Mathematical Biosciences and Engineering

MBE, 19(7): 6841–6859.
DOI: 10.3934/mbe.2022322
Received: 22 November 2021
Revised: 28 February 2022
Accepted: 10 April 2022
Published: 07 May 2022

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

Research article

ACRnet: Adaptive Cross-transfer Residual neural network for chest X-ray images discrimination of the cardiothoracic diseases

Boyang Wang and Wenyu Zhang*

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

* **Correspondence:** Email: zhangwenyu8518@ustl.edu.cn, zhangwenyu8518@126.com.

Abstract: Cardiothoracic diseases are a serious threat to human health and chest X-ray image is a great reference in diagnosis and treatment. At present, it has been a research hot-spot how to recognize chest X-ray image automatically and exactly by the computer vision technology, and many scholars have gotten the excited research achievements. While both emphysema and cardiomegaly often are associated, and the symptom of them are very similar, so the X-ray images discrimination for them led easily to misdiagnosis too. Therefore, some efforts are still expected to develop a higher precision and better performance deep learning model to recognize efficiently the two diseases. In this work, we construct an adaptive cross-transfer residual neural network (ACRnet) to identify emphysema, cardiomegaly and normal. We cross-transfer the information extracted by the residual block and adaptive structure to different levels in ACRnet, and the method avoids the reduction of the adaptive function by residual structure and improves the recognition performance of the model. To evaluate the recognition ability of ACRnet, four neural networks VGG16, InceptionV2, ResNet101 and CliqueNet are used for comparison. The results show that ACRnet has better recognition ability than other networks. In addition, we use the deep convolution generative adversarial network (DCGAN) to expand the original dataset and ACRnet's recognition ability is greatly improved.

Keywords: convolutional neural network; chest X-ray image; deep convolution generative adversarial networks; cardiothoracic diseases

1. Introduction

Cardiothoracic diseases mainly include lung disease and heart disease. The lung is connected with the outside world and it is the main organ of the respiratory system, and the blood of the whole body flows through it. The heart plays a major role in the body's circulatory system, and it pushes blood around the body and carries away metabolic waste. Both lung and heart are important organs for us. According to the data from World Health Organization, lung disease is the third most lethal disease in the world, while heart disease is the first one. All over the world, cardiothoracic diseases cause more than ten million deaths every year.

There are many kinds of cardiothoracic diseases, among which emphysema and cardiomegaly are two common diseases. The characteristic pathological symptoms of emphysema are shown as the increase of the lung volume, the excessive dilation of the bronchial, the decrease of the bronchial airway elasticity and the destruction of the airway walls. The disease is usually caused by smoking, industrial dust, chronic bronchitis and pulmonary inflammation. Emphysema can be sub-divide into senile one, compensatory one, interstitial one and obstructive one so on. The emphysema physiological symptoms include chronic cough, sputum, tightness of chest and shortness of breath. In the chest radiograph of emphysema, bronchial swelling or lung bubble, can be obviously observed. Cardiomegaly is shown as an increase of the cardiac load caused mainly by circulatory and respiratory diseases, and it can be subdivided into physiological one and pathological one. Cardiomegaly symptoms are typically similar to ones of emphysema, including cough, chest tightness, heart palpitations, and dyspnea. The coronary heart disease, hypertension and corpulmonale, are the main pathological causes of cardiomegaly. The chest radiograph of cardiomegaly is often shown as enlargement of the right heart with pulmonary hypertension. In fact, cardiomegaly is often secondary to lung diseases including emphysema, and the symptoms of them are similar, so there is a close relationship between emphysema and cardiomegaly. While emphysema and cardiomegaly are two diseases that require different medications and treatment regimens, so it is important to accurately diagnose them.

Chest X-ray is the most commonly used medical technology for the diagnosis of cardiothoracic diseases, including emphysema and cardiomegaly. It is very effective in identifying and detecting cardiothoracic, pulmonary and interstitial diseases, and plays an important role in the treatment of cardiothoracic diseases [1]. Accurately analyzing patients' condition is a huge challenge for radiologists, and computer-aided diagnosis can reduce differences of the radiologists' conclusions and provide the valid reference for clinicians in abnormal cases. In recent years, we have been using deep learning in many fields which includes medical research and computer vision, and have gotten significant progress [2–6]. Related studies include the detection of pneumonia [7–9], pneumothorax [10], tuberculosis [11], pulmonary nodules [12,13], lung cancer [14–16], and COVID-19 [17–19]. In this work, we construct a valid machine learning model to recognize cardiomegaly, emphysema and normal by the chest X-ray images. Our contributions are as follows.

1) We construct an adaptive cross-transfer residual neural network to identify three kinds of chest X-ray images. To identify better, we alternately transmit the feature information extracted by adaptive structure and residual block to different layers.

2) We use VGG16, InceptionV2, ResNet101 and CliqueNet to compare the recognition ability with ACRnet. The results show that the recognition performance of ACRnet is better than that of the four classical models. In addition, compared with state-of-the-art convolutional neural networks, ACRnet

has better recognition ability too.

3) To further improve the recognition ability of ACRnet, we use DCGAN to expand the original dataset. Compared with the previous results, the recognition ability of ACRnet is improved further, when ACRne is trained by the expanded dataset.

The paper is structured as follows: in Section 2, related works about the deep learning model for the X-ray images recognition are introduced. In Section 3, the design and tuning of the ACRnet are illustrated. Experiment result and analysis are given in Section 4. Finally, conclusions are stated in Section 5.

2. Related works

Here we summarize the current literature on deep learning technology for medical image detection. In the early researches, lacking data is a big problem for the application of deep learning in chest X-ray image diagnosis. To solve this problem, Qasim et al. [20] used data expansion technology to increase the chest X-ray images of COVID-19 from 219 to 1000, and the new image dataset was used to train models to recognize chest X-ray images. By using these synthetic images to expand the existing data, the diversity of training data was increased and the classification performance was improved. Loey et al. [21] used AlexNet, GoogleNet and ResNet18 to detect chest X-ray images of COVID-19, normal, bacterial and viral pneumonia. In order to improve the recognition accuracy, they used the GAN network to expand the original dataset. In their research, AlexNet reaches the highest detection accuracy of 85.2%. Although GAN was helpful to improve the detection accuracy of the model, for the shallow and simple networks, it was hard to get a satisfactory performance. In order to improve the identification accuracy, Lu et al. [22] used the image feature fusion technology to diagnose COVID-19 in lung window computed tomography (CT). Initially, ResNet18 and ResNet50 were selected as the backbone deep networks to generate corresponding image representations from the CT images. Secondly, the representative information extracted from the two networks was fused by discriminant correlation analysis to obtain refined image features. Thirdly, three randomized neural networks (RNNs): extreme learning machine, Schmidt neural network and random vector functional-link net were trained using the refined features, and the predictions of the three RNNs were integrated to get a more robust classification performance. In another study, Lu et al. [23] utilized transfer learning to obtain the image-level representation (ILR) based on the backbone deep convolutional neural network. Then, a novel neighboring aware representation (NAR) was proposed to exploit the neighboring relationships between the ILR vectors. To obtain the neighboring information in the feature space of the ILRs, an ILR graph was generated based on the k-nearest neighbor algorithm, in which the ILRs were linked with their k-nearest neighboring ILRs. Afterward, the NARs were computed by the fusion of the ILRs and the graph. On the basis of this representation, a novel end-to-end COVID-19 classification architecture called neighboring aware graph neural network (NAGNN) was proposed and achieved good results. To detect better tuberculosis, Duong et al. [24] conceptualized a framework by adopting three recent deep neural networks as the main classification engines, and the three networks were called modified EfficientNet, modified original Vision Transformer, and modified Hybrid EfficientNet with Vision Transformer. Moreover, they empowered the learning process by various augmentation techniques. To further study the proposed approach, they compared it with two state-of-the-art systems. The results showed that their method had the best detection effect. Ferreira

et al. [25] used the DenseNet model to detect cardiomegaly in chest X-ray images and they used a public dataset to train the neural network, and a private dataset network to verify the generalization ability of the model. DenseNet obtained an AUC value of 0.818 on the public dataset and 0.809 on the private dataset. In the detection of lung diseases, some scholars made more efforts to achieve better detection results, some advanced methods were used, and some classical neural network models were transformed into more complex structures, unfortunately, the results were still dissatisfactory. To further improve the ability to identify lung diseases, Wu et al. [26] proposed a vision transformer (ViT) model to classify the emphysema subtypes via CT images. The average accuracy of the pre-trained ViT model achieves 95.95% in their lab's own dataset, which is higher than that of AlexNet, Inception-V3, MobileNet-V2, ResNet34, and ResNet50. In some recent studies, Rajpurkar et al. [27] developed the CheXNeXt network model by using Chest X-ray 8 dataset, and compared the AUC values obtained by the model with those obtained by cardiothoracic radiologists. The AUC values of cardiomegaly and emphysema obtained by radiologists were 0.888 and 0.911 respectively, while those obtained by CheXNeXt were 0.831 and 0.704. CheXNeXt's performance in differentiating cardiomegaly and emphysema is far from that of radiologists. For the Chest X-ray 14 dataset, Angelica et al. [28] proposed GraphXNET, and achieved good results on it. Wang et al. [29] proposed TieNet. Compared with the previous classical neural networks, TieNeT significantly improved the recognition ability of chest X-ray images (average AUC increase by 6 %). Zhao et al. [30] proposed AM_DenseNet for chest X-ray image classification. This model used a dense connection network and added a attention module at the end of each dense block to optimize the feature extraction ability. They used the Focal Loss function to solve the data imbalance problem. The average AUC of AM_DenseNet in detecting the chest diseases reached 0.8537. For different types of lung diseases, some scholars had proposed their own state-of-the-art methods, but the detection effect still was expected to further increase. Further research found both the poor model structure and the noise of the data might lead to low accuracy of the detection of lung diseases. Some scholars have mentioned that although the total number of the Chest X-ray 14 dataset exceeds 100,000, a large number of chest X-ray images have more than two diseases [31]. If the data is directly classified without screening, it will cause great interference to the model and seriously affect the discriminant result. So in this work, we select only cardiomegaly, emphysema and normal from the Chest X-ray 14 dataset for research. In previous studies, scholars have proposed a large number of methods and models to identify chest X-ray images of cardiothoracic diseases and achieved good results. While due to strong relation and similarity between emphysema and cardiomegaly diseases, it is difficult to better distinguish them, so it is still necessary to build a more high precision recognition model in order to decrease the misdiagnosis rate. In this work, we construct a network model, ACRnet, for the identification of cardiomegaly, emphysema and normal. Compared with previous research, ACRnet improves comprehensively the identification performance of cardiomegaly and emphysema in a variety of evaluation indexes.

3. Proposed approaches

3.1. Design

This section describes the overall process of using the expanded dataset to train ACRnet. As shown in Figure 1, we firstly use DCGAN to synthesize cardiomegaly and emphysema chest X-ray

images, and then, send the expanded dataset into ACRnet. After several rounds of feature extraction, the input images will be finally recognized by the model and the category label will be output. We call module A as a block, and ACRnet is composed of multiple blocks in series.

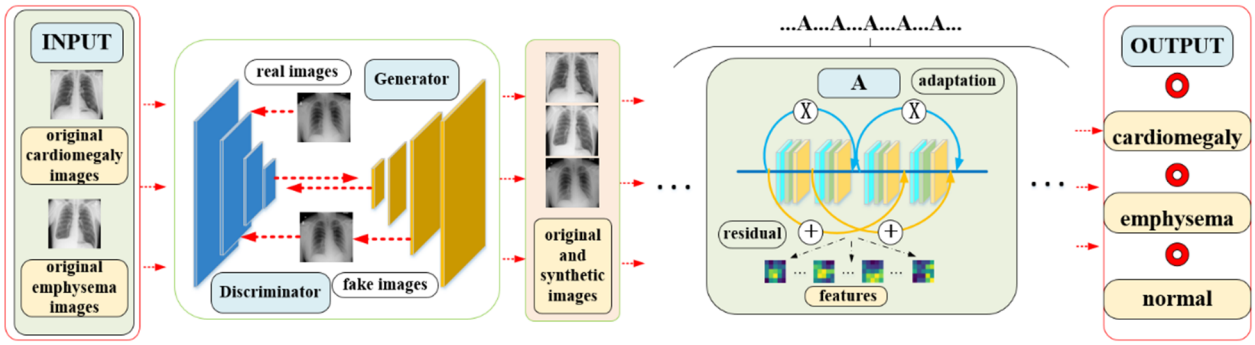

Figure 1. The flowchart of the image synthesis, training and recognition of the ACRnet.

3.2. Data description

The images in our study come from the Chest X-ray 14 dataset (Download address: https://www.kaggle.com/nih-chest-xrays/data), and each image has single or multiple pathological markers. Radiological reports show that the accuracy of pathological markers is more than 90% [31]. We select images with cardiomegaly or emphysema from Chest X-ray 14 dataset, and add normal images as the dataset. We divide all data into 80% for training and 20% for test. Table 1 shows the number of chest X-ray images screened from the Chest X-ray 14 dataset for training and test. To get better performance, we used DCGAN to produce 2000 images of cardiomegaly and emphysema, respectively. Since there are more than 60,000 normal images in the chest X-ray 14 dataset, we select 2000 normal images together with the synthesized images to expand the original train set. In order to meet the requirements of the model for the input size, we resize both the original 1024 × 1024 pixels image and the 256 × 256 pixels image synthesized by DCGAN into 224 × 224 pixels image.

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

Disease	Train set	Test set
cardiomegaly	840	210
emphysema	720	180
normal	800	200

3.3. Deep convolution generative adversarial network

Deep convolutional generative adversarial network (DCGAN) is a special deep learning model which combines convolutional neural network (CNN) and generative adversarial network (GAN) [32]. DCGAN consists of two different types of networks, one is a generator network for synthesizing images, the other is a discriminator network for identifying real and synthetic images. In the process of training DCGAN, the discriminator and generator are trained simultaneously. The

generator network used in the study consists of 7 transposed convolution layers, 6 Rectified Linear Units (ReLU) layers, 3 batch normalization layers and 1 Tanh layer on the end. The discriminator network consists of 7 convolution layers, 6 ReLU layers, 3 batch normalization layers and a sigmoid layer on the end. Each convolutional layer has 16 filters which are the same window size of 4×4 pixel. DCGAN network can solve the problem of insufficient training caused by small dataset. DCGAN is trained by using all the chest X-ray images of emphysema and cardiomegaly from the original dataset. After 1000 iterations, we use DCGAN to produce 2000 chest X-ray images of emphysema and cardiomegaly to expand the training dataset. Table 2 shows the model structure and parameters of generator and discriminator in DCGAN.

Table 2. The model structure and parameters of generators and discriminator in DCGAN.

Generator			Discriminator		
Layer name	Input	output	Layer name	Input	output
ConvTranspose2d	100	16 × 32	Conv2d	3	16
BatchNorm2d			LeakyReLU		
ReLU			Conv2d	16	16 × 2
ConvTranspose2d	16 × 32	16 × 16	LeakyReLU		
BatchNorm2d			Conv2d	16 × 2	16 × 4
ReLU			LeakyReLU		
ConvTranspose2d	16 × 16	16 × 8	Conv2d	16 × 4	16 × 8
BatchNorm2d			BatchNorm2d		
ReLU			LeakyReLU		
ConvTranspose2d	16 × 8	16 × 4	Conv2d	16 × 8	16 × 16
ReLU			BatchNorm2d		
ConvTranspose2d	16 × 4	16 × 2	LeakyReLU		
ReLU			Conv2d	16 × 16	16 × 32
ConvTranspose2d	16 × 2	16	BatchNorm2d		
ReLU			LeakyReLU		
ConvTranspose2d	16	3	Conv2d	16 × 32	1
Tanh			Sigmoid		

3.4. Overview of convolutional neural networks

All of the CNNs are trained in a feedforward & feedback fashion. Input information is transmitted feedforward from the initial layer to the output layer, and errors are propagated feedback from the last layer. In this section, we will introduce 4 neural networks used in the experiments. VGG is a typical and effective neural network and it has 6 different structures, in which VGG16 and VGG19 are commonly used [33]. The VGG model usually consists of 5 blocks, and each contains several convolution layers and a pooling (Pooling Function) layer. VGG16 includes 13 convolution layers, 3 fully connected layers and 5 pooling layers. The convolution layer of VGG16 adopts 3 × 3 convolution kernels, and the pooling layer adopts 2 × 2 pooling kernels.

InceptionV2 is an upgraded version of GoogleLeNet (InceptionV1) [34]. InceptionV1 has a 22-layer network structure with four channels, and convolution kernels of 1 × 1, 3 × 3 and 5 × 5 sizes are used at the same time. Since convolution kernels with different sizes have different receptive

fields, InceptionV1 can better learn different scales features. In addition, in InceptionV1, visual information is aggregated at different scales, so that the subsequent network level can extract different scales features. InceptionV2 firstly use batch normalization to accelerate network training and prevent gradient from disappearing [35]. InceptionV2 also replaces every convolution kernel of 5 × 5 in InceptionV1 with two 3 × 3 convolution kernels, and the change reduces the parameters and enhances the nonlinear processing capability when the receptive field is kept the same.

ResNet is a widely used neural network with 5 structures of 18-layer, 32-layer, 50-layer, 101-layer and 152-layer [36]. The residual block in ResNet uses the jump connection structure to alleviate the gradient disappearance. Using residual structure not only improves the accuracy of the model, but also makes the model to converge faster.

In the block of CliqueNet [37], each layer is the input, is also the output of other layers, and all these layers constitute a ring structure and update alternately. This structure makes high-level visual information be sent back to the previous level for feature reuse. In addition, the introduction of attention mechanism inhibits the irrelevant neurons representing background and noise, which further improves the recognition ability of the model.

3.5. Constitution of ACRnet

Compared with the shallow neural network model, the deep model can better extract image features and has better recognition ability. However, with the increase of network layers, it becomes difficult to train, and the accompanying gradient disappearance problem will reduce the accuracy of the model, while the residual structure is a good chance to solve this problem. In this study, we construct a deep neural network model with residual structure, and at the same time, we add adaptive structure to the model. In order to avoid the reduction effect of residual structure on adaptive structure when two different structures are put together, we adjust the feature transfer mode of residual and adaptive, and construct a residual and adaptive alternate transfer structure. The structure not only achieves high accuracy but also can converge faster. The experimental results show that our model structure has a good ability to recognize emphysema and cardiomegaly. When we compared ACRnet with four classical neural network models and state-of-the art methods, it is found that ACRnet has the best recognition effect in the identification of emphysema and cardiomegaly.

This section introduces the structure of ACRnet, which is a 97-layer neural network. In ACRnet, we construct an adaptive cross-transfer residual structure to extract feature information to improve the recognition efficiency of the model. Adaptive learning is to obtain the importance of different features through autonomous learning, and then suppress secondary features and enhance main features according to importance [38]. In the adaptive module, the adaptive global average pooling layer (Adaptive Avg Pool) is used to compress the feature along the spatial dimension, and each two-dimensional feature is transformed into a figure which has global sensing ability and reflects the feature distribution. And then we use 1 × 1 convolution kernel (input channel is a, output channel is a/4) for dimension reduction, then use ReLU activation function to increase nonlinearity, and finally use another 1 × 1 convolution kernel (input channel is a/4, output channel is a) for dimension increase, so as to reduce the computational cost. In the last layer of the adaptive structure, the sigmoid function is used to generate a figure between 0 and 1 as the output result. The adaptive structure generates a characteristic adaptive coefficient between 0 and 1, and the residual structure generates a matrix. When the adaptive figure is passed to the later network hierarchy, it means that

the eigen-matrix is multiplied by a weight coefficient. If both adaptive and residual structures transfer feature information from the same layer to the same subsequent network layer, it means that the feature matrix generated by the residual structure will be added to a matrix multiplied by the adaptive coefficient, so the result of the adaptive function will be weakened by the residual structure. In order to extract image feature better, we combine adaptive and residual structure to construct a deep neural network model. In ACRnet, the residual structure transmits the features of layer 0 and layer 1 to the end of layer 3 and 4. The adaptive structure transmits the features of layer 0 and layer 2 to the end of layer 2 and layer 4. This structural design is valid to avoid the reduction effect of residual structure on adaptive function. Figure 2 demonstrates ACRnet's adaptive cross-transfer residual block.

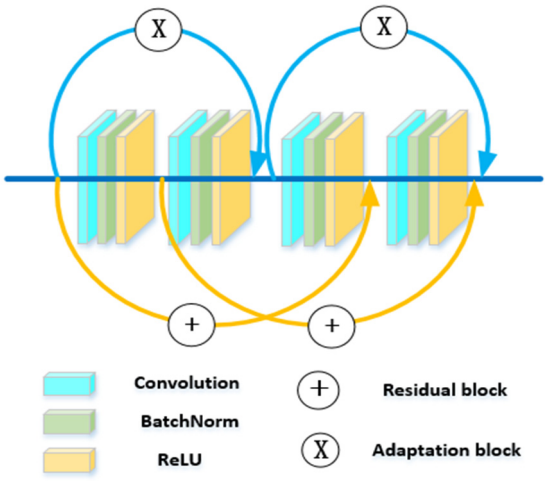

Figure 2. Adaptive cross-transfer residual block in ACRnet.

Table 3. Structure and parameters of ACRnet.

layer	Output size	Kernel size	channel
Input (224 × 224 × 3)			
block × 4	112 × 112	3	60
block × 4	56 × 56	3	120
block × 4	28 × 28	3	240
bBlock × 6	14 × 14	3	480
block × 6	7 × 7	3	960
AvgPool (kernel size = 7, stride = 1)			
Linear (output = 3)			

3.6. Hyper-parameters tuning

In our work, we try to estimate the influence of various hyper-parameters on the performance of the depth model. We divide all the parameters into external parameters and internal parameters and we fine-tune the parameters respectively. The external parameters of the model include optimization algorithm, learning rate and batch size, which affect model performance very much. We find that Adam is significantly better than Ada-delta or SGD, so it is selected as the optimization algorithm.

Learning rate affects the convergence speed of the model, but excessive learning rate will reduce the accuracy of the model. We tried seven different learning rates {1e-3,1e-4,1e-5,3e-3,3e-4,3e-5,3e-6} and find it's better on the condition of 1e-4 or 3e-4. Through experimenting repeatedly, we select 32 as batch size from {16,24,32,48,64}. The internal parameters in the neural network include the size of convolution kernel, stride, the size of pooling layer window and channel [39]. In the VGG16, InceptionV2, ResNet101 and CliqueNet, we choose the original internal parameters. Table 3 shows the structure and parameters of ACRnet model.

4. Experiments

4.1. Performance evaluation

We choose Inception Score (IS) [40] and Fréchet Inception Distance (FID) [41] as indexes to evaluate the quality of synthetic images. For the universal standard, both of IS and FID should be gotten by the InceptionV3 model pre-trained on ImageNet dataset [42,43]. The higher IS, the lower FID, and the better image quality. We introduce accuracy (ACC) [44], Precision, Recall and F1-score [45] as indicators to evaluate the performance of the detection model. The formulas for these indicators are as follows.

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

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

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

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

Here TP, FP, TN and FN represent the number of true positive, false positive, true negative and false negative, respectively. In addition, in order to further evaluate the model performance, we use the confusion matrix [46–48], plot the receiver operating characteristic (ROC) curve and get the area under the curve (AUC). The higher the AUC, the better performance.

4.2. Results

This section introduces the result of recognizing cardiomegaly, emphysema and normal by ACRnet and other four classical neural networks: VGG16, InceptionV2, ResNet101 and CliqueNet. In order to improve the recognition ability of ACRnet, we use DCGAN to expand the original train-dataset. The original images are all 1024 × 1024 pixels, and the images generated by GAN are all 256 × 256 pixels. Before we use DCGAN to generate the image, the original image is converted into 256 × 256 pixels. We convert all images into 224 × 224 pixels before feeding them into the networks for final disease identification. In addition, due to the high quality of the original image, we do not carry out de-noising, sharpening or clipping before the original image is input into the neural network. All of our experiments are done on a personal computer with an NVIDIA GTX2080Super

(8 G) GPU and we use the PyTorch framework. The training process of The DCGAN and all of the CNNs is finished on the 1000 training epochs. The size of the trained ACRnet model is 11,780 K. Saving the trained model can make it easier to import the model into a computer for disease detection in practical applications. Figure 3 shows the original images from the dataset and the synthetic images generated by DCGAN, which are very similar. We calculate the IS scores of them respectively, and find that the scores are very close, and it means that the quality of the synthetic images is good. This conclusion is proved after calculating the FID scores, because we sent them into the detection model simultaneously and obtain a relatively low FID value. Table 4 shows the IS and FID scores.

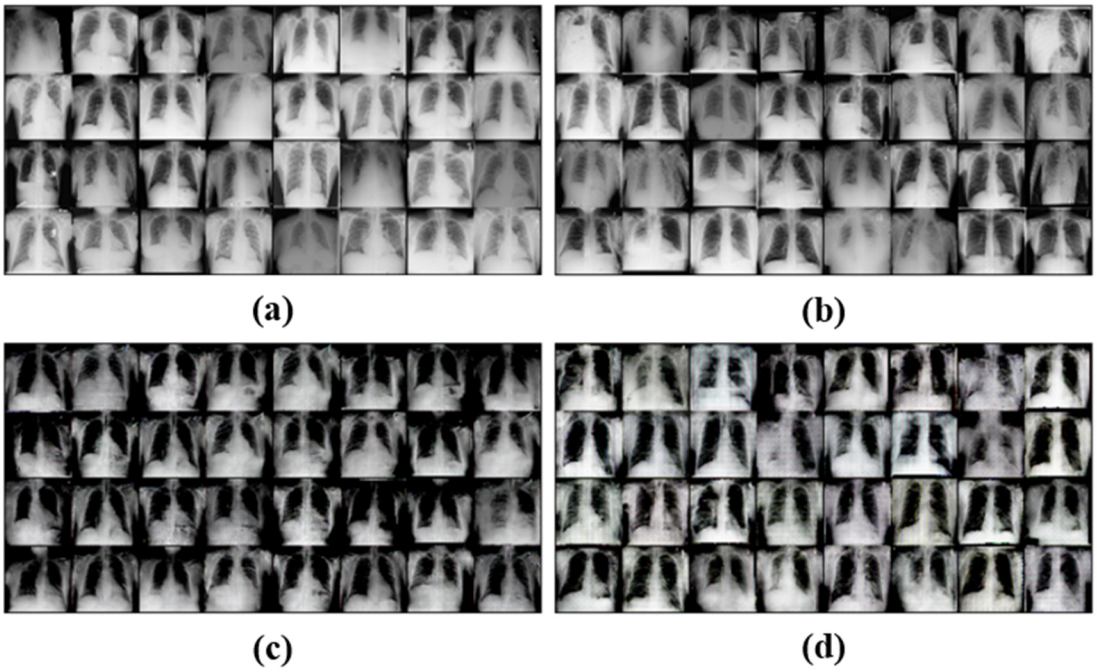

Figure 3. Original and the synthetic chest X-ray images: (a) original cardiomegaly images, (b) original emphysema images, (c) synthetic cardiomegaly images, (d) synthetic emphysema images.

Table 4. IS and FID scores.

	original images	synthetic images
IS	3.72	3.54
FID	77.6	

Table 5 shows the experimental results of different neural networks. Models are trained on all the available data in each training set, and global assessment is presented by 5-fold cross validation on the whole dataset. When using four classical neural networks to identify cardiomegaly, the recall index obtained by CliqueNet is at a high level. The precision index obtained by InceptionV2 and ResNet101 are better than that obtained by VGG16 and CliqueNet. The F1-scores obtained by the four classical neural networks are close. The AUC index of InceptionV2 is lower than the other three networks. When using four classical neural networks to identify emphysema, CliqueNet achieves the

optimal precision score. The recall and F1-score obtained by InceptionV2 and ResNet101 are significantly better than CliqueNet and VGG16. The recall index obtained by CliqueNet is the lowest. The AUC index obtained by InceptionV2 and CliqueNet are higher than the other two neural networks. When identifying normal, the F1-score and AUC index obtained by the four classical neural networks are relatively close. VGG16 achieves the best precision index and ResNet101 achieves better recall index than other networks. Overall, the existing four classical neural networks: VGG16, InceptionV2, ResNet101 and CliqueNet have their respective advantages in recognizing cardiomegaly, emphysema and normal chest X-ray images. But none of them are very good at identifying all three types of chest X-ray images. ACRnet is a deep network model with cross residual structure and the adaptive mechanism, and can extract image features better. Compared with the four classical network structures, ACRnet has higher precision, recall, F1-Score and AUC indexes in recognizing the three types of chest X-ray images. After data expansion by DCGAN, the recognition ability of ACRnet is further improved. Figure 4 shows intuitively the detection result of all models for three types of chest X-ray images.

Table 5. The mean precision, recall, F1-score and AUC values obtained by ACRnet, VGG16, InceptionV2, ResNet101 and CliqueNet in detecting cardiomegaly, emphysema and normal using 5-fold cross validation.

Model	Disease	Precision	Recall	F1-score	AUC
ACRnet +	Cardiomegaly	0.8532	0.8857	0.8691	0.95
DCGAN	Emphysema	0.8788	0.8056	0.8507	0.93
	Normal	0.9372	0.9700	0.9533	0.99
ACRnet	Cardiomegaly	0.8045	0.8429	0.8233	0.92
	Emphysema	0.8137	0.7278	0.7684	0.90
	Normal	0.9282	0.9700	0.9486	0.99
VGG16	Cardiomegaly	0.7093	0.7667	0.7369	0.89
	Emphysema	0.6590	0.6333	0.6456	0.84
	Normal	0.9789	0.8857	0.9300	0.99
InceptionV2	Cardiomegaly	0.8143	0.6476	0.7214	0.90
	Emphysema	0.6473	0.8056	0.7178	0.87
	Normal	0.9548	0.9500	0.9524	1.00
ResNet101	Cardiomegaly	0.8086	0.6762	0.7365	0.88
	Emphysema	0.6872	0.7444	0.7147	0.83
	Normal	0.8995	0.9850	0.9437	1.00
CliqueNet	Cardiomegaly	0.6768	0.8476	0.7537	0.90
	Emphysema	0.7121	0.5222	0.6025	0.87
	Normal	0.9692	0.9450	0.9569	0.96

Table 6 shows mean ACC and average-AUC index obtained by ACRnet and four classical neural networks using 5-fold cross validation. The result shows that the ACC and average-AUC values of four classical neural networks in identifying cardiomegaly, emphysema and normal are close. The overall recognition performance of ACRnet trained by the original dataset is better than that of other 4 classical neural networks. The recognition ability of ACRnet is further improved on the expanded dataset. Figure 5 shows the confusion matrixes obtained by neural networks using

5-fold cross validation. Comparing the confusion matrixes, we find that VGG16 has better recognition ability for cardiomegaly than InceptionV2 and ResNet101, and worse recognition ability for cardiomegaly than InceptionV2 and ResNet101. Compared with the other three classical neural networks, CliqueNet has the best recognition ability for cardiomegaly, but the worst recognition ability for emphysema. Confusion matrix shows that ACRnet has a good and relatively balance recognition ability in identifying three types of chest X-ray images: cardiomegaly, emphysema and normal. The overall performance of ACRnet is better than that of the existing four classical neural networks. On the expanded dataset, the recognition ability of ACRnet is further improved. Figure 6 shows the ROC curves obtained by neural networks using 5-fold cross validation. The results show that the recognition ability of ACRnet for emphysema is better than the existing four classical neural networks, and the recognition ability for cardiomegaly and normal is also at a high level. After data expansion, ACRnet's ability to recognize cardiomegaly and emphysema is significantly improved.

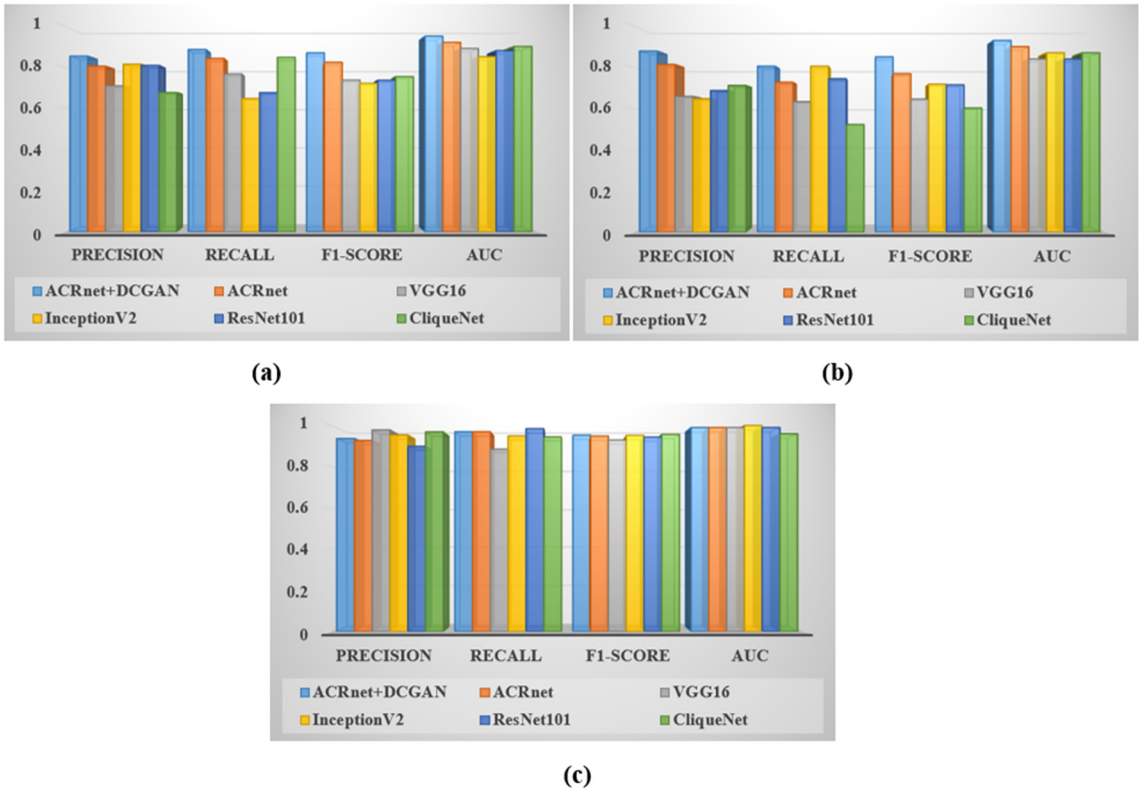

Figure 4. Mean precision, recall, F1-score and AUC values obtained by different models using 5-fold cross validation: (a) cardiomegaly, (b) emphysema, (c) normal.

Table 6. Mean ACC and average-AUC obtained by convolutional neural networks.

Model	ACC	average-AUC
ACRnet+DCGAN	0.8898	0.956
ACRnet	0.8508	0.936
VGG16	0.7813	0.906
InceptionV2	0.7983	0.923
ResNet101	0.8017	0.903
CliqueNet	0.7814	0.910

Analyzing furtherly the confusion matrix in Figure 5, for emphysema and cardiomegaly, in (a) the overall misdiagnosis rate is calculated as 15.1%, (b) 21%, (c) 30%, (d) 27.9%, (e) 29.2%, (f) 33.6%. It is obviously that the misdiagnosis rate of ACRnet is lower than the other four classical neural networks. When the expanded dataset is used, the misdiagnosis rate of ACRnet has been declined to 15.1%. With the decrease of the misdiagnosis rate, patients will have gotten more valid and timely treatment. So ACRnet has good practical applied value.

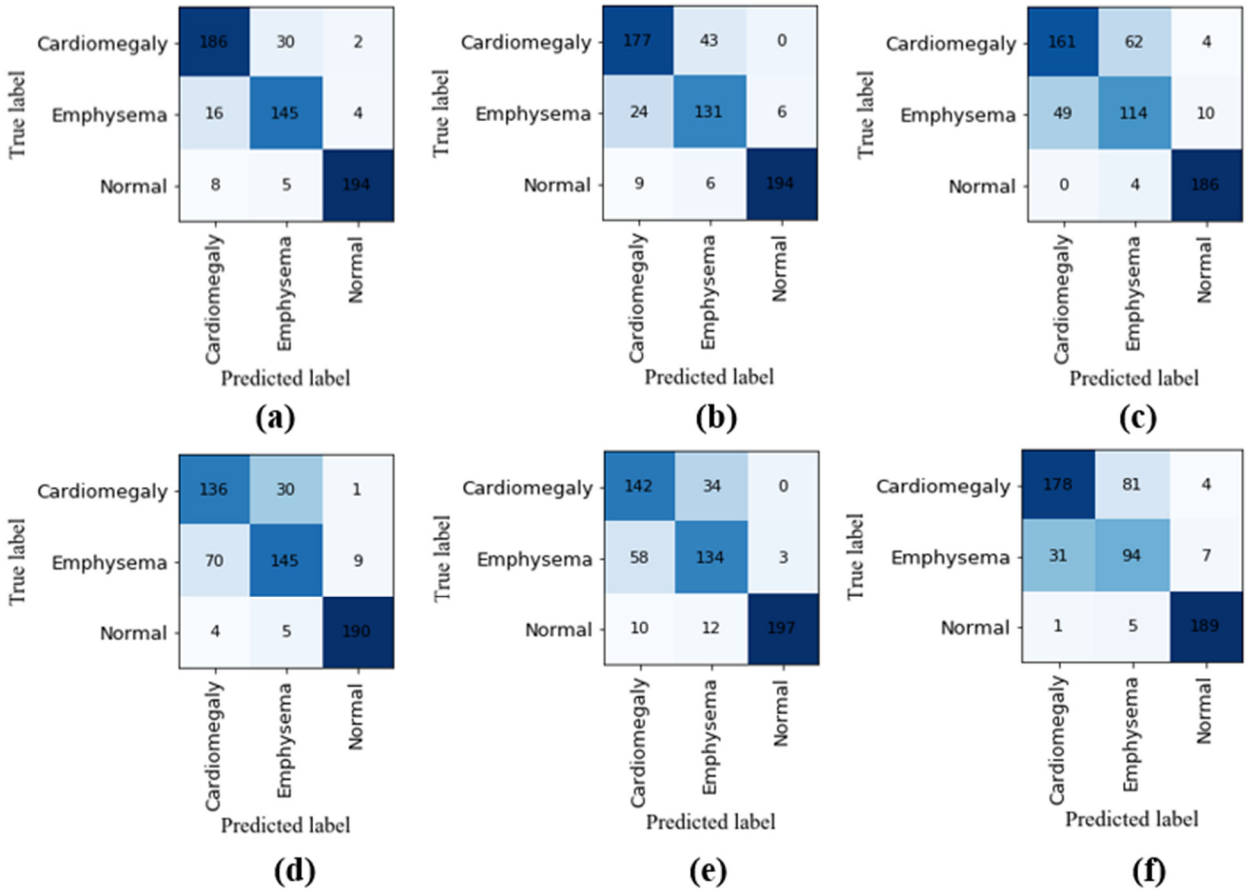

Figure 5. Confusion matrixes obtained by neural networks: (a) ACRnet trained in dataset which expanded by DCGAN, (b) ACRnet trained in original dataset, (c) VGG16 trained in original dataset, (d) InceptionV2 trained in original dataset, (e) ResNet101 trained in original dataset, (f) CliqueNet trained in original dataset.

By analyzing ROC curves, it is not difficult to find that the ROC curves obtained by ACRnet are all relatively smooth. Compared with other models, ACRnet obtains the best AUC index for recognizing three types of chest X-ray images. The result further prove that ACRnet has good recognition ability. In addition, the ACRnet's identification performance will be further improved when the expanded dataset is used to train the model.

To comprehensively evaluate the performance of our method for identifying cardiomegaly, emphysema and normal, we compare our method with several most advanced methods. Table 7 shows the AUC indices of our method and the other 4 state-of-the-art convolutional neural networks: CheXNet [27], GraphXNet [28], TieNet [29] and AM_DenseNet [30]. We can find that our method

has excellent performance in identifying the three types of chest X-ray images.

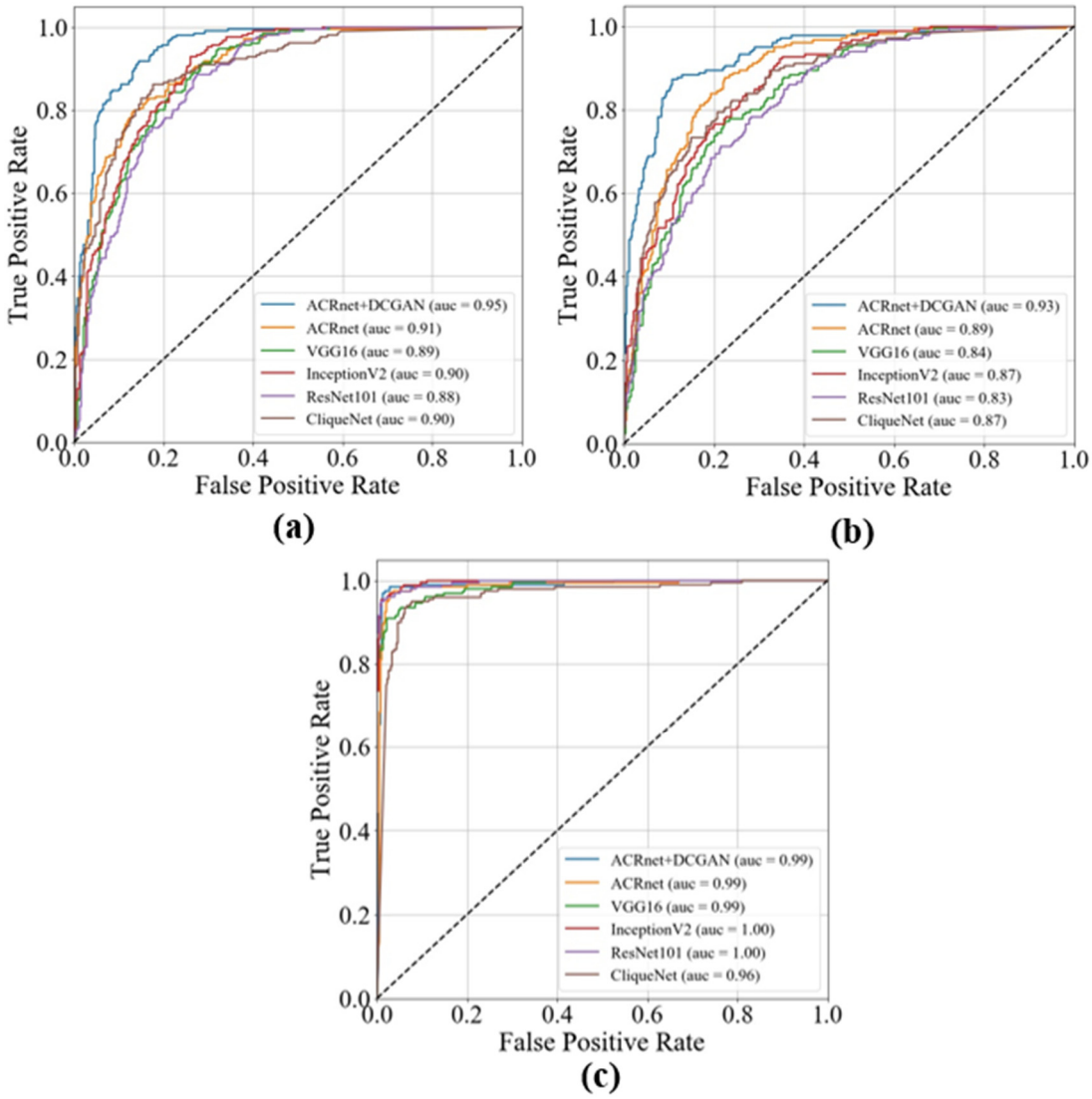

Figure 6. ROC curves obtained by neural networks: (a) cardiomegaly, (b) emphysema, (c) normal.

Table 7. Comparison of the AUC among our method and 4 state-of-the-art convolutional neural networks.

methods	cardiomegaly	emphysema	normal
CheXNet	0.92	0.93	-
GraphXNet	0.88	0.84	-
TieNet	0.84	0.86	0.70
AM_DenseNet	0.91	0.94	-
ACRnet	0.91	0.89	0.99
ACRnet+DCGAN	0.95	0.93	0.99

5. Conclusions

Cardiothoracic diseases are a serious threat to human health, and chest radiography interpretation is very important for timely detection of cardiothoracic diseases. However, this time-consuming work is usually completed by radiologists, and subjective differences among different people will lead to missed diagnosis and misdiagnosis. In this way, using deep learning technology to build neural network model can reduce clinicians' subjective errors, which is of great significance for disease diagnosis and treatment. In this paper, we construct ACRnet to recognize cardiomegaly, emphysema and normal chest X-ray images, and achieve good recognition effect. Using DCGAN technology to expand the existing dataset also can further improve the accuracy of the model. In future research, we hope that our method can be used to distinguish more types of cardiothoracic diseases, so as to help clinicians diagnose cardiothoracic diseases more accurately.

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. D. Brenner, J. McLaughlin, R. Hung, Previous lung diseases and lung cancer risk: A systematic review and meta-analysis, *PLoS One*, **6** (2011). https://doi.org/10.1371/journal.pone.0017479
2. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolution al neural networks, *Commun. ACM*, **60** (2017), 84–90. https://doi.org/10.1371/journal.pone.001747910.1145/3065386
3. 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. https://doi.org/10.1007/s11263-015-0816-y
4. M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: A retrospective, *Int. J. Comp. Vision*, **111** (2014), 98–136, https://doi.org/10.1007/s11263-014-0733-5
5. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft COCO: Common objects in context, in *Computer Vision – ECCV 2014*, Springer, **8693** (2014), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
6. 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. https://doi.org/10.1007/s12539-021-00458-z
7. G. Liang, L. Zheng, A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Comput. Methods Programs Biomed.*, **187** (2020). https://doi.org/10.1016/j.cmpb.2019.06.023

8. X. Wei, Y. Chen, Z. Zhang, Comparative experiment of convolutional neural network (CNN) models based on pneumonia X-ray images detection, in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, (2020), 449–454. https://doi.org/10.1109/MLBDBI51377.2020.00095
9. 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. https://doi.org/10.1109/IT51528.2021.9390137
10. 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). https://doi.org/10.1371/journal.pmed.1002697
11. 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. https://doi.org/10.1007/978-3-030-14802-7_34
12. R. Zhang, M. Sun, S. Wang, K. Chen, Computed Tomography pulmonary nodule detecti on method based on deep learning, US 10937157B2, L. Infervision Medical Technology, 2021. Available from: https://patentimages.storage.googleapis.com/9c/00/cc/4c302cd759496a/US10937157.pdf.
13. 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. https://doi.org/10.1109/JSAC.2020.3020657
14. L. J. Hyuk, S. H. Young, P. Sunggyun, K. Hyungjin, H. E. Jin, G. J. Mo, 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), 687–696. https://doi.org/10.1148/radiol.2020201240
15. 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). https://doi.org/10.1371/journal.pmed.1002711
16. 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. https://doi.org/10.3390/s21030748
17. 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. https://doi.org/10.1109/TMI.2020.2994459
18. H. T. Qing, K. Mohammad, M. Mokhtar, P. GholamReza, T. Karim, T. A. Rashid, 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). https://doi.org/10.1016/j.bspc.2021.102764
19. 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). https://doi.org/10.1016/j.compeleceng.2020.106960

20. Y. Qasim, B. Ahmed, T. Alhadad, H. A. Sameai, O. Ali, The impact of data augmentation on accuracy of COVID-19 detection based on X-ray images, in *Innovative Systems for Intelligent Health Informatics, Lecture Notes on Data Engineering and Communications Technologies*, Springer, **72** (2021), 1041–1049. https://doi.org/10.1007/978-3-030-70713-2_93
21. M. Loey, F. Smarandache, N. E. M. Khalifa, Within the lack of chest COVID-19 X-ray dataset: A novel detection model based on GAN and deep transfer learning, *Symmetry*, **12** (2020). https://doi.org/10.3390/sym12040651
22. S. Y. Lu, D. Wu, Z. Zhang, S. H. Wang, An explainable framework for diagnosis of COVID-19 pneumonia via transfer learning and discriminant correlation analysis, *ACM Trans. Multimedia Comput. Commun. Appl.*, **17** (2021), 1–16. https://doi.org/10.1145/3449785
23. S. Y. Lu, Z. Q. Zhu, J. M. Gorriz, S. H. Wang, Y. D. Zhang, NAGNN: Classification of COVID-19 based on neighboring aware representation from deep graph neural network, *Int. J. Intell. Syst.*, **37** (2021), 1572–1598. https://doi.org/10.1002/int.22686
24. L. T. Duong, N. H. Le, T. B. Tran, V. M. Ngo, P. T. Nguyen, Detection of tuberculosis from chest X-ray images: boosting the performance with vision transformer and transfer learning, *Expert Syst. Appl.*, **184** (2021), 115519. https://doi.org/10.1016/j.eswa.2021.115519
25. J. R. F. Junior, D. A. Cardona, R. A. Moreno, M. F. S. Rebelo, J. E. Krieger, M. A. Gutierrez, A general fully automated deep-learning method to detect cardiomegaly in chest x-rays, in *Progress in Biomedical Optics and Imaging 2021: Computer-Aided Diagnosis*, 2021. https://doi.org/10.1117/12.2581980
26. Y. Wu, S. Qi, Y. Sun, S. Xia, Y. Yao, W. Qian, et al., A vision transformer for emphysema classification using CT images, *Phys. Med. Biol.*, **66** (2021), 245016. https://doi.org/10.1088/1361-6560/ac3dc8
27. 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). https://doi.org/10.1371/journal.pmed.1002686
28. A. I. A. 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. https://doi.org/10.1007/978-3-030-32226-7_56
29. 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. https://doi.org/10.1109/CVPR.2018.00943
30. J. Zhao, M. Li, W. Shi, Y. Miao, Z. Jiang, B. Ji, A deep learning method for classificat ion of chest X-ray images, *J. Phys. Conf. Ser.*, **1848** (2021). https://doi.org/10.1088/1742-6596/1848/1/012030
31. 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. https://doi.org/10.1109/ACCESS.2020.3020802
32. Y. Xiao, M. Lu, Z. Fu, Covered face recognition based on deep convolution generative adversarial networks, in *Lecture Notes in Computer Science*, (2020), 133–141. https://doi.org/10.1007/978-3-030-57884-8_12

33. 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. https://doi.org/10.1109/acpr.2015.7486599
34. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al, Going deeper with convolutions, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2015), 1–9. https://doi.org/10.1109/cvpr.2015.7298594
35. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *32nd International Conference on Machine Learning*, (2015), 448–456. Available from: http://proceedings.mlr.press/v37/ioffe15.pdf.
36. 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*, (2016), 770–778. https://doi.org/10.1109/cvpr.2016.90
37. 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. https://doi.org/10.1109/CVPR.2018.00256
38. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372
39. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, preprint, arXiv:1207.0580v1.
40. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs, *Adv. Neural Inf. Process. Syst.*, (2016), 2234–2242. Available from: https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf.
41. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, *Adv. Neural Inf. Process. Syst.*, (2017), 6627–6638. Available from: https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf.
42. F. F. Li, J. Deng, K. Li, ImageNet: Constructing a large-scale image database, *J. Vision*, **9** (2009). https://doi.org/10.1167/9.8.1037
43. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2818–2826. https://doi.org/10.1109/cvpr.2016.308
44. 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). https://doi.org/10.1115/1.4043033
45. 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.
46. T. Fawcett, An introduction to ROC analysis, *Pattern Recognit. Lett.*, **27** (2006), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
47. 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. https://doi.org/10.1007/3-540-44886-1_25

48. G. Zeng, On the confusion matrix in credit scoring and its analytical properties, in *Communications in Statistics - Theory and Methods*, **49** (2020), 2080–2093. https://doi.org/10.1080/03610926.2019.1568485

AIMS Press