Mathematical Biosciences and Engineering

MBE, 19(1): 331–350.
DOI: 10.3934/mbe.2022017
Received: 21 August 2021
Accepted: 25 October 2021
Published: 15 November 2021

http://www.aimspress.com/journal/MBE

Research article

MARnet: multi-scale adaptive residual neural network for chest X-ray images recognition of lung diseases

Boyang Wang and Wenyu Zhang*

School of Computer Science and Software Engineering, University of Science and Technology, Liaoning 114044, China

* **Correspondence:** Email: zhangwenyu8518@126.com.

Abstract: Chest X-ray image is an important clinical diagnostic reference to lung diseases that is a serious threat to human health. At present, with the rapid development of computer vision and deep learning technology, many scholars have carried out the fruitful research on how to build a valid model for chest X-ray images recognition of lung diseases. While some efforts are still expected to improve the performance of the recognition model and enhance the interpretability of the recognition results. In this paper, we construct a multi-scale adaptive residual neural network (MARnet) to identify chest X-ray images of lung diseases. To make the model better extract image features, we cross-transfer the information extracted by residual block and the information extracted by adaptive structure to different layer, avoiding the reduction effect of residual structure on adaptive function. We compare MARnet with some classical neural networks, and the results show that MARnet achieves accuracy (ACC) of 83.3% and the area under ROC curve (AUC) of 0.97 in the identification of 4 kinds of typical lung X-ray images including nodules, atelectasis, normal and infection, which are higher than those of other methods. Moreover, to avoid the randomness of the train-test-split method, 5-fold cross-validation method is used to verify the generalization ability of the MARnet model and the results are satisfactory. Finally, the technique called Gradient-weighted Class Activation Mapping (Grad-CAM), is adopted to display significantly the discriminative regions of the images in the form of the heat map, which provides an explainable and more direct clinical diagnostic reference to lung diseases.

Keywords: lung disease; chest X-ray image; convolutional neural network; heat map

1. Introduction

According to statistical data from the World Health Organization, lung diseases rank the third among all causes of death worldwide [1]. Lung diseases lead to the death of more than five million people worldwide each year [2]. As an important organ of the respiratory system, the lungs connect with the outside and the whole blood flows, so the lungs are vulnerable to internal and external microbial attack. Pneumonia is an inflammation of the lung parenchyma caused by pathogenic microbiological, physical and chemical factors, immune damage, allergies and drugs [3]. Bacterial pneumonia is the most common one, and its chest X-ray image shows pulmonary exudation or solid shadow. Lung infiltration is there are some abnormal cells or foreign body in the lungs. Lung infiltration is caused by bacteria, fungi and viruses, and infiltrated shadow can be seen on chest X-ray images. Exudative lesions are acute inflammation, mostly caused by bacterial, viral or fungal infections, and its chest X-ray images will appear cloudy shadows. Atelectasis is a common complication of respiratory diseases which is divided into congenital atelectasis and secondary atelectasis. The symptoms include lung compression and decrease of ventilation-flow ratio in lungs [4]. The causes of congenital atelectasis are mostly childbirth injury, and the causes of secondary atelectasis include tuberculosis and drowning. The X-ray features of atelectasis include low-texture region, cavity, high density areas and fibrosis so on. For the lung nodule, there are some opaque nodular sections with clear edge, surrounded by air-bearing lung tissue, which diameter is less than or equal to 3 cm, without symptoms of atelectasis, hilar enlargement or pleural effusion. Tumor, infectious granuloma and congenital lesions can cause pulmonary nodules. In chest X-ray images, nodules show small, localized, round, and high-density shadows. Since malignant tumors have nodular lesions, effective detection of nodules plays an important role in the diagnosis of lung cancer [5].

Chest X-ray image is the most commonly medical imaging technology in clinical diagnosis of lung diseases, which is very effective in the identification and detection of cardiothoracic, pulmonary and interstitial diseases and plays an important role in the treatment of lung diseases [6]. Accurate analysis of patient's health information is a great challenge for radiologists. So computer-aided diagnosis system is proposed to control the differences among radiologists and provide reference for clinicians. In recent years, deep learning has been widely used in medical image field. Compared with manual feature extraction, deep learning technology can automatically learn features from data by training large-scale dataset, and has made significant progress in computer vision [7–11]. Deep learning has been increasingly applied in the field of medical image, especially image classification [12–14]. Related researches mainly include the automatic classification of lung tuberculosis [15], lung nodules detection [16,17], lung cancer detection [18–20], pneumonia detection [21–23], pneumothorax detection [24] and COVID-19 prediction [25–27]. Here, we summarize the related work of deep learning technology for lung disease detection.

Recent studies showed that deep neural networks can automatically learn features from training data, that show higher efficiency and discrimination accuracy. Kim et al. [28] used a six-layer convolutional neural network (CNN) consisting of four convolutional layers and two fully connected layers to classify interstitial lung diseases. Comparing with the results of support vector machine (SVM), they found that the accuracy of CNN classifier is 6–9% higher than SVM classifier. Sivaramakrishnan et al. [29] compared the effects of chest X-ray image in crowd screening, who found that the pre-trained and customized deep models are much better than the shallow models in performance. In order to detect nodules early, Pinheiro et al. [30] trained convolutional neural networks

using swarm intelligence algorithms, and verified whether this method is more effective than traditional training algorithms such as back propagation and gradient descent. The results showed that in the lung image database consortium image collection (LIDC-IDRI), the model achieved an accuracy of 93.71%, which was superior to the back propagation model. In the application of deep neural network in lung tumor prediction, Xu et al. [31] evaluated the role of deep neural network in predicting clinical results by analyzing CT images of patients with locally advanced non-small cell lung cancer (NSCLC), and achieved good results. In order to screen lower respiratory tract diseases in children, Chen et al. [32] proposed a computer-aided chest X-ray diagnostic program for the classification of bronchiolitis/bronchitis, bronchopneumonia/interstitial pneumonia, lobar pneumonia, pneumothorax and normal children. What's more, Grad-GAM technique was used to show the results of the model. Since they used the existing ResNet34 and DenseNet169 as discriminant models, the recognition performance was expected to be improved.

Rajpurkar et al. [33] developed a model called CheXNet and tested 14 types of lung diseases. They compared the detection performance of CheXNet with 4 professional radiologists, and found that CheXNet exceeds average radiologist level on the F1-score. For Chest X-ray 14 dataset, Angelica et al. proposed GraphXNET [34] and achieved good results. Wang et al. proposed TieNet [35]. Compared with the previous classical neural networks, TieNet has significantly improved the recognition ability of chest X-ray images (AUCs increase by 6% on average). Xu et al. [36] designed CXNet-m1 and proposed a new loss function sin-loss. In addition, they optimized the convolution kernel of CXNet-m1 to achieve better classification accuracy. The experimental results showed that CXNet-m1 with sin-loss function achieves better promotion in the index of accuracy rate, recall rate, F1-score, and AUC value. Zhao et al. [37] proposed AM_DenseNet for chest X-ray image classification. This model used a dense connection network and added an attention module after each dense block to optimize the ability of the model to extract features. They used the Focal Loss function to solve the data imbalance problem. The average AUC of AM_DenseNet detection for 14 kinds of chest diseases was 0.8537. Ho et al. [38] proposed different knowledge distillation strategies for 14 diseases, and used heat map techniques to visualize the abnormal characteristics of X-ray images. As them mentioned in their paper, although the total amount of chest X-ray 14 dataset exceeds 100,000, it is extremely unbalanced among various categories and there are annotation errors. In addition, some diseases of different categories have great similarity or related characteristics, directly classifying these data will cause great interference to the model and lead to poor results.

At present, there are still several limitations in using deep learning to study chest X-ray images. First of all, most studies focus on a certain type of disease, and two-class classification lacks universality. In multi-classification problems, the similarity of pathological features of some diseases leads to poor recognition [32]. What's more, in previous researches about chest X-ray images, some models cannot well identify the morphological and scale characteristics of different lung diseases on chest X-ray images, so the recognition results are poor. In addition, although a large number of convolutional neural networks recognize chest X-ray images accurately, they do not explain how and why the algorithm gets the final results to the medical staff. In order to display the output results more intuitively, target detection technology has been widely used. This technology uses the bounding box to circle the diseased parts, so as to visualize the classification and positioning of chest X-ray images [39]. However, the use of target detection technology needs to establish a special training database which records the disease area coordinates marked by experts. This greatly limits the further development of deep learning technology in the field of X-ray image detection. Our

contributions are as follows:

1) We construct a relatively large and balanced dataset, including four types of chest X-ray images of nodules, atelectasis, infection and healthy individuals. Firstly, we isolate six types of images which suffer from only one disease from chest X-ray 14. Since the three types of images have high similarity, we classify them as a class called infection. Then we add healthy individual chest X-ray images. Finally, we constitute the dataset for test MARnet performance.
2) According to the characteristics of different lung diseases on chest X-ray images, we construct MARnet to identify chest X-ray images with different symptoms. In order to identify four types of chest X-ray images better, we alternately transmit the feature information extracted by adaptive structure and residual block to different layers. Experimental studies have shown that MARnet has improved greatly the ability to identify chest X-ray images compared with other models. Finally, we use 5-fold cross-validation to test the generalization of MARnet.
3) In order to solve the problem that the identification of neural network is difficult to be understood by human beings, we use Grad-CAM technology to visualize the disease area in the form of heat map. Heat maps make the results of MARnet more intuitive, providing a more valuable reference for clinicians to diagnose pulmonary diseases.

2. Materials and methods

2.1. Design

This section describes the overall process of MARnet for chest X-ray image recognition. As shown in Figure 1, chest X-ray images are first sent to block A. Image features are extracted from block A and the information is introduced into a series of blocks B. Then feature information by block B is introduced into the following block A and another series of blocks B. The process of image processing through 1 block A and 4 or 6 blocks B is called one round feature extraction. At the overall process, the input image is extracted by 4 rounds of feature extraction. The heat map is used to display the disease area of the image. Finally, the result is archived.

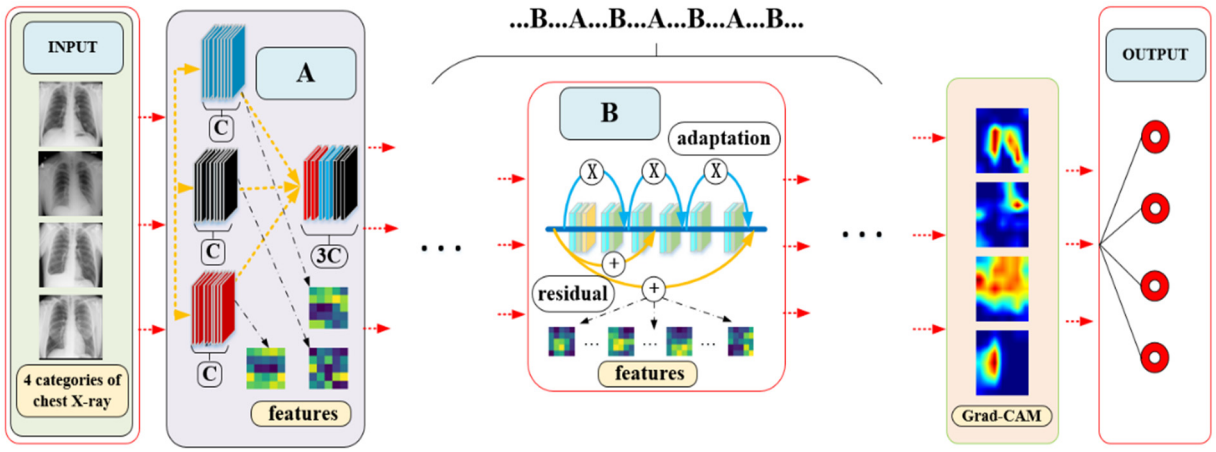

Figure 1. The flowchart of MARnet identifies chest X-ray images.

In Figure 1, the parameter C in block A represents the number of features extracted from the convolution layer. The color squares in block A and B represent the feature information extracted from the convolution kernels.

2.2. Data description

The chest X-ray 14 dataset selected in our experiment is the largest publicly available one in the world, which contains 112,120 X-ray images from 30, 805 unique patients. Each image is marked with a single or multiple pathological labels, and radiological reports shows that the accuracy rate is more than 90% [38]. Considering the large number of patients with multiple pulmonary diseases in the dataset, we select images from the dataset with only infiltration, effusion, pneumonia, nodule, and atelectasis in order to eliminate interference. There is big similarity in the chest X-ray images among infiltration, effusion and pneumonia, so medical examination is required to confirm the diagnosis. We combine these three types into the infection. Finally, we study the chest X-ray images of 13,382 patients with nodules, atelectasis, infection and normal. Figure 2 shows the randomly selected chest X-ray images of nodule, atelectasis, normal and infection.

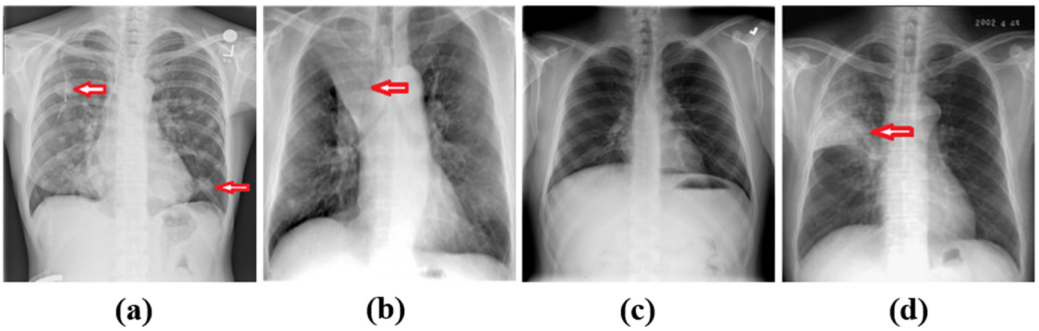

Figure 2. Randomly selected four types of chest X-ray images. (a) nodule, (b) atelectasis, (c) normal, (d) infection. Red arrow points to the disease area.

The datasets are divided into train sets, test sets and validation sets in a ratio of 7:2:1. Table 1 shows the number and category of chest X-ray images.

Table 1. Number of the chest X-ray images in training, test and validation.

Disease	Train set	Test set	Val set
Nodule	1890	550	270
Atelectasis	2940	840	420
Normal	2200	612	310
Infection	2360	670	320

To inspect the generalization of MARnet, we use 5-fold cross-validation. We process the dataset in Table 1. First, we combine all the data, and then split the data into five equal pieces. One is selected as the test set and the rest as the train set in turn. Finally, MARnet is trained and tested five times. We resize images from 1024 × 1024 pixels to 224 × 224 pixels to meet the input size requirements of the model.

2.3. Constitution of MARnet

This section introduces the structure of MARnet, which is a 137-layer network. In MARnet, the adaptive structure and residual block alternately transmit the extracted feature information to the subsequent layer with high efficiency.

2.3.1. Residual structure

The deeper the neural network is, the more semantic information can be extracted from different levels and the more abstract meaning can be expressed. However, when the level of neural network increases to a certain amount, gradient dispersion or gradient explosion will occur, so that the accuracy is no longer improved or even decreased. The residual structure can be used to train a deeper neural network. Therefore, we added residual structures to the multi-scale and adaptive modules of MARnet.

2.3.2. Adaptive structure

The feature map in convolutional neural network reflects the feature information extracted by convolutional kernel to a certain extent. Adaptive learning is to automatically obtain the importance degree of each feature. Then, according to this importance degree, enhancing the main feature and suppressing the secondary feature [40]. In the adaptive structure, the adaptive global average pooling layer (AdaptiveAvgPool) is used to compress the feature along the spatial dimension, and each two-dimensional feature map is transformed into a real number with global sensing ability, which can reflect the feature distribution. Then a 1 × 1 convolution kernel (input channel number is a, output channel number is a/4) is used to reduce dimension. ReLU activation function is used to increase nonlinearity. Another 1 × 1 convolution kernel (input channel number is a/4, output channel number is a) is used to increase dimension and reduce computational cost. In the last layer of the adaptive structure, Sigmoid function is used to generate a real number between 0 and 1 as the output result. The adaptive number is transmitted to the behind layer of the network, in fact which means a weight coefficient is multiplied to the feature matrix to enhance the main feature and suppress the secondary feature.

2.3.3. Multi-scale residual block

The chest X-ray images of nodule, atelectasis, normal and infection have different feature scales. Considering that convolution kernels of different sizes have different receptive fields in neural networks, we use 3 parallel convolution kernels of different scales to extract feature information, and fuse them together. This multi-scale structure shows good performance in our experiments. The MARnet has three-group parallel multi-scale residual blocks and each group is composed of 4 convolution layers. We exploit the residual structure to transmit the information of the 0th layer to the end of the last layer. In the multi-scale residual block, we use 3 × 3 uniform convolution kernel in the first group. In the second group, we replace the first two 3 × 3 convolution layer with 5 × 5 convolution kernel on the basis of group 1. In the third group, we replace the first three convolution layer of 3 × 3 with 7 × 7 convolution kernel.

2.3.4. Adaptive residual block

In order to better extract the image's feature information, we add the residual structure in the adaptive block to construct a deeper neural network. The adaptive structure generates a feature adaptive coefficient between 0 and 1, and the residual structure creates a matrix. If the adaptive and residual structures transmit the feature information from the same layer to the same subsequent layer, it means the feature matrix generated by the residual structure will be added to another matrix multiplied by the adaptive coefficient, that lead to the weakening of the adaptive function due to the residual structure. Based on the reason, in MARnet, the residual structure transmits the 0th layer feature to the end of the 3rd and 6th layers. Meanwhile, the adaptive structure transmits the 0th, 2nd and 4th layer feature to the end of the 2nd, 4th and last layer, respectively. This structure efficiently avoids the weakening effect of residual block on adaptive structure. Figure 3 demonstrates MARnet's adaptive residual block.

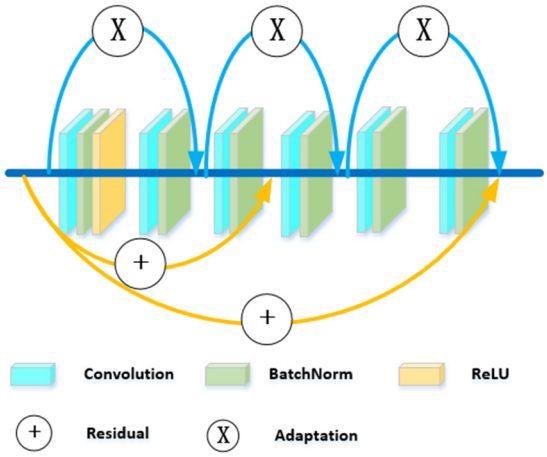

Figure 3. Adaptive residual block in MARnet.

2.4. Hyper-parameters tuning

In our exploration, we try to estimate the influence of various hyper-parameters on the performance of the neural network. The parameters include batch size, learning rate and activation functions, which have an important impact on the model performance. Five different batch size parameters, {24, 36, 32, 48, 64}, are balanced by lots of contrast experiments, and finally the model performance is best when this parameter is got as 32. The learning rate affects the convergence speed of the model, and excessive learning rate will reduce the accuracy. We attempt 8 kinds of learning rate, {1e-3, 1e-4, 1e-5, 3e-3, 3e-4, 3e-5, 3e-6, 3e-7}, to find the result is best when the learning rate is set to 1e-4 or 3e-4. In terms of learning algorithm, we find that Adam is significantly better than SGD or Adadelta. The internal parameters in the neural network include convolution kernel size, stride, kernel size of the pooling layer, channel and dropout rate. Table 2 shows the structure and parameters of MARnet.

2.5. Grad-CAM visualization

In terms of image recognition, if the local part of the image determining the decision classification is highlighted, a good visual result will be given to improve the interpretation of recognition. Gradient-weighted class activation map (Grad-CAM) uses gradient information flowing into the last convolution layer of the neural network to infer the importance of each neuron, and gives the corresponding results in the form of heat map which helps to visualize the prediction results of the network [41]. We use Grad-CAM technology to display the disease area on chest X-ray images, which provides a more intuitive reference to doctors.

Table 2. Structure and parameters in MARnet.

layer	Output size	Kernel size	channel
Input(224 × 224 × 3)			
A	112 × 112	3, 5, 7	45
AdaptiveAvgPool (2 × 2, stride = 1)			
B × 4	56 × 56	3	45
A	28 × 28	3, 5, 7	135
B × 4	28 × 28	3	135
A	14 × 14	3, 5, 7	405
B × 6	14 × 14	3	405
A	7 × 7	3, 5, 7	1215
B × 6	7 × 7	3	1215
AvgPool (kernel_size = 7, stride = 1)			
Linear (output = 4)			

3. Experiments

3.1. Performance evaluation

In order to comprehensively evaluate the performance of our model, we introduce accuracy (ACC) [42], Precision, Recall and F1-score as the evaluation indicator [43]. The formulas for these indicators are shown below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{2}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

$$F1-score = 2 \times \frac{\mathrm{Precision \times Recall}}{Precision+Recall} \tag{4}$$

Here TP, FP, TN and FN represent the number of true positive, false positive, true negative and false

negative, respectively. In addition, we draw the receiver operating characteristic (ROC) curve and calculate the area under curve (AUC) to further evaluate the performance of the model [44,45]. The higher the AUC value, the better the performance of the model.

3.2. Ablation study

We conduct ablation study on MARnet to evaluate the effectiveness of such design. In ablation study, Multi-scale Residual convolutional neural network (MRnet), adaptive residual convolutional neural network (ARnet) and fusion adaptive convolutional neural network (FAnet) are used. MRnet is a 43-layer neural network that uses three sets of residual neural networks with similar structures in parallel. ARnet is a 43-layer neural network that links 13 adaptive residual modules. In the ARnet, each adaptive residual module contains three layers of convolution, and the residual and adaptive structure pass the 0th layer information to the end of the last layer respectively. The FAnet has 75 layers and consists of four groups of similar structures, and each group uses a block A followed by 4 block B that is the same as the block in ARnet.

3.3. Results

Table 3. The precision, recall, F1-score and AUC values obtained by MARnet, MRnet, ARnet and FAnet in detecting nodule, atelectasis, normal and infection.

Model	Disease	Precision	Recall	F1-score	AUC
MARnet	Nodule	69.85 ± 0.1	58.55 ± 0.1	59.46 ± 0.1	0.90
	Atelectasis	75.88 ± 0.1	85.00 ± 0.1	76.49 ± 0.1	0.93
	Normal	98.50 ± 0.1	86.90 ± 0.9	70.50 ± 0.4	0.99
	Infection	91.03 ± 0.1	98.51 ± 0.1	75.04 ± 0.1	1.00
MRnet	Nodule	63.21 ± 0.1	24.44 ± 0.1	35.97 ± 0.1	0.85
	Atelectasis	64.69 ± 0.1	91.12 ± 0.2	75.59 ± 0.2	0.89
	Normal	98.79 ± 0.2	80.87 ± 0.1	88.85 ± 0.1	0.67
	Infection	85.91 ± 0.1	98.72 ± 0.1	91.38 ± 0.1	0.95
ARnet	Nodule	55.10 ± 1.4	55.60 ± 1.2	55.41 ± 1.4	0.86
	Atelectasis	72.20 ± 0.1	71.78 ± 0.2	72.04 ± 0.1	0.90
	Normal	96.68 ± 0.1	75.46 ± 0.1	84.83 ± 0.1	0.90
	Infection	81.57 ± 0.1	97.92 ± 0.1	89.04 ± 0.2	0.98
FAnet	Nodule	64.84 ± 0.2	55.02 ± 0.2	59.01 ± 0.1	0.88
	Atelectasis	73.13 ± 0.1	80.12 ± 0.1	76.46 ± 0.1	0.90
	Normal	98.30 ± 0.1	86.85 ± 0.2	92.12 ± 0.2	0.94
	Infection	89.75 ± 0.1	98.41 ± 1.0	92.82 ± 0.3	0.99

Table 3 shows the test results of MARnet compared with MRnet, ARnet and FAnet for nodule, atelectasis, normal and infection. The results of ablative study show that the recognition performance of MARnet can not be achieved only by using multi-scale residual structure, only using adaptive residual structure or using adaptive residual structure (as ARnet). In general, the design of MARnet is effectiveness.

Table 4 shows the test results of MARnet compared with AlexNet [7], VGG16 [46], InceptionV2 [47], ResNet101 [12] and CliqueNet [48]. All experimental results are obtained by multiple experiments on the test dataset.

Table 4. The precision, recall, F1-score and AUC values obtained by AlexNet, VGG16, InceptionV2, ResNet101, CliqueNet and MARnet in detecting nodule, atelectasis, normal and infection.

Model	Disease	Precision	Recall	F1-score	AUC
MARnet	Nodule	69.85 ± 0.1	58.55 ± 0.1	59.46 ± 0.1	0.90
	Atelectasis	75.88 ± 0.1	85.00 ± 0.1	76.49 ± 0.1	0.93
	Normal	98.50 ± 0.1	86.90 ± 0.9	70.50 ± 0.4	0.99
	Infection	91.03 ± 0.1	98.51 ± 0.1	75.04 ± 0.1	1.00
AlexNet	Nodule	46.31 ± 0.1	12.54 ± 0.1	14.79 ± 0.1	0.69
	Atelectasis	61.49 ± 0.1	89.52 ± 0.1	65.39 ± 0.1	0.87 ± 0.01
	Normal	93.79 ± 0.1	66.67 ± 0.1	50.65 ± 0.1	0.89
	Infection	74.91 ± 0.1	96.72 ± 0.1	61.98 ± 0.1	0.97
VGG16	Nodule	59.50 ± 2.6	34.00 ± 2.2	37.20 ± 1.4	0.87
	Atelectasis	67.20 ± 0.1	90.00 ± 0.1	70.49 ± 0.1	0.91 ± 0.01
	Normal	97.78 ± 0.1	72.06 ± 0.1	58.23 ± 0.1	0.98
	Infection	83.51 ± 0.1	95.97 ± 0.1	70.14 ± 0.1	0.98
InceptionV2	Nodule	58.94 ± 0.1	42.55 ± 0.1	42.01 ± 0.1	0.88 ± 0.01
	Atelectasis	69.03 ± 0.1	81.19 ± 0.1	67.86 ± 0.1	0.91 ± 0.01
	Normal	95.50 ± 1.0	73.00 ± 1.2	57.80 ± 0.7	0.97
	Infection	80.25 ± 0.2	97.00 ± 1.0	66.80 ± 0.4	0.98
ResNet101	Nodule	55.00 ± 1.2	12.00 ± 0.7	16.20 ± 0.6	0.87 ± 0.01
	Atelectasis	63.06 ± 0.1	95.71 ± 0.1	69.91 ± 0.1	0.92
	Normal	97.20 ± 0.1	73.50 ± 0.1	56.50 ± 0.3	0.97
	Infection	80.91 ± 0.1	98.06 ± 0.1	65.50 ± 0.1	0.99
CliqueNet	Nodule	57.20 ± 1.5	31.50 ± 1.6	33.00 ± 1.5	0.85
	Atelectasis	66.67 ± 0.1	86.19 ± 0.1	67.14 ± 0.1	0.91 ± 0.01
	Normal	94.79 ± 0.1	68.30 ± 0.1	54.46 ± 0.1	0.84
	Infection	77.73 ± 0.1	96.87 ± 0.1	65.00 ± 0.1	0.96

From Table 4, we can see that the shallow and simple AlexNet model cannot well identify four types of chest X-ray images. For example, when AlexNet identifies nodule, the recall, F1-score and AUC are 12.5, 14.8% and 0.69, respectively, indicating that the identification effect is not good. Compared with AlexNet, the overall recognition performance of the deeper network VGG16 for four types of images is improved, but the recall index is 34.0 ± 2.2% and the F1-score is 37.2 ± 1.4% when recognizing nodule, which still cannot meet the needs of practical application. ResNet101 with deeper structure and residual block, and CliqueNet with complex feature transfer mechanism improve furtherly the recognition effect of atelectasis, normal and infection. ResNet101 achieves the highest recall of 95.7% when identifying atelectasis, but these two models still fail to solve the problem of low recognition rate of nodule. InceptionV2 has the characteristics of multi-scale structure and improves the recognition ability of nodule, the precision, recall and F1-score reach 58.9, 42.5 and 42.0%,

respectively. However, due to the shallow network layer, the recognition efficiency of InceptionV2 for the other three types of images cannot be improved. MARnet is a deep convolutional neural network with multi-scale characteristics. When MARnet recognizes nodule, the precision, recall and F1-score reach 69.8, 58.5 and 59.4%, respectively. The recognition ability of MARnet to atelectasis, normal and infection is also improved furtherly. Except that the recall index of atelectasis recognition is lower than ResNet101, the other indicators reach the highest level. Overall, MARnet performs best in all models. Figure 4 shows the ROC curves obtained by MARnet and five classical neural methods in identifying nodule, atelectasis, normal and infection on the test set. ROC curve shows that the recognition performance of MARnet on nodule is significantly better than that of other networks. In the identification of atelectasis, although the identification capabilities of these six models are relatively close, MARnet is still at a high level. When identifying normal, different types of neural networks show significant differences in recognition performance, and MARnet shows better recognition ability than other models. In the identification of infection, the traditional five neural networks have shown good recognition results, while the recognition performance of MARnet is significantly better than that of the others.

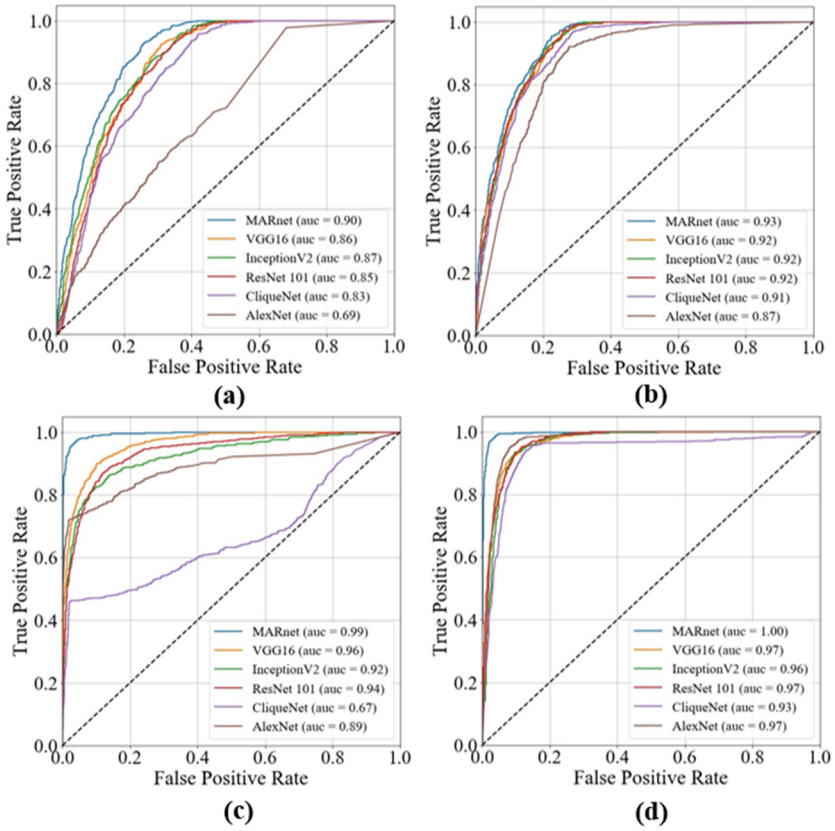

Figure 4. ROC curves obtained by MARnet and other five classical neural networks in identifying chest X-ray images: (a) nodule, (b) atelectasis, (c) normal and (d) infection.

Figure 5 reveals the ACC and average AUC obtained by MARnet and other neural networks on the test set. From this figure, we can find that ACC obtained by MARnet reaches 83.3%, which is 7.9% higher than that of the second (VGG16), and the average AUC index reaches 0.970, which is 3.0% higher than that of the second (VGG16, InceptionV2 and ResNet101). It is clear that MARnet is

significantly better than the classical five neural networks in ACC and average AUC.

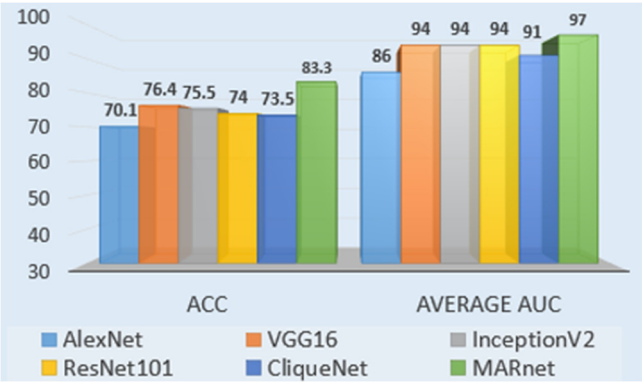

Figure 5. ACC and average AUC obtained by MARnet and other five networks on the test set.

Table 5 shows the results of comparing MARnet with 4 state-of-the-art convolutional neural networks: CheXNet [33], GraphXNet [34], TieNet [35] and AM_DenseNet [37]. It shows that MARnet works best in identifying four types of chest X-ray images. Since other datasets do not merge infiltration, effusion, pneumonia into infection as we do, the infection here takes the average value of infiltration, effusion and pneumonia.

Table 5. Comparison of the AUC between MARnet with 4 state-of-the-art convolutional neural network.

Disease	CheXNet	GraphXNet	TieNet	AM_DenseNet	MARnet
Nodule	0.78	0.71	0.69	0.81	0.90
Atelectasis	0.81	0.72	0.73	0.83	0.93
Normal	-	-	0.70	-	0.99
Infection	0.79	0.76	0.71	0.80	1.00

To evaluate comprehensively the performance of MARnet we use 5-fold cross-validation to test the recognition effect of the model on chest X-ray images. Figure 6 shows the precision, recall and F1-score obtained by MARnet. The identification results of MARnet for normal and infection are relatively stable. The results of MARnet in identifying nodule and atelectasis fluctuate greatly, which may be due to the large interference of the data itself.

Figure 7 shows the ROC curve of MARnet at 5-fold cross-validation. In the identification of nodule and atelectasis, the results obtained by MARnet under different data partitions are slightly different. When it comes to identifying normal and infection, MARnet gets roughly the same results on different data partitions. Above all, how to part the data set has little effect on the classification performance of MARnet that show the model has good generalization ability. The ROC curve obtained by MARnet at 5-fold cross-validation shows strongly correlation with the precision, recall and F1-score. If accuracy, recall and F1 score fluctuate, the ROC curve is slightly different. While, when accuracy, recall and F1 score are relatively stable, ROC curve is relatively stable. Overall, ROC curves show that MARnet has relatively stable recognition performance.

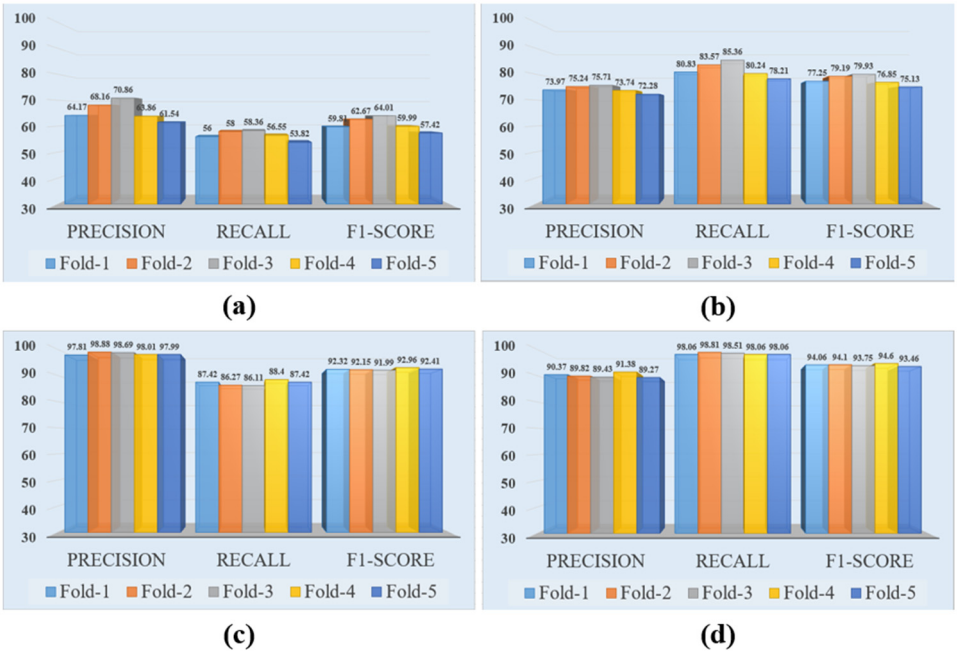

Figure 6. The precision, recall and F1-score obtained by MARnet in 5-fold cross-validation: (a) nodule, (b) atelectasis, (c) normal, and (d) infection.

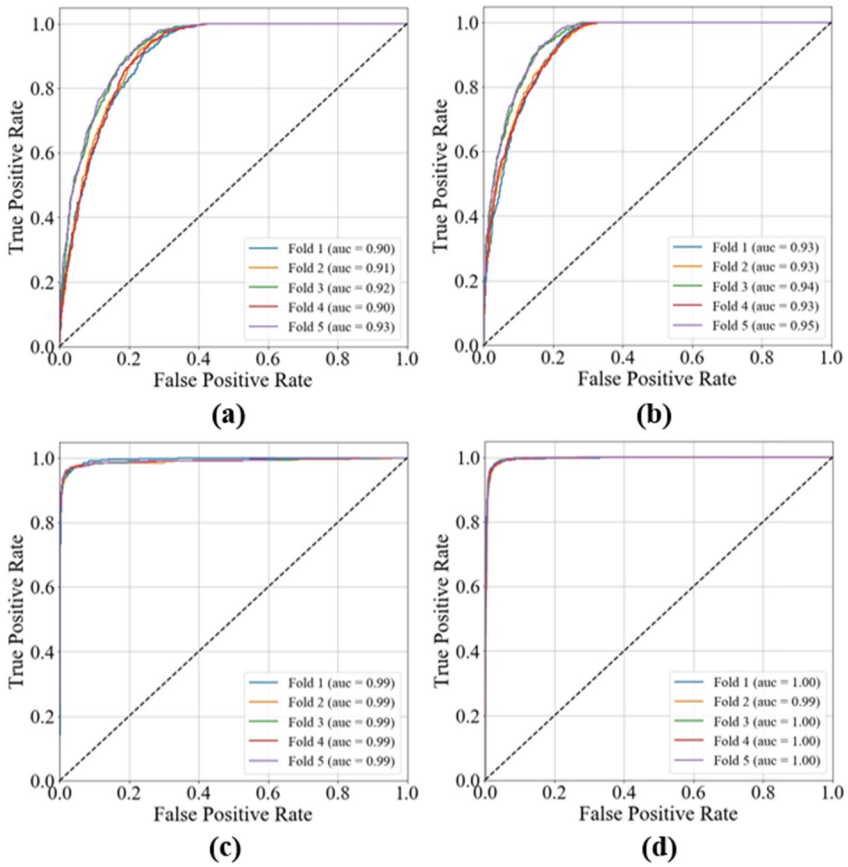

Figure 7. ROC curve for 5-fold cross-validation: (a) nodule, (b) atelectasis, (c) normal, and (d) infection.

Figure 8 shows the confusion matrix [49] obtained by using MARnet in 5-fold cross-validation. MARnet achieves the highest ACC value of 83.50%, the lowest ACC value of 80.99%, and the average ACC value of 82.46%. In general, MARnet has relatively stable recognition of chest X-ray images on all dataset.

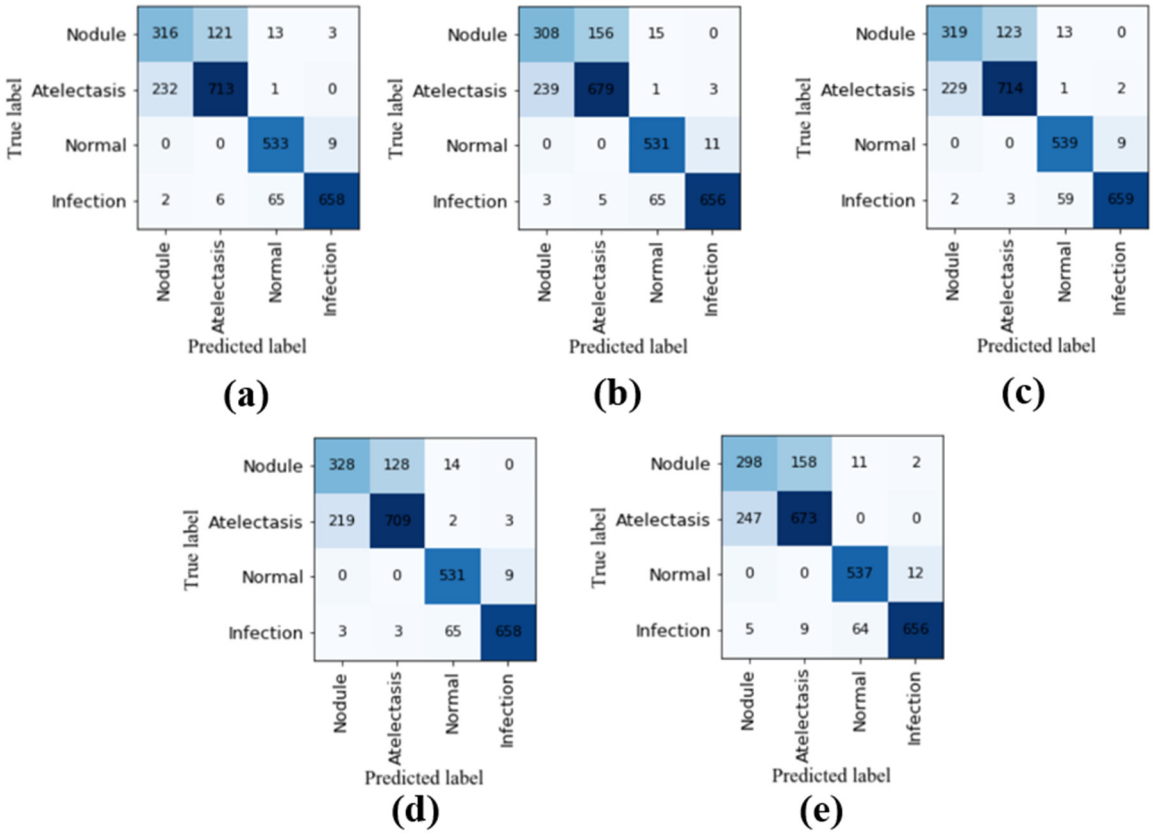

Figure 8. Confusion matrix for 5-fold cross-validation: (a) Fold 1, (b) Fold 2, (c) Fold 3, (d) Fold 4, (e) Fold 5.

In order to explain MARnet discriminant basis for chest X-ray images, we use Grad-CAM technology to display the discriminant results of MARnet in the form of heat map. Figure 9 shows the heat maps of MARnet for distinguishing four kinds of chest X-ray images: nodule, atelectasis, normal and infection. When identifying chest X-ray images with nodule, the disease area identified by the model contains a round-like high-density shadow at the bottom of the left lungs, which is the symptom of nodule. In the recognition of atelectasis, the model identifies the upper left lung area as location of atelectasis symptom. And when recognizing normal, MARnet does not find abnormalities in the bilateral lungs. In identifying infection, our model identifies the cloudy shadows on both sides of the lung, which is the lung infection caused by infiltration or effusion. Heat map makes the discriminant results of MARnet more directly and convenient to understand.

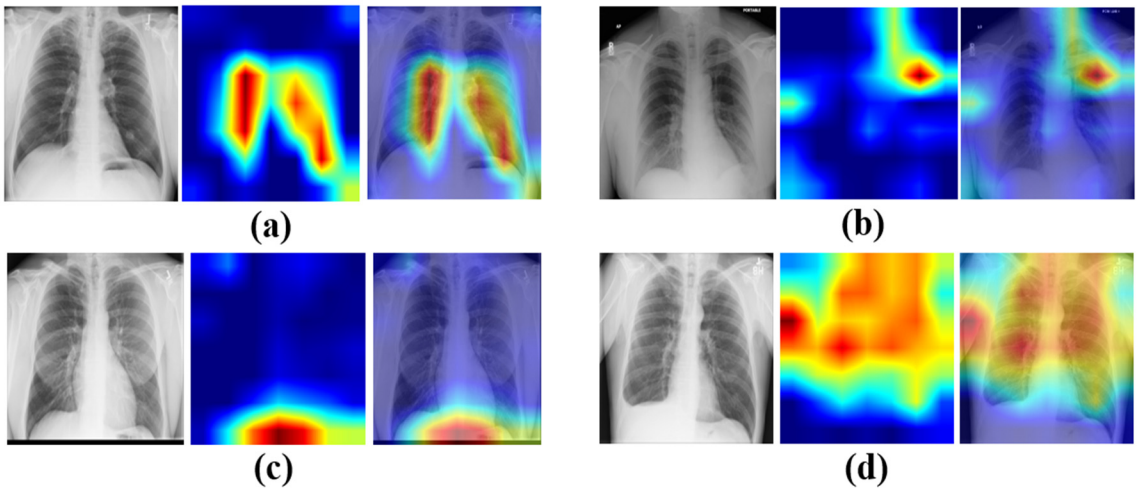

Figure 9. The heat maps generated by MARnet in identifying chest X-ray images: (a) nodule, (b) atelectasis, (c) normal, and (d) infection. In the three images of each group, the order from left to right is the chest X-ray image, the heat map obtained from the chest X-ray image, and the result of putting the heat map into the corresponding position of the chest X-ray image.

4. Discussion and conclusions

In recent years, convolutional neural networks have been widely used in supporting medical staff to diagnose lung diseases. According to the characteristics of chest X-ray images, we construct MARnet to recognize 4 types of chest X-ray images including nodule, atelectasis, normal and infection, and achieve better recognition results than other models. In addition, we use the Grad-CAM technology to reveal the discriminant results of the MARnet, which provides a more intuitive reference to doctor.

However, MARnet is not effective in recognizing nodule. The confusion matrix in Figure 8 shows that MARnet misjudges over 40% of nodules as atelectasis, and misjudges approximately 15% of atelectasis as nodules. In practice, many patients diagnosed with nodule are accompanied by varying degrees of atelectasis. The factor makes nodule and atelectasis have similar characteristics on chest X-ray images and result in a large number of misjudgments.

In order to further study the reasons for MARnet's poor performance in identify nodule and atelectasis, we identify nodule, atelectasis, normal with ResNet-152. Figure 10 shows the confusion matrix obtained by using ResNet-152 to classify nodule and atelectasis, classify nodule and normal, and classify nodule and normal, respectively. The results reveal that there is a high misjudgment rate in identifying nodule and atelectasis. But in the classification of nodule and normal, all of nodule image are judged correctly, and there are only 4 normal are misjudged as nodule. In the classification of atelectasis and normal, there is only 1 atelectasis is misjudged as normal, and 2 normal are misjudged as atelectasis. Thus, the data itself is the cause that it is difficult to identify nodule and atelectasis.

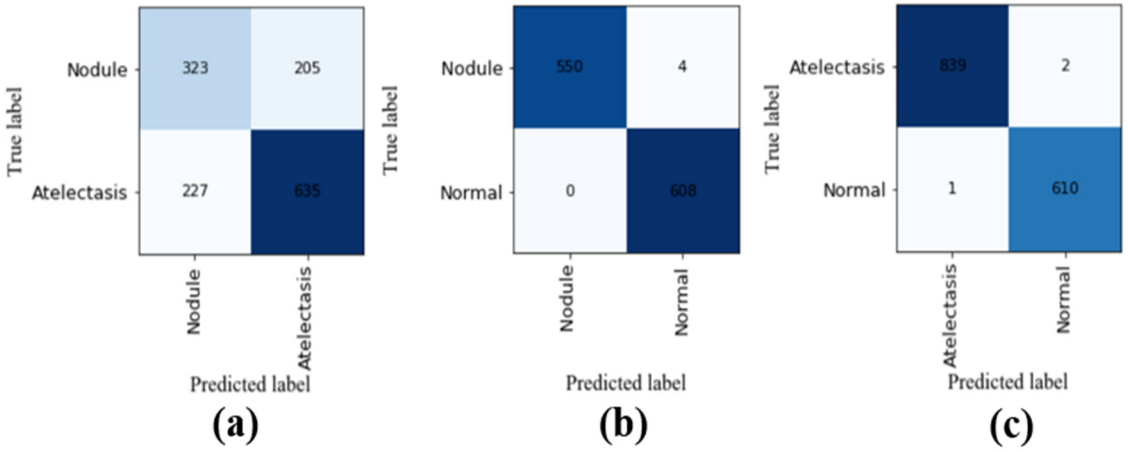

Figure 10. Confusion matrix of using ResNet-152 to classify chest X-ray images: (a) nodule and atelectasis, (b) nodule and normal, (c) atelectasis and normal.

Further studies find that in the chest X-ray 14 dataset, many patients suffer from multiple diseases such as nodule, atelectasis and even infection. To avoid interference, we extract chest X-ray images which has only nodule or atelectasis symptom from the original dataset. However, due to the specificity of the disease itself, the symptom is often complicated and associated. For example, the chest X-ray images marked as nodule sometimes have a certain degree of atelectasis symptoms, and the chest X-ray images marked as atelectasis sometimes have a certain degree of nodule symptoms. The factor causes great interference to MARnet for distinguishing nodule and atelectasis. In future work, we need to build more accurate dataset. A more reasonable classification scheme is needed to classify diseases with similar symptoms to avoid unnecessary interference to the model. What's more, in the dataset of chest X-ray 14 with more than 100,000 images, there are 65,000 healthy images, and the remaining 14 kinds of disease is extremely unbalanced, which seriously affects the application of deep learning in distinguish them. Finally, due to the lack of disease annotations from chest X-ray 14, we cannot determine whether the abnormal parts obtained by MARnet are accurate when using the heat map to display the discriminant results. It means that we cannot compare the discriminant results with the real situation. This limits our ability to reversely correct our model from the visual results. Therefore, we need to construct a certain number of datasets that accurately label the location of disease information to improve MARnet in future.

Acknowledgments

The work was supported by the Natural Science Foundation of Liaoning Province of China (No.20180551011).

Conflict of interest

The authors declare no conflict of interest.

References

1. A. Bhandary, G. A. Prabhu, V. Rajinikanth, K. P. Thanaraj, S. C. Satapathy, D. E. Robbins, et al., Deep-learning framework to detect lung abnormality a study with chest X-Ray and lung CT scan images, *Pattern Recognit. Lett.*, **129** (2020), 271–278. doi: 10.1016/j.patrec.2019.11.013.
2. T. Han, V. X. Nunes, L. F. De Freitas Souza, A. G. Marques, I. C. L. Silva, M. A. A. F. Junior, et al., Internet of medical things-based on deep learning techniques for segmentation of lung and stroke regions in CT scans, *IEEE Access*, **8** (2020), 71117–71135. doi: 10.1109/ACCESS.2020.2987932.
3. S. Kumar, P. Singh, M. Ranjan, A review on deep learning based pneumonia detection systems, in *Proceedings-International Conference on Artificial Intelligence and Smart Systems, ICAIS*, (2021), 289–296. doi: 10.1109/ICAIS50930.2021.9395868.
4. W. S. U. S. Krimsky, *Induced Atelectasis and Pulmonary Consolidation Systems and Methods*, 2019. Available from: https://patentimages.storage.googleapis.com/d1/1e/27/edb84321a9bb25/US10448886.pdf.
5. C. A. de Pinho Pinheiro, N. Nedjah, L. de Macedo Mourelle, Detection and classification of pulmonary nodules using deep learning and swarm intelligence, *Multimedia Tools Appl.*, **79** (2020), 15437–15465. doi: 10.1007/s11042-019-7473-z.
6. D. Brenner, J. McLaughlin, R. Hung, Previous lung diseases and lung cancer risk: a systematic review and meta-analysis, *PLoS One*, **6** (2011). doi: 10.1371/journal.pone.0017479.
7. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84–90. doi: 10.1145/3065386.
8. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vision*, **115** (2015), 211–252. doi: 10.1007/s11263-015-0816-y.
9. L. Zhang, P. Yang, H. Feng, Q. Zhao, H. Liu, Using network distance analysis to predict lncRNA-miRNA interactions, *Interdiscip Sci.*, **13** (2021), 535–545. doi: 10.1007/s12539-021-00458-z.
10. P. P. Sun, Y. B. Chen, B. Liu, Y. X. Gao, Y. Han, F. He, et al., DeepMRMP: a new predictor for multiple types of RNA modification sites using deep learning, *Math. Biosci. Eng.*, **16** (2019), 6231–6241. doi: 10.3934/mbe.2019310.
11. X. Y. Wang, H. Wang, S. Z. Niu, J. W. Zhang, Detection and localization of image forgeries using improved mask regional convolutional neural network, *Math. Biosci. Eng.*, **16** (2019), 4581–4593. doi: 10.3934/mbe.2019229.
12. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, (2016), 770–778. doi: 10.1109/CVPR.2016.90.
13. P. Wang, E. Fan, P. Wang, Comparative analysis of image classification algorithms based on traditional machine learning and deep learning, *Pattern Recognit. Lett.*, **141** (2021), 61–67. doi: 10.1016/j.patrec.2020.07.042.
14. S. Zeng, Y. Cao, Q. Lin, Z. Man, T. Deng, R. Wang, Deep learning SPECT lung perfusion image classification method based on attention mechanism, *J. Phys. Conf. Ser.*, **1748** (2021). doi: 10.1088/1742-6596/1748/4/042050.
15. T. K. K. Ho, J. Gwak, O. Prakash, J. I. Song, C. M. Park, Utilizing pretrained deep learning models for automated pulmonary tuberculosis detection using chest radiography, in *Intelligent Information and Database Systems*, Springer, **11432** (2019), 395–403. doi: 10.1007/978-3-030-14802-7_34.

16. R. Zhang, M. Sun, S. Wang, K. Chen, *Computed Tomography Pulmonary Nodule Detection Method Based on Deep Learning*, 2021. Available from: https://patentimages.storage.googleapis.com/9c/00/cc/4c302cd759496a/US10937157.pdf.
17. C. Tong, B. Liang, Q. Su, M. Yu, J. Hu, A. K. Bashir, et al., Pulmonary nodule classification based on heterogeneous features learning, *IEEE J. Sel. Areas Commun.*, **39** (2021), 574–581. doi: 10.1109/JSAC.2020.3020657.
18. J. H. Lee, H. Y. Sun, S. Park, H. Kim, E. J. Hwang, J. M. Goo, et al., Performance of a deep learning algorithm compared with radiologic interpretation for lung cancer detection on chest radiographs in a health screening population, *Radiology*, **297** (2020). doi: 10.1148/radiol.2020201240.
19. A. Hosny, C. Parmar, T. P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, et al., Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study, *PLoS Med.*, **15** (2018). doi: 10.1371/journal.pmed.1002711.
20. M. Masud, N. Sikder, A. A. Nahid, A. K. Bairagi, M. A. Alzain, A machine learning approach to diagnosing lung and colon cancer using a deep learningbased classification framework, *Sensors (Basel)*, **21** (2021), 1–21. doi: 10.3390/s21030748.
21. G. Liang, L. Zheng, A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Comput. Methods Programs Biomed.*, **187** (2020). doi: 10.1016/j.cmpb.2019.06.023.
22. X. Wei, Y. Chen, Z. Zhang, Comparative experiment of convolutional neural network (CNN) models based on pneumonia X-ray images detection, in *Proceedings-2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI*, (2020), 449–454. doi: 10.1109/MLBDBI51377.2020.00095.
23. L. Račić, T. Popovic, S. Caki, S. Sandi, *Pneumonia Detection Using Deep Learning Based on Convolutional Neural Network*, in *2021 25th International Conference on Information Technology*, (2021), 1–4. doi: 10.1109/IT51528.2021.9390137.
24. A. G. Taylor, C. Mielke, J. Mongan, Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study, *PLoS Med.*, **15** (2018). doi: 10.1371/journal.pmed.1002697.
25. S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, et al., Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, *IEEE Trans. Med. Imaging*, **39** (2020), 2676–2687. doi: 10.1109/TMI.2020.2994459.
26. T. Hu, K. Mohammad, M. Mokhtar, P. Gholam-Reza, T. K. Sarkhel H, R. Tarik A, Real-time COVID-19 diagnosis from X-Ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm, *Biomed. Signal Process. Control*, **68** (2021). doi: 10.1016/j.bspc.2021.102764.
27. M. A. Khan, S. Kadry, Y. D. Zhang, T. Akram, M. Sharif, A. Rehman, et al., Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine, *Comput. Electr. Eng.*, **90** (2021). doi: 10.1016/j.compeleceng.2020.106960.
28. G. B. Kim, K. H. Jung, Y. Lee, H. J. Kim, N. Kim, S. Jun, et al., Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease, *J. Digit. Imaging*, **31** (2018), 415–424. doi: 10.1007/s10278-017-0028-9.
29. R. Sivaramakrishnan, S. Antani, S. Candemir, Z. Xue, J. Abuya, M. Kohli, et al., Comparing deep learning models for population screening using chest radiography, in *SPIE Medical Imaging 2018: Computer-Aided Diagnosis*, Houston, Texas, USA, **10575** (2018). doi: 10.1117/12.2293140.

30. C. A. de Pinho Pinheiro, N. Nedjah, L. de Macedo Mourelle, Detection and classification of pulmonary nodules using deep learning and swarm intelligence, *Multimed. Tools Appl.*, **79** (2019), 15437-15465. doi: 10.1007/s11042-019-7473-z.
31. X. Yiwen, H. Ahmed, Z. Roman, P. Chintan, C. Thibaud, F. Idalid, et al., Deep learning predicts lung cancer treatment response from serial medical imaging, *Clin. Cancer Res.*, **25** (2019). doi: 10.1158/1078-0432.CCR-18-2495.
32. K. C. Chen, H. R. Yu, W. S. Chen, W. C. Lin, Y. C. Lee, H. H. Chen, et al., Diagnosis of common pulmonary diseases in children by X-ray images and deep learning, *Sci. Rep.*, **10** (2020), 17374. doi: 10.1038/s41598-020-73831-5.
33. P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, et al., Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists, *PLoS Med.*, **15** (2018). doi: 10.1371/journal.pmed.1002686.
34. A. I. Aviles-Rivero, N. Papadakis, R. Li, P. Sellars, Q. Fan, R. T. Tan, et al., GraphX NET-chest X-Ray classification under extreme minimal supervision, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019-22nd International Conference*, (2019). doi: 10.1007/978-3-030-32226-7_56.
35. X. Wang, Y. Peng, L. Lu, Z. Lu, R. M. Summers, TieNet: text-image embedding network for common thorax disease classification and reporting in chest X-rays, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2018), 9049–9058. doi: 10.1109/CVPR.2018.00943.
36. S. Xu, H. Wu, R. Bie, CXNet-m1: anomaly detection on chest X-rays with image-based deep learning, *IEEE Access*, **7** (2019), 4466–4477. doi: 10.1109/ACCESS.2018.2885997.
37. J. Zhao, M. Li, W. Shi, Y. Miao, Z. Jiang, B. Ji, A deep learning method for classification of chest X-ray images, *J. Phys. Conf. Ser.*, **1848** (2021). doi: 10.1088/1742-6596/1848/1/012030.
38. T. K. K. Ho, J. Gwak, Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities, *IEEE Access*, **8** (2020), 160749–160761. doi: 10.1109/ACCESS.2020.3020802.
39. I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, B. Ibragimov, Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database, *Comput. Electr. Eng.*, **78** (2019), 388-399. doi: 10.1109/ACCESS.2020.3020802.
40. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 2011–2023. doi: 10.1109/TPAMI.2019.2913372.
41. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, *Int. J. Comput. Vision*, **128** (2020), 336–359. doi: 10.1109/ICCV.2017.74.
42. N. L. Ramo, K. L. Troyer, C. M. Puttlitz, Comparing predictive accuracy and computational costs for viscoelastic modeling of spinal cord tissues, *J. Biomech. Eng.*, **141** (2019). doi: 10.1115/1.4043033.
43. D. M. Powers, Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation, *J. Mach. Learn. Technol.*, **2** (2011), 2229–3981. Available from: http://hdl.handle.net/2328/27165.
44. T. Fawcett, An introduction to ROC analysis, *Pattern Recognit. Lett.*, **27** (2006), 861–874. doi: 10.1016/j.patrec.2005.10.010.

45. C. X. Ling, J. Huang, H. Zhang, AUC: a better measure than accuracy in comparing learning algorithms, in *Advances in Artificial Intelligence, 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003*, Halifax, Canada, 2003. doi: 10.1007/3-540-44886-1_25.
46. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings*, (2015). arXiv:1409.1556.
47. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, *IEEE Comput. Soc.*, (2015), 1–9. doi: 10.1109/CVPR.2015.7298594.
48. Y. Yang, Z. Zhong, T. Shen, Z. Lin, Convolutional neural networks with alternately updated clique, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2018), 2413–2422. doi: 10.1109/CVPR.2018.00256.
49. G. Zeng, On the confusion matrix in credit scoring and its analytical properties, *Commun. Stat. Theory Methods*, **49** (2020), 2080–2093. doi: 10.1080/03610926.2019.1568485.

AIMS Press