IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2018, accepted November 25, 2018, date of publication December 13, 2018,
date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885997

CXNet-m1: Anomaly Detection on Chest X-Rays With Image-Based Deep Learning

SHUAIJING XU, HAO WU, AND RONGFANG BIE
College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

Corresponding author: Rongfang Bie (rfbie@bnu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2016NT14, in part by the National Natural Science Foundation of China under Grant 61601033 and Grant 61571049, and in part by the Inter-discipline Research Funds of Beijing Normal University under Grant BNUXKJC1825.

ABSTRACT Detecting anomaly of chest X-ray images by advanced technologies, such as deep learning, is an urgent need to improve the work efficiency and diagnosis accuracy. Fine-tuning existing deep learning networks for medical image processing suffers from over-fitting and low transfer efficiency. To overcome such limitations, we design a hierarchical convolutional neural network (CNN) structure for ChestX-ray14 and propose a new network CXNet-m1, which is much shorter, thinner but more powerful than fine-tuning. We also raise a novel loss function sin-loss, which can learn discriminative information from misclassified and indistinguishable images. Besides, we optimize the convolutional kernels of CXNet-m1 to achieve better classification accuracy. The experimental results show that our light model CXNet-m1 with sin-loss function achieves better accuracy rate, recall rate, F1-score, and AUC value. It illustrates that designing a proper CNN is better than fine-tuning deep networks, and the increase of training data is vital to enhance the performance of CNN.

INDEX TERMS Chest X-Rays image, anomaly detection, deep neural network, self-adapting loss function.

I. INTRODUCTION

The chest X-ray is one of the most commonly accessible examinations for screening and diagnosis of many lung diseases including Infiltration, Effusion, Atelectasis, Nodule, Mass, Pneumothorax, Consolidation, Pleural Thickening, Cardiomegaly, Emphysema, Edema, Fibrosis, Pneumonia and Hernia [1]. In the USA alone, over 35 million chest X ray images are taken every year and radiologists have to read more than 100 X-ray studies in a day [2]. In China, there are much more patients and chest X-ray images due to the large population and increasing health consciousness. Advanced technologies and automated algorithms can assist radiologists to diagnose disease effectively and efficiently. Trained tools can classify normal and abnormal chest x-ray images into two categories automatically and radiologists could pay more attention to these abnormal images. Besides, trained tools can classify chest x-ray images into more categories according to different diseases. What's more, trained tools can also help to localize disease and visualize them.

Theoretically, there are dozens of classical classification algorithms and their improved versions. Naive Bayes [3] is one of the most efficient inductive learning algorithms for classification due to its simple structure and incremental construction [4], [5]. As an efficient paradigm, Support Vector Machine (SVM) has a strongest mathematical model for classification and regression [6], [7]. Various improvements of SVM has appeared over the past few decades, such as Lagrangian SVM, least square SVM and twin SVM [6], [8]–[11]. Random Forests is one of the most traditional ensemble learning methods and has been successfully applied to various classification tasks [13]–[15]. Practically, these methods are not capable to process large number of chest x-ray images. On the one hand, most of them can only get good results on small datasets, such as SVM and Naive Bayes. On the other hand, researchers have to manually extract image features including Local Binary Pattern, Histogram of Oriented Gradient and Haar-like before classification, which is complex and difficult [16], [17].

Deep learning, especially convolutional neural networks architecture classification approach, has gained popularity in recent years due to their ability to learn representative image features through automatic back propagation [18]. We have conducted thorough research on CNN and show it solely in the part of Related Work. Based on the background of

big data, CNN is a powerful tool to train a robust classifier for images. However, large training sets are generally not available in the medical domain due to the privacy protection of patients. ImageNet contains 1.3 million natural images to train large deep CNN architectures, but most marked medical image databases only contain at most less than 10 thousand images in total [19]. The largest public dataset of chest X-rays was OpenI, containing 3,955 radiology reports from the Indiana Network for Patient Care and 7,470 associated chest x-rays from the hospitals picture archiving and communication system (PACS) [1], [20]. Reference [1] released a much larger database ChestX-ray14 which contains more than 30,000 patients, 112,120 labeled chest x-ray images last year. Although imbalanced and flawed, ChestX-ray14 is large to train and thus we choose ChestX-ray14 as our dataset.

For classification task, transfer learning, where the source and target domains should be different but related, has been the first choice of many papers to process medical images including chest x-ray images [21]. Researchers train ChestX-ray14 through fine-tuning existing deep networks trained on ImageNet. The only peer-reviewed published work fine tuned four standard CNN architectures (AlexNet, VGGNet GoogLeNet and ResNet) and ResNet achieved the best result [1], [22]–[24]. Others can be found on arXiv. Reference [25] utilized a 121-layer DenseNet architecture and made little modification on this existing network. Reference [26] combined a variant of DenseNet and Longshort Term Memory Networks (LSTM) to exploit the dependencies between abnormalities.

However, fine-tuning existing deep networks is not always the best choice for low transfer efficiency caused by the dissimilarity between medical images and natural images. In addition, excess parameters may cause over-fitting and unnecessary space waste. ChestX-ray14 is large enough for us to train a new smaller CNN without taking too much time and memory. In this paper, we propose a new CNN architecture CXNet-m1 and train it from scratch on ChestX-ray14.

There are two main contributions in our paper. (1)After designing a hierarchical CNN structure for ChestX-ray14 dataset, we propose light model CXNet-m1 to classify chest X-ray images into normal and abnormal categories. CXNet-m1 reduces unnecessary parameters, adjusts the order of some layers, and takes some advantages of classic networks to learn details. It is light and therefore easy to train and store. (2) We also propose a loss function sin-loss used to train CXNet-m1 to improve classification accuracy. Compared with classic loss function, Sin-loss can learn more from misclassified and indistinguishable images through multiplying a self-adapting coefficient.

II. RELATED WORK

Deep learning is a sort of representation-learning method which connects layers and non-linear module to obtain multiple levels of representation. It is clever at coping with high-dimensional data and wisely used in many

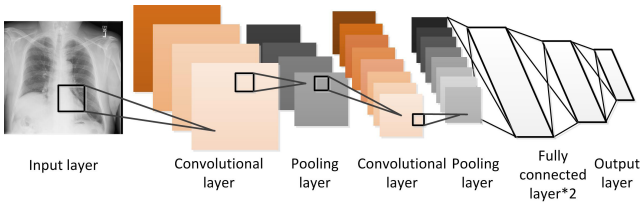

FIGURE 1. Basic architecture of CNN.

domains such as image process, speech recognition and robot manipulators [27]–[31]. Reference [27] made use of Bayesian convolutional neural networks and active learning for hyperspectral image classification. Reference [28] described the latest version of Microsoft's conversational speech recognition system, which added a CNN-BLSTM acoustic model to previous architectures. References [29]–[31] concentrated on robot manipulators and reference [29] proposed a novel robust zeroing neural-dynamics (RZND) model for solving the inverse kinematics problem of mobile robot manipulators. With the development of graphics processing units, deep learning is more and more popular and successful to process data in large amounts because it requires very little engineering by hand.

Convolutional neural networks are capable to process data in the form of multiple arrays such as sequences (1D), images (2D) and videos (3D) [32]–[35]. There are two main characteristics of CNN, local connectivity and shared weights. The two characteristics not only ensure the affine invariance of CNN, but also reduce the number of parameters, which is the reason that CNN is wisely used to learn from complex data, such as images and videos. The basic architecture of CNN is demonstrated as Fig. 1. The first few stages of CNN contain convolutional layers and pooling layers. Convolutional layers are used to detect local conjunctions of features from the previous layer, and pooling layers behind are designed to merge similar features and reduce computational complexity. After extracting features, there are more convolutional and fully-connected layers to classify data. In order to input images on any scales, some scientists replace fully connected layers with convolutional layers, called fully convolutional networks (FCN). Traditional loss functions include 0-1 loss function, square loss function, hinge loss function and log loss function [36]–[39]. Among them, log loss function is most common for its property of convex function, which can avoids the local minimum. Loss function and backpropagating gradients allow all the weights in all filters to be trained and could converge gradually [40], [41].

In recent years, many robust CNN frameworks have been designed including VGGNet, ResNet, Inception-Resnet and DenseNet [22]–[24], [42]. There are two versions of VGGNet, VGGNet-16 and VGGNet-19, the difference between which is the number of layers. In order to learn details, the receptive fields of convolutional layer are very small in VGGNet, such as 3*3 and even 1*1. After several convolutions and poolings, there are three fully-connected

layers. Besides, Relu is used in every hidden layer for the first time and the result has improved because of it. Ideally, CNN should be trained on much large data set to ensure reliability, and thus be deeper to avoid under-fitting. In order to solve the problem caused by deep and complex model, Resnet propose a thought of residual block by adding an identity mapping. The experiment result shows the 50-layer and deeper Resnet can achieve better accuracy than VGGnet. Inception-Resnet makes the network wider based on Resnet and also gets better result. Inspired by Resnet, Densenet creates short paths from early layers to later layers by concatenation to relieve vanishing-gradient, which can be deeper till several hundred layers. All of them are trained on Imagenet, which contains 1.3 million train images and 1000 categories including human, animals, plants, living goods and natural scene.

Training a deep convolutional neural network (CNN) from scratch is not easy because of the property that requires numerous labeled training data and a great deal of expertise to ensure convergence [43]. A promising alternative is to fine-tune networks pre-trained on Imagenet or other large-scale datasets. Reference [44] fine-tuned all layers by backpropagation through the whole FCN-AlexNet, FCN-VGG16, and FCN-GoogLeNet for semantic segmentation. Reference [45] performed category-specific object segmentation in weakly labeled videos by a self-paced fine-tuning network (SPFTN)-based framework. Reference [46] presented a fine-tuning algorithm to update the network pre-trained on images of urban scenes and it transfers semantic features to a different environment successfully. As for medical images much different from natural images, papers hold different ideas. Reference [47] demonstrated that fine-tuning clearly outperformed feature extraction from scratch in multi-class grade assessment of knee osteoarthritis. However, [48] showed that using a new CNN to extract features outperformed fine-tuning in cytopathology image classification.

III. CNN METHOD

There are two parts in this section and the second part is the main idea of this paper. In the first part, we suggest a hierarchical structure to cope with ChestX-ray14 after analyzing the global distribution of it. Compared with designing a classifier containing several parallel outputs, we believe that hierarchical structure is more scientific for it holds more specific and detailed categories and avoids low efficiency caused by imbalance. In the second part, we design the first classifier in the hierarchical structure, CXNet-m1, to cope with the first problem, normal image or not. There are two main contributions in this part, improving the loss function and designing a lighter model architecture. Inspired by the alternating current of electromagnetism, we design a self-adapting factor to adjust the learning process to focus on the features of indistinguishable and misclassified images. Then, considering some disadvantages of fine tuning and the scale of ChestX-ray14, we decide to design a new architecture, which is thinner, lighter and better, and train it from scratch to achieve binary classification. In the Experiment section, we verified that no matter the sin-loss or the architecture of CXNet-m1 or the combination of the loss and architecture can obtain good results and the combination option is most successful.

A. HIERARCHICAL STRUCTURE

It is difficult to access annotated large-scale medical image database. ChestX-ray14 is largest and the most appropriate database accessible at present to train a chest x-ray image classifier. References [25] and [26] construct classifiers whose last layer has 14 parallel outputs to classify the different abnormal images. However, they did not take normal images into consideration or differentiate multi-label and single-label images. ChestX-ray14 contains 112,120 labeled chest x-ray images and many kinds abnormal images. As shown in Fig. 7, the chest x-ray image numbers of each category in Chest X-ray14 is extremely imbalanced. Besides, there are not only 15 categories (normal images and 14 kind abnormal images), but also many multi-label data. For instance, if an image is labeled as Infiltration, Effusion and Atelectasis, it belongs to a new category rather than any of the three categories alone. According to the characteristics of ChestX-ray14 shown in detail in Section IV, we design a hierarchical structure to train images in this dataset. As shown in Fig. 2, training images from chest X-ray14 are firstly labeled as normal and abnormal to train a binary classifier CXNet-m1. Then abnormal images are labeled as multi-label and single-label to train a binary classifier CXNet-m2. After that, CXNet-m3 and CXNet-m4 are trained on respective training images and final results are obtained.

This hierarchical structure is scientific because it not only takes more information of this dataset (normal images and multi-label images), but also eases the problem of imbalance. After all, it is not rational to train a parallel classifier with more than 60 thousand normal images and around 200 hernia images.

B. CXNET-M1

According to the hierarchical structure above, four classifiers CXNet-m1, CXNet-m2, CXNet-m3 and CXNet-m4 should be respectively designed. This paper focuses on the first classifier CXNet-m1, a binary classifier to differentiate normal and abnormal chest x-ray images.

1) PROBLEM FORMULATION

The anomaly detection task of chest x-ray images is a binary classification problem, where the input is a chest X-ray image I and the output is a binary label $y \in \{0, 1\}$ indicating the absence or presence of disease respectively. For images in Chest X-ray14, we use the softmax cross entropy loss function to optimize:

$$C = -\frac{1}{n}\sum [y_i \ln p_j + (1 - y_i) \ln(1 - p_j)] \qquad (1)$$

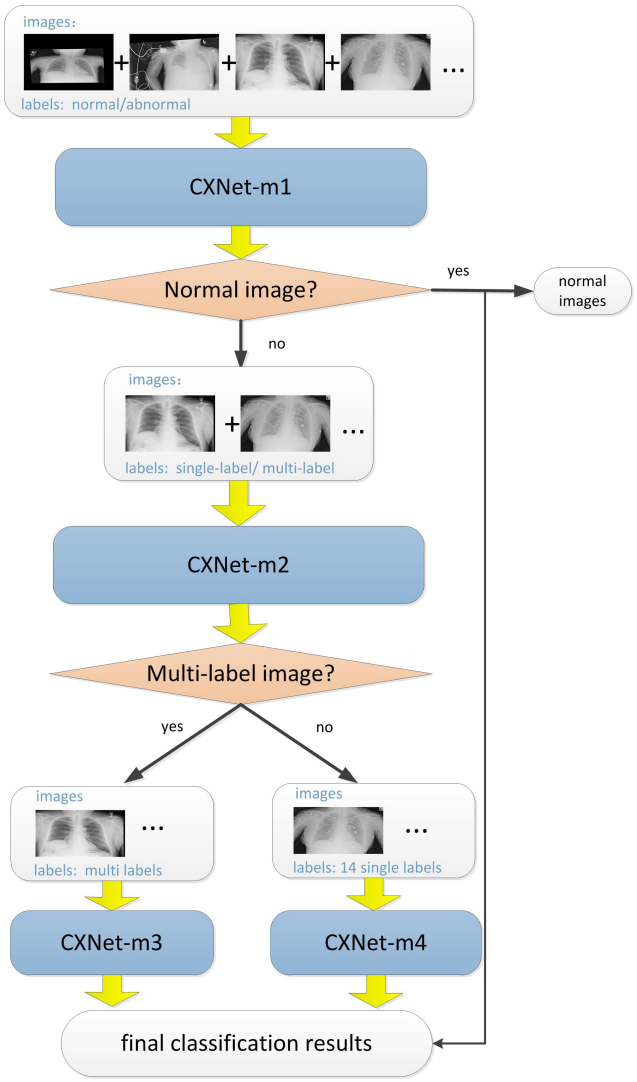

FIGURE 2. Hierarchical model structure to avoid the shortcomings of Chest Xray14 database, there are 4 classifiers to finish different tasks hierarchically.

Where n is the number of training images and $p_j \in [0, 1]$ is defined as :

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{2}$$

Where Z is the input of softmax layer, K is the number of categories and $j \in [0, K-1]$, here K=2. When processing training image I, the output of the last layer P_i is mapped by the network model as (3):

$$p_i = M(I|\theta_f), p_{ij} \in P_j \tag{3}$$

where M is the whole non-linear model and θ_f is the vector of parameters of all layers. The aim of the training is to find out the best parameter combinations through adjusting them to make $C_i(P_i, y_i|\theta_f)$ minimal, as shown in (4):

$$argmin_{\theta_f \in \theta_F} \frac{1}{n} \sum C_i(P_i, y_i|\theta_f) \tag{4}$$

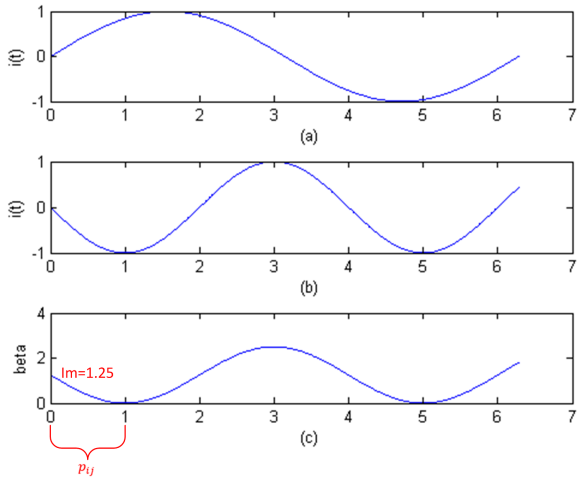

FIGURE 3. The shape of i_t. (a) is the basic shape, (b) is a variant of (a) by setting $\omega = \frac{\pi}{2}$ and $ji_0 = \pi$, (c) is a variant of (b) by adding I_m and $I_m = 1.25$.

Where θ_F is the parameter space and n is the number of images.

It is not easy to differentiate normal and abnormal chest x-ray images especially when the normal images are noisy and disease area of abnormal images are inconspicuous. Therefore, we try to multiply loss by a self-adapting coefficient β_i when processing training image I, as shown in (5):

$$C_i(I, y_i|\theta_f) = \beta_i[y_i \ln p_{ij} + (1 - y_i) \ln(1 - p_{ij})] \tag{5}$$

What we want is a proper β_i whose basic shape is monotone decreasing and convex in the interval of [0,1] and we are inspired by alternating current function. In electromagnetism [49], the instantaneous value of the alternating current is sinusoidal, as shown in

$$i_t = I_m \sin(\omega t + ji_0) \tag{6}$$

Where I_m is amplitude, ω is angular frequency and ji_0 is initial phase. Given $\omega = 1$ and $ji_0 = 0$, the function is shown in Fig. 3(a); Given $\omega = \frac{\pi}{2}$ and $ji_0 = \pi$, the shape of i_t is shown as Fig. 3(b). It can be found that the convex curve decreases from 0 to -1 when $t \in [0, 1]$.

Inspired by alternating current i_t, the curve in Fig. 3(b) can be moved up by stride 1 to get β_i to learn much more from indistinguishable and misclassified chest x-ray images. β_i is defined as (7):

$$\beta_{i1} = I_m \sin(\frac{\pi}{2} p_{ij} + \pi) + I_m \tag{7}$$

Where $p_{ij} \in [0, 1]$ and $I_m = 1.25$ in our paper, and the shape of β_{i1} is shown in Fig. 3(c). In this interval, the curve is convex function. According to the principle that the product of two convex functions is still a convex function, the new cross entropy loss function multiplied by β_i is convex and can avoid local optimum.

The β_i can be further improved according to the comparison of loss shape in Fig. 4. Fig. 4(a) is basic cross entropy

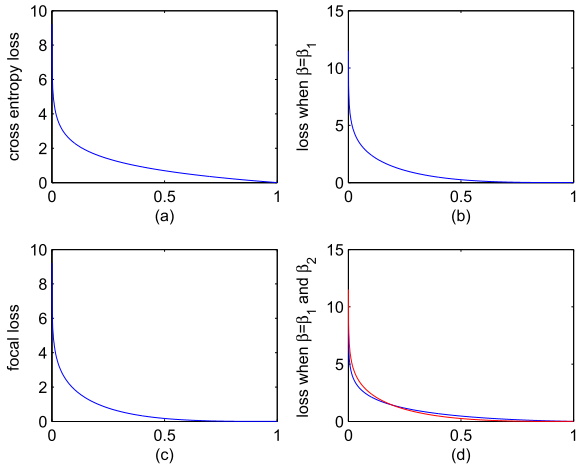

FIGURE 4. Loss curve based on different value of β_i, (a) is the basic cross entropy loss when β_i =1, (b) is a variants of (a) when β_i is set as formula (7), (c) is a kind of focal loss proposed by literature [50], in (d), the blue curve is a variants of (a) when β_i is set as formula (8) and the red one is the one that in (b).

loss when $\beta_i = 1$ and blue curve in Fig. 4(b) is a variant of Fig. 4(a) when β_i is set as formula (7). It can be found that the blue curve in Fig. 4(b) decreases more sharply than Fig. 4(a) when $p_{ij} < 0.5$, which means loss value is larger when p_{ij} is small. And this is the reason why blue curve in Fig. 4(b) can focus on misclassified images to learn. Fig. 4(c) shows a kind of focal loss proposed by literature [50] and the shape of it is similar with the blue curve in Fig. 4(b), which indicates the rationality of our thinking. When $p_{ij} > 0.5$, the value of blue curve in Fig. 4(b) is very small and decreases slowly. However, we also want to learn more from indistinguishable images and the loss value should be larger when p_{ij} is around 0.5. The blue curve in Fig. 4(d), where β_i is defined as formula (8), is eligible because it is larger than the blue curve in Fig. 4(b) when p_{ij} is around 0.5 and smaller than the curve in Fig. 4(a) when $p_{ij} > 0.75$.

$$\beta_{i2} = 1 - I_n p_{ij} \tag{8}$$

Two curves in Fig. 4 (d) cross at point C, where the red curve is the one that in Fig. 4 (b) and $I_n = 0.65$. The optimal combination is the blue curve when p_{ij} is small and the red curve when p_{ij} is large. Therefore, β_i is redefined as formula (9):

$$\beta_i = \begin{cases} I_m \sin(\frac{\pi}{2} p_{ij} + \pi) + I_m, & p_{ij} \le p_c. \\ 1 - I_n p_{ij}, & otherwise. \end{cases} \tag{9}$$

As the result, the sin-loss function is as (10):

$$C_{new}(P_i, y_i|\theta_f) = -\frac{1}{n}\sum \beta_i C_i(P_i, y_i|\theta_f) \tag{10}$$

The parameters θ_f can be updated and converged by optimizing C_{new} with backpropagation as (11):

$$\theta_f = \theta_f - \mu \frac{\partial C_{new}}{\partial \theta_f} \tag{11}$$

After θ_f initialized randomly, $\frac{\partial C_{new}}{\partial \theta_f}$ is acquired through back propagation and then θ_f is updated using gradient descent method, where μ is the learning rate and $\frac{\partial C_{new}}{\partial \theta_f}$ is the gradient of $C_{new}(P_i, y_i|\theta_f)$. In order to find out the best parameters θ_f to make $C_{new}(P_i, y_i|\theta_f)$ minimal, (11) is repeated automatically and both the value of θ_f and $C_{new}(P_i, y_i|\theta_f)$ keep changing till $C_{new}(P_i, y_i|\theta_f)$ converges to the minimum.

We use back propagation to train our networks in this paper. Compared with some classic search algorithms such as PSO [51], ACO [52] and BAS [53] for optimization, back-propagation based algorithm is more proper when training deep learning networks. To get proper weights and biases, back propagation provides approximate partial derivatives of the error related to them while search algorithms lead to the increase of computation by a factor of the population size. There are some novel algorithms taking advantages of both back propagation and search algorithms [54], [55], which we would like to discuss in future works. Running updates (11) makes the network pay more attention to misclassified and indistinguishable images and leads to the emergence of more discriminative features to improve the accuracy rate of classification.

2) MODEL ARCHITECTURE

To train medical images, most papers have tried to fine tune existing CNN networks, such as VGGNet, ResNet and DenseNet [22], [23], [42]. However, transfer learning is not an optimal idea when source domain and target domain are dissimilar with each other. Chest x-ray images in Chest X-ray14 are 1024*1024 gray images, which are totally different from natural images in Imagenet. With the development of computer performance, training a targeted classifier for specific domain or database from scratch is implementable. In this paper, we therefore design a new CNN architecture called CXNet-m1 (**C**hest **X**-ray **N**etwork-**m**odel 1) to train. As shown in Fig. 5, there are only three cascaded blocks (a convolutional layer and a pooling layer after that in each block) to extract image features. The numbers of convolution kernels are 32, 64, 128, respectively. In order to focus on small lesion area and learn much more details, the size of convolution kernels should be small and we set as 3*3. After that, there are a convolutional layer conv-4, a dropout layer drop-1, a batchnorm layer bn-1, a global pooling layer glpool-1 and a convolutional layer conv-5 as the output layer. conv-4, glpool-1 and conv-5 are designed instead of fully connected layer to construct less parameters, and drop-1 and bn-1 are designed to overcome overfitting.

Different from most popular networks, CXNet-m1 consists of only 5 convolutional layers (3 of them extract image features), which is far less than Vgg-net(16 layers), Resnet (50 layers) and Densenet (121 layers) [22], [23], [42]. Besides, it cascades a dropout layer, a batchnorm layer bn-1 and a global pooling between the last 2 convolutional layers in order to overcome overfitting, which is not used in other networks. During the training phase, training images are input

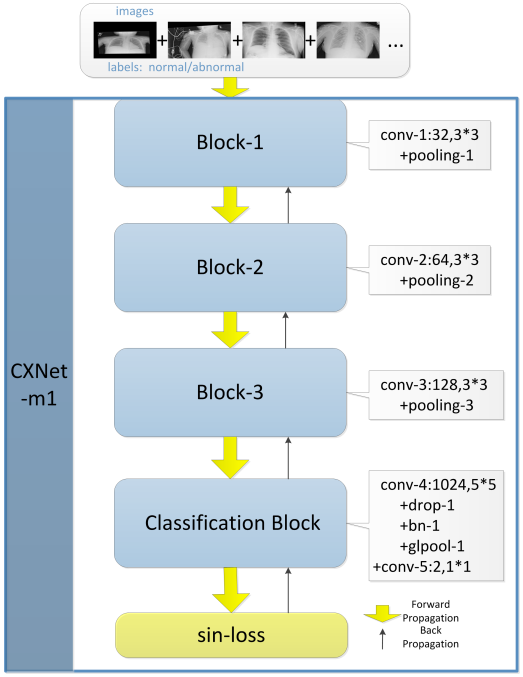

FIGURE 5. CXNet-m1 architecture, there are three convolutional layers to extract image feature and 2 convolutional layers to classify images, between which a dropout layer, a batchnorm layer and a global pooling are cascaded; Yellow arrows and black arrows show the forward propagation and back propagation respectively, optimizing sin-loss function.

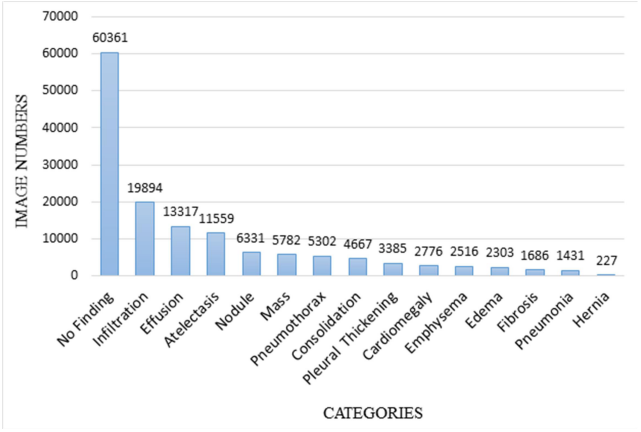

FIGURE 6. Image numbers of 15 categories in ChestX-ray14, from left to right are 60361, 19894, 13317, 11559, 6331, 5782, 5302, 4667, 3385, 2776, 2516, 2303, 1686, 1431 and 227 in turn.

into CXNet-m1 and parameters are updated through back propagation to minimize the sin-loss function value.

IV. EXPERIMENTS

There are five parts in this section to show the experiment result and demonstrate the experiment analysis. The first part is the deep analysis of ChestX-ray14's scale and distribution. The result shown in Fig. 6 and Fig. 7 motivates us to design the hierarchical structure in section III. In the second part, we introduce the metrics to compare experimental results including accuracy rate, precision rate, recall rate, F-measure

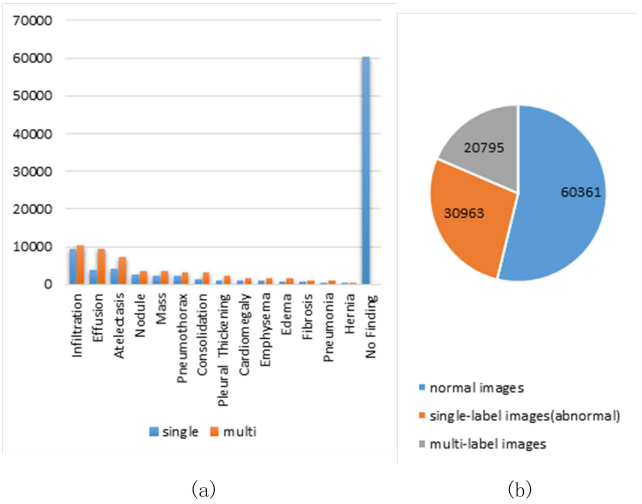

FIGURE 7. Numbers of single-label and multi-label images in ChestX-ray14, in (a), blue bar shows number of single-label images in each category and orange bar shows that of multi-label images; in (b), the pie chart demonstrate the proportion of normal images (blue part), single-label abnormal images (orange part) and multi-label images (gray part) and their corresponding numbers.

value and AUC score. F-measure value is very important because it combines precision rate and recall rate. F-measure value would be high only when both precision rate and recall rate are high. Then we list some CNN settings to show our experiment details. In the forth part, we conduct several experiments to verify the performance of CXNet-m1. Firstly, following the official patient-wise split, we compare the value of above metrics between VGGNet-16, VGGNet-16-DCNN, ResNet-50, ResNet-50-DCNN, Inception-ResNet, Inception-ResNet-DCNN and CXNet-m1. Secondly, after making some cross validations, we decide to adjust the proportion of training data and test data, the input images size and some corresponding modification of CXNet-m1 architecture. CXNet-m1(v0) is the combination of original CXNet-m1 architecture and standard cross entropy loss function, CXNet-m1(v1) is the combination of original CXNet-m1 architecture and sin-loss loss function, CXNet-m1(v2) is the combination of new CXNet-m1 architecture and sin-loss loss function. Then, we analyze the reason of the experiment results in detail in the fifth part. We find that ResNet-50-DCNN achieves the best results among the fine-tuning models, but still worse than CXNet-m1. The experiment results also show the capacity of CXNet-m1 architecture and sin-loss loss function and the importance of more details to learn.

A. DATASET

ChestX-ray14 is the most appropriate database accessible at present to train a chest x-ray image classifier. It contains more than 30,000 patients, 297,541 labeled chest x-ray images and 14 kinds abnormal images including Infiltration, Effusion, Atelectasis, Nodule, Mass, Pneumothorax, Consolidation, Pleural Thickening, Cardiomegaly, Emphysema, Edema, Fibrosis, Pneumonia and Hernia.

TABLE 1. ChestX-ray14 characteristics and corresponding operation to utilize advantages and avoid disadvantages.

ChestX-ray14 Characteristics	Operation and Solution
large-scale	training from scratch
imbalanced	hierarchical structure
multi-labeled	hierarchical structure

TABLE 2. 4 Evaluation results and corresponding symbols.

Symbols	Descriptions
TP	The result is positive and the prediction is true
TN	The result is negative and the prediction is true
FP	The result is positive and the prediction is false
FN	The result is negative and the prediction is false

However, the chest x-ray image numbers of each category in Chest X-ray14 is extremely imbalanced, as shown in Fig. 6. The numbers of images from left to right are 60361, 19894, 13317, 11559, 6331, 5782, 5302, 4667, 3385, 2776, 2516, 2303, 1686, 1431 and 227. The number of normal images, namely 'No finding', is almost as 260 times as the 'hernia' images. What's more, some images are multi-label data, which means there are much more than 15 categories (normal images and 14 kind abnormal images). Figure 7(a) shows the number of single-label and multi-label images in different categories. Figure 7(b) shows the proportion of normal images, single-label images (abnormal) and multi-label images where the number of labels ranges from 2 to 14. Pointing at the drawbacks of ChestX-ray14, we design a hierarchical structure to classify images and CXNet-m1 is the first important part of it to recognize abnormal images.

As summarized in Table 1, The biggest advantage of ChestX-ray14 is that the large scale can support us to train our own model from scratch instead of fine tuning from networks pre-trained by natural images, which are dissimilar with X-ray images. The serious disadvantage of ChestX-ray14 is that the imbalanced sample may lead to wrong learning and bad classifier. Apart from improving loss function and expanding database, our designed hierarchical structure can help to ease the problem. Another Characteristics of ChestX-ray14, multiple labels, is not watched by researchers. This can also be solved by designed hierarchical structure.

B. METRICS

There are 4 kind results of test images and the specific definitions are shown in Table 2. The accuracy rate A is defined as (12) [56]:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

The precision rate P is defined as (13) [56]:

$$P = \frac{TP}{TP + FP} \tag{13}$$

The recall rate R is defined as (14) [56]:

$$R = \frac{TP}{TP + FN} \tag{14}$$

In most cases, the higher the recall rate, the lower the accuracy rate and vice versa. F-measure value is defined to take both P and R into consideration (15) [56]:

$$F = \frac{(\alpha^2 + 1) * P * R}{\alpha^2(P + R)} \tag{15}$$

Where α^2 is weight factor, and when $\alpha^2 = 1$, P and R are equally-weighted.

The AUC is defined as the Area Under the ROC Curve. Obviously, the value of this Area will not be greater than 1. Since ROC curve is generally above the line y=x, the value range of AUC is between 0.5 and 1. Using the AUC value as the evaluation standard is more clear and direct than ROC Curve. As a numeric value, if the AUC is larger, the classifier is better.

C. TRAINING

We trained our network four times on 80458 of 112120 training images (setting aside 6056 for validation and 25596 for test) and 84090 of 112120 training images (setting aside 11212 for validation and 16818 for test) on Chest x-ray14 dataset in the first time and the last three times, respectively. The size of input images were set as 224*224 and 512*512 in the first three times and the last time, respectively. Theoretically, our model can handle images of any size because they are all full convolutional networks. However, the input size is constrained by the scale of GPU memory used to store the outputs of intermediary layers.

The experimental environment was an ubuntu linux server with 2 GeForce GTX 1080 Ti GPUs and the models were implemented using Tensorflow (GPU and ubuntu version) slim framework [57]. We firstly converted the images as TFRecord format, the unique format to input into the slim framework. Then we ignored the preprocessing step, such as clipping, which may cause bad training samples. The networks were trained end-to-end using stochastic gradient descent (SGD) with standard cross entropy loss and our sin-loss. Due to the limit of GPU memory, the batch size = 32 is set as a constant. According to experience and the validation results, we finally chose a proper initial learning rate of 0.01, decayed by a factor of 2 or 5 or 10 manually through monitoring the loss curve in TensorBoard. The networks were validated against the validation set after every 3000 iterations to monitor convergence and overfitting. If the validation results are on the increase and then go down, the inflection point is exactly the timing of early stop. Some CNN settings above are listed in TABLE 3.

D. BINARY CLASSIFICATION RESULTS

To classify images in Chest Xray14, DCNN [1] takes the weights from pre-trained models, replacing fully-connected layers and final classification layers with transition layers and prediction layers. transition layers are for localization and prediction layers are for classification, and both of them are trained from scratch. In order to compare our model CXNet-m1 with DCNN, we perform the same network surgery

TABLE 3. Some detailed settings.

Server	GPU type	GPU number	Framework	Batch size	Optimization method	Initial learning rate	Decay factor	Tricks
Ubuntu linux	GeForce GTX 1080 Ti	2	Tensorflow	32	SGD	0.01	2/5/10	early stop

on the pre-trained models except adding transition layers. Firstly, we split the dataset into 86524 training samples and 25596 test samples, following the previous work on ChestX-ray14. Among training samples, 6056 images are selected as validation set to help monitor the training process, select proper hyper-parameters and determine the timing of early stop. Thus, the proportion of training set, validation set and test set is 70%:7%:23%. Then we train our model CXNet-m1 and compare it with VGGNet-16, ResNet-50, Inception-ResNet and corresponding DCNN versions [1], which are used for binary classificaion here. In order to achieve convergence, we train not only the last prediction layer of ResNet-50-DCNN. Table 4 demonstrates the accuracy rate, precision rate, recall rate, F-measure value and AUC score of different models. The performance varies greatly, in which CXNet-m1 achieves the best results.

Although the performance of CXNet-m1 is much better, it is still not satisfactory because the value of accuracy rate, precision rate, recall rate, F-measure and AUC score are all less than 75%. Reference [58] redistributes training data and test data and could get better results than following the official patient-wise split. In order to check if it is the problem of data split, we randomly split test data into 4 parts on average and then randomly choose 2 parts of them as training data to train several iterations when the training accuracy is already more than 90% and the loss is close to 0. ensuring the data chosen random, we repeat the above operation for 24 times using four models including VGGNet-16-DCNN, ResNet-50-DCNN, Inception-resnet-DCNN and CXNet-m1. Fig. 8 is the result of one of the 6 experiments on VGGNet-16-DCNN and it can be representative because the similar experimental phenomenon happened in other 23 experiments. As shown in Figure 8, training accuracy and loss value are decreased and increased sharply when above models are trained on images from test dataset. It can be inferred from Fig. 8 and other 23 cross validations that CNN models need learn more features from more varied training samples. As a result, Original test images are redistributed randomly and the proportion of training data, validation data and test data is now about 75% : 10%:15%. CXNet-m1 and DCNN-Resnet-50 whose performance in Table 4 are top two are trained again on the new dataset.

Furthermore, in order to learn more information, we adjust the input image size into 512*512 instead of 224*224 and reset corresponding convolution kernels. As shown in 9, the basic architecture is unaltered, the size of convolution kernels are larger and a convolutional layer is added in block 3.

Except for Chest X-ray14, OpenI is currently the largest public database that we can access. Testing on both OpenI and Chest X-ray14 is more convincing than testing on a single one. As what Wang [1] did, we use 2,435 chest x-ray images

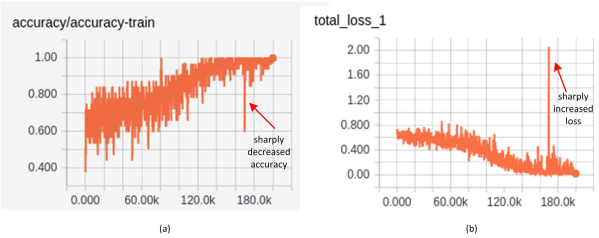

FIGURE 8. Training accuracy and loss curve, in (a), the accuracy is decreased to 0.6 when train the images from test dataset; in (b), the corresponding loss value rise to 2.0, which is much far from 0.

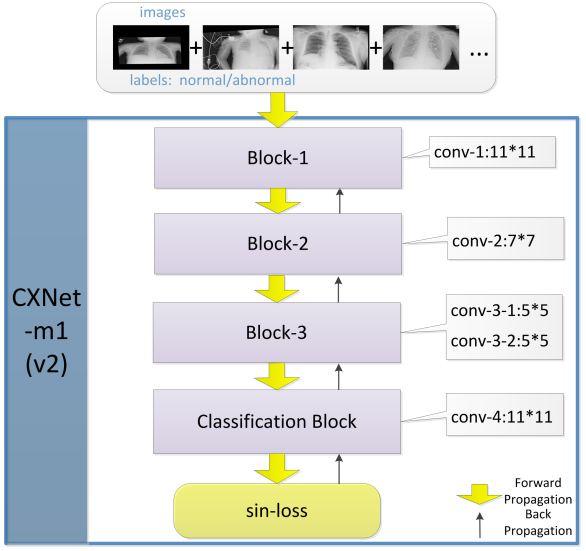

FIGURE 9. CXNet-m1(v2) architecture, there are larger convolution kernels and one more convolutional layer, the size of convolution kernels in block 1 is large in order to learn lower-dimensional feature from higher resolution (512*512) images.

(1379 are normal images and others are abnormal) in OpenI to test CXNet-m1(v1), CXNet-m1(v0) and ResNet-50-DCNN. We did not test CXNet-m1(v2) due to the image sizes in OpenI are not large enough to input into CXNet-m1(v2).

As shown in Table 5, CXNet-m1(v1) is a standard model in Fig. 5, CXNet-m1(v0) has the same architecture of CXNet-m1(v1) but replaces the loss function with standard cross entropy, CXNet-m1(v2) has the same loss function of CXNet-m1(v1) but uses a new architecture of CXNet-m1, as shown in Fig. 9. Table 6 shows the evaluation of binary classification results on redistributed Chest X-ray14. Table 7 shows the evaluation of binary classification results on OpenI.

To further prove the rationality of data split and validity of our model, we check the patient ID of test and training images. Chest X-ray 14 ensures that there is no patient overlap between the splits, but there is a little overlap in our new

TABLE 4. Result evaluation using A, P, R, F1-score and AUC on published data split.

metrics / networks	accuracy rate (%)	precision rate (%)	recall rate (%)	F-measure value (%)	AUC (%)
VGGNet-16	37.9	48.1	12.8	20.2	51.2
VGGNet-16-DCNN	56.4	64.5	64.6	64.6	60.9
ResNet-50	45.9	72.2	19.6	30.8	53.9
ResNet-50-DCNN	66.1	72.5	72.3	72.4	64.2
Inception-ResNet	38.6	**78.4**	0.5	1.0	50.0
Inception-ResNet-DCNN	61.6	65.0	65.1	65.1	61.1
CXNet-m1	**67.6**	73.6	**73.8**	**73.7**	**65.8**

TABLE 5. Three versions of CXNet-m1.

CXNet-m1(v0)	CXNet-m1(v1)	CXNet-m1(v2)
original CXNet-m1 model + cross entropy loss	original CXNet-m1 model + sin-loss	new CXNet-m1 model + sin-loss

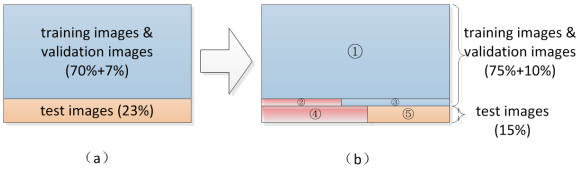

FIGURE 10. The redistribution of training data and test data on Chest X-ray14.

split because we randomly select some test images to train, as shown in Fig. 10. In Fig. 10(a), 77% images are used for training and validation (70%+7%) and 23% images are used for test. In Fig. 10(b), part ①②③ are new training and validation images and ④⑤ are new test images. ② and ④ are chest x-ray images that from the same patients. We select images from ①③ to train and test the model on images from ④. The average test accuracy rate is 90% (>84.4%), which means that the good result owes to proper data split and powerful model, rather than patient overlap.

E. DISCUSSION

Table 4 compares accuracy rate, precision rate, recall rate, F-value and AUC value between VGGNet-16, VGGNet-16-DCNN, ResNet-50, ResNet-50-DCNN, Inception-ResNet, Inception-ResNet-DCNN and CXNet-m1. Among them, VGGNet-16, ResNet-50 and Inception-resnet are pre-trained models on Imagenet and not trained again using chest x-ray image. The performance of these three models are all terrible, which means transfer learning from natural images to chest x-ray images without any surgery is not a good choice. The recall rate of Inception-resnet is only 0.5%, where only 76 abnormal images are classified correctly, other 15663 abnormal images are considered as normal images. The performance of ResNet-50 is better than VGGNet-16 and Inception-resnet, which means ResNet-50 is more suitable here for transfer learning. Compared with them, the accuracy rate, F-value and AUC value of VGGNet-16-DCNN, ResNet-50-DCNN and Inception-resnet-DCNN are much higher, which means discriminatory features can be learned from chest x-ray images through fine tuning pre-trained models. Among them, ResNet-50-DCNN has the best performance. However, CXNet-m1 achieves better results than ResNet-50-DCNN with only 5 convolutional layers, indicating the effectiveness of CXNet-m1. The numeric value in Table 4 demonstrates that CXNet-m1 is more appropriate than fine tuning from existing networks. It is thinner, lighter but more capable than classic networks. Although getting the second high value in Table 4, ResNet-50-DCNN holds far more layers and weights, resulting in the waste of time and space.

Training data and test data are redistributed to learn more information and Table 6 is formed to show the performance of ResNet-50-DCNN, CXNet-m1(v0), CXNet-m1(v1) and CXNet-m1(v2), which are trained and tested again on the new data split. Although achieving the highest precision rate, the accuracy rate, recall rate and F-value of ResNet-50-DCNN is much lower than the other three models. The recall rate, accuracy rate and F-value of CXNet-m1(v1) are higher than those of CXNet-m1(v0), which means the sin-loss function is better than standard cross entropy loss function based on the same model architecture. Using the self-adapting factor β, sin-loss function guides the training process to learn more indistinguishable information from misclassified images and it improves the performance of CXNet-m1. CXNet-m1(v2) gets the highest accuracy rate, recall rate, F value and AUC score. Compared with CXNet-m1(v1), it keeps the highest recall rate and improves the precision rate and other metrics value, which means learning more details really benefits the final classification results. In a word, the performance of CXNet-m1(v1) illustrates the capacity of sin-loss function, the performance of CXNet-m1(v2) demonstrates the importance of learning details and the performance of CXNet-m1(v0), CXNet-m1(v1) and CXNet-m1(v2) shows that CXNet-m1 can achieve the best performance with less parameters and less layers.

To make the conclusion more convincing, we test CXNet-m1(v1), CXNet-m1(v0) and ResNet-50-DCNN on OpenI and Table 7 is therefore formed. The performance of both versions of CXNet-m1 are better than ResNet-50-DCNN in general and CXNet-m1(v1) achieves higher accuracy rate, recall rate, F-value and AUC. The results further prove that CXNet-m1 is good at classifying chest x-ray images and sin-loss improves the performance of CXNet-m1(v0).

TABLE 6. Result evaluation using A, P, R, F1-score and AUC on new data split.

metrics / networks	accuracy rate (%)	precision rate (%)	recall rate (%)	F-measure value (%)	AUC (%)
ResNet-50-DCNN	76.3	**91.2**	72.8	81.0	78.5
CXNet-m1(v0)	79.4	86.8	82.9	86.7	77.3
CXNet-m1(v1)	81.7	83.1	**92.3**	86.9	75.2
CXNet-m1(v2)	**84.4**	86.2	**92.3**	**89.1**	**79.5**

TABLE 7. Result evaluation using A, P, R, F1-score and AUC on OpenI.

metrics / networks	accuracy rate (%)	precision rate (%)	recall rate (%)	F-measure value (%)	AUC (%)
ResNet-50-DCNN	90.5	90.0	87.8	88.9	87.1
CXNet-m1(v0)	93.1	**93.9**	89.9	91.9	85.9
CXNet-m1(v1)	**93.6**	92.4	**93.0**	**92.7**	**87.3**

V. CONCLUSION

Chest X-ray is the most popular mean to detect lung lesion and deep learning is a good tool to assist the diagnosis. For classification tasks of chest X-ray images, it is promising to fine tune existing deep networks due to limits of data size, labeling and computer hardware. However, it may lead to low transfer efficiency, overfitting and other problems when chest X-ray dataset and source dataset are totally different. To avoid these problems, we analyze chest X-ray14 database, design a hierarchical classification structure and present a newly-designed convolutional neural network CXNet-m1 to detect anomaly of chest X-ray images. In addition, we also propose a novel loss function sin-loss to promote the performance of CXNet-m1. Furthermore, we slightly adjust the CXNet-m1 architecture to learn more details and extract more discriminative information. The experiment result shows that CXNet-m1, which contains less layers and parameters, can achieve better accuracy than fine-tuning, no matter with or without sin-loss. It also demonstrates that CXNet-m1 with sin-loss can learn more useful information and therefore acquires better performance. Besides, it indicates that training images with higher resolution and larger number can promote the classification performance. In a word, the key of good result is making every effort to learn more useful and accurate features, no matter through designing a more proper CNN or utilizing more data.

In the future, we will continue to explore other three models, CXNet-m2, CXNet-m3 and CXNet-m4. According to different classification tasks, models and loss functions should be designed respectively. For instance, CXNet-m3 may use LSTM to learn the relationship between multiple labels and CXNet-m3 should learn more features for 14 categories. CNN is a promising method of image process and the loss function and architecture are still the emphasis of our further research.

ACKNOWLEDGMENT

The authors particularly appreciate Dr. Chao Zhang, Mr. Zhiyuan Cao and Mr. Liangchi Li for their generous help in the completion of this paper. Dr. Chao Zhang give them many instructive advise and useful suggestions on the draft, and Mr. Zhiyuan Cao and Mr. Liangchi Li offered lots of technical support to complete the experiment.

REFERENCES

[1] X. Wang, Y. Peng, Z. Lu, M. Bagheri, R. M. Summers, and L. Lu, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.

[2] S. I. Kamel, D. C. Levin, V. M. Rao, and L. Parker, "Utilization trends in noncardiac thoracic imaging, 2002–2014," *J. Amer. College Radiol.*, vol. 14, no. 3, pp. 337–342, 2017.

[3] H. Zhang, "The optimality of Naive Bayes," *AA*, vol. 1, no. 2, p. 3, 2004.

[4] N. B. Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs decision trees in intrusion detection systems," in *Proc. ACM Symp. Appl. Comput.*, 2004, pp. 420–424.

[5] W. Chen, X. Yan, Z. Zhao, H. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, Naive Bayes and RBFNetwork models for the Long County area (China)," *Bull. Eng. Geol. Environ.*, vol. 77, pp. 1–20, Mar. 2018.

[6] R. Eskandarpour and A. Khodaei, "Leveraging accuracy-uncertainty tradeoff in SVM to achieve highly accurate outage predictions," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1139–1141, Jan. 2018.

[7] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *Int. J. Inf. Technol.*, vol. 2, pp. 1–11, Jan. 2018.

[8] O. L. Mangasarian and D. R. Musicant, "Lagrangian support vector machines," *J. Mach. Learn. Res.*, vol. 1, pp. 161–177, Mar. 2001.

[9] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, 2010.

[10] D. Niu *et al.*, "Sustainability evaluation of power grid construction projects using improved TOPSIS and least square support vector machine with modified fly optimization algorithm," *Sustainability*, vol. 10, no. 1, p. 231, 2018.

[11] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, May 2007.

[12] Z. Liu, W. Guo, W. Ma, and J. Hu, "A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM," *ISA Trans.*, 2017, vol. 66, pp. 249–261.

[13] Y. Zhang, G. Cao, B. Wang, and X. Li, "Cascaded random forest for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1082–1094, Apr. 2018.

[14] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, "Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model," *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1081–1091, May 2018.

[15] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," *Statist. Comput.*, vol. 27, no. 3, pp. 659–678, 2017.

[16] Z. Hu, J. Tang, B. P. Patlolla, and P. Zhang, "Identification of bruised apples using a 3-D multi-order local binary patterns based feature extraction algorithm," *IEEE Access*, vol. 6, pp. 34846–34862, 2018.
[17] Z. Xiang, H. Tan, and W. Ye, "The excellent properties of a dense grid-based HOG feature on face recognition compared to Gabor and LBP," *IEEE Access*, vol. 6, pp. 29306–29319, 2018.
[18] Y. Bar, I. Diamant, H. Greenspan, and L. Wolf, "Deep learning with non-medical training used for chest pathology identification," *Proc. SPIE*, vol. 9414, p. 94140V, Mar. 2015.
[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[20] H. C. Shin *et al.*, "Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2497–2506.
[21] R. K. Sevakula *et al.*, "Transfer learning for molecular cancer classification using deep neural networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published.
[22] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[23] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. CVPR*, 2015, pp. 1–9.
[24] K. He, X. Zhang, J. Sun, and S. Ren, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Jun. 2016, pp. 770–778.
[25] P. Rajpurkar *et al.* (Dec. 2017). "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning." [Online]. Available: https://arxiv.org/abs/1711.05225
[26] L. Yao, E. Poblenz, B. Covington, D. Bernard, K. Lyman, and D. Dagunts. (Oct. 2017). "Learning to diagnose from scratch by exploiting dependencies among labels." [Online]. Available: https://arxiv.org/abs/1710.10501
[27] J. M. Haut, M. E. Paoletti, J. Li, and J. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.
[28] W. Xiong *et al.*, "The Microsoft 2016 conversational speech recognition system," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2018, pp. 5934–5938.
[29] D. Chen and Y. Zhang, "Robust zeroing neural-dynamics and its time-varying disturbances suppression model applied to mobile robot manipulators," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4385–4397, Sep. 2018.
[30] D. Chen, Y. Zhang, and S. Li, "Tracking control of robot manipulators with unknown models: A jacobian-matrix-adaption method," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3044–3053, Jul. 2018.
[31] D. Chen and Y. Zhang, "A hybrid multi-objective scheme applied to redundant robot manipulators," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1337–1350, Jul. 2017.
[32] T. Sun, B. Zhou, J. Pei, and L. Lai, "Sequence-based prediction of protein protein interaction using a deep-learning algorithm," *BMC Bioinf.*, vol. 18, no. 1, p. 277, 2017.
[33] J. Saltz *et al.*, "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images," *Cell Rep.*, vol. 23, no. 1, pp. 181–193, 2018.
[34] Y.-G. Jiang, Z. Wu, Z. Li, X. Xue, S.-F. Chang, and J. Tang, "Modeling multimodal clues in a hybrid deep learning framework for video classification," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3137–3147, Nov. 2018.
[35] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[36] L. Rosasco, E. Vito, A. Caponnetto, M. Piana, and A. Verri, "Are loss functions all the same?" *Neural Comput.*, vol. 16, no. 5, pp. 1063–1076, May 2004.
[37] F. Xia, T.-Y. Liu, W. Zhang, H. Li, and J. Wang, "Listwise approach to learning to rank: Theory and algorithm," in *Proc. 25th ACM Int. Conf. Mach. Learn.*, 2008, pp. 1192–1199.
[38] L. Bottou, "Stochastic gradient descent tricks," *Neural Networks, Tricks of the Trade, Reloaded*. Berlin, Germany: Springer, 2012, pp. 421–436.
[39] R. Rastogi and P. Saigal, "Tree-based localized fuzzy twin support vector clustering with square loss function," *Appl. Intell.*, vol. 47, no. 1, pp. 96–113, 2017.
[40] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[41] K. A. Wilmes, J. H. Schleimer, and S. Schreiber, "Spike-timing dependent inhibitory plasticity to learn a selective gating of backpropagating action potentials," *Eur. J. Neurosci.*, vol. 45, no. 8, pp. 1032–1043, 2017.
[42] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, no. 2, 2017, p. 3.
[43] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[44] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[45] D. Zhang, L. Yang, X. Xu, J. Han, and D. Meng, "Spftn: A self-paced fine-tuning network for segmenting objects in weakly labelled videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 4429–4437.
[46] W. Zhou *et al.*, "Transferring visual knowledge for a robust road environment perception in intelligent vehicles," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
[47] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in *Proc. 23rd Int. Conf. IEEE Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1195–1200.
[48] E. Kim, M. Corte-Real, and Z. Baloch, "A deep semantic mobile application for thyroid cytopathology," *Proc. SPIE*, vol. 9789, p. 97890A, Apr. 2016.
[49] J.-M. Jin, *Theory and Computation of Electromagnetic Fields*. Hoboken, NJ, USA: Wiley, 2011.
[50] M. Dorigo and B. Birattari, "Ant colony optimization," *Scholarpedia*, vol. 2, no. 3, p. 1461, 2007.
[51] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, Jun. 2007.
[52] X. Jiang and S. Li. (Oct. 2017). "BAS: Beetle antennae search algorithm for optimization problems." [Online]. Available: https://arxiv.org/abs/1710.10724
[53] T.-Y. Lin, P. Goyal, K. He, P. Dollár, and R. Girshick. (Aug. 2017). "Focal loss for dense object detection." [Online]. Available: https://arxiv.org/abs/1708.02002
[54] F. Gaxiola, P. Melin, F. Valdez, J. R. Castro, and O. Castillo, "Optimization of type-2 fuzzy weights in backpropagation learning for neural networks using GAs and PSO," *Appl. Soft Comput.*, vol. 38, pp. 860–871, Jan. 2016.
[55] L. Zhang, Z. Sun, F. Dong, P. Wei, and C. Zhang, "Numerical investigation of the dynamic responses of long-span bridges with consideration of the random traffic flow based on the intelligent ACO-BPNN model," *IEEE Access*, vol. 6, pp. 28520–28529, 2018.
[56] S.-J. Yen and Y.-S. Lee, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5718–5727, 2009.
[57] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16. 2016, pp. 265–283.
[58] S. Guendel *et al.* (Mar. 2018). "Learning to recognize abnormalities in chest X-rays with location-aware dense networks." [Online]. Available: https://arxiv.org/abs/1803.04565

SHUAIJING XU received the B.S. degree from Beijing Normal University, Beijing, China, in 2015, where she is currently pursuing the Ph.D. degree with the College of Information Science and Technology. Her major is computer application technology. Her major interests are big data, computer vision, and deep learning.

HAO WU received the B.E. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively. From 2013 to 2015, he was a Research Associate with Lawrence Berkeley National Laboratory. He is currently a Post-Doctoral Research Fellow with the College of Information Science and Technology, Beijing Normal University. Until now, he takes charge of some related research projects at the Lawrence Berkeley National Laboratory. He joined the Center for Big Data Mining and Knowledge Engineering, in 2015. He has published about 20 papers in international journals or magazines, such as *Neurocomputing*, *Visual Computer*, *Multimedia Tools and Applications*, *Journal of Visual Communication and Image Representation*, and *IET Computer Vision*.

RONGFANG BIE received the M.S. and Ph.D. degrees from the College of Information Science and Technology, Beijing Normal University, in 1993 and 1996, respectively. In 2003, she was a Visiting Faculty Member with the Computer Laboratory, University of Cambridge. She is currently a Professor with the College of Information Science and Technology, Beijing Normal University. She has authored or co-authored more than 100 papers. Her current research interests include knowledge representation and acquisition for the Internet of Things, dynamic spectrum allocation, and big data analysis and application.