IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 15, 2020, accepted April 28, 2020, date of publication May 21, 2020, date of current version June 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2996217

Cxnet-M3: A Deep Quintuplet Network for Multi-Lesion Classification in Chest X-Ray Images Via Multi-Label Supervision

SHUAIJING XU, XIAOYILEI YANG, JUNQI GUO, (Member, IEEE), HAO WU, GUANGZHI ZHANG, AND RONGFANG BIE, (Member, IEEE)
School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

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

This work was supported by the National Natural Science Foundation of China under Grant 61571049, Grant 61601033, and Grant 61977006.

ABSTRACT Medical image analysis is motivated by the success of deep learning, where annotations are usually expensive and not easy to obtain. In this paper, we propose a deep quintuplet network CXNet-m3, where the classification of lesion type of chest x-ray images (CXRs) could benefit from easily accessible annotations like patient age, gender, identity and view position. To improve classification performance, a novel loss function combining both deep metric learning and deep learning is first designed based on multiple labels. Then, a deep model based on transfer learning is built to optimize the loss function. To solve the problem of slow convergence, a quintuplet mining algorithm is presented to provide valuable training samples for the proposed classification model. The experimental results on Chest X-ray14 database show that our classification method outperforms some state-of-art models under Area Under Curve (AUC) score, reaching 0.824 on an average. Besides, our proposal achieves more than 0.9 AUC values in the case of Infiltration, Atelectasis, Cardiomegaly and Nodule.

INDEX TERMS Medical image, chest X-ray image classification, deep neural network, deep metric learning

I. INTRODUCTION

Many chest lesions such as nodules and emphysema are early manifestations of lung cancer, the leading cause of death in the world [1]–[3]. Some lesions shown on chest X-ray images (CXRs) are also useful biomarkers associated with severe heart failure and respiratory diseases [4]–[6]. Therefore, diagnosing chest lesions is essential for reducing morbidity and mortality from lung, heart and respiratory diseases.

Chest X-ray is the most commonly used radiology exam for screening and diagnosing chest lesions. With growing population and increasing health awareness, demand for chest readings is growing. In the United States of America (USA) alone, over 35 million CXRs are taken every year and radiologists have to read more than 100 CXRs in a day [7]. Meanwhile, manual method has problems with providing expert readings and correct diagnosis for CXRs. According to a report, within 12 months, up to 23,000 CXRs were not formally reviewed by radiologists at Queen Alexandra Hospital alone [8]. Therefore, advanced technologies are urgently needed to assist radiologists, improve the work efficiency and enhance the diagnosis accuracy.

With the development of computer computing power and the advent of the era of big data, deep learning (DL) technology based on artificial neural networks has been a great success in many fields including image processing [9]–[12]. Compared with traditional machine learning methods such as support vector machine, K-nearest neighbor method and random forest, deep learning method does not have to manually extract image features including Local Binary Pattern, Histogram of Oriented Gradient and Haar-like [13]–[17]. In contrast, deep convolutional neural networks (CNNs) obtain multiple levels of image features automatically by end-to-end training [18]–[20]. The local connectivity and shared weights make CNN to be the leading computational intelligence for image processing and classification. However, the training of CNNs should be supervised by expert annotations, generally expensive and not easy to obtain in the field of medical

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiwei Gao.

imaging. Therefore, the value of labels like patient gender and identity, easy to obtain but usually ignored by researchers, should also be explored.

In this paper, a classification model, CXNet-m3, is proposed based on convolutional neural network to provide auxiliary diagnosis for CXRs in ChestX-ray14 database. In CXNet-m3, the classification of lesion types benefits from easily accessible annotations like patient age, gender, identity and view position. Taking advantage of transfer learning, CXNet-m3 is built with quintuplet inputs and trained by both classification losses and embedding distances between quadruplets of CXRs. These quintuplets are generated using quintuplet mining (QM) algorithm, where quintuplets are also filtered based on information from multiple labels. Therefore, the main contributions in this paper are listed as follows:

1) A novel idea of using easily available label information is proposed to improve the classification performance of CNN model for lesion types.

2) A novel loss function with the help of deep metric learning for classification is proposed to make use of multiple label information.

3) A quintuplet mining algorithm based on multiple labels is presented to provide valuable training samples for the proposed classification model.

II. RELATED WORK

The recent success of deep learning in image processing tasks has led to rich applications in medical image field including the classification of chest x-ray images. Using 5,232 CXRs as training set, Yadav *et al.* trained a linear support vector machine, a fine-tuned convolutional neural network, and a capsule neural network to classify CXRs into bacterial pneumonia, viral pneumonia and disease-free CXRs [21], [22]. Experimental results prove that deep learning algorithms are superior to traditional machine learning algorithm, and the fine-tuning method is better than training from scratch. Lakhani *et al.* took a dataset containing 1,007 chest X-ray images as the research object, and classified tuberculosis based on convolutional neural networks including AlexNet and GoogLeNet [23]–[25]. Among them, AlexNet won the championship in the ImageNet image classification competition in 2012, far exceeding the second place. GoogLeNet introduced inception module to improve the expressive ability of CNN without increasing the amount of calculation. Shin *et al.* used a recurrent neural network (RNN) model to read chest X-ray images and conducted experiments on the open database OpenI containing 3,955 radiological reports [26].Kieu *et al.* proposed a multi-CNN model combining with fusion rules to detect abnormal chest radiographs [27]. Anavi *et al.* used age and gender to visualize patients and improve deep learning frameworks for chest X-ray image retrieval [28]. Although the above researches have achieved good results, the study of deep learning methods in the field of chest X-ray imaging is still restricted because of the limited scale of dataset.

In 2017, National Institutes of Health (NIH) of USA released one of the world's largest accessible labeled chest X-ray image archive, ChestX-ray14. ChestX-ray14 contains 112,120 chest x-ray images from 30,805 patients [29]. Due to the large scale, it triggers a considerable attention in deep learning community. Xu *et al.* trained a two-class deep model from scratch to detect abnormal chest radiographs in ChestX-Ray14 [30]. Yao *et al.* and Xu *et al.* made use of image features and dependencies between labels by combining CNN and RNN to detect multiple lesions in a single image [31], [32]. Most scholars conducted researches to classify CXRs in ChestX-ray14 into 14 kind of lesion types and most of them only used the label of lesion type to supervise the training process of convolutional neural network. X. Wang *et al.* fine-tuned four standard CNN architectures including AlexNet, VGGNet, GoogLeNet and ResNet [24], [25], [29], [33], [34]. Compared to AlexNet, the small convolutional kernels and stacked convolutional layers are two improvements of VGGNet. Such a design improves the ability of extracting features, reduces network complexity, and facilitates training convergence. To solve the problem of vanishing gradient caused by deep layers, ResNet connects a skip connection between the input and output of two stacked convolutional layers, which also reduces the time complexity of training. Among these four classic CNN models, ResNet achieved the best result for multi-class classification in the research work of X.Wang *et al.*. Li *et al.* presented a model for ChestX-ray14 that simultaneously performed the classification of lesion type and the localization of lesion based on Resnet and a simple recognition network [35]. P. Rajpurkar *et al.* utilized a 121-layer DenseNet architecture with little modification to detect pneumonia using ChestX-ray14 [36], [37]. Compared with Resnet, DenseNet further establishes a skip connection not only between the residual blocks, but also between each layer. Yao *et al.* introduced an architecture that learned at multiple resolutions and used a learnable lower bound adaptability to parameterize the pooling function. They achieved satisfactory classification and recognition results for up to 9 lesion types while generating high-resolution saliency maps [38]. Aviles *et al.* proposed a graph-based semi-supervised learning method for chest X-ray image classification. They introduced a new loss function to strengthen the synergy between a limited number of labels and a large amount of unlabeled data. They obtained good results on the Chest X-ray14 database while greatly reducing the need for annotated data [39]. Different from their method, Baltruschat *et al.* tried to use as many label information as possible. Their research is currently the only one that uses view position, patient gender, and patient age information besides image information to train classification model for Chest X-ray14 [13]. They abstracted the three label into a 3-dimensional feature vector and concatenated it with 2048-dimensional feature vector of training image. Although excellent work, low-dimensional non-image features may be hard to play a really powerful role after being concatenated with high-dimensional image features. However, the success

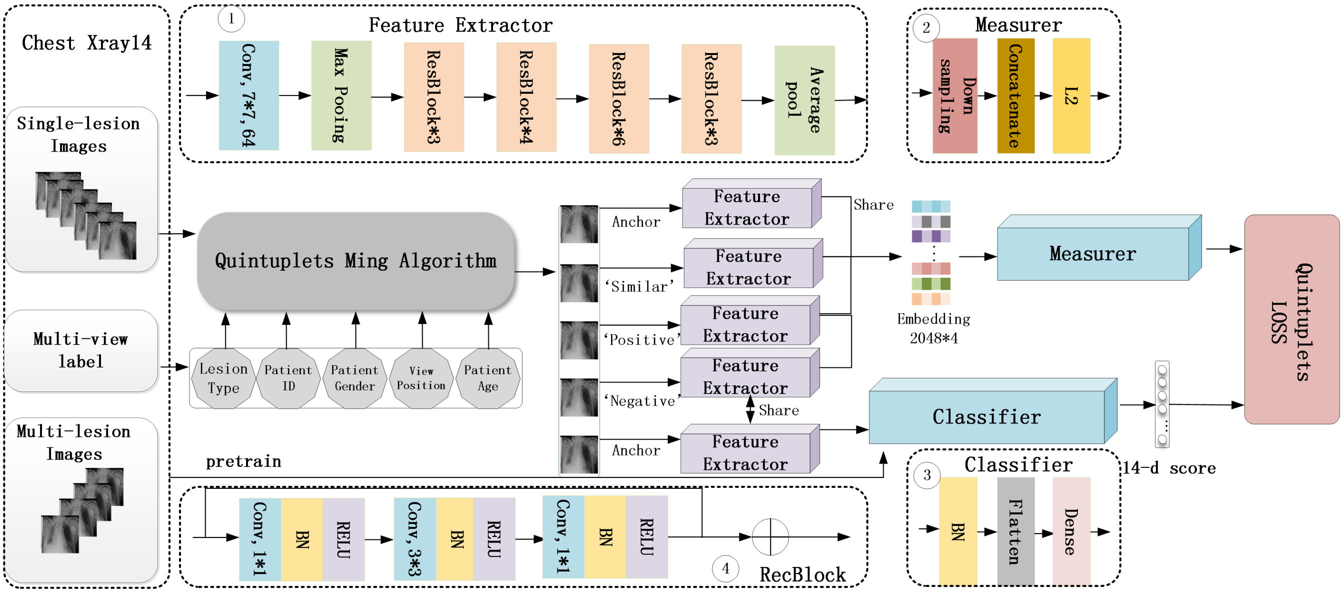


FIGURE 1. The outline of our overall method for the classification of CXR lesions. Quintuplets are first mined from Chest X-ray 14 database, then 5 2048-dimensional feature vectors are extracted by parameter-shared feature extractors. Four of them are sent to the measurer and the rest one is sent to the classifier. ①, ②, ③ and ④ shows the details of Extractor, Measurer, Classifier and RecBlock. The whole model is optimised by the proposed quintuplet loss during training.

of their method has revealed that patient information and view position information, usually ignored by researchers, are also very useful. Inspired by their work, we propose a multi-label supervision method that makes full use of non-image information in this paper. Different from the research work of Li *et al.*, Our model is trained without the supervision of lesion location, expensive and difficult to obtain. In ChestX-ray14 database, only 0.8% of chest X-ray images were labeled by lesion location. we are committed to dig out more value from other labels, easily accessible but usually ignored by researchers. Rather than transforming non-image information into features like what Baltruschat did, we take advantage of them by combination of deep learning and deep metric learning to implement the supervision of multiple labels.

Compared with classic metric learning, deep metric learning can make non-linear mapping of input features, and has been widely used in the field of computer vision, such as image clustering and image retrieval [40]–[45]. Deep metric learning learns the mapping of samples to features through a loss function. Under this mapping, the metric between features can reflect the degree of similarity between samples. Taking the advantages of feature extraction of deep learning, contrastive loss mapped the original input space to Euclidean space, directly constraining the feature distance of samples [46]. Triplet loss further considered the relative relationship between intra-class pairs and inter-class pairs [47]. By optimizing triplet loss, the distance between features of intra-class (anchor-negative) is longer than that between features of intra-class (anchor-positive). Triplet loss has a good performance for extreme classification tasks such as face recognition and person retrieval [48], [49]. In this paper, the thought of triplet loss of constructing positive and negative sample pairs is transferred and improved to classify CXRs based on multi-label information.

It can be found that most of the above medical image classification researches make use of CNN by fine-tuning the existing deep learning models such as AlexNet, VGGNet, GoogLeNet, ResNet and DenseNet. Fine-tuning is a kind of transfer learning method, proposed to overcome problems caused by training models directly on relatively small-scale dataset, such as over-fitting and poor robustness [50]–[53]. In the field of medical images, datasets as large as Imagenet are very difficult to obtain because of expensive expert annotations [24]. Therefore, despite of the success of natural image processing, the performance of deep CNNs trained directly on medical images is limited. Transfer learning solves to some extent the contradiction between the use of deep learning methods and limited-scale medical data set. Except for the research of Xu *et al.* of training a two-classifier from scratch, above research teams trained classification models for ChestX-ray14 by transferring parameters trained on ImageNet [13], [29], [30], [35]–[37]. These researches prove the effectiveness of deep transfer learning from natural domain to CXRs. Different from them, this paper involves not only the transfer between different domains, but also the transfer between the same domain.

III. PROPOSED CXNet-m3

The outline of our overall method for the classification of CXR lesion types is shown in Fig.1. Taking advantage of multiple labels, including lesion type, patient identity (ID), gender, age and view position, the quintuplet mining algorithm is first presented to mine quintuplets from

ChestX-ray14 database. After being extracted by parameter-shared feature extractors, four feature vectors are sent to the measurer and the rest one is sent to the classifier. The initial parameters of the feature extractor are transferred from ImageNet, and the parameters of classifier are pre-trained before the formal training. The whole model is optimised by the proposed quintuplet loss during training. Proposed loss function and mining algorithm are discussed in the first subsection and the model architecture is described in the seconde subsection.

A. PROBLEM FORMULATION

1) CLASSIFICATION LOSS

To aid diagnosis, deep learning can be used to train an end-to-end multi-lesion classification model. Each input of the model is a chest X-ray image I and the output is K-dimensional predictions, where K is the number of lesion types. The location of the largest probability value ranging from 0 to 1 represents the type of lesion predicted by the model. In order to prevent local optimization, the softmax-based cross-entropy loss function is used to optimize model parameters, as shown in (1):

$$C_{cla} = -\frac{1}{n}\sum y_i \ln p_i \tag{1}$$

where n is the number of training images, y_i is lesion type label, and $p_i \in [0, 1]$ is defined as (2):

$$p_i = \frac{e^{z_i}}{\sum_{k=0}^{(K-1)} e^{z_k}} \tag{2}$$

where $z_i \in Z$ is the input of softmax layer. We use M to donate the whole non-linear model and θ_f to donate the parameter vector of M. The aim of the training of M is to find out the best parameter combinations in the parameter space θ_F through optimising C_{cla}, as shown in (3):

$$argmin_{\theta_f \in \theta_F} \frac{1}{n} \sum C_i(M(I|\theta_f), y_i) \tag{3}$$

where *argmin* means "make it minimal", n is the number of training chest x-ray images,I donates a training image, y_i is the label of image I and C_i is the loss of image I, where $C_{cla} = \sum C_i(M(I|\theta_f), y_i)$.

2) DML-BASED LOSS

a: LESION-WISE LOSS

Deep metric learning (DML) implements classification by optimizing the distance of features in the embedding layer. Among them, triplet loss is widely used in the field of face detection, such as FaceNet [47]. FaceNet sets two face pictures belonging to the same person as anchor sample *anc* and positive sample *pos*, and sets pictures not belonging to this person as negative sample *neg*. The idea of triplet loss is that $d(anc, pos)$, the distance between *anc* and *pos* should be less than $d(anc, neg)$, the distance between *anc* and *neg*. Guided by this thought, deep model can be trained by optimizing triplet loss function as shown in (4).

$$C_{tri} = \sum_{j=1}^{T} max(d(anc, pos) - d(anc, neg) + margin, 0) \tag{4}$$

The idea of triplet loss can be transferred to classification task of chest radiographs. Assume that there is an ideal model that can correctly classify all the chest radiographs. This model should be able to accurately capture the discriminative features of each lesion. Therefore, the extracted features between CXRs belonging to the same lesion should be highly similar, while the extracted features between CXRs belonging to different lesion types should be highly different. It can be set as an optimization goal to train the parameters of the multi-lesion classification model we build, just like (4). The only difference is that *anc*, *neg* and *pos* are selected based on the type of lesion rather than the person's ID.

b: PATIENT-WISE LOSS

In above subsections, model is supposed to be optimized from the supervision of lesion labels. In fact, patient ID is also a kind of important label because medical images from the same patient are possibly more similar. Therefore, patient-wise split is often required to construct training, validation and test data sets. Such requirements eliminate the bad effects that the model's performance may be biased by seeing images of the same patient ID from different subsets.

However, instead of avoiding the problems caused by the same patient ID, it is better to utilise the patient ID for optimizing. Ideally, suppose that there are some chest radiographs containing lesion type x and a 100% accurate multi-lesion classification model. The model extracts powerful features and accurately classify them into the same class x. In this case, patient-wise differences are no longer obvious and it can be set as a part of optimization goal to train the classification model. We set up a similar sample *sim* with the same lesion type and patient ID as the anchor sample *anc*, and a positive sample *pos* with the same lesion but different patient ID as the anchor sample *anc*. The distance between *anc* and *pos*, $d(anc, pos)$ should be close to the distance between *anc* and *sim*, $d(anc, sim)$. Guided by this thought, the following loss function is formulated, shown as(5).

$$C_{pti} = \sum_{j=1}^{T} |d(anc, sim) - d(anc, pos)| \tag{5}$$

Taking advantage of lesion-wise loss and patient-wise loss, a DML-based loss is constructed as (6):

$$C_{dml} = \beta * C_{tri} + (1 - \beta) * C_{pat} \tag{6}$$

As shown in Figure 2, this DML-based loss minimizes the distance between an anchor and a positive CXR, and maximizes the distance between the anchor and a negative CXR. At the same time, this DML-based loss makes the distance between the anchor and positive CXR and the distance between the anchor and similar CXR as close as possible.

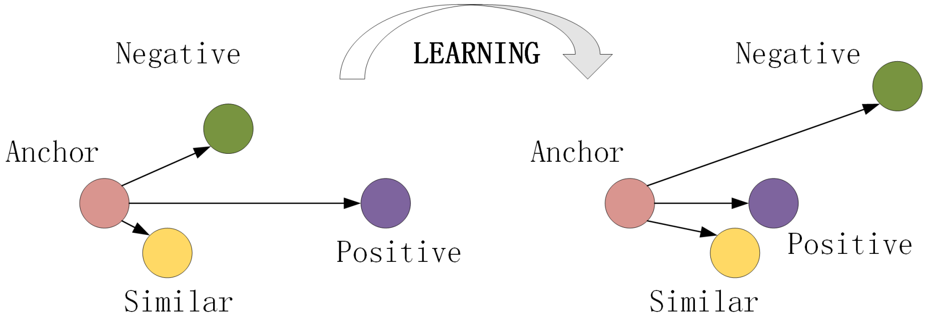


FIGURE 2. The changes of distances between CXRs through proposed metric learning.

3) QUINTUPLET LOSS

In summary, C_{cla} is a direct classification loss function, while C_{dml} provides an optimization target for the multi-lesion model from lesion type and patient ID. Our strategy is to perform a weighted sum of C_{dml} and C_{cla}. Although metric learning is not main task, the optimization of C_{dml} could guide the model to learn discriminative features for classification. Taking advantage of classification loss and DML-based loss, the Quintuplet loss is formulated as (7):

$$C_{qui} = \alpha * C_{cla} + (1 - \alpha) * C_{dml} \tag{7}$$

According to the loss function, every 5 CXRs including anchor sample * 2, a positive sample * 1, a negative sample * 1 and a similar sample*1 need to be put in model each time to train. And this is why formula (7) is called Quintuplet loss.

4) QUINTUPLET MINING ALGORITHM

In addition to lesion type and patient ID, there are also patient gender label ∈ {0, 1}, view position label ∈ {0, 1}, and patient age information of each CXR in ChestX-ray14 database. In order to make use of them, a strategy is to treat them as three-dimensional features connected to high-dimensional image features [13]. However, under the contrast of high-dimensional image features, these low-dimensional features are difficult to play an important role. Another strategy is to integrate them into the quintuplet loss function and then provide online supervision for model along with the lesion type. However, it involves more hyper-parameters and slower convergence in our experiment. Finally, we decide to use these information to mine the quintuplets, achieving off-line selection to accelerate model convergence.

Generating all possible quintuplets would result in super large-scale data pairs that are easily fulfill the constraint in formula (4) and formula (5). These quintuplets would not contribute to the training but slows down the convergence of model. It is crucial to select relatively hard quintuplets according to multi-view label.

First, symbols are used to define some relationships, as shown in Table (1). According these relationships, total three constraints are added to the pairs that appear in formula (4) and formula (5). As a hard pair, the distance between the anchor sample and the positive sample should be as far as possible except for the type of lesion. The relationship between anchor sample and positive sample should satisfy the constraint shown in formula (8):

TABLE 1. Symbols and their descriptions.

Symbols	Descriptions
L_s	Two CXRs contain the same lesion.
L_d	Two CXRs contain different lesions.
P_s	Two CXRs come from the same patient.
P_d	Two CXRs come from different patients.
G_s	Two CXRs come from patients whose gender are the same.
G_d	Two CXRs come from patients whose gender are different.
V_s	Two CXRs are taken from the same view position.
V_d	Two CXRs are taken from different view positions.
A_c	Age difference of two patients of corresponding CXR

$$Constraint_{ap} = \{L_s \& P_d \& G_d \& V_d \& largeA_c\} \tag{8}$$

In contrast, the distance between anchor sample and negative sample should be as close as possible. Although the type of lesion is different, gender and view position should be the same. The relationship between anchor sample and negative sample should satisfy the constraint shown in formula (9):

$$Constraint_{an} = \{L_d \& P_s || L_d \& P_d \& G_s \& V_s \& smallA_c\} \tag{9}$$

The relationship between anchor sample and similar sample should satisfy the constraint shown in formula (11):

$$Constraint_{as} = \{L_s \& P_s\} \tag{10}$$

Based on these constraints, the quintuplet mining algorithm is proposed for quintuplet selection, as shown in Algorithm1:

B. MODEL ARCHITECTURE

In the problem formulation subsection, a method is proposed to jointly supervise the deep model using classification loss and embedding layer distance. Accordingly, an end-to-end deep model with five inputs is designed, as shown in Figure1.

After comparing the current multi-lesion classification methods for ChestX-ray14, we find that transfer learning is widely used and more effective than training from scratch. Further, Baltruschat et.al. compared some existing deep convolutional network models that could be used for transfer learning on ChestX-ray14. They found that Resnet

Algorithm 1 Quintuplet Mining (QM) Algorithm

```
Input:
    Chest X-ray14 dataset;
Output:
    Quintuplets generated from Chest X-ray14 dataset;
Sort CXRs according to patient ID;
for each $img1$, $img2$,$img3$,$img4$ in CXR dataset do
    Define empty list: $Obj$, $Pos$, $Neg$;
    if $img1$ & $img2$ : $Constraint_{as}$, $img1$ & $img3$ : $Constraint_{ap}$,
    $img1$ & $img4$ : $Constraint_{an}$ then
        Append $img1$ and $img2$ to $Obj$;
        Append $img3$ to $Pos$;
        Append $img4$ to $Neg$;
    end if
    Constraint the length of list $Obj$, $Pos$, $Neg$ by random selection
    while loop time do
        Select $anc$ & $sim$ from $Obj$, $pos$ from $Pos$, $neg$ from $Neg$ randomly;
        Put $anc*2$, $sim$, $pos$, $neg$ together into a quintuplet;
    end while
end for
```

had the best transferability [13]. We therefore choose Resnet-50 as the base-bone of our model. As shown in Figure 1, the model is expanded into three parts based on Resnet-50: a feature extractor, a classifier, and a measurer. Among them, the network structure of the feature extractor is the same as that of Resnet-50 except for class layers, shown in Figure 1①④. Such a network structure can benefit from transfer learning by directly loading the parameters trained on ImageNet. The classifier consists of a BanchNorm layer, a Flatten layer, and a Dense layer, as shown in Figure 1③. The BanchNorm layer normalizes the learned features, and the Dense layer outputs the predicted classification results. The measurer first uses the Concatenate layer to connect the down-sampled features of anchor sample, similar sample, positive sample and negative sample. After L2 regularization, the output feature vectors are measured, shown in Figure 1②. The classifier and the measurer are not connected to each other, while they are both connected to the last layer of the parameter-shared feature extractors.

The parameters of classifier and the measurer are separately optimized by two different loss functions from quintuplet loss, while the design of parameter sharing allows them to jointly optimize feature extractors. Although classification performance is our only concern, the measurer is used to further enhance the capabilities of the feature extractors.

IV. EXPERIMENT AND RESULTS

A. DATASET

The dataset used in this study is ChestX-ray14, established by the researchers from the National Library of Medicine and Clinical Center of NIH. Chest X-ray14 is the largest publicly accessible chest x-ray database, downloaded through https ://nihcc.app.box.com/v/ChestXray-NIHCC. Its recent release triggered research on chest radiographs by the deep learning community. Chest X-ray14 contains more than 30,000 patients, 112,120 labeled chest x-ray images labeled by 14 kinds of lesion types including Infiltration, Effusion, Atelectasis, Nodule, Mass, Pneumothorax, Consolidation, Pleural Thickening (PT), Cardiomegaly, Emphysema, Edema, Fibrosis, Pneumonia and Hernia. In Chest X-ray14, 60,361 CXRs are lesion-free, while other 51,759 CXRs are abnormal. Although a large scale, Table 2 shows that the lesion distribution in Chest X-ray14 is imbalanced.

Each CXR with a resolution of 1024 * 1024 has a unique image identity. In addition to the type of lesion, CXRs are also labeled by other basic information such as patient ID (30805), gender(female or male), view position(posterio anterior (PA) or anterior posterio (AP)), and age. The vast majority of chest radiographs are concentrated in the 20 to 80 age group, with the largest number in the 50 to 60 age group. Table 6 shows that the gender-wise and view position-wise distribution are relatively uniform. Therefore, the impact of gender and view position on classification accuracy cannot be ignored.

B. METRICS

AUC, accuracy, recall, precision, and F-value are important and commonly-used metrics for classification tasks in the field of machine learning. There are 4 quantities are first defined:

- true positive as TP: The prediction is positive and the prediction is true.
- true negative as TN: The prediction is negative and the prediction is true.
- false positive as FP: The prediction is positive and the prediction is false.
- false negative as FN: The prediction is negative and the prediction is false

1) ACCURACY [54]

The classification accuracy rate refers to the proportion of correctly classified samples in total samples. The accuracy rate A is defined as (11):

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

2) PRECISION [55]

The classification precision rate is the ratio of the number of positive samples classified correctly to the number of samples determined by the classifier as positive samples. The precision rate P is defined as (12):

$$P = \frac{TP}{TP + FP} \tag{12}$$

TABLE 2. Overview of lesion distributions in the ChestX-ray14 dataset.

Lesion	Infiltration	Effusion	Atelectasis	Nodule	Mass	Pneumothorax	Consolidation
Number	19,894	13,317	11,559	6,331	5,782	5,302	4,667
Ratio (%)	17.74	11.88	10.31	5.65	5.16	4.73	4.16

Lesion	Pleural Thickening	Cardiomegaly	Emphysema	Edema	Fibrosis	Pneumonia	Hernia
Number	3,385	2,776	2,516	2,303	1,686	1,431	227
Ratio (%)	3.02	2.48	2.24	2.05	1.50	1.28	0.20

3) RECALL [56]

The classification recall rate refers to the ratio of the number of positive samples that are correctly classified to the number of samples that are truly positive. The recall rate R is defined as (13):

$$R = \frac{TP}{TP + FN} \tag{13}$$

4) F-VALUE [57]

In most cases, the higher the recall rate, the lower the precision rate and vice versa. Therefore, using either P or R cannot fully measure the performance of classification model. F-measure value is defined to take both P and R into consideration (14):

$$F = \frac{(\alpha^2 + 1) * P * R}{\alpha^2(P + R)} \tag{14}$$

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

5) AUC [58]

The AUC is defined as the area under the receiver operating characteristic (ROC) curve, which has typically horizontal axis as False Positive Rate and vertical axis as True Positive Rate. True Positive Rate (Sensitivity) is computed as TP/(TP+FN) and False Positive Rate is defined as FP/(TN+FP). Using the AUC value as the evaluation standard is more clear and direct than ROC Curve. Larger AUC means the classification performance is better.

C. EXPERIMENTAL SETUP

1) EXPERIMENT ENVIRONMENT

The experiments are conducted on an ubuntu linux server with 32G random access memory (RAM) and a 16-core central processing unit (CPU). Both the quintuplet mining procedure and the model architecture are developed with Python and deep learning libraries (e.g., Keras and Tensorflow). The whole model is trained using 2 GeForce GTX 1080 Ti graphics processing units (GPUs).

2) TRAINING DETAILS

In our experiments, the CXRs in the Chest Xray14 database are divided into training set, validation set and test set at a ratio of 8:1:1.

TABLE 3. Overview of gender-wise and VP-wise distributions in the ChestX-ray14 dataset.

Gender	Female	Male
Number	63,340	48,780
Ratio (%)	56.55	43.44

View position	PA	AP
Number	67,310	44,810
Ratio (%)	60.10	39.90

During training, the feature extractor's parameters of Resnet-50, which were pre-trained from Imgenet, are first transferred to our model. Then parameters of the classifier are trained by freezing feature extractor and optimizing formula (1), using all lesion-labeled CXRs in Chest X-ray14 except the test set and validation set. After that, the feature extractors' parameters trained by ImageNet and the classifier's parameters trained by Chest X-ray14 are loaded into the whole model. In the last step, classifier, measurer, and high-level features of feature extractor in CXNet-m3 are trained by parameter transfer between the same domain and optimising formula (7).

3) HYPERPARAMETER SETTING

In the quintuplet mining algorithm, A_c in $Constraint_{ap}$ is set as larger than 10, while A_c in $Constraint_{ap}$ is set as equal to 0.

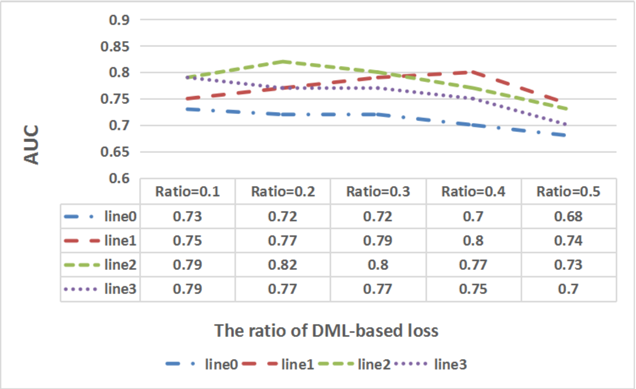
In the training procedure, the weight of the classification loss in formula (7) α is set as 0.8. In formula (6), the proportions of patient-wise distance and lesion-wise distance are each set to 0.5. According to experience and the validation results, we choose the different initial learning rates for two losses, decayed by 10 manually through monitoring the loss curve.

D. EXPERIMENTAL RESULTS AND DISCUSSION

Ling *et al.* discussed about AUC in their paper and they conclude that AUC is a better measure than accuracy based on formal definitions of discriminancy and consistency [58]. The implicit goal of AUC is to deal with situations where there is a skewed sample distribution and over-fit to a single class should be avoided. The paper recommends using AUC as a "single number" measure to over accuracy when evaluating and comparing classifiers.

TABLE 4. Result evaluation and comparison on AUC.

Lesion	Reference [29]	Reference [35]	Reference [39]	Reference [38]	Reference [13]	ours
atelectasis	0.716	0.80	0.719	0.733	0.763	**0.908**
cardiomegaly	0.807	0.87	0.880	0.858	0.869	**0.904**
effusion	0.784	**0.87**	0.792	0.806	0.822	**0.870**
infiltration	0.609	0.70	-	0.673	0.694	**0.902**
mass	0.706	0.83	0.809	0.777	0.820	**0.837**
nodule	0.671	0.75	0.711	0.718	0.747	**0.906**
pneumonia	0.633	0.67	**0.766**	0.684	0.714	0.626
pneumothorax	0.806	**0.87**	0.837	0.805	0.840	0.808
consolidation	0.708	**0.80**	0.734	0.711	0.716	0.723
edema	0.835	**0.88**	0.802	0.806	0.846	0.791
emphysema	0.815	**0.91**	0.841	0.842	0.895	0.86
fibrosis	0.769	0.78	0.803	0.743	**0.816**	0.792
PT	0.708	0.79	0.757	0.724	**0.763**	0.757
hernia	0.767	0.77	0.872	0.775	**0.937**	0.844
A.V.G	0.738	0.807	0.789	0.761	0.806	**0.824**



	Ratio=0.1	Ratio=0.2	Ratio=0.3	Ratio=0.4	Ratio=0.5
line0	0.73	0.72	0.72	0.7	0.68
line1	0.75	0.77	0.79	0.8	0.74
line2	0.79	0.82	0.8	0.77	0.73
line3	0.79	0.77	0.77	0.75	0.7

FIGURE 3. The average AUC values under different parameter choices. The ordinate is AUC values. The abscissa is the weight factor of C_{dml}, ranging from 0.1 to 0.5. Four lines adopt different sample selection strategies.

As shown in Table 2, the sample distribution in the Chest X-ray14 dataset is extremely uneven. For example, the category with the most samples is Infltration, with 19,894 CXRs, while the category with the least samples is Hernia, with only 227 CXRs. Therefore, AUC is used as a measure for each class in our paper and the values are compared with experimental results of state-of-art algorithms for multi-lesion classification for this dataset.

As shown in Figure 3, the average AUC values are compared in our experiments under different parameter choices. the ratio of C_{dml} in formula (7) is set as 0.1,0.2,0.3,0.4 and 0.5, respectively. the difference between line 0, line 1, line 2, and line 3 is the quintuplet sample selection strategies, as shown in Table 5. QM refers to whether the proposed quintuplet mining algorithm is used. S-N, P-N, and N-N represent the number of similar samples, the number of positive samples, and the number of negative samples, respectively. Q-N-A donates the number of quintuplets of each Anchor sample. Here, since the similar samples are very similar, the number of all similar samples is set to 1. Since the convergence rate of C_{dml} and C_{cla} is different, the total number of samples should not be too much. Otherwise, C_{dml} will not recognize any pattern when C_{cla} has converged. In this case, if the training process is continued, the over-fitting phenomenon will be exacerbated. If the training process is stopped, C_{dml} would not contribute to the model training.Therefore, Q-N-A is set three values of 4, 8, and 16, where Q-N-A is the product of S-N, P-N, and N-N. Compared with positive samples, the distribution of negative samples in the feature space is more scattered. Therefore, when the Q-N-A is limited, the number of negative samples N-N is set to a larger value.

In Figure 3, the AUC values of line 0 is much lower than those of other lines. It means that the quintuplet mining algorithm is effective for improving model performance. The use of quintuplet mining algorithm also speeds up the convergence of the model during training. It takes about a day to converge without quintuplet mining algorithm, while the loss value no longer changes significantly only after about 5 hours with training data selection. With the increase of the ratio, the AUC values of line 1 and line 2 tend to increase first and then decrease. However, the AUC values of line 3 decreases as the ratio becomes larger. Compared with line 1 and line 2, the Q-N-A value of line 3 is higher, which is 16. It means that the quintuplet mining algorithm allocates 16 quintuplets for each eligible Anchor sample. The model thus converges more slowly during training. The larger the weight of C_{dml}, the slower the model will converge. Line 1 and line 2 reach their peaks at a ratio of 0.2 and 0.4, respectively. Both of them obtain high AUC values. It can be inferred from Figure 3 that the selection of samples and the proportion of two losses are very important influencing factors. When selecting samples, the quintuplet mining algorithm is effective, and the total sample size should not be too large. Besides, 0.2 to 0.4 is an optimal range of weight factor of C_{dml} for our task.

As shown in Table 4, our work is compared with the work of Wang, Z. Li, A. I. Aviles-Rivero, Y. L, and I. M. Baltruschat

TABLE 5. Different sample selection strategies.

Settings Line Number	QM	S-N	P-N	N-N	Q-N-A
line 0	No	1	2	4	8
line 1	Yes	1	2	2	4
line 2	Yes	1	2	4	8
line 3	Yes	1	4	4	16

TABLE 6. Result evaluation on Accuracy and F-score.

Subsets Metrics	Training set	Validation set	Test set
Accuracy	0.925	0.770	0.749
F-score	0.950	0.866	0.853

[13], [29], [35], [38], [39]. Among them, Wang *et al.* first released the Chest Xray14 dataset in 2017 and used deep learning models to classify 14 lesions, which was published on IEEE Conference on Computer Vision and Pattern Recognition [29]. Li *et al.* used a simple recognition network to assist the training of Resnet. Although only a few hundred chest radiographs labeled with lesion location, the auxiliary task greatly improved the classification effect. Their research was published on IEEE Conference on Computer Vision and Pattern Recognition in 2018 [35]. Although a newly published paper in 2019, the citation rate of I. M. Baltruschat et al's paper is relatively high [13]. They used view position, gender, and age information as feature to train deep model and obtained good results. The work of Reference [39] and Reference [38] are also very new, printed in arXiv in 2019 an 2018, relatively.

Table 4 shows that our method achieves an average AUC value of 0.824, which is higher than other research work. Compared with the initial work of Reference [29], the average AUC values of later research work are all increased. In addition to our work, the average AUC values of Reference [35] and Reference [13] are also relatively high. Compared with other research, both of our proposal and their work use auxiliary information, such as lesion location, patient ID, gender and view position. It can be inferred that the use of auxiliary information can to some extent contribute to classification.

In addition to average performance, our model shows better performance for about half of the lesions, such as Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass and Nodule. Although not the best, our results of other lesions, such as Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, PT and Hernia, are not bad. However, in terms of AUC values, the model has the worst classification effect on Pneumonia. One possible reason is that our model does not obtain the strong features of pneumonia from small sample set.

Table 4 also illustrates that our method has outstanding classification performance for Atelectasis, Cardiomegaly, Infiltration, Nodule, getting AUC values higher than 0.9. Comparing Table 2 and Table4, we find that sample size of lesions with high AUC values is relatively large. For Infiltration, the AUC value has improved by nearly 30% when comparing with the earliest research work in Reference [29]. Deep learning encourages a large number of training samples to learn more robust features. Although the sample size is the largest, the AUC value of Infiltration has not always been high before our work. One possible reason is that Infiltration's classification features are difficult to learn. Our method is specifically designed to enhance the learning performance of classification features, thus improving the classification performance of large sample classes.

For Chest X-ray14 with uneven sample distribution, all state-of-art methods only use AUC as the evaluation metric, as compared in Table4. As auxiliary measures, we also verify the overall model performance under accuracy and F-score which takes both precision and recall into consideration. When the accuracy and F-score of the training set reach 0.925 and 0.950, respectively, the accuracy and F-score of the validation set reach the maximum. At this time, the accuracy and F-score of the test set are average 0.749 and 0.853, respectively. Generally, as a robust evaluation metric based on all cutoff values, the value of AUC is smaller than the value of accuracy. In our experiments, the accuracy value is relatively small, which may be caused by a large cutoff threshold.

In addition to the classification performance, the complexity of CXNet-m3 model is also analyzed and compared to the existing results. In neural networks, the spatial complexity is determined by the number of parameters, related to the number of convolution kernels, the number of output channels, and the number of layers. The more the number of parameters, the higher the spatial complexity of the model. Floating-point operations (FLOPs) is used to measure the time complexity of the model, related to the number of feature maps, the number of convolution kernels, the number of output channels and the number of layers [59]. FLOPs refers to the number of additions and multiplications in the model. The larger the FLOPs value, the higher the time complexity of the model. Table 7 shows the comparison of parameter number (Params) and FLOPs between CXNet-m3 and other methods. Among them, Reference [29]-1 to Reference [29]-4 are improved based on AlexNet, GoogLeNet, VGGNet and ResNet, respectively. Correspondingly, the model size and FLOPs value are similar to respective base model. Although based on ResNet, References [35], [13] and our method have different spatial complexities and time complexities due to different improvement strategies. The model size and FLOPs value of Reference [13] and our method are similar to standard ResNet, while the method in Reference [35] needs more parameters and larger FLOPs because of changed image scale and added recognition network. Classification in Reference [38] in designed based on both standard ResNet and standard DenseNet, which means parameter number and FLOPs should be close to the sum of ResNet and DenseNet. It can be seen from Table 7 that parameters in our model is the second least, which means relatively low space

TABLE 7. Complexity analysis and comparison on model size and FLOPs.

Methods	Params(x10^8)	FLOPs(x10^{10})
Reference [29]-1	0.610	1.454
Reference [29]-2	0.070	1.144
Reference [29]-3	1.384	3.095
Reference [29]-4	0.256	0.766
Reference [35]	0.351	1.380
Reference [13]	0.257	0.767
Reference [38]	0.336	1.357
ours	0.236	0.765

complexity. Table 7 also shows that FLOPs of our model is similar to Reference [29]-4 and Reference [13], but less than Reference [29]-1, Reference [29]-2, Reference [29]-3, Reference [35] and Reference [38]. FLOPs is related to the size of the input image. Except that Reference [35] sets the input image size to 299 * 299, other FLOPs in Table 7 are calculated based on a 224 * 224 input image. During training, CXNet-m3 model is more time-consuming than single-input models because it needs to process 5 images at each time. However, only one input is open during the inference process of CXNet-m3 model, which is really time-saving according to Table 7.

V. CONCLUSION

Chest X-rays are the most common imaging examination tool used to detect lesions related to heart, lungs, and respiratory system. In this paper, a deep multi-lesion classification model CXNet-m3 for CXRs in Chest X-ray14 is proposed to aid diagnosis. In CXNet-m3, easily accessible labels are explored to assist the classification of lesion types. To enhance the classification performance, a DML-based loss function is first constructed using labels of lesion type and patient ID. Then, a deep model taking advantage of transfer learning is built with quintuplet inputs to optimize both DML-based loss function and the classification loss function. To overcome the problem of slow convergence, a quintuplet mining algorithm for the selection of training sample is proposed based on labels of lesion type, view position, patient ID, patient age, and patient gender. The experiment results show that our method can achieve better AUC values than some state-of-art methods for the classification of multiple lesions in Chest X-ray14. The analysis of the experimental results also shows that our method has a significant effect on improving the classification performance of large sample categories. The disadvantage of this method is that CXNet-m3 involves a lot of hyper-parameters. How to make a reasonable or adaptive selection of the best hyper-parameters to make further improvements will be the focus of our future work.

REFERENCES

[1] T. R. Harwood, D. R. Gracey, and H. Yokoo, "Pseudomesotheliomatous carcinoma of the lung: A variant of peripheral lung cancer," *Amer. J. Clin. Pathol.*, vol. 65, no. 2, pp. 159–167, Feb. 1976.

[2] N. K. Sangani and S. M. Naliath, "Pseudomesotheliomatous type of sarcomatoid squamous cell lung cancer presenting with hemothorax," *Ann. Thoracic Surg.*, vol. 106, no. 4, pp. 201–203, Oct. 2018.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[4] G. L. Snider *et al.*, "The definition of emphysema: Report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases workshop," *Amer. Rev. Respiratory Disease*, pp. 182–185, 1985.

[5] B. D. Hobbs, "Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis," *Nature Genet.* vol. 49, no. 3, p. 426, 2017.

[6] A. Sforza *et al.*, "A case of pulmonary edema: The critical role of lung-heart integrated ultrasound examination," *Monaldi Arch. Chest Disease*, vol. 88, no. 3, p. 982, 2018.

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

[8] (2017). *Queen Alexandra Hospital Quality Report*. [Online]. Available: https://www.cqc.org.uk/location/RHU03

[9] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[10] J. Sarangapani, *Neural Network Control of Nonlinear Discrete-Time Systems*. Boca Raton, FL, USA: CRC Press, 2018.

[11] R. Rahimilarki, Z. Gao, A. Zhang, and R. Binns, "Robust neural network fault estimation approach for nonlinear dynamic systems with applications to wind turbine systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6302–6312, Dec. 2019.

[12] A. Santoro, "A simple neural network module for relational reasoning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4967–4976.

[13] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 6381.

[14] Y. Zhang, G. Cao, X. Li, and B. Wang, "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.

[15] 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.

[16] Z. Hu, J. Tang, P. Zhang, and B. P. Patlolla, "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] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017.

[19] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, Apr. 2018.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[21] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.

[22] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019.

[23] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017.

[24] 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.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[26] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2497–2506.
[27] P. N. Kieu, H. S. Tran, T. H. Le, T. Le, and T. T. Nguyen, "Applying multi-CNNs model for detecting abnormal problem on chest X-ray images," in *Proc. 10th Int. Conf. Knowl. Syst. Eng. (KSE)*, Nov. 2018, pp. 300–305.
[28] Y. Anavi, "Visualizing and enhancing a deep learning framework using patients age and gender for chest X-ray image retrieval," *Proc. SPIE Med. Imag. Comput.-Aided Diagnosis*. vol. 9785, Jul. 2016, Art. no. 978510.
[29] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "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. 2097–2106.
[30] S. Xu, H. Wu, and R. Bie, "CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2019.
[31] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," 2017, *arXiv:1710.10501*. [Online]. Available: http://arxiv.org/abs/1710.10501
[32] S. Xu, "CXNet-m2: A deep model with visual and clinical contexts for image-based detection of multiple lesions," in *Proc. Int. Conf. Wireless Algorithms, Syst. Appl.* Cham, Switzerland: Springer, 2019.
[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[35] Z. Li, "Thoracic disease identification and localization with limited supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2018, pp. 8290–8299.
[36] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: http://arxiv.org/abs/1711.05225
[37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4780.
[38] L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, "Weakly supervised medical diagnosis and localization from multiple resolutions," 2018, *arXiv:1803.07703*. [Online]. Available: http://arxiv.org/abs/1803.07703
[39] A. I. Aviles-Rivero *et al.*, "GraphX:Chest X-ray classification under extreme minimal supervision," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 504–512.
[40] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 31–38.
[41] Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, "No fuss distance metric learning using proxies," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 360–368.
[42] P.-E. Danielsson, "Euclidean distance mapping," *Comput. Graph. Image Process.*, vol. 14, no. 3, pp. 227–248, Nov. 1980.
[43] R. A. Melter, "Some characterizations of city block distance," *Pattern Recognit. Lett.*, vol. 6, no. 4, pp. 235–240, Sep. 1987.
[44] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 709–720.
[45] Y. Wen, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 499–512.
[46] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Dec. 2006, pp. 1735–1742.
[47] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[48] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille, "NormFace: L2Hypersphere embedding for face verification," in *Proc. ACM Multimedia Conf.*, 2017, pp. 1041–1049.
[49] Y. Sun, "Beyond part models: Person retrieval with refined part pooling and a strong convolutional baseline," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. 2018, pp. 480–496.
[50] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[51] K. Bardovi-Harlig and Rex A. Sprouse, "Negative versus positive transfer," in *Proc. TESOL Encyclopedia English Lang. Teaching*, 2018, pp. 1–6.
[52] F. Gao, H. Yoon, T. Wu, and X. Chu, "A feature transfer enabled multi-task deep learning model on medical imaging," *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 112957.
[53] H. Hermessi, O. Mourali, and E. Zagrouba, "Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning," *Expert Syst. Appl.*, vol. 120, pp. 116–127, Apr. 2019.
[54] C.-J. Huang, Y.-J. Yang, D.-X. Yang, and Y.-J. Chen, "Frog classification using machine learning techniques," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3737–3743, Mar. 2009.
[55] S. Akcay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1057–1061.
[56] H. Bhandari, "Generic text summarization using probabilistic latent semantic indexing," in *Proc. 3rd Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2008, pp. 1–8.
[57] B. Chen, J. Shi, S. Zhang, and F.-X. Wu, "Identifying protein complexes in protein-protein interaction networks by using clique seeds and graph entropy," *Proteomics*, vol. 13, no. 2, pp. 269–277, Jan. 2013.
[58] C. X. Ling and J. H. H. Zhang, "AUC: A statistically consistent and more discriminating measure than accuracy," in *Proc. IJCAI*. vol. 3, 2003, pp. 519–524.
[59] N. Ma, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. 2018, pp. 116–131.

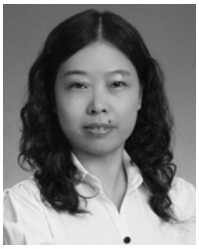
SHUAIJING XU received the B.S. degree from Beijing Normal University, Beijing, China, in 2015. She is currently pursuing the Ph.D. degree in computer application technology with the College of Artificial Intelligence, Beijing Normal University. Her major interests include big data, computer vision, deep learning, and medical image.

XIAOYILEI YANG received the B.S. degree from Beijing Normal University, Beijing, China, in 2018. She is currently pursuing the master's degree with the College of Artificial Intelligence of Beijing Normal University. Her major interests include computer vision and deep learning.

JUNQI GUO (Member, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2010. He is currently an Associate Professor and the Deputy Dean for Undergraduate Teaching with the School of Artificial Intelligence, Beijing Normal University, Beijing, China. He has published about more than 40 articles on international journals and academic conferences in the past five years, such as *Sensors*, *Computer Networks*, *Personal and Ubiquitous Computing*, *Multimedia Tools and Applications*, IEEE-related conferences, and so on. His research interests include artificial intelligence, intelligent signal and information processing, image processing and applications in smart education, and computer-aided diagnosis.

HAO WU received the B.E. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively. From October 2013 to April 2015, he worked as a Research Associate with the Lawrence Berkeley National Laboratory. Until now, he still takes charge of some related research projects in Lawrence Berkeley National Laboratory. He joined the Center for Big Data Mining and Knowledge Engineering, in December 2015. He is currently a Postdoctoral Research Fellow with the College of Information Science and Technology, Beijing Normal University. He has published about 20 articles including many international journals or magazines, such as *Neuro Computing, the Visual Computer, Multimedia Tools and Applications, the Journal of Visual Communication and Image Representation, the IET computer vision, and so on.*

GUANGZHI ZHANG is currently pursuing the Ph.D. degree with Beijing Normal University. He devotes himself to the research of knowledge intelligent computing and medical big data.



RONGFANG BIE (Member, IEEE) received the M.S. and Ph.D. degrees from Beijing Normal University, in June 1993 and June 1996, respectively. She was a Visiting Faculty with the Computer Laboratory, University of Cambridge, from March 2003, for one year. She is currently a Professor with the College of Information Science and Technology, Beijing Normal University. She is the author or coauthor of more than 100 articles. Her current research interests include knowledge representation and acquisition for the Internet of Things, dynamic spectrum allocation, big data analysis and application, and so on.