

# A Hybrid Deep Learning Approach for Lung Nodule Classification

Cheng Ren and Shouming Hou

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo Henan, 454003, China

---

**Abstract:** Lung cancer has the highest morbidity and mortality rates worldwide. Pulmonary nodules are an early manifestation of lung cancer. Therefore, accurate classification of pulmonary nodules is of great significance for the early diagnosis and treatment of lung cancer. However, the classification of lung nodules is a complex and time-consuming task requiring extensive image reading and analysis by expert radiologists. Therefore, using deep learning technology to assist doctors in detecting and classifying pulmonary nodules has become a current research trend. A lightweight classification model named Res-VGG is proposed for classifying lung nodules as benign or malignant. The Res-VGG model improves on VGG16 by reducing the use of convolutional and fully connected layers. To reduce overfitting, residual connections are introduced. The training of the model was performed on the LUNA16 database, and a ten-fold cross-validation method was used to evaluate the performance of the model. In addition, the Res-VGG model was compared with three other common classification networks, and the results showed that the Res-VGG model outperformed the other models in terms of accuracy, sensitivity, and specificity.

**Keywords:** Lung Nodule Benign and Malignant Classification; Deep Learning; ResNet; Convolutional Neural Network; VGGNet.

---

## 1. Introduction

In the medical field, the detection and classification of pulmonary nodules is an important and challenging task. Pulmonary nodules are small pieces of lung tissue that appear on lung computed tomography (CT) scans, and they may be benign or malignant [1]. Most lung cancers originate from small malignant nodules. However, automatic identification of benign and malignant pulmonary nodules in CT images becomes a challenge due to the wide range of shape and texture variations of pulmonary nodules, as well as the visual similarity between benign and malignant pulmonary nodules [2].

Deep learning, as an important branch of machine learning, has achieved remarkable success in various fields, including natural language processing, speech recognition, computer vision, etc. Deep learning models, such as convolutional neural networks (CNN), have been widely used in medical image analysis, including the detection and classification of pulmonary nodules. However, although deep learning has made significant progress in these fields, some challenges still exist. For example, some deep learning models, such as AlexNet, VGGNet, Microsoft ResNet, DenseNets, etc., may face the problem of decreased accuracy when increasing the network depth.

To solve this problem, this paper proposes a new deep learning model named Res-VGG. This model combines two different deep learning models, namely VGG16 and ResNet.

The rest of the paper is organized as follows. Section 2 analyzes the related works. Section 3 presents the proposed methodology for the classification of lung nodules. The experimental results obtained are discussed in Section 4. The conclusion of the paper can be found in Section 5.

## 2. Related Works

After lung nodules are successfully detected, the next key task is to determine whether these detected nodules are benign

or malignant. The diagnostic part of the pulmonary nodule Computer aided diagnosis (CAD) system can automatically distinguish between benign and malignant nodules based on characteristics such as size, shape, and appearance of the nodules. To evaluate the performance of CAD systems on binary classification tasks, the receiver operating characteristic (ROC) curve is usually used for evaluation. The area under the ROC curve (AUC) is a commonly used evaluation index, which can quantitatively reflect the performance of a classifier under different thresholds. The malignancy of pulmonary nodules is closely related to their geometric size, shape, and appearance description. Therefore, the pulmonary nodule CAD system diagnosis part automatically extracts effective features such as texture, shape, and growth rate of pulmonary nodules from CT images to distinguish pulmonary nodules. nature.

CNN and big data technology have also achieved remarkable results in the diagnosis and classification of pulmonary nodules. In fact, classification is the most widely used field of CNN. Starting from AlexNet, ResNet, DenseNet, VGG network, etc., most deep learning networks originating from the field of computer vision have been applied in medical image classification.

First, based on the traditional CNN structure, classification can be achieved through a series of convolution operations [3]. The study by Yang et al. [3] showed that simple geometric features cannot capture important features of lung nodules that support classification using original images and nodule masks containing rich nodule information. Hua et al. [4] introduced deep belief networks and CNN in the context of nodule classification. Experimental results show that deep learning methods can achieve better recognition results and have broad application prospects in the field of CAD diagnostic applications. Li et al. [5] designed a network consisting of three convolutional layers, a max pooling layer, and three fully connected layers. They identified input nodule images as solid, semi-solid, and ground-glass opacity (GGO) nodules or non-nodules. On the LIDC dataset, the sensitivity

of this method is 87.1%. Instead of using 3D volumes, Sahu et al. [6] used a lightweight network to process cross-sections of nodules, used compact representations and convolutions for nodule classification, and achieved an average of 93.18% on the LIDC dataset Accuracy.

Residual Networks (ResNet) [7] is another commonly used network, which is easier to converge during the training phase due to skip connections. Gong et al. [8] used a CNN model based on residual learning similar to the ResNet structure to predict the possibility of ground glass nodules being invasive adenocarcinoma (IA). The AUC of this method for classifying IA and non-IA ground-glass nodules was  $0.92 \pm 0.03$ . Xia et al. [9] proposed a recursive residual CNN based on U-Net to segment nodules. They combined deep learning and radiomics features to classify non-IA nodules and IA nodules, and the obtained AUC value was  $0.90 \pm 0.03$ . Wu et al. [10] used a 50-layer ResNet network structure as the initial model and combined residual learning and transfer learning to achieve an average accuracy of 98.23%.

The advantage of DenseNet is that it improves the information flow and gradient of the entire network. Each layer can directly obtain the gradient from the loss function and the original input signal, thus forming implicit deep supervision and making the network easy to train. However, this network takes up more GPU memory and requires more training samples. For pulmonary nodule classification, Li et al. [11] used a multi-task learning deep neural network based on 3D DenseNet and achieved 88.8% results in identifying benign or malignant pulmonary nodules. Baldwin et al. [12] used a similar DenseNet structure to design a lung cancer prediction CNN (LCP-CNN) to predict malignant tumors in pulmonary nodules, with an AUC of 89.6%.

Ren et al. [13] are based on the encoder and decoder structure and use this structure to learn the manifold of the original nodule image and classify it based on Fully Convolutional Networks (FCNs) from the latent space. The classification accuracy was 0.90, the sensitivity was 0.81, and the specificity was 0.95. Lakshmanprabu et al. [14] used LDA to reduce the feature dimension and used an improved gravity search algorithm for classification. The algorithm has a diagnostic sensitivity of 95.3% and a specificity of 96.2% for malignant tumors.

In order to study different network structures, Song et al. [15] compared three types of deep neural networks (such as CNN, DNN, and stacked encoders and decoders) for benign and malignant pulmonary nodules classification. The results show that the CNN network achieved the best performance, with an accuracy of 84.15%, a sensitivity of 83.96%, and a specificity of 84.32%. Overall, the ResNet structure outperformed CNN and DenseNet, as well as autoencoder networks for nodule classification. Part of the reason is that residual learning can effectively retain the information in the original feature map and learn the task through residual blocks, which is more efficient than the convolution of the original feature map. On the other hand, autoencoders compress the original input and feature maps into low-dimensional vectors, which have the advantage of facilitating statistical modeling, but will lose a lot of image information, and may require a large number of convolution kernels to store the information learned from training samples. Information".

### 3. Materials and Methods

#### 3.1. Residual Networks

Residual Networks (ResNet) was proposed by Kaiming He and four other Chinese from Microsoft Research [7]. They successfully trained a 152-layer neural network and won the championship in the ILSVRC2015 competition. The error rate in the top 5 was 3.57%. At the same time, the number of parameters was lower than that of VGGNet, and the effect was very outstanding. The principle of ResNet is that based on the original CNN, a residual block is introduced. Each residual block contains multiple convolution layers and normalization layers, and the output and input are added through cross-layer connections. combination of forms. This cross-layer connection allows the model to be optimized during training, thereby improving model accuracy.

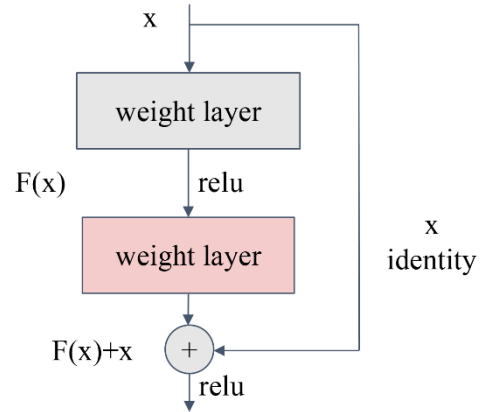


Fig 1. The residual block

Residual learning is used for every few stacked layers. The structure of the residual module is shown in Figure 1. The module is defined as:

$$y = F(x, \{W_i\}) + x \quad (1)$$

Here  $x$  and  $y$  are the input and output vectors of the layer under consideration. The function  $F(x, \{W_i\})$  represents the residual mapping to be learned. For the two-layer example in Figure 1,  $F = W_2 \sigma(W_1 x)$ , where  $\sigma$  represents ReLU [16], the bias is omitted to simplify the notation. The operation  $F + x$  is performed via shortcut connection and element addition. Take the second nonlinearity after addition, which is  $\sigma(y)$ .

The shortcut connection in equation (1) neither introduces additional parameters nor increases computational complexity. In formula 1 the dimensions of  $x$  and  $F$  must be equal. If this is not the case (e.g. when changing input or output channels), linear projection  $W_s$  can be performed via a shortcut connection to match the dimensions:

$$y = F(x, \{W_i\}) + W_s x \quad (2)$$

The structure of ResNet mainly consists of the following parts: the first construction layer, which is constructed of a common convolution layer and a maximum pooling layer. The second construction layer consists of 3 residual modules. The third, fourth, and fifth construction layers all start with the downsampling residual module, followed by 3, 5, and 2 residual modules. There are two different residual structures,

one is used for fewer network layers (34 layers), as shown in Figure 2(a), and the other is used for 50/101/152 layers, as shown in Figure 2(b).

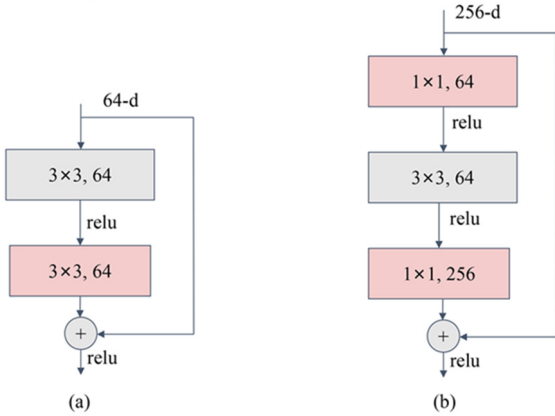


Fig 2. The structure of ResNet

The main contribution of ResNet is not to improve the expressive ability of the model but to optimize the training process of the model. Although ResNet has the same solution space as the forward network of the same depth, ResNet effectively improves the stability and efficiency of network training by introducing residual connections. Therefore, the core advantage of ResNet lies in its optimization of the deep network training process. ResNet can train very deep neural networks, avoid the vanishing gradient problem, and improve the expression ability and performance of the model. Using residual connections can preserve the original features, making the network learning smoother and more stable, and further improving the accuracy and generalization ability of the model. In addition, using ResNet can avoid gradient disappearance and gradient explosion problems during training and accelerate network convergence.

In recent years, deep learning technology has shown excellent performance in different fields of medical image analysis. Several deep learning architectures have been proposed and used for computational pathology classification, segmentation, and detection tasks. Due to its simple modular structure, most downstream applications still use ResNet and its variants.

### 3.2. VGG Net

The VGG network, proposed by the Visual Geometry Group (VGG) of Oxford University [17], is a deep convolutional neural network architecture. In the 2014 ImageNet image classification competition, the VGG network performed well with its balance of depth and performance, winning the runner-up in the classification task and the championship in the positioning task, and has been widely used in many computer vision tasks.

The main contribution of the VGG network is to prove that increasing the depth of the network can affect the final performance of the network to a certain extent. To achieve this goal, the VGG network uses three 3×3 convolution kernels to replace the 7×7 convolution kernel, and two 3×3 convolution kernels to replace the 5×5 convolution kernel. The main purpose of this is to improve the depth of the network while ensuring the same receptive field, thereby improving the effect of the neural network to a certain extent.

The entire VGG network uses the same convolution kernel size (3×3) and maximum pooling size (2×2). In addition, the VGG network can improve performance by continuously

deepening the network structure. As shown in Figure 3, the VGG network has two main structures, namely VGG16 and VGG19. VGG16 contains 16 hidden layers (13 convolutional layers and 3 fully connected layers), while VGG19 contains 19 hidden layers (16 convolutional layers and 3 fully connected layers). There is no essential difference between the two structures, but the network depth is different.

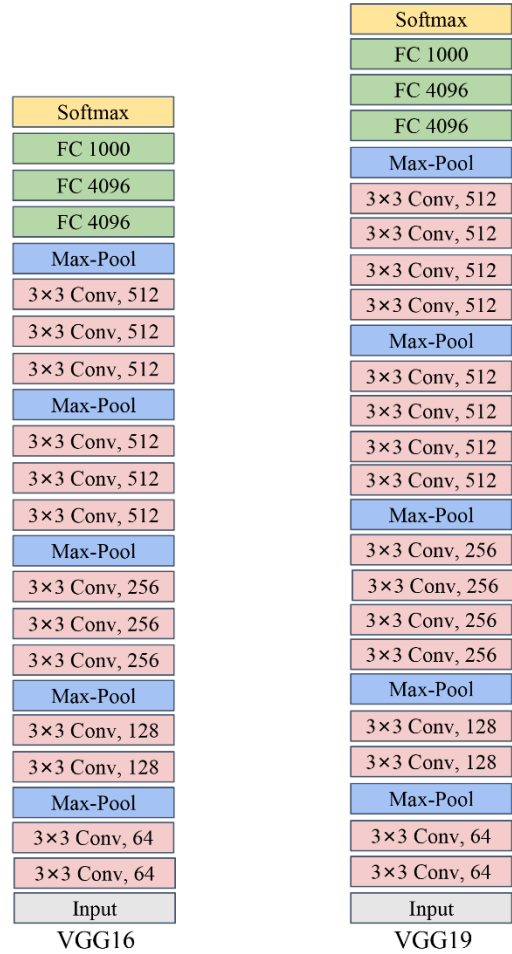


Fig 3. The structure of the VGG network

The VGG network also solves some optimization problems, such as lowering the learning rate, adjusting the initialization method of parameters, adjusting the standardization method of input data, modifying the Loss function, adding regularization, standardizing middle layer data, and using dropout. These optimization strategies further improve the performance of the VGG network.

In medical image classification, the VGG network is also widely used. For example, pre-trained VGG models can be used to identify and classify medical images through transfer learning. This method takes advantage of the rich feature representation obtained by the VGG network trained on large-scale image data sets, and can effectively improve the performance of medical image classification.

### 3.3. Loss Functions

Cross-entropy loss function, derived from the concept of cross-entropy in information theory, is a commonly used loss function and is mainly used in binary classification and multi-classification problems. Its core idea is to measure the difference between two probability distributions, which are usually the distribution of true labels and the distribution of model prediction results. In machine learning, especially deep learning, the cross-entropy loss function is widely used as an

optimization target and used to train various classification models, such as logistic regression, deep neural networks, etc. By minimizing the cross-entropy loss function, the prediction results of the model can be made as close as possible to the real labels, thereby improving the performance of the model.

For a binary classification problem, the expression of the cross-entropy loss function is:

$$L = -\frac{1}{N} \sum_i [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (3)$$

Among them,  $N$  is the number of samples,  $y_i$  represents the true label (0 or 1) of sample  $i$ , and  $p_i$  represents the probability that sample  $i$  is predicted to be a positive class.

For multi-classification problems, the expression of the cross-entropy loss function is:

$$L = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (4)$$

Among them,  $N$  is the number of samples,  $M$  is the number of categories,  $y_{ic}$  is the sign function (0 or 1), taking 1 if the true category of sample  $i$  is equal to  $c$ , otherwise taking 0,  $p_{ic}$  is the predicted probability that the observed sample  $i$  belongs to category  $c$ .

The cross-entropy loss function has the following characteristics:

When the prediction result is correct, the greater the probability of the correct category (the closer  $p_{ic}$  is to 1), the smaller the cross-entropy loss function. When the prediction result is correct, but the probability of the correct category is not large enough ( $p_{ic}$  is smaller), the cross-entropy loss function is large. When the prediction result is wrong, the cross-entropy loss function is also large.

In a classification problem, the goal is to make the model's predictions as close as possible to the true labels. The cross-entropy loss function exactly meets this requirement. It can effectively measure the difference between the model's prediction results and the real labels, and guide the optimization of the model. In addition, the cross-entropy loss function has good mathematical properties, making the optimization process more stable and efficient. Therefore, the cross-entropy loss function is usually used as the optimization objective in classification networks.

The classification of benign and malignant pulmonary nodules is a binary classification problem. This section will use the binary cross-entropy function as the loss function. In this problem, one probability distribution represents the true label (i.e., whether the pulmonary nodule is benign or malignant), and the other probability distribution represents the model's predictions.

### 3.4. Data Acquisition

The experimental training and testing data sets in this article come from LUNA16. LUNA16 is refined from the publicly available Lung Image Database Consortium image collection (LIDC-IDRI) [18]. The American College of Radiology recommends the use of thin-section CT scans for nodule detection and classification. Therefore, scans with slice thickness greater than 3 mm, inconsistent slice spacing, or missing slices were discarded. In addition, nodules smaller

than 3 mm were removed. After screening, 888 CT scans and 1004 nodules were selected for training and testing (450 malignant nodules and 554 benign nodules).

In LIDC-IDRI, each CT scan is accompanied by an annotated XML file provided by four experienced radiologists [18]. Annotations in the dataset were collected during a two-stage image annotation process by at least 3 radiologists. It includes almost all defining characteristics, such as the location, size, and malignancy of the nodule (from 1 to 5). Level 1 indicates a very small likelihood of cancer, Level 2 indicates a moderate likelihood of cancer, Level 3 indicates a moderate likelihood of cancer, Level 4 indicates a moderate suspicion of cancer, and Level 5 indicates a very strong suspicion of cancer. The first two grades are considered benign and the last two are malignant.

For benign and malignant classification of pulmonary nodules, binary classification was performed by calculating the average grade of each selected nodule. If the final average grade equals 3, then the judgment of benign or malignant nodules is considered uncertain and is therefore eliminated. For nodules with a final mean grade greater than 3, they were labeled as malignant nodules [19]. Otherwise, they are considered benign. All annotations were obtained from the LIDC-IDRI database.

## 4. Experiments and Results

In this section, the experimental environment and algorithm evaluation indicators are first introduced. Then the model was trained on the LUNA16 data set, and the proposed benign and malignant pulmonary nodule classification method was compared with the current state-of-the-art methods to verify the effectiveness of the proposed method.

### 4.1. Experiment Setup

This experiment configured Python for training and testing based on the Anaconda environment on the Windows 10 system, and used the PyTorch deep learning framework and CUDA for GPU accelerated training and verification of model feasibility. PyTorch provides many extension libraries and interfaces for easy modular design, network construction, and customization. In addition, PyTorch supports efficient parallel computing on CPU and GPU, and also supports multi-machine distributed computing, which can greatly improve Training speed and model performance.

### 4.2. Algorithm Evaluation Metrics

In order to evaluate the performance of the algorithm, Accuracy, Sensitivity, Specificity, and AUC indicators are used in the experiments. Accuracy represents the proportion of correctly predicted results in the total sample. Sensitivity and specificity measure the proportion of correctly identified malignant nodules and benign nodules respectively. AUC represents the area under the ROC curve. The larger the AUC value, the better the classifier effect. ROC is a curve obtained by using FPR (False Positive Rate) as the abscissa and TPR (True Positive Rate) as the ordinate. TPR is Sensitivity. The calculation formulas for Accuracy, Sensitivity, and Specificity are as follows:

$$Accuracy = \frac{(TP + TN)}{TP + FN + FP + FN} \quad (5)$$

$$Sensitivity = TPR = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity} = \text{TNR} = \frac{TN}{(TN + FP)} \quad (7)$$

Among them,  $TP$  represents a positive sample predicted as positive by the model, the prediction is 1 and the prediction is correct, that is, the actual value is 1.  $FP$  represents a negative sample predicted as positive by the model, the prediction is 1 and the prediction is wrong, that is the actual value is 0.  $FN$  represents the positive sample predicted as negative by the model, the prediction is 0 and the prediction is wrong, that is, the actual value is 1.  $TN$  represents the negative sample predicted as negative by the model, the prediction is 0, and the prediction Correct, that is, it is 0.

There are two main methods for calculating AUC. One is to directly calculate the area under the ROC curve, which is obtained by approximately calculating the area of each small trapezoid under the ROC curve; the other is to count the proportion of positive and negative sample pairs. The idea of this method is to randomly select one sample from each of the positive and negative samples, and calculate the probability that the predicted value of the positive sample is greater than the negative sample. The computational complexity of these two methods differs. The time complexity of directly calculating the area under the ROC curve is  $O(P * N)$ , while the time complexity of counting the proportion of positive and negative sample pairs is  $o((P + N) * \log(P + N))$ , where  $P$  and  $N$  are the numbers of positive and negative samples respectively.

This section mainly uses the second method to calculate AUC: first, sort all samples from low to high according to the predicted value. For each positive sample, calculate its position in the sorting, recorded as  $r_i$ . Then calculate the position sum of all positive samples, denoted as  $\sum r_i$ . Calculate the number of all positive samples, denoted as  $P$ , and the number of negative samples, denoted as  $N$ . The final AUC calculation formula is:

$$\text{AUC} = \frac{\sum r_i - P * (P + 1) / 2}{P * N} \quad (8)$$

Among them,  $\sum r_i - P * (P + 1) / 2$  represents the number of positive and negative sample pairs whose predicted values of all positive samples are greater than the predicted values of negative samples.

### 4.3. Proposed Architecture

The model Res-VGG proposed in this chapter combines two different CNN models, namely VGG16 and ResNet. The model uses 10 convolutional layers, 2 fully connected layers, and 1 softmax layer, for a total of 13 layers. Additionally, max pooling and average pooling functions are used, as well as different convolution kernel sizes, and pooling layers are implemented. Figure 4 is the structure diagram of Res-VGG.

This study is improved based on VGG16, which contains 13 convolutional layers and 3 fully connected layers. First, the convolutional layers of VGG16 are reduced to 10, which is intended to reduce the complexity of the model and improve computational efficiency. Every two convolutional layers form a pair, and except for the second pooling layer, all other layers use average pooling. This design helps to better retain feature information while reducing the number of parameters of the model. In addition, by reducing one pooling layer and one fully connected layer, the complexity of the model is further reduced and helps reduce the possibility of overfitting. Next, the kernel size of the second convolutional layer was modified to  $5 \times 5$  to better capture local features, resulting in an improved 13-layer VGG16 network.

On this basis, in order to solve the vanishing gradient problem in deep networks and improve the training efficiency of the model, a residual connection is added between the input and output of each pair of convolutional layers. When the input and output dimensions are the same, the connection method of equation (1) is used. When the dimensions increase, the connection method of equation (2) is used to match the size. Fully connected layers are used to extract information, while the last layer of the network is a SoftMax layer for predicting two categories of lung nodules (benign and malignant).

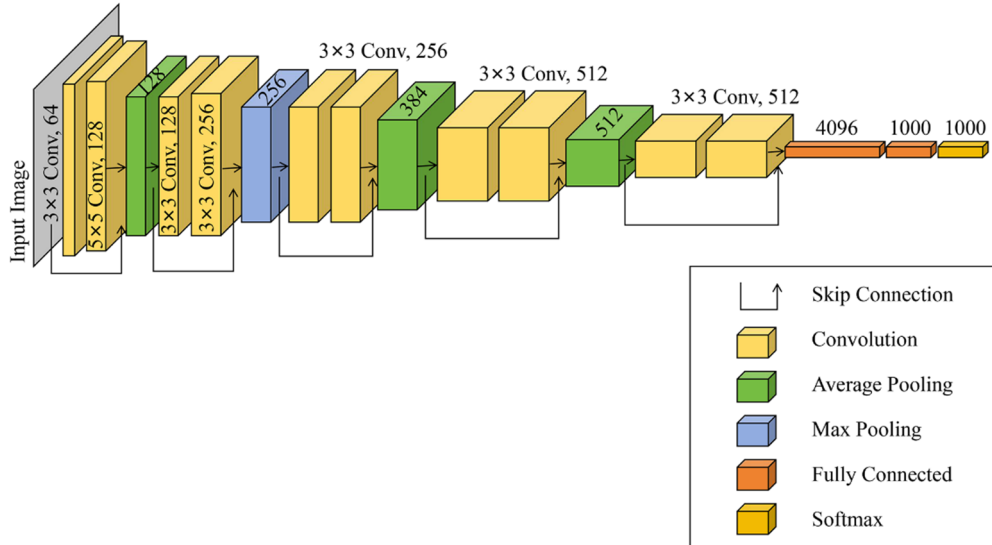


Fig 4. The structure of the Res-VGG

#### 4.4. Results and Analysis

During the training process, data sets with a benign and malignant degree of 3 (indicating uncertainty) were discarded. It is only necessary to judge whether benign or malignant, so the benign and malignant degrees of 1 and 2 are classified as benign (Malignant), and the benign and malignant degrees of 4 and 5 are classified as malignant (Benign). Through the data preprocessing in the previous section, 450 malignant nodules and 554 benign nodules were finally obtained.

The Res-VGG proposed in this chapter was evaluated on the LUNA16 database containing 1004 nodules (554 benign and 450 malignant). All image data were randomly divided into 10 subsets for ten-fold cross validation. For each fold, nine subsets were selected for training and one subset for testing.

This experiment was conducted under the Windows 10 operating system with NVIDIA GeForce RTX 4070 12GB GPU. The learning process is optimized using the stochastic gradient descent (SGD) algorithm with a momentum of 0.9. In order to reduce overfitting, the network uses regularization technology and dropout. Additionally, due to the 12GB GPU memory limit, the batch size is limited to 8. The learning model was trained for 200 epochs. The initial learning rate is 0.01, which decreases to 0.001 after 60 epochs and to 0.0001 after 120 epochs. The weight decay is set to  $1 \times 10^{-4}$ . During the optimization process, it needs to use a higher learning rate to move quickly, but when it reaches a point close to the relative optimal point, the learning rate must be reduced so as not to miss the optimal point. Therefore, the learning rate is reduced by a decay factor in each epoch. The parameter settings of this experiment are shown in Table 1.

**Table 1.** Experimental parameter configuration

Parameters	Batch size	Epoch	Learning rate	Weight decay
Value	8	200	0.01	0.0001

The final performance of the method is the average performance of ten test folds, evaluated by measuring Accuracy, Sensitivity, Specificity, and AUC. The proposed benign and malignant pulmonary nodule classification

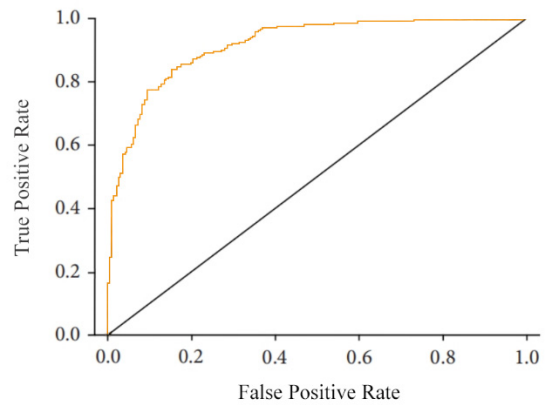
network Res-VGG was compared with other networks VGG16, ResNet, and DenseNet used for pulmonary nodule classification. The experimental results are shown in Table 2.

**Table 2.** Comparison experiment

Models	Accuracy/%	Sensitivity/%	Specificity/%	AUC
Res-VGG	84.35	83.86	84.52	0.920
VGG16	83.36	82.53	82.88	0.895
ResNet	82.35	80.67	83.80	0.876
DenseNet	82.58	83.96	81.45	0.882

As shown in Table 2, compared with DenseNet, ResNet is slightly lower in both accuracy and sensitivity, but has a better impact on specificity, which is 83.80%. Good specificity means that more malignant pulmonary nodules can be detected in the same data set, which may be more helpful for the early diagnosis of pulmonary nodules. Res-VGG has the best accuracy, with an accuracy rate of 84.35%, a sensitivity of 83.86%, and a specificity of 84.52%. The accuracy of VGG16 was 83.36%, the sensitivity was 82.53%, and the specificity was 82.88%. The reason why the Res-VGG model can achieve the best results is mainly because it combines the advantages of VGG and ResNet. It not only retains the characteristics of the VGG deep convolution network, but also introduces the residual connection of ResNet, effectively solving the depth problem. Vanishing gradient and representation bottleneck problems in networks. In addition, Res-VGG improves the expression ability of the model by increasing the depth and width of the network, making it perform well in the task of classifying benign and malignant pulmonary nodules.

The network performance of the proposed method is measured using the ROC curve, and the ROC curve of Res-VGG is shown in Figure 5.



**Fig 5.** The structure of the Res-VGG

Figure 6 provides an example of benign and malignant nodule classification results. Malignant and benign nodules correctly classified by the proposed method are shown in the first and second rows, respectively, and the numbers below the nodules are the predicted malignant probabilities. The average degree of malignancy among four expert radiologists was compared with the predicted probability of malignancy for each nodule. It was observed that the predicted malignancy probability of each nodule was consistent with the average diagnostic level.

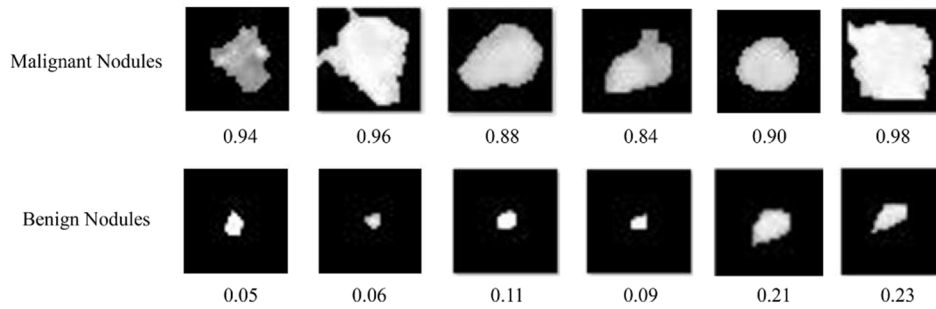


Fig 6. Example of classification results

## 5. Conclusion

This paper proposes a new classification model, Res-VGG, which combines two different CNN models to classify benign and malignant pulmonary nodules. The experimental content is introduced in detail, including steps such as data preprocessing, network training, and performance evaluation. All experiments were conducted on the LUNA16 database, and a ten-fold cross-validation method was used to evaluate the performance of the model. Experimental results show that the Res-VGG model achieves good performance on the LUNA16 database, proving its effectiveness in the classification task of benign and malignant pulmonary nodules. Compared with current methods based on traditional features and deep features, the accuracy of the Res-VGG model has been significantly improved. In addition, the proposed model is compared with three other common classification networks, and the results show that the Res-VGG model outperforms other models due to the use of different sizes of convolution kernels and maximum pooling and average pooling.

## References

- [1] M. Marianna Sockrider, What is a Lung Nodule?, *American Journal of Respiratory and Critical Care Medicine*, 193 (2016) 1.
- [2] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, *IEEE transactions on medical imaging*, 38 (2018) 991-1004.
- [3] S. Liu, Y. Xie, A. Jirapatnakul, A.P. Reeves, Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks, *Journal of Medical Imaging*, 4 (2017) 041308-041308.
- [4] K.-L. Hua, C.-H. Hsu, S.C. Hidayati, W.-H. Cheng, Y.-J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, *OncoTargets and therapy*, (2015) 2015-2022.
- [5] W. Li, P. Cao, D. Zhao, J. Wang, Pulmonary nodule classification with deep convolutional neural networks on computed tomography images, *Computational and mathematical methods in medicine*, 2016 (2016).
- [6] P. Sahu, D. Yu, M. Dasari, F. Hou, H. Qin, A lightweight multi-section CNN for lung nodule classification and malignancy estimation, *IEEE journal of biomedical and health informatics*, 23 (2018) 960-968.
- [7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [8] J. Gong, J. Liu, W. Hao, S. Nie, B. Zheng, S. Wang, W. Peng, A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images, *European Radiology*, 30 (2020) 1847-1855.
- [9] X. Xia, J. Gong, W. Hao, T. Yang, Y. Lin, S. Wang, W. Peng, Comparison and fusion of deep learning and radiomics features of ground-glass nodules to predict the invasiveness risk of stage-I lung adenocarcinomas in CT scan, *Frontiers in oncology*, 10 (2020) 418.
- [10] P. Wu, X. Sun, Z. Zhao, H. Wang, S. Pan, B. Schuller, Classification of lung nodules based on deep residual networks and migration learning, *Computational intelligence and neuroscience*, 2020 (2020).
- [11] X. Li, B. Hu, H. Li, B. You, Application of artificial intelligence in the diagnosis of multiple primary lung cancer, *Thoracic cancer*, 10 (2019) 2168-2174.
- [12] D.R. Baldwin, J. Gustafson, L. Pickup, C. Arteta, P. Novotny, J. Declerck, T. Kadir, C. Figueiras, A. Sterba, A. Exell, External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules, *Thorax*, 75 (2020) 306-312.
- [13] Y. Ren, M.-Y. Tsai, L. Chen, J. Wang, S. Li, Y. Liu, X. Jia, C. Shen, A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification, *International journal of computer assisted radiology and surgery*, 15 (2020) 287-295.
- [14] S. Lakshmanaprabu, S.N. Mohanty, K. Shankar, N. Arunkumar, G. Ramirez, Optimal deep learning model for classification of lung cancer on CT images, *Future Generation Computer Systems*, 92 (2019) 374-382.
- [15] Q. Song, L. Zhao, X. Luo, X. Dou, Using deep learning for classification of lung nodules on computed tomography images, *Journal of healthcare engineering*, 2017 (2017).
- [16] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.
- [17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, (2014).
- [18] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans, *Medical physics*, 38 (2011) 915-931.
- [19] G. Zhang, Z. Yang, L. Gong, S. Jiang, L. Wang, H. Zhang, Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations, *La radiologia medica*, 125 (2020) 374-383.