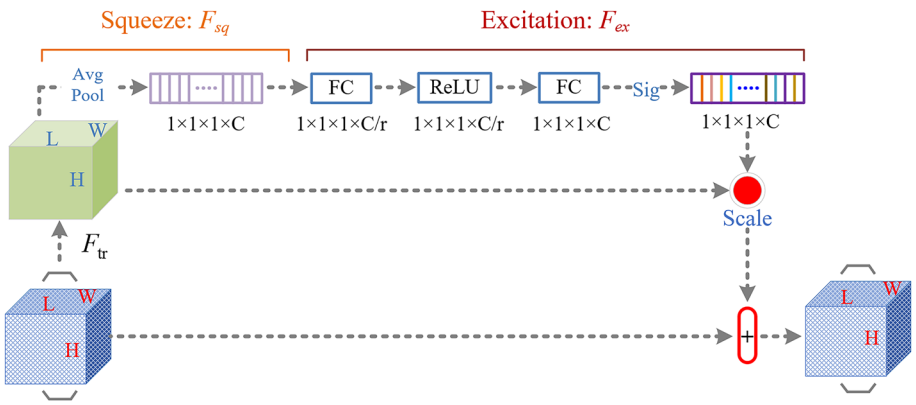

Fig. 4 Structure of a SENet module

The squeeze operation is proposed to generate channel-wise statistics which can be defined as

$$Z_n = F_{\mathrm{sq}}\left(U_n^{\mathrm{res}}\right) = \frac{1}{L \times W \times H} \sum_{i=1}^{L} \sum_{j=1}^{W} \sum_{k=1}^{H} U_n^{\mathrm{res}}(i,j,k) \tag{2}$$

where Z_n represents the nth element of $Z = \left[Z_1, Z_2, Z_3 \ldots Z_n\right]$ and $U_n^{\mathrm{res}} \in \mathbb{R}^{L \times W \times H}$ represent the nth channel from the residual feature U^{res}. n is the number of channels of the residual mapping, L, W, and H represent the spatial length, width, and height of U_n^{res}.

To utilize the aggregated information acquired in the squeeze operation efficiently, the excitation operation is employed to fully capture channel-wise dependencies. To address this objective, a gating mechanism is proposed, which can be described as

$$S = F_{\mathrm{ex}}(Z, W) = \sigma\left(W_2 \delta\left(W_1 Z\right)\right) \tag{3}$$

The gating mechanism is parameterized by two FC layers to reduce the model complexity and enhance generalization, which are dimensionality-reduction layer with parameters $\mathbf{W}1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and reduction ratio r=16, a ReLU δ and a dimensionality-increasing layer with parameters $\mathbf{W}2 \in \mathbb{R}^{C \times \frac{C}{r}}$. Besides, the σ represents sigmoid activation. After excitation operation, the final output of the SE block $\tilde{X}_n^{\mathrm{res}}$ is obtained by rescaling the residual feature map U^{res} with the activations:

$$\tilde{X}_n^{\mathrm{res}} = F_{\mathrm{scale}}\left(U_n^{\mathrm{res}}, S_n\right) = S_n \cdot U_n^{\mathrm{res}} \tag{4}$$

where $\tilde{X}^{\mathrm{res}} = \left[\tilde{X}_1^{\mathrm{res}}, \tilde{X}_2^{\mathrm{res}}, \ldots, \tilde{X}_n^{\mathrm{res}}\right]$ represents the final calibrated residual feature, and F_{scale} is the channel-wise multiplication. S_n is the scale value which represents the importance degree of the nth channel. The output feature of a SE-ResNeXt module Y can be acquired by the operation of shortcut connection and element-wise addition with the activation function δ:

$$Y = \delta\left(\tilde{X}^{\mathrm{res}} + X\right) \tag{5}$$

The 3D SE-ResNeXt module makes it more effective to optimize the deep network to boost feature discriminability of lung nodules. Moreover, it integrates the advantages of feature reuse, selectively emphasizing informative nodule features and suppressing less efficient ones. Therefore, the used 3D SE-ResNeXt is reliable and effective in learning high discriminative features to classify malignant and benign nodules from raw CT images automatically.

Nodule classification network architecture

The network for lung nodule classification based on 3D SE-ResNeXt is shown in Fig. 5. The input of the network is 32 × 32 × 32 image patches. Firstly, a convolutional level (3 × 3×3) is applied to calculate the feature map. Then, four 3D SE-ResNet blocks interleaved with four 3D max-pooling layers are employed. There are three, four, six, and three SE-ResNet modules in the first to fourth SE-ResNet block, respectively. Each block is followed by a max-pooling layer. The network ends with a global average pooling layer, a FC layer, and softmax. For a binary classification problem, a cross-entropy loss function is adopted in this work. Finally, the predicted malignant probability (PMP) of each lung nodule can be acquired after the softmax.

Experiments and results

The proposed method was evaluated on the LUNA16 database with 1004 (554 benign and 450 malignant) nodules. All the image data were randomly and equally split into 10 subsets for tenfold cross-validation. For each fold, nine subsets were selected for training and one subset for testing.

For nodule-level experiments, a binary classification was performed by calculating the average level for each selected nodule and defined that if the final average level is equal to 3, it is considered to be uncertain about malignant or benign nodules and hence is removed. For the nodules with final average level greater than 3, they are labeled as malignant nodule. Otherwise, they are considered to be benign. All the annotations were acquired from the LIDC–IDRI database. Besides, the malignant probability was also predicted to help the radiologists make a decision. All experiments

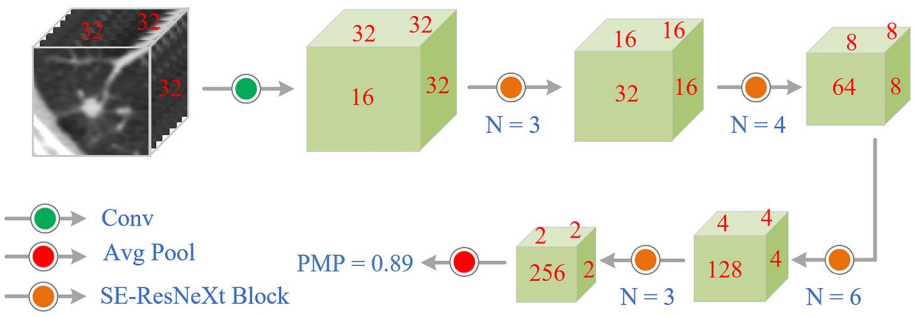

Fig. 5 Architecture of the proposed network for lung nodule classification

were carried out under a Linux OS with 6 NVIDIA TITAN Xp 12 GB GPUs.

Implementation details

The raw image data acquired from the LUNA16 database were screened by different CT scanners with slightly different spatial resolutions. Therefore, all the CT images were resampled to a specific resolution of $1 \times 1 \times 1$ mm before training. The image pixel values from the cropped patch were clipped to $[-1200, 600]$ and normalized to $[0, 1]$. Additionally, data augmentation was introduced to training dataset to alleviate the overfitting problem by left–right flipping and rotating the patches randomly. The learning process was optimized by stochastic gradient descent (SGD) algorithm with a momentum of 0.9, and the weight decay was set to 1×10^{-4}. Furthermore, the batch size was limited to 24 because of the 12 GB GPU memory constraint. The learning model was trained for 150 epochs. The initial learning rate was 0.01, reduced to 0.001 after 60 epochs and 0.0001 after 120 epochs. The final performance of the proposed method was the average performance on the ten test folds, which was evaluated by measuring the AUC score and accuracy. The accuracy is the percentage of lung nodules that are correctly classified, which is defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{FN}} \tag{6}$$

where the TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Implementation of comparison networks

Different variants of networks were trained and evaluated for comparison. The SE-ResNeXt we adopted in this work combined the ResNeXt and SENet. Therefore, the abilities of ResNeXt and SENet in nodule classification were explored to illustrate the potential advantage of SE-ResNeXt. Additionally, to prove the performance of ResNeXt was superior to ResNet, contrast experiments based on ResNet and ResNeXt were also designed. It should be noted that the ResNet, ResNeXt, SE-ResNet, and SE-ResNeXt were 50-layer. The classification accuracies of selected learning models were shown in Fig. 6. Compared to ResNet (88.64%), both ResNeXt and SENet achieved better accuracies of 89.04% and 89.67%. Besides, the SE-ResNeXt also achieved a better performance than SE-ResNet with the best classification accuracy of 91.67%.

An additional way of measuring the performance of the proposed method was used. A receiver operating characteristic (ROC) curve indicated the true positive rate as a function of the false positive rate (as shown in Fig. 7). This provided

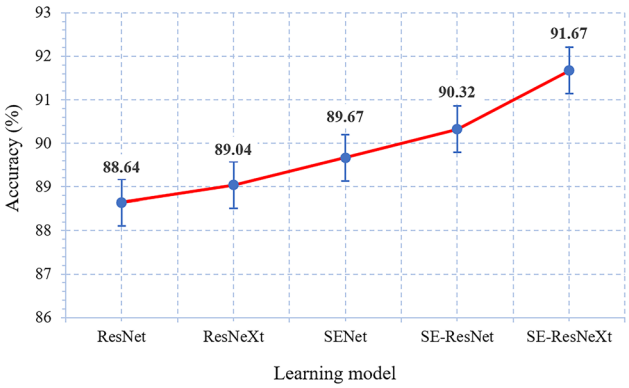

Fig. 6 Classification accuracies of different frameworks

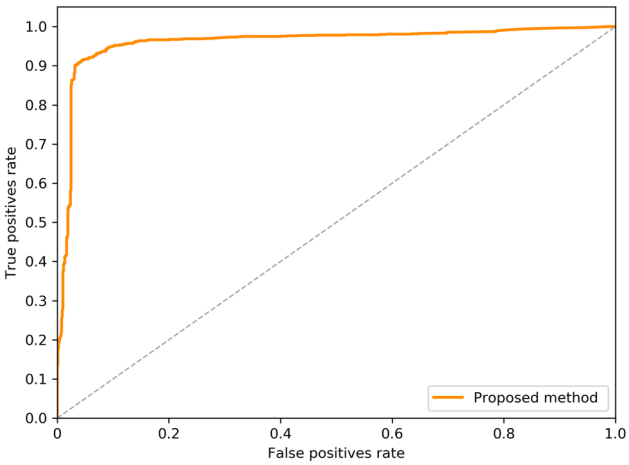

Fig. 7 ROC curve for the proposed method

evidence that the proposed method achieved a promising AUC of 0.9563.

Figure 8 provided many examples of nodule classification results. The malignant and benign nodules, which were classified correctly by our methods, are presented in the first and second row, respectively. Furthermore, a comparison between the average malignancy level from four professional radiologists and the predicted malignant probability of each nodule was made. It was observed that the predicted malignant probability of each nodule was greatly consistent with the average diagnosis level.

Discussion

In this paper, a novel CADx scheme for lung nodule classification with CT images is presented. The method employs 3D SE-ResNeXt to learn high discriminative features for malignant and benign lung nodule classification. Experimental results on the LUNA16 dataset demonstrate that the

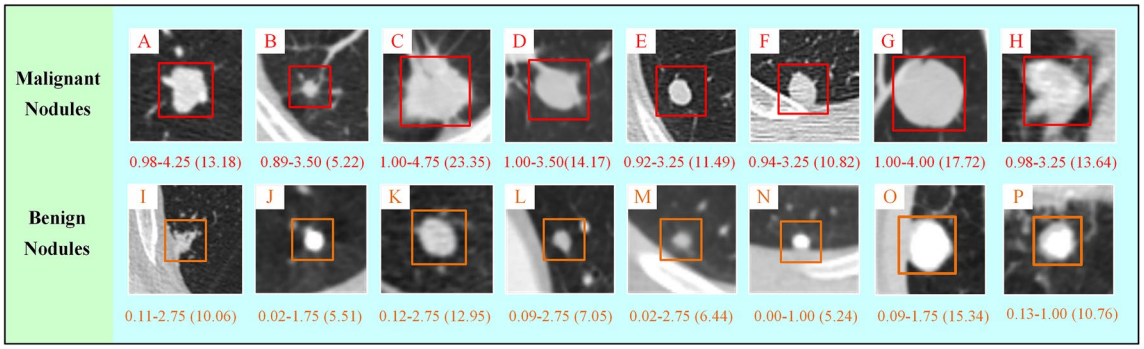

Fig. 8 Classification results visualization of benign and malignant nodule. The number below each nodule is shown as predicted malignant probability, the average malignancy level of four professional radiologists and the nodule diameter (mm)

proposed method achieves promising performance with a classification accuracy of 91.67% and AUC of 0.9563. To further illustrate the superiority of the proposed method in the diagnosis of lung nodules, several existing state-of-the-art methods based on handcrafted features and deep features are selected for comparison as follows.

Comparison of the proposed method with other methods based on handcrafted features

Table 2 provides a comparison of our method with state-of-the-art methods based on handcrafted features and various classifiers. The work in [7, 13, 14, 15] used morphology, intensity, geometry, and texture features to characterize lung nodule, respectively. Though the classification accuracies achieved were promising, only using one kind of handcrafted feature is unwise, which might affect the generalization ability of classifiers, especially when worked on different public databases. Moreover, only 84 CT scans and 120 images were used in [14, 15] for training and testing, the results were less persuasive. Besides, the AUC value was not mentioned in [7, 13]. Both Li et al. [9] and Gong et al. [10] used the same kinds of handcrafted features to represent lung nodules and acquired similar accuracies of 90%. But the training and testing database in [10] was private, which prevented the replication of the results for comparison. Similarly, Wu et al. [11] also used intensity, shape, and texture features to describe nodules and achieved a higher accuracy of 92.37%. Ferreira et al. [12] achieved an AUC of 0.858 and accuracy of 80.0%, which were lower than ours. The works in [35, 36, 37] were based on hybrid features, which fused handcrafted features and deep features. These methods described lung nodules systematically, but the computation load of algorithm would also increase. Besides, few training data were used in [36] and the AUC value was not given.

Table 2 Comparison of the proposed with other methods based on handcrafted features

Authors	Year	Feature (s)	Classifier (s)	Database	Samples	AUC	Accuracy (%)
Akram et al. [14]	2015	I	SVM	LIDC	84 Scans	0.9967	96.54
de Carvalho et al. [13]	2016	S	SVM	LIDC–IDRI	1405 nodules	–	93.19
de Carvalho et al. [7]	2017	T	SVM	LIDC–IDRI	1405 nodules	–	92.52
Ferreira et al. [12]	2017	Mg + T	RF	LIDC	1171 nodules	0.858	80.0
Farahani et al. [15]	2018	Mp	MLP, KNN and SVM	LIDC	120 images	–	96.5
Li et al. [9]	2018	I + S + T	RF	LIDC and GHGMC	1300 nodules	0.95	90
Gong et al. [10]	2018	I + S + T	SVM, NBC, LDA	Private	562 nodules	0.93, 0.93, 0.95	90.12, 90.12, 90.12
Xie et al. [35]	2018	T + S + D	BPNN	LIDC–IDRI	2669 nodules	0.9665	89.53
Kaya et al. [36]	2018	T + S + D	CCs	LIDC	439 nodules	–	84.70
Paul et al. [37]	2018	D + CR	RF	NLST	18,792 samples	0.78	76.79
Wu et al. [11]	2019	I + S + T	RF	LIDC	952 nodules	0.9702	92.37
Our work	2019	SE-ResNeXt		LUNA16	1004 nodules	0.9563	91.67

T texture feature, *W* wavelet feature, *S* shape (geometry) feature, *D* deep feature, *I* intensity feature, *Mg* margin feature, *Mp* morphology feature, *Sz* size feature, *CR* classical radiomics feature

Though handcrafted features (texture, shape, and intensity features, etc.) can reflect the structure of nodules more visualize and provide useful information to describe the nodules in CT images, but there still exist many limitations, such as low generalization ability due to the variety of nodules, low level of automation from accurate nodule segmentation. As shown in Fig. 8, for the malignant nodule "*E,*" the average malignant level was 3.25 according to the annotations and the proposed method correctly classified the nodule with a predicted malignant probability of 0.92. However, it had clear border and round shape, which are indications of benignant nodule. As well as benign nodule "*I,*" the average malignant level was 2.75 and the malignant probability we predicted was 0.11. But the border was unclear and the shape was irregular, which are all indications of malignant nodule. Furthermore, nodules "*K,*" "*O,*" and "*P*" had larger sizes than nodule "*B,*" but had a lower cancer probability. Therefore, it is unwise to depends enormously on texture, shape, or size handcrafted features, which might affect the classification accuracy. Though several proposed methods achieved promising classification accuracies, the performance was uncertain when using different databases.

Comparison of the proposed method with other methods based on deep features

The major advantage of deep learning techniques is the ability to learn high discriminative features from various raw CT image automatically. Table 3 provides a comparison with other methods based on deep features in lung nodule classification. The 3D MC-CNN proposed in [23] was based on CNN and it was optimized with the multi-crop pooling strategy which cropped different regions from convolutional feature maps and then applied max-pooling different times. They achieved an AUC of 0.93 with an accuracy of 87.14%. Tajbakhsh et al. [21] proved that the MTANNs had a better performance than CNNs in nodule classification. However, both MTANNs and CNNs adopted were based on 2D image data. The proposed 3D CNNs in [33], by contrast, achieved a high AUC than 2D CNNs and MTANNs, because 3D CNNs can encode richer spatial information from CT images to learn more distinguishable features. The 2D ResNet used in [25] focused on feature reusage to reduce the feature redundancy with the skip connection, but it was limited to explore new features. The 3D DenseBTNet proposed in [28] was based on DenseNet, and it was optimized with the center-crop operation. It was better at exploring new features but it might suffer from feature redundancy due to its densely connected mechanism. Therefore, the classification accuracies achieved by ResNet and DenseBTNet were lower than DPN, which integrated the advantages of ResNet for feature reuse and DenseNet for exploring new features. However, the 3D DPN used in [30] had a higher complexity, which resulted in time-consuming. Furthermore, the performance of [34] were less persuasive due to the less training data, though the classification accuracy acquired was higher than ours. Qiang et al. [31] proposed the 2D DSDAE for nodule classification, but the experiments were performed on private database, which prevented the replication of the results for comparison.

Analysis of the proposed method's performance

In the handcrafted feature-based CADx systems, extracting and choosing features are time-consuming and complex. Moreover, the performance is still uncertain when evaluating on different databases because the handcrafted feature-based CADx systems are very sensitive to small variations of nodules types. Therefore, a high-quality feature set of lung nodules is becoming more and more necessary. Deep learning techniques have strong ability to boost feature discriminability of lung nodules automatically, and the learned features have the potential in maintaining a sustainable and reliable performance on the ever-changing database. In this study, we proposed an effective lung nodule classification method based on 3D DCNN framework with SE-ResNeXt modules. The SE-ResNeXt integrated the advantages of ResNeXt for feature reuse and SENet for feature recalibration, showed a

Table 3 Comparison of the proposed with other methods based on deep features

Authors	Year	Method (s)	Database	Samples	AUC	Accuracy (%)
Shen et al. [23]	2016	3D MC-CNN	LIDC–IDRI	2618 nodules	0.93	87.14
Tajbakhsh et al. [21]	2016	2D MTANNs, CNNs	Private	489 nodules	0.8806, 0.7755	–
Nibali et al. [25]	2017	2D ResNet	LIDC–IDRI	831 nodules	0.9459	89.90
Qiang et al. [31]	2017	2D DSDAE	Private	2810 nodules	0.94	92.81
Zhu et al. [30]	2018	3D DPN	LUNA16	1004 nodules	–	90.44
Mao et al. [34]	2018	2D DAE	ELCAP	50 scans	–	93.9
Dai et al. [33]	2018	3D CNNs	LIDC	1011 nodules	0.969	91.47
Liu et al. [28]	2018	3D DenseBTNet	LIDC–IDRI	2001 nodules	0.9360	89.50
Our work	2019	3D SE-ResNeXt	LUNA16	1004 nodules	0.9563	91.67

promising performance in lung nodule classification. Considering the heterogeneity of lung nodules, 3D networks can capture more rich spatial information and extract more representative features than 2D networks. Although 3D networks are memory consumption limited by GPU memory, they are more suitable for the lung nodule classification task. Moreover, the 3D SE-ResNeXt module used in this work allows optimization of the deep network more effectively to boost feature discriminability of lung nodules. Besides, the proposed method works on raw CT images without nodule or lung segmentation, which enhances the level of automation. From Fig. 6, it was obvious that the ResNeXt achieved a high classification than ResNet, and the SE-ResNeXt was also superior to SE-ResNet. It mainly benefited from introducing the "cardinality" operation to ResNeXt, which is more effective than going deeper or wider in classification task. Benefitting from SENet, SE-ResNeXt, and SE-ResNet architecture achieved a 2.63% and 1.68% improvement in classification accuracy compared the ResNeXt and ResNet, respectively. Experimental results demonstrate that the proposed method is robust in the nodule diagnosis task and it has the potential to help radiologists to interpret diagnostic data and make decision.

Conclusion

In this paper, we develop a novel CADx scheme for lung nodule malignancy classification with CT images. The network employs 3D SE-ResNeXt to learn high discriminative features to classify malignant and benign lung nodules. Experimental results on the LUNA16 database demonstrate that the proposed method achieves a promising performance with an AUC of 0.9563 and accuracy of 91.67%, which are more accurate than current methods based on traditional features and deep features. Our future work will aim to optimize existing methods and develop deep learning techniques to characterize lung nodules for classification.

Acknowledgements This study was funded by the National Natural Science Foundation of China (Grant Number 81871457), National Natural Science Foundation of China (Grant Number 51811530310), National Natural Science Foundation of China (Grant Number 51775368) and Science and Technology Planning Project of Guangdong Province, China (Grant Number 2017B020210004) and the Science and Technology Project of Tianjin (Grant Number 18YFZCSY01300).

Compliance with ethical standards

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics. CA-Cancer J Clin 68:7–30
2. Bach PB, Mirkin JN, Oliver TK et al (2012) Benefits and harms of CT screening for lung cancer a systematic review. JAMA-J Am Med Assoc 307:2418–2429
3. Rasmussen JF, Siersma V, Pedersen JH et al (2014) Healthcare costs in the Danish randomized controlled lung cancer CT-screening trial: a registry study. Lung Cancer 83:347–355
4. Awai K, Murao K, Ozawa A et al (2004) Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. Radiology 230:347–352
5. Awai K, Murao K, Ozawa A et al (2006) Pulmonary nodules: estimation of malignancy at thin-section helical CT: effect of computer-aided diagnosis on performance of radiologists. Radiology 239:276–284
6. Metz S, Ganter C, Lorenzen S et al (2015) Multiparametric MR and PET imaging of intratumoral biological heterogeneity in patients with metastatic lung cancer using voxel-by-voxel analysis. PLoS ONE 10:e0132386
7. de Carvalho AS, Silva AC, de Paiva AC et al (2017) Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM. J Digit Imaging 30:812–822
8. Orozco HM, Villegas OOV, Sanchez VGC et al (2015) Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. Biomed Eng Online 14:9
9. Li XX, Li B, Tian LF et al (2018) Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. IET Image Process 2:1253–1264
10. Gong J, Liu JY, Sun XW et al (2018) Computer-aided diagnosis of lung cancer: the effect of training datasets on classification accuracy of lung nodules. Phys Med Biol 63:035036
11. Wu WH, Hu HH, Jing G et al (2019) Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis. Phys Med Biol 64:035017
12. Ferreira JR, Oliveira MC, de Azevedo-Marques PM (2018) Characterization of pulmonary nodules based on features of margin sharpness and texture. J Digit Imaging 31:451–463
13. de Carvalho AO, Silva AC, de Paiva AC et al (2017) Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm, and SVM. Med Biol Eng Comput 55:1129–1146
14. Akram S, Javed MY, Hussain A et al (2015) Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques. J Exp Theor Artif Intell 27:737–751
15. Farahani FV, Ahmadi A, Zarandi MHF (2018) Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning. Math Comput Simul 149:48–68
16. Lakshmanaprabu SK, Mohanty SN, Shankar K et al (2019) Optimal deep learning model for classification of lung cancer on CT images. Future Gener Comput Syst 2:374–382

17. Kaya A, Can AB (2015) A weighted rule based method for predicting malignancy of pulmonary nodules by nodule characteristics. J Biomed Inform 56:69–79
18. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60:84–90
19. Ren S, He K, Girshick R et al (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal 39:1137–1149
20. Hua KL, Hsu CH, Hidayati HC et al (2015) Computer-aided classification of lung nodules on computed tomography images via deep learning technique. Onco Targets Ther 8:2015–2022
21. Tajbakhsh N, Suzuki K (2017) Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: mTANNs vs. CNNs. Pattern Recogn 63:476–486
22. Kang GX, Liu K, Hou BB et al (2017) 3D multi-view convolutional neural networks for lung nodule classification. PLoS ONE 12:e0188290
23. Shen W, Zhou M, Yang F et al (2017) Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn 61:663–673
24. Telgarsky M (2016) Benefits of depth in neural networks. arXiv :1602.04485
25. Nibali A, He Z, Wollersheim D (2017) Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol 12:1799–1808
26. He KM, Zhang XY, Ren SQ et al (2016) Deep Residual Learning for Image Recognition. CVPR 770-778
27. Huang G, Liu Z, van der Maaten L et al (2017) Densely connected convolutional networks. In: CVPR, pp 2261–2269
28. Liu YJ, Hao PY, Zhang P et al (2018) Dense convolutional binary-tree networks for lung nodule classification. IEEE Access 6:49080–49088
29. Chen Y, Li JN, Xiao HX et al (2017) Dual path networks. arXiv :1707.01629
30. Zhu WT, Liu CC, Fan W et al (2018) DeepLung: deep 3D dual path nets for automated pulmonary nodule detection and classification. In: WACV, pp 673–681
31. Qiang Y, Ge L, Zhao X et al (2017) Pulmonary nodule diagnosis using dual-modal supervised autoencoder based on extreme learning machine. Expert Syst 34:e12224
32. Luckehe D, von Voigt G (2018) Evolutionary image simplification for lung nodule classification with convolutional neural networks. Int J Comput Assist Radiol 13:1499–1513
33. Dai YJ, Yan SJ, Zheng B et al (2018) Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification. Phys Med Biol 63:245004
34. Mao KM, Tang RJ, Wang XQ et al (2018) Feature representation using deep autoencoder for lung nodule image classification. Complexity 2018:1–11. https://doi.org/10.1155/2018/3078374
35. Xie YT, Zhang JP, Xia Y et al (2018) Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. Inform Fusion 42:102–110
36. Kaya A (2018) Cascaded classifiers and stacking methods for classification of pulmonary nodule characteristics. Comput Meth Prog Bio 166:77–89
37. Paul R, Hawkins SH, Schabath MB et al (2018) Predicting malignant nodules by fusing deep features with classical radiomics features. J Digit Imaging 5:011021
38. Setio AAA, Traverso A, de Bel T et al (2017) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 42:1–13
39. McNitt-Gray MF, Armato SG, Meyer CR et al (2007) The lung image database consortium (LIDC) data collection process for nodule detection and annotation. Acad Radiol 14:1464–1474
40. Armato SG, McLennan G, Bidaut L et al (2011) The lung image database consortium, (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys 38:915–931
41. Kazerooni EA, Austin JHM, Black WC et al (2014) ACR–STR practice parameter for the performance and reporting of lung cancer screening thoracic computed tomography (CT). J Thorac Imaging 29:310–316
42. MacMahon H, Naidich DP, Goo JM et al (2017) Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner society. Radiology 284:228–243
43. Aberle DR, Adams AM, Berg CD et al (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365:395–409
44. Xie SN, Girshick R, Dollar P et al (2017) Aggregated residual transformations for deep neural networks. In: CVPR, pp 5987–5995