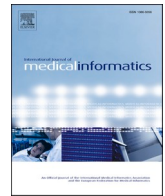




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms

Morteza Heidari<sup>a,\*</sup>, Seyedehnafiseh Mirniaharikandehei<sup>a</sup>, Abolfazl Zargari Khuzani<sup>b</sup>, Gopichandh Danala<sup>a</sup>, Yuchen Qiu<sup>a</sup>, Bin Zheng<sup>a</sup>

<sup>a</sup> School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA

<sup>b</sup> Department of Electrical and Computer Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA

### ARTICLE INFO

#### Keywords:

Coronavirus  
Convolution neural network (CNN)  
Disease classification  
VGG16 network  
COVID-19 diagnosis  
Computer-aided diagnosis

### ABSTRACT

**Objective:** This study aims to develop and test a new computer-aided diagnosis (CAD) scheme of chest X-ray images to detect coronavirus (COVID-19) infected pneumonia.

**Method:** CAD scheme first applies two image preprocessing steps to remove the majority of diaphragm regions, process the original image using a histogram equalization algorithm, and a bilateral low-pass filter. Then, the original image and two filtered images are used to form a pseudo color image. This image is fed into three input channels of a transfer learning-based convolutional neural network (CNN) model to classify chest X-ray images into 3 classes of COVID-19 infected pneumonia, other community-acquired no-COVID-19 infected pneumonia, and normal (non-pneumonia) cases. To build and test the CNN model, a publicly available dataset involving 8474 chest X-ray images is used, which includes 415, 5179 and 2,880 cases in three classes, respectively. Dataset is randomly divided into 3 subsets namely, training, validation, and testing with respect to the same frequency of cases in each class to train and test the CNN model.

**Results:** The CNN-based CAD scheme yields an overall accuracy of 94.5 % (2404/2544) with a 95 % confidence interval of [0.93,0.96] in classifying 3 classes. CAD also yields 98.4 % sensitivity (124/126) and 98.0 % specificity (2371/2418) in classifying cases with and without COVID-19 infection. However, without using two preprocessing steps, CAD yields a lower classification accuracy of 88.0 % (2239/2544).

**Conclusion:** This study demonstrates that adding two image preprocessing steps and generating a pseudo color image plays an important role in developing a deep learning CAD scheme of chest X-ray images to improve accuracy in detecting COVID-19 infected pneumonia.

### 1. Introduction

From the end of 2019, a new coronavirus namely COVID-19, was confirmed in human bodies as a new category of diseases that cause dangerous respiratory problems, heart infection, and even death. To more effectively control COVID-19 spread and treat patients to reduce mortality rate, medical images can play an important role [1]. In current clinical practice, chest X-ray radiography and computed tomography (CT) are two imaging modalities to detect COVID-19, assess its severity, and monitor its prognosis (or response to treatment). Although CT can achieve higher detection sensitivity, chest X-ray radiography is more commonly used in clinical practice due to the advantages, including low cost, low radiation dose, easy-to-operate and wide accessibility in the general or community hospitals [2]. However, pneumonia can be caused

by many different types of viruses and bacterial. Thus, it may be time-consuming and challenging for general radiologists in the community hospitals to read a high volume of chest X-ray images to detect subtle COVID-19 infected pneumonia and distinguish it from other community-acquired non-COVID-19 infected pneumonia. It is because there are many similarities between pneumonia infected by COVID-19 and other types of viruses or bacteria. Thus, this is a clinical challenge faced by the radiologists in this pandemic [3].

To address this challenge, developing computer-aided detection or diagnosis (CAD) schemes based on medical image processing and machine learning has been attracting broad research interest, which aims to automatically analyze disease characteristics and provide radiologists valuable decision-making supporting tools for more accurate or efficient detection and diagnosis of COVID-19 infected pneumonia. To this aim,

\* Corresponding author.

E-mail address: [Morteza.heidari@ou.edu](mailto:Morteza.heidari@ou.edu) (M. Heidari).

<https://doi.org/10.1016/j.ijmedinf.2020.104284>

Received 22 June 2020; Received in revised form 17 September 2020; Accepted 21 September 2020

Available online 23 September 2020

1386-5056/© 2020 Elsevier B.V. All rights reserved.

studies may involve following steps of preprocessing images, segmenting regions of interest (ROIs) related to the targeted diseases, computing and identifying effective image features, and building multiple-feature fusion-based machine learning models to detect and classify cases. For example, one study [4] computed 961 image features from the segmented ROIs depicting chest X-ray images. After applying a feature selection algorithm, a KNN classification model was built and yielded an accuracy of 96.1 % to classify between COVID-19 and non-COVID-19 cases.

However, due to the difficulty in identifying and segmenting subtle pneumonia-related disease patterns or ROIs on chest X-ray images, recent studies have demonstrated that developing CAD schemes based on deep learning algorithms without segmentation of suspicious ROIs and computing handcrafted image features is more efficient and reliable than the use of the classical machine learning methods. As a result, many deep learning models have been reported recently in the literature to detect and classify COVID-19 cases [2,5–13]. Although some deep learning convolution neural network (CNN) models are applied to CT images [5,6], more studies applied CNN models to detect and classify COVID-19 cases using chest X-ray images. They include different existing CNN models (i.e., Resnet50 [2,7], MobileNetV2 [8], CoroNet [9], Xception + ResNet50V2 [10]) and several new special CNN models (i.e., DarkCovidNet [11], COVID-Net [12] and COVIDX-Net [13]). These studies used different image datasets with a varying number of COVID-19 cases (i.e., from 25 to 224) among the total number of cases from 50 to 11,302. The reported sensitivity to detect COVID-19 cases ranged from 79.0%–98.6%.

Despite the promising results reported in previous studies, many issues have not been well investigated regarding how to train deep learning models optimally. For instance, whether applying image preprocessing algorithms can help to improve the performance and robustness of the deep learning models. To better address some of the challenges or technical issues, we in this study develop and test a new deep learning based CAD scheme of chest X-ray radiography images. The scheme can detect and classify images into 3 classes namely, COVID-19 infected pneumonia, the other community-acquired non-COVID-19 infected pneumonia, and normal (non-pneumonia) cases. The hypothesis in this study is that instead of directly using the original chest X-ray images to train deep learning models, we can apply image processing algorithms to remove the majority of diaphragm regions, normalize image contrast and reduce image noise, and generate a pseudo color image to feed in 3 input channels of the existing deep learning models that were pre-trained using color (RGB) images in the transfer learning process. It may help significantly improve model performance and robustness in detecting COVID-19 cases and distinguishing them from other community-acquired non-COVID-19 infected pneumonia cases. To test this study hypothesis and demonstrate the potential advantages of new approaches, we assemble a relatively large chest X-ray image dataset with 3 class cases. Then, we select a well-trained VGG16 based CNN model as a transfer learning model used in our CAD scheme. The details of the study design and data analysis results are reported in the following sections of this article.

## 2. Materials and method

### 2.1. Dataset

In this study, we utilize and assemble a dataset of chest X-ray radiography (CXR) images that are acquired from several different publicly available medical repositories [14–18]. These repositories were initially created and examined by the Allen Institute for AI in partnership with the Chan Zuckerberg Initiative, Georgetown University’s Center for Security and Emerging Technology, Microsoft Research, and the National Library of Medicine - National Institutes of Health, in coordination with The White House Office of Science and Technology Policy. Specifically, the dataset used in this study includes 8474 2D X-ray images in the

posteroanterior (PA) chest view. Among them, 415 images depict with the confirmed COVID-19 disease, 5179 with other community-acquired non-COVID-19 infected pneumonia, and 2880 normal (non-pneumonia) cases.

### 2.2. Image preprocessing

Fig. 1 shows examples of three chest X-ray images acquired in three classes of normal, community-acquired non-COVID-19 infected pneumonia and COVID-19 pneumonia cases (from top to bottom). It shows that the bottom part of images includes a diaphragm region with high-intensity (or bright pixels), which may have a negative effect on distinguishing and quantifying lung disease patterns using deep learning models. Hence, an image pre-processing algorithm is applied to identify and remove diaphragm regions. Specifically, the algorithm detects the maximum (the brightest -  $V_{max}$ ) and minimum (the darkest -  $V_{min}$ ) pixel value of the image, then uses a threshold  $T = V_{min} + 0.9 \times (V_{max} - V_{min})$  to segment the original image into a binary image as shown in Fig. 1(b). Next, after labeling all connected regions in the binary image, CAD scheme detects the biggest region, fills the holes in this region, and deletes all other small regions (if any) as shown in Fig. 1(c). This detected region locates in the diaphragm. Then, morphological filters are applied to smooth the boundary of the region as shown in Fig. 1(d). Last, the processed binary image is mapped back to the original image, CAD scheme removes overlapped pixels in the corresponding locations on the original image as shown in Fig. 1(e). Images after this step are named ( $I_p$ ).

In the next step, we convert the segmented grayscale images ( $I_p$ ) to 3-channel images suitable for fine-tuning an existing CNN model pre-trained using color (RGB) images. To do so, we apply an image noise filtering method and a contrast normalization method to preprocess the image after removing the diaphragm region. First, since the X-ray images often include additive noise, we apply a bilateral low-pass filter ( $BF$ ) to  $I_p$ . This filter is a non-linear filter and highly effective at noise removal while preserving textural information compared to the other low pass filters. In other words, this filter analyzes intensity values locally and considers the intensity variation of the local area to replace the intensity value of each pixel with the averaged intensity value of the pixels in the local area. To calculate the weights, we apply a Gaussian low-pass filter in the space domain. This step generates a noise-reduction image. Based on our experimental results, we select the following parameters in the bilateral filtering ( $\epsilon = 9$  and  $\sigma = 75$ ). Second, chest X-ray images may have different image contrast or brightness due to the difference in patient body size and/or variation of X-ray dose. To compensate such a potentially negative impact, we apply a histogram equalization ( $HE$ ) method to normalize  $I_p$  images. This filter can enhance lung tissue patterns and characteristics associated with COVID-19 infection. Then, as shown in Fig. 2, three preprocessed images namely,  $I_p$ ,  $I_b = BF(I_p)$  and  $I_{eq} = HE(I_p)$  form a pseudo color image that is fed into 3 input (RGB) channels of the CNN model.

### 2.3. Transfer learning

In this study, we adopt a transfer learning approach since the previous studies have shown in order to avoid either overfitting or underfitting consequences using a small training dataset, a better approach is to take advantage of a CNN initially trained using a large-scale dataset [19]. Currently, many CNN models have been previously developed and are available for different engineering applications. In this study, we select a VGG16 model, which was pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) using a large dataset with 14 million images [20]. VGG16 model won the first place on image localization task and second place on image classification task in the 2014 ILSVRC challenge [21]. As shown in Fig. 3, the VGG16 model has 13 convolutions, 5 max pooling and 3 fully connection layers in 6 blocks,

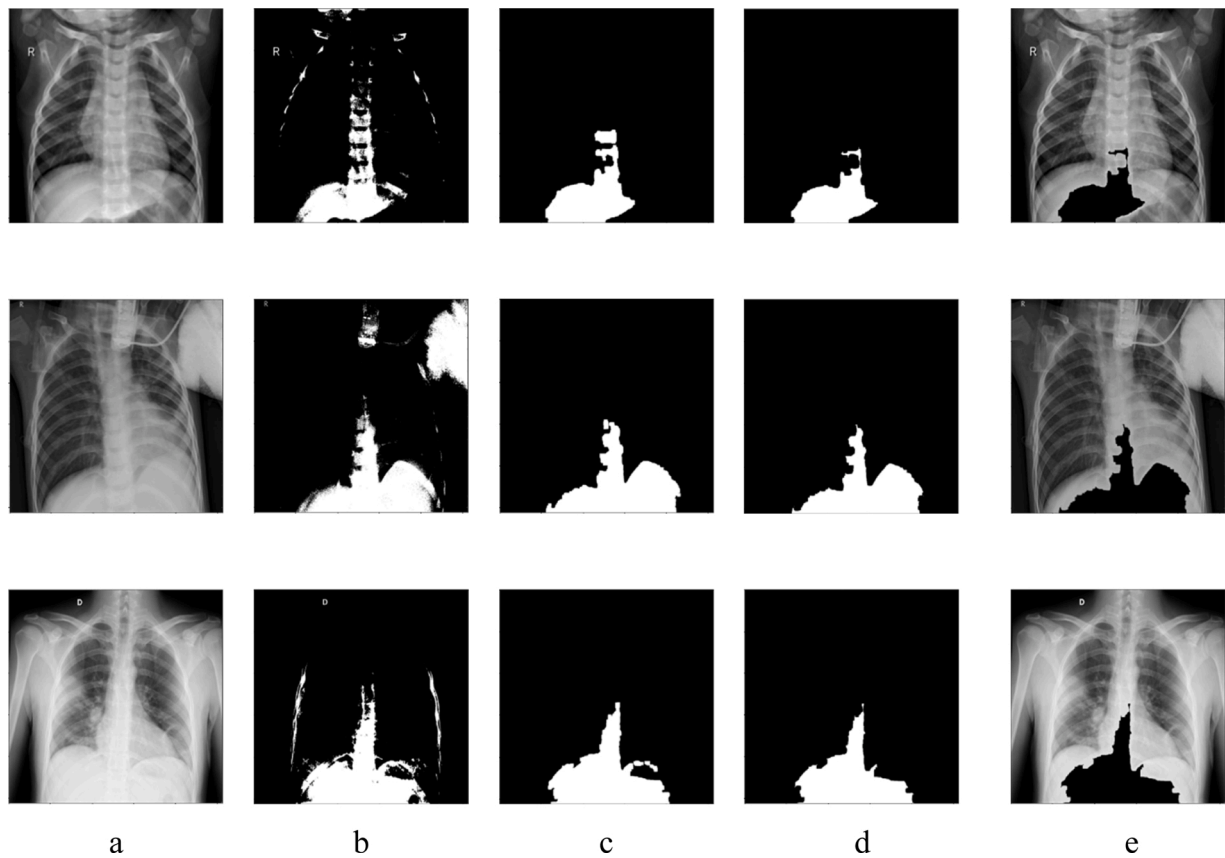


Fig. 1. Example of chest X-ray images in three classes (top – normal, middle – community-acquired non-COVID-19 pneumonia, and bottom – COVID-19 infected pneumonia case). The figure also shows (a) the original Images, (b) the binary images after threshold, (c) images after selecting the biggest segmented region, (d) images after applying morphological filtering and (e) the original image after removing the majority part of diaphragm region ( $I_p$ ).

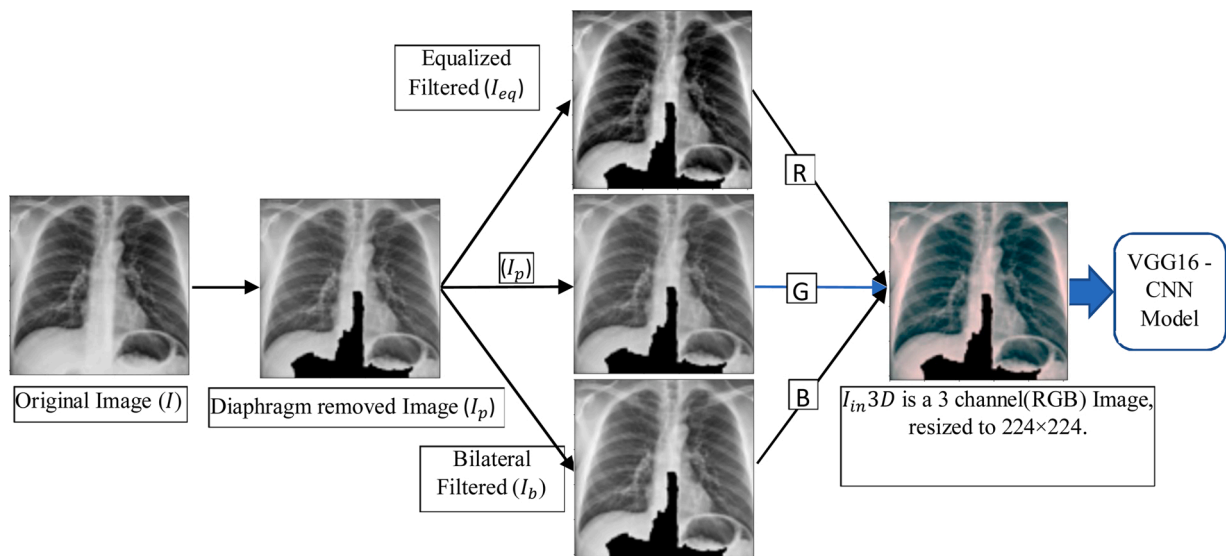


Fig. 2. A flow diagram to illustrate image pre-processing steps to generate input of a CNN model, where ( $I$ ) is the original Image in the dataset. ( $I_p$ ) is the diaphragm removed image. ( $I_{eq}$ ) is an image after applying histogram equalization on ( $I_p$ ), and ( $I_b$ ) is an image after applying bilateral filtering on ( $I_p$ ). Three images ( $I_p$ ), ( $I_b$ ), and ( $I_{eq}$ ) are fed into three channels of the CNN model to simulate the RGB image.

which include over 138 million trainable parameters.

In our transfer learning, the weights between all connected nodes in front or low layers of the VGG16 based CNN model maintain unchanged (blocks 1–5 as shown in Fig. 3). Next, block 6 in the model is modified by replacing with one flatten layer and two fully connected layers, which

include 256 and 128 nodes, respectively. In these layers, the rectified linear unit (ReLU) [22] is used as their activation function. Then, all trainable weights in all connection nodes of the whole modified VGG16 model are fine-tuned using chest X-ray image data. In this fine-tuning process, a small learning rate (learning rate =  $10^{-5}$ ) is used to make a

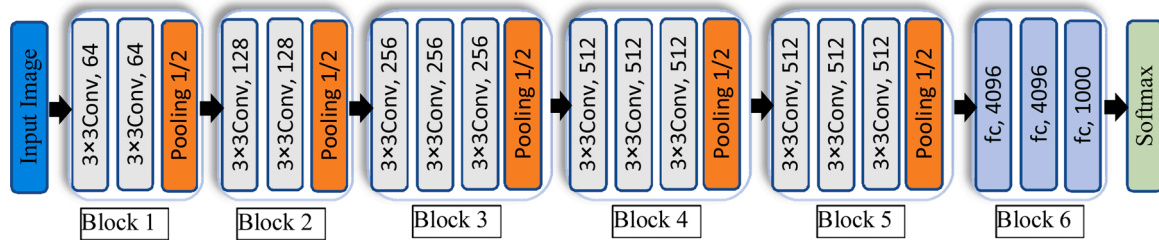


Fig. 3. Illustration of the architecture of VGG16 based CNN model.

small variation to the pre-trained parameters. In this way, we will preserve the valuable parameters as much as possible by avoiding dramatic changes on the pre-trained parameters and let the model learn the special characteristics of chest X-ray images. Finally, in the last classification layer, Softmax is used as the activation function. As a result, a new transfer learning model is built to fulfill a three-class classification task. The complete CNN model is compiled with Adam [23] optimizer with a batch size of 4, max epoch = 200, initial learning rate =  $10^{-5}$ , and monitoring validation loss for reducing the learning rate every 5 epochs with a factor of 0.8. Table 1 shows the complete architecture of the transfer learning VGG16 model built in this study.

2.4. Model training and testing

First, the original chest X-ray image has  $1024 \times 1024$  pixels, while the VGG16 model was pre-trained using images of  $224 \times 224$  pixels. Thus, each chest X-ray image is down-sampled to  $224 \times 224$  pixels to fit the VGG16 model. Then, for training and evaluating the proposed VGG16 based transfer leaning CNN model, we randomly split the entire image dataset of 8474 cases into 3 independent subsets of training, validation, and testing. Overall, 10 % of cases (848) are assigned to test subset. On the remaining 7626 cases, 10 % cases are assigned to the validation subset (757), while 90 % case (6869) are formed as the training subset. To maintain the same case partition ratios for three classes of COVID-19 infected pneumonia, other community-acquired non-COVID-19 infected pneumonia and normal cases, the case partition or assignment is done on three classes independently. Table 2 shows the number of cases in each subset.

Second, there are different available techniques to deal with imbalanced data [24]. In this study, the class weight technique, as one possible way, is applied during training to reduce the potential consequences of imbalance data. In the class weigh technique, we adjust

Table 1  
The architecture of the new VGG16 model after transfer learning with new layers (19 to 22).

Number	Layer	Size	Activation
0	Input Image	$224 \times 224 \times 3$	—
2	2× Convolution (3 × 3)	$224 \times 224 \times 64$	ReLU
3	Max Pooling	$112 \times 112 \times 64$	ReLU
5	2× Convolution (3 × 3)	$112 \times 112 \times 128$	ReLU
6	Max Pooling	$56 \times 56 \times 128$	ReLU
9	3× Convolution (3 × 3)	$56 \times 56 \times 256$	ReLU
10	Max Pooling	$28 \times 28 \times 256$	ReLU
13	3× Convolution (3 × 3)	$28 \times 28 \times 512$	ReLU
14	Max Pooling	$14 \times 14 \times 512$	ReLU
16	3× Convolution (3 × 3)	$14 \times 14 \times 512$	ReLU
18	Max Pooling	$7 \times 7 \times 512$	ReLU
19	Flattening	25,088	—
20	Fully Connected	256	ReLU
21	Fully Connected	128	ReLU
22	Fully Connected	3	SoftMax

Table 2  
Distribution of cases in three subsets.

Image Data Subset	Training	Validation	Testing
COVID-19 cases	366	37	42
Other pneumonia cases	4201	460	518
Normal cases	2332	260	288
Total number of cases	6899	757	848

weights inversely proportional to class frequencies in the input data [25]. The weight,  $\omega_i$  in class  $i$  is computed using the following equation.

$$\omega_i = \frac{(\text{Total number of cases})}{(\text{number of classes}) \times (\text{number of cases in class } i)} \quad (1)$$

The weights of the classes are utilized while fitting the model. Hence, in the loss function, we assign higher values to the instances of smaller classes. Therefore, the calculated loss will be a weighted average, where the weights of each sample corresponding to each class during loss calculation are specified with  $\omega_i$ .

Additionally, in the training data of minority cases (COVID-19 cases), a common augmentation technique [26] is applied to increase the training sample size. First, using shearing factors ( $\leq 0.2$ ), image intensity is sheared based on the shearing angle in a counter-clockwise direction. Second, using zooming factors ( $\leq 0.2$ ), images are randomly magnified. Third, using rotation factors (within  $\pm 20^\circ$ ), images are randomly rotated. Fourth, using a shift factor ( $\leq 0.2$ ), images are randomly shifted in 4 directions (up, left, down right). Last, images are flipped horizontally in a random base to generate as much augmented data as possible.

Multiple iteration or epochs are applied to train the VGG16 based CNN model. The model is first trained using the data in the training subset and validated using validation subset. During the training process, the optimizer tries to force the architecture to learn more and more information to reduce the performance gap between training and validation. To control overfitting and maintain training efficiency, we limit model training epochs to 200. Hence, at the end of 200 training epoch, the trained model is saved and then tested using the data in the testing subset, which does not involve in the model training and validation process.

To reduce the risk of potential bias in data partition into three subsets of training, validation, and testing, we repeat this model training and testing process three times by randomly dividing all cases into training, validation and testing subsets three times using the same case ratios or numbers as shown in Table 2. In addition, during these three times of case partition, the cases assigned to the validation and testing subsets are totally different (no duplication). Three trained models are tested using totally different testing cases. Thus, the total number of testing cases increases (as shown in Table 2) to 2544 ( $848 \times 3$ ). Fig. 4 shows a schematic diagram that illustrates the complete architecture of this VGG16 transfer learning CNN model, as well as the training, validation, and testing phase.

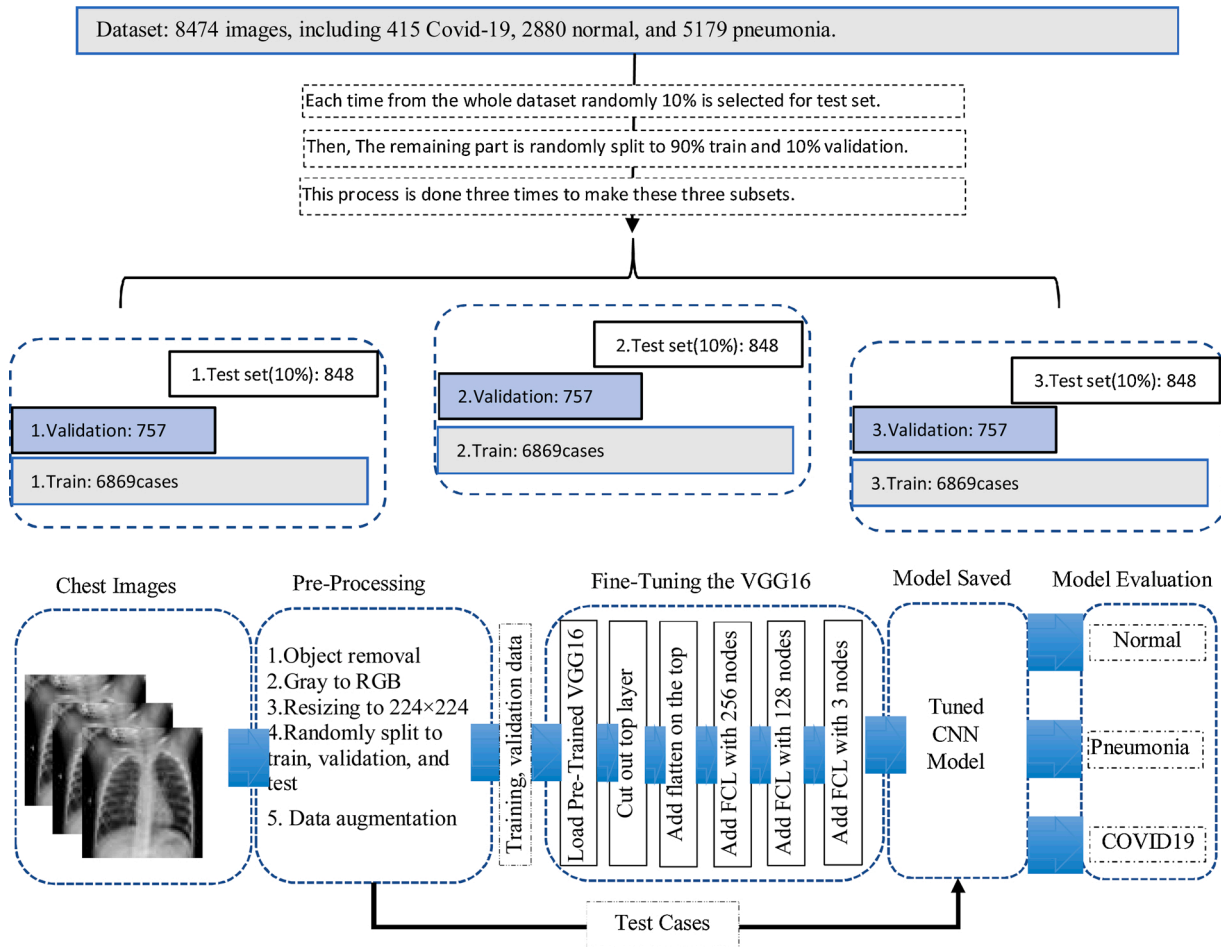


Fig. 4. schematic representing training and validation phase of the proposed scheme.

### 2.5. Performance assessment

We perform experiments to analyze two different accuracies. The first one is accuracy for a three-class classification to distinguish between COVID-19 infected pneumonia, community-acquired pneumonia, and normal (non-pneumonia) cases. We compute accuracy values in detecting images in 3 classes. We also calculate (1) a macro averaging, which is the average of 3 accuracy values of 3 classes without considering the proportion of the number of the cases in each class ( $A_{mac} = (A_1 + A_2 + A_3)/3$ ), and (2) a weighted averaging, which is the weighted average of 3 accuracy values weighted with respect to the proportion of the classes ( $A_w = w_1A_1 + w_2A_2 + w_3A_3$ ), where  $A_1, A_2, A_3$  are accuracy values of 3 classes, while  $w_1, w_2, w_3$  are weighting factors of 3 classes representing the ratios of cases in 3 classes. Then, for the three-class classification, a confusion matrix is generated from which several evaluation indices, including precision, recall, F1-score, and Cohen's Kappa [27] values are computed to evaluate CAD performance. The value of Cohen's kappa coefficients (ranging from zero and one) indicates the possibility of the predicted results occurring by chance. The lower Kappa value shows the more randomness of the results, while the higher value shows a better similarity and higher robustness.

The second accuracy evaluation refers to classification between the COVID-19 and non-COVID19 cases (including both normal and community-acquired pneumonia cases). In this circumstance, we compute true positive (TP) for the cases correctly identified as COVID-19, false negative (FN) for the COVID-19 cases being incorrectly classified as normal or community-acquired pneumonia cases, true negative (TN) for the cases correctly identified as non-COVID-19 cases, and false positive (FP) for the normal and community-acquired pneumonia cases

being incorrectly classified as COVID-19 by the CNN model. Then, the accuracy, sensitivity, specificity, recall, and F1-scores of model classification are computed and tabulated.

### 3. Results

Fig. 5(a–c) presents trend curves of training and validation accuracy of the new transfer learning VGG16 based CNN model in three experiments using different training and validation subsets in the left column. Then, by applying the trained models on the corresponding testing subsets, three confusion matrices of the models on the testing subsets are shown in the right column. All three curves show that as the increase of training iteration epochs during the training process, the prediction accuracy of the validation subset varies greatly (with big oscillation) initially, and then gradually converges to a higher accuracy level with much small oscillation. Thus, for all three subsets after epoch 75, validation accuracy is following the training accuracy, which indicates that learning is happening during different epochs. The trend graph also shows that the proposed technique does not suffer significant overfitting or underfitting in our transfer learning model. Then, by combining three confusion matrixes of the three independent testing subsets, as shown in second column of Fig. 5(a–c), Fig. 5(d) displays a combined 3-class confusion matrix of 2544 ( $3 \times 848$ ) cases.

First, based on three confusion matrices as shown in Fig. 5(a–c), the overall 3-class classification accuracy levels are 93.9 % (796/848), 94.7 % (803/848), and 94.9 % (805/848), respectively. The difference is approximately 1%. Then, based on the confusion matrix of the combined data as shown in Fig. 5(d), we compute the precision, recall rate, F1-score, and prediction accuracy of the new transfer learning VGG16

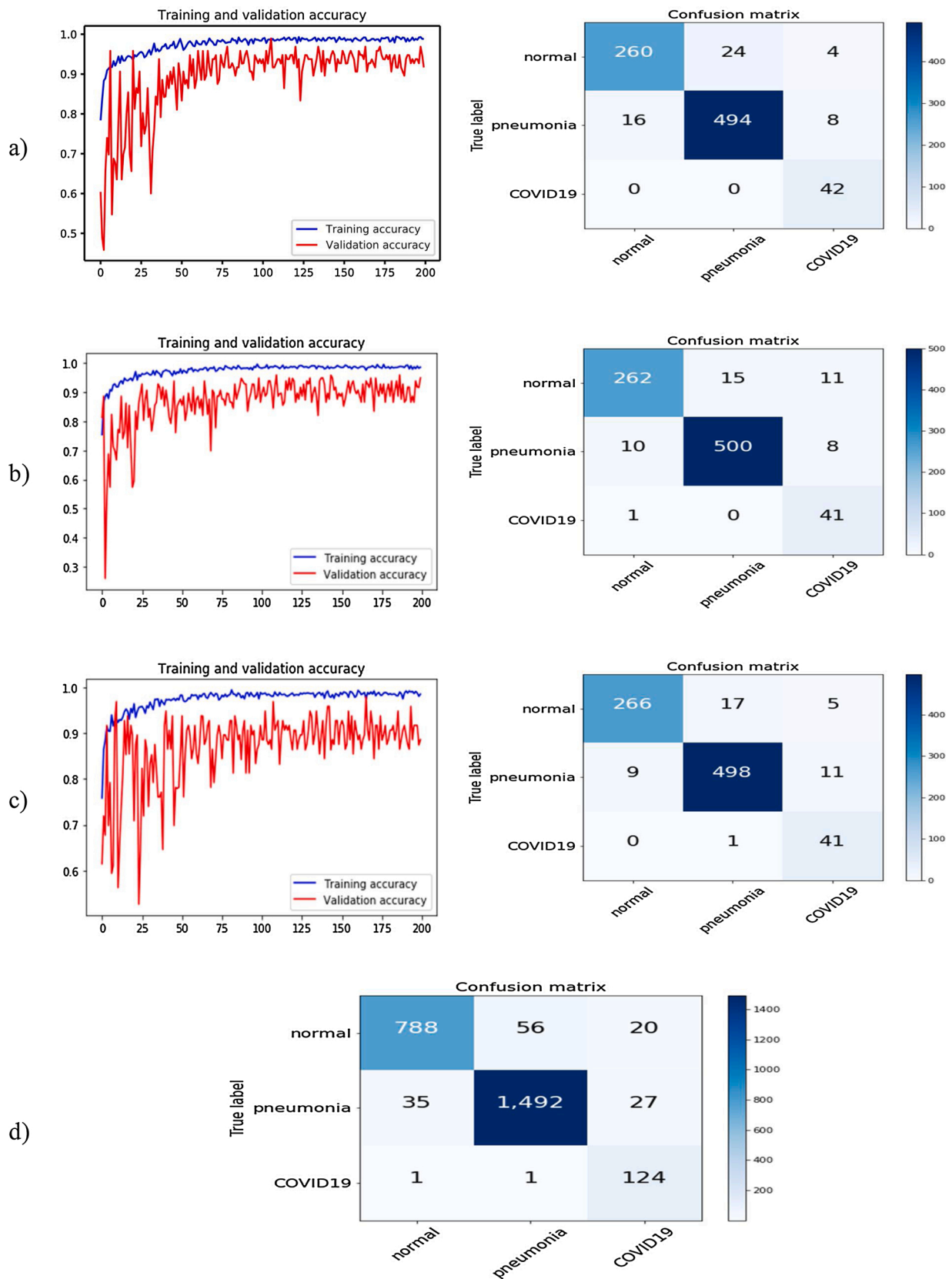


Fig. 5. (a-c) Left column show three sets of performance curves applying to the training and validation subsets for 3 experiments in 200 training epochs, respectively. The horizontal axis shows the number of epochs, and the vertical axis shows the accuracy. The right column show three confusion matrices of the corresponding testing results are shown on the right. (d) A combined confusion matrix of applying three trained models to three independent testing data subsets with total 2544 cases.

based CNN model, as shown in Table 3. Among 2544 testing cases, 2404 are correctly detected and classified into 3 classes. The overall accuracy is 94.5 % (2404/2544) with 95 % confidence interval of [0.93, 0.96]. In addition, the computed Cohen's kappa coefficient is 0.89, which confirms the reliability of the proposed approach to train this new deep transfer learning model to do this classification task.

To further evaluate the performance of our CAD scheme in detecting the COVID19 infected pneumonia cases using chest X-ray images, we place both normal and community-acquired pneumonia images into the negative class and COVID-19 infected pneumonia cases into the positive class. Combining the data in the confusion matrix, as shown in Fig. 5(d), the CAD scheme yields 98.4 % detection sensitivity (124/126) and 98.0 % specificity (2371/2418). The overall accuracy is 98.1 % (2495/2544).

Next, Table 4 shows and compares (1) confusion matrixes generated by four models trained and tested using different input images and three data subsets generated from the data partition, as well as (2) overall classification accuracy and 95 % confidence intervals. The results indicate that (1) without using the data augmentation technique, the model accuracy on data of the testing subset drops to 82.3 % with the kappa score of 0.71. (2) Without applying image preprocessing and directly feeding the original chest X-ray images into the VGG16 based CNN model ("simple model"), classification accuracy is 88.0 % with a Cohen's kappa score of 0.75. (3) Using image filtering and pseudo color images without removing the majority part of diaphragm regions, the "filter-based model" yields 91.2 % accuracy and a Cohen's kappa score of 0.83. All three models yield lower classification accuracy than the proposed model involving data augmentation technique and two steps of image preprocessing.

In addition, Table 5 compares our transfer learning VGG16 based CNN model and 10 state-of-art models recently reported in the literature to detect and classify COVID-19 cases. The Table shows the number of cases in the training and testing data subsets, imaging modality (CT or X-ray radiography), and reported classification performance including either 3-class or 2-class classification for these studies. Although the reported performance of these studies cannot be directly compared due to the use of different image dataset and testing methods, the presented data clearly demonstrate that our model is tested using relatively large dataset and yields very comparable classification performance as comparing to the state-of-art models developed and tested in this research field.

#### 4. Discussion

In this study, we developed and tested a novel deep transfer learning CNN model to detect and classify chest X-ray images depicting COVID19 infected pneumonia. This study has several unique characteristics as compared to the previously reported studies in this field and produces several new interesting observations. First, since the deep learning CNN model includes a considerable number of parameters that need to be trained and determined, a large and diverse image dataset is required to produce robust results [28]. Although we used a relatively large image dataset of 8474 chest X-ray images, the dataset is unbalanced in 3 classes of images, and the number of the COVID-19 infected pneumonia cases (415) remains small. Thus, in order to build a robust deep learning model, we apply a class weight technique during the training process

**Table 3**  
Classification report of the proposed method.

	Precision	Recall	F1-score	Support cases
Normal	0.96	0.91	0.93	864
Other Pneumonia	0.96	0.96	0.96	1554
COVID19	0.73	0.98	0.84	126
Accuracy	—	—	<b>0.95</b>	2544
Macro avg	0.88	0.95	0.91	2544
Weighted avg	0.95	0.94	0.94	2544

and select a well-trained VGG16 model and apply a transfer learning approach. Specifically, the original VGG16 model includes over 138 million parameters. These parameters have been trained and determined using a large ImageNet database over 14 million images. It is difficult to train so many parameters from scratch robustly using a dataset of 8474 images. Thus, we retrain or fine-tune the pre-trained VGG16 (as shown in Fig. 3) to reduce the overfitting risk. Study results demonstrate that this transfer learning approach can yield higher performance with the overall accuracy of 94.5 % (2404/2544) in the classification of 3 classes and 98.1 % (2495/2544) in classifying cases with and without COVID-19 infection, as well as the high robustness with a Cohen's kappa score of 0.89.

Second, unlike the regular color photographs, chest X-ray images are gray-level images. Thus, in order to fully use the pre-trained VGG16-based CNN model, we generate two new gray-level images. Then, instead of applying the original chest X-ray image to the CNN model directly, 3 different gray-level images are fed into 3 input (RGB color) channels of the CNN model. Specifically, we apply a bilateral low-pass filter to generate a noise-reduced image and a histogram equalization method to generate a contrast normalized image. Comparing two approaches of using only original chest X-ray image and 3 different images to generate a pseudo color image as an input image to the CNN model, our study results show that using a pseudo color image approach, overall classification accuracy increases 3.6 % from 91.2%–94.5%, and Cohen's kappa score increases 7.2 % from 0.83 to 0.89, respectively. The results demonstrate the advantage of using our new approach to fully use 3 input channels of the CNN model pre-trained using color images because these two filtered gray-level images contain additional information, which can enhance image classification capability.

Third, since in the area of medical imaging, generally, disease's patterns are not comparable to the other existing patterns in the image, preprocessing steps are noteworthy [29]. Hence, we apply an image preprocessing algorithm to automatically detect and remove the majority part of the diaphragm region from the chest X-ray images. Comparing the approaches with and without removing the diaphragm regions, classification performance of the CNN model changes from 94.5%–88.0% and 0.89 to 0.75 for the overall classification accuracy and Cohen's kappa coefficients, respectively, which indicates a 7.4 % increase in the overall classification accuracy and 18.7 % increase in Cohen's kappa coefficient by removing the majority of diaphragm regions. Thus, although skipping segmentation of the suspicious disease regions of interest is one important characteristic of deep learning, our study demonstrates that applying an image preprocessing and segmentation algorithm to remove irrelevant regions on the image can also play an important role in increasing performance and robustness of deep learning models.

In addition, we observe and confirm that applying data augmentation in the training data is also essential. Without data augmentation to increase training dataset size, the overall classification accuracy of the CNN model significantly reduces to around 82.3 %. In summary, we in this paper present a new deep transfer learning model to detect and classify the COVID-19 infected pneumonia cases, as well as several unique image preprocessing approaches to optimally train the deep learning model using the limited and unbalanced medical image dataset. The similar learning concept and image preprocessing approaches can also be adopted to develop new deep learning models for other medical images to detect and classify other types of diseases (i.e., cancers [30, 31]).

Despite encouraging results, this study also has limitations. First, although we used a publicly available dataset of 8474 cases, including 415 COVID-19 cases, due to the diversity or heterogeneity of COVID-19 cases, the performance and robustness of this CAD scheme need to be further tested and validated using other large and diverse image databases. Second, this study only investigates and tests two image preprocessing methods to generate two filtered images, which may not be the best or optimal methods. New methods should also be investigated



**Table 4**

Confusion matrix of four CNN models on X-ray Images. 95 % confidence interval (CI) for the accuracy is shown in the last column.

		Normal	Pneumonia	COVID19	Accuracy	95 % CI
Proposed Model	Normal	788	56	20	94.5 %	[0.93,0.96]
	Pneumonia	35	1492	27		
	COVID 19	1	1	124		
Filter-based model	Normal	750	89	25	91.2 %	[0.90,0.92]
	Pneumonia	64	1452	38		
	COVID19	2	6	118		
Simple model	Normal	701	123	40	88.0 %	[0.86,0.89]
	Pneumonia	72	1431	51		
	COVID19	6	13	107		
No-augmentation	Normal	653	158	53	82.3 %	[0.80,0.84]
	Pneumonia	124	1346	74		
	COVID19	8	23	95		

**Table 5**

Comparison accuracy results of the proposed method with the other deep learning methods on COVID-19 diagnosis.

Approach	Data Type	Cases number (including COVID-19 cases)	Method utilized	2 classes accuracy (%)	3 classes accuracy (%)	COVID-19 detection Sensitivity (%)
Narin et al. [2]	X-ray	100 (50)	ResNet50	98.0	—	96.0
Sethy et al. [7]	X-ray	50 (25)	ResNet50+SVM	95.4	—	97.0
Ioannis et al. [8]	X-ray	1427 (224)	MobileNetV2	96.7	93.5	98.6
Wang et al. [5]	CT	237 (119)	M-Inception	82.9	—	81.0
Tulin et al. [11]	X-ray	1127 (127)	DarkCovidNet	98.08	87.02	90.6
Khan et al. [9]	X-ray	221 (29)	CoroNet (Xception)	98.8	94.52	95.0
Rahimzadeh & attar [10]	X-ray	11,302 (31)	Xception + ResNet50V2	99.5	91.4	80.53
Wang et al. [12]	X-ray	300 (100)	COVID-Net	96.6	93.3	91.0
Ying et al. [6]	CT	57 (30)	DRE-Net (ResNet50)	86	—	79.0
Hemdan et al. [13]	X-ray	50 (25)	COVIDX-Net	90	—	—
Our new method	X-ray	2544 (126)	VGG16	98.1	94.5	98.4

and compared in future studies. Third, to further improve model performance and robustness, it also needs to develop new image processing and segmentation algorithms to more accurately remove the diaphragm and other regions outside lung areas in the images. Therefore, more research work is needed to overcome these limitations in the future studies.

**5. Conclusion**

In this study, we proposed and investigated several new approaches

to develop a transfer deep learning CNN model to detect and classify COVID-19 cases using chest X-ray images. Study results demonstrate the added value of performing image preprocessing to generate better input image data to build deep learning models. These include removing irrelevant regions, normalizing image contrast-to-noise ratio, and generating pseudo color images to feed into all three channels of the CNN models in applying the transfer learning method. The reported high classification performance is also promising, which provides a solid foundation to further optimize the deep learning models to detect COVID-19 cases and validate its performance and robustness using large

**Summary points**

**What was Already Known on the topic**

- Due to the low cost, low radiation, wide accessibility, chest X-ray radiography is a good imaging modality to detect COVID-19. However, its sensitivity is lower than CT.
- Developing deep learning model based CAD schemes of chest X-ray images may play a useful role in facilitating detection and diagnosis of COVID-19.
- A few deep learning models using chest X-ray images to detect COVID-19 have been reported using small datasets. The models were trained using the original images only.

**What this study adds to our knowledge**

- Due to the diversity of image contrast and noise, adding image preprocessing steps is important and can help improve deep learning model performance.
- In transfer learning, one should not only use original images. It should add two additional filtered images to fill in 3 input channels of the deep learning model, which can enhance information learning and improve model performance.
- The deep learning CAD scheme can achieve high performance in detecting and classifying not only between COVID-19 cases and healthy (non-pneumonia) cases, and also between COVID-19 infected pneumonia and other community-acquired non-COVID-19 infected pneumonia cases.
- Our model is tested using a larger dataset as comparing to previous studies reported in the literature, which further supports the feasibility of this CAD approach.

and diverse image datasets in future studies.

### Credit authorship contribution statement

Morteza Heidari and Bin Zheng design the study. Morteza Heidari implements the idea and write the required computer scheme coding link to VGG16 model. Abolfazl Zargari and Seyedehnafiseh Mirniaharikandehei collect the dataset and help test image filtering and normalization algorithms. Gopichandh Danala helps designing and testing image pre-processing and blob discovery strategy. Morteza Heidari drafts the manuscript. Bin Zheng and Yuchen Qiu review and revise the manuscript.

### Declaration of Competing Interest

The authors report no declarations of interest.

### Acknowledgements

This work is support in part by the grant from National Cancer Institute (R01 CA197150). The authors also thank the research support from Stephenson Cancer Center, University of Oklahoma, which helps establishment of our Computer-aided Diagnosis Laboratory.

### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijmedinf.2020.104284>.

### References

- [1] P. Lei, Z. Huang, G. Liu, et al., Clinical and computed tomographic (CT) images characteristics in the patients with COVID-19 infection: what should radiologists need to know, *J. Xray Sci. Technol.* 28 (3) (2020) 369–381.
- [2] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, *arXiv preprint arXiv:2003.10849* (2020).
- [3] W. Dai, H. Zhang, J. Yu, et al., CT imaging and differential diagnosis of COVID-19, *Can. Assoc. Radiol. J.* 71 (2) (2020) 195–200.
- [4] E.M. Abd, K.M. Hosny, A. Salah, et al., New machine learning method for image-based diagnosis of COVID-19, *PLoS One* 15 (6) (2020) e0235187.
- [5] S. Wang, B. Kang, J. Ma, et al., A deep learning algorithm using CT images to screen for corona virus disease (COVID-19), *MedRxiv* (2020), <https://doi.org/10.1101/2020.02.14.20023028>.
- [6] S. Ying, S. Zheng, L. Li, et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, *medRxiv* (2020).
- [7] P.K. Sethy, S.K. Behera, Detection of Coronavirus Disease (COVID-19) Based on Deep Features, *Preprints 2020030300*, 2020.
- [8] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, *Phys. Eng. Sci. Med.* (2020), <https://doi.org/10.1007/s13246-020-00865-4>.
- [9] K.A. Iqbal, J.L. Shah, M.M. Bhat, Coronet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, *Comput. Methods Programs Biomed.* (2020), 105581.
- [10] M. Rahimzadeh, A. Attar, A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2, *Inform. Med. Unlocked* (2020) 100360.
- [11] O. Tulin, M. Talo, E.A. Yildirim, et al., Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Comput. Biol. Med.* (2020), 103792.
- [12] W. Linda, A. Wong, COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-Ray images, *arXiv preprint arXiv:2003.09871* (2020).
- [13] H.E. El-Din, M.A. Shouman, M.E. Karar, Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images, *arXiv preprint arXiv:2003.11055* (2020).
- [14] D. Kermany, K. Zhang, M. Goldbaum, Large dataset of labeled optical coherence tomography (OCT) and chest X-Ray images, *Mendeley Data* (2018), <https://doi.org/10.17632/rscbjbr9sj.3>.
- [15] M. Chowdhury, T. Rahman, A. Khandakar, et al., Can AI help in screening viral and COVID-19 pneumonia? *arXiv preprint arXiv:2003.13145* (2020). <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>.
- [16] N. Chen, M. Zhou, X. Dong, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (2020) 507–513.
- [17] C.J. Paul, P. Morrison, L. Dao, COVID-19 image data collection, *arXiv preprint arXiv:2003.11597* (2020). <https://github.com/ieee8023/covid-chestxray-dataset>.
- [18] D.S. Kermany, M. Goldbaum, W. Cai, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172 (5) (2018) 1122–1131.
- [19] S. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2009) 1345–1359.
- [20] O. Russakovsky, J. Deng, H. Su, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
- [21] P. Myeongsuk, S. Kim, A review of deep learning in image recognition, in: 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), IEEE, 2017, pp. 1–3.
- [22] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (2010) 807–814.
- [23] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [24] M.A. Kassem, K.M. Hosny, M.M. Fouad, Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning, *IEEE Access* 8 (2020) 114822–114832.
- [25] K. Gary, L. Zeng, Logistic regression in rare events data, *Political Anal.* 9 (2) (2001) 137–163.
- [26] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, *arXiv preprint arXiv:1712.04621* (2017).
- [27] M.L. McHugh, Interrater reliability: the kappa statistic, *Biochem. Med.* 22 (3) (2012) 276–282.
- [28] M. Heidari, A. Khuzani, A.B. Hollingsworth, et al., Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm, *Phys. Med. Biol.* 63 (3) (2018), 035020.
- [29] M. Heidari, S. Mirniaharikandehei, W. Liu, et al., Development and assessment of a new global mammographic image feature analysis scheme to predict likelihood of malignant cases, *IEEE Trans. Med. Imaging* 39 (4) (2020) 1235–1244.
- [30] X. Zhao, S. Qi, B. Zhang, Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning, *J. Xray Sci. Technol.* 27 (4) (2019) 615–629.
- [31] K. Wang, B.K. Patel, L. Wang, et al., A dual-mode deep learning transfer learning (D2TL) system for breast cancer detection using contrast enhanced digital mammograms, *Iise Trans. Healthc. Syst. Eng.* 9 (4) (2019) 357–370.