

## Article

# Pneumonia Detection from Chest X-ray Images Based on Convolutional Neural Network

Dejun Zhang <sup>1</sup>, Fuquan Ren <sup>1,\*</sup>, Yushuang Li <sup>1</sup>, Lei Na <sup>2</sup> and Yue Ma <sup>1</sup>

<sup>1</sup> School of Science, Yanshan University, Qinhuangdao 066004, China; ZhangDejun1997@163.com (D.Z.); yushuangli@ysu.edu.cn (Y.L.); mayue13203525596@163.com (Y.M.)

<sup>2</sup> Affiliated Hospital of Chengde Medical College, Chengde 067000, China; nalei625@163.com

\* Correspondence: renfu\_quan@ysu.edu.cn

**Abstract:** Pneumonia has caused significant deaths worldwide, and it is a challenging task to detect many lung diseases such as like atelectasis, cardiomegaly, lung cancer, etc., often due to limited professional radiologists in hospital settings. In this paper, we develop a straightforward VGG-based model architecture with fewer layers. In addition, to tackle the inadequate contrast of chest X-ray images, which brings about ambiguous diagnosis, the Dynamic Histogram Enhancement technique is used to pre-process the images. The parameters of our model are reduced by 97.51% compared to VGG-16, 85.86% compared to Res-50, 83.94% compared to Xception, 51.92% compared to DenseNet121, but increased MobileNet by 4%. However, the proposed model's performance (accuracy: 96.068%, AUC: 0.99107 with a 95% confidence interval of [0.984, 0.996], precision: 94.408%, recall: 90.823%, F1 score: 92.851%) is superior to the models mentioned above (VGG-16: accuracy, 94.359%, AUC: 0.98928; Res-50: accuracy, 92.821%, AUC, 0.98780; Xception: accuracy, 96.068%, AUC, 0.99623; DenseNet121: accuracy, 87.350%, AUC, 0.99347; MobileNet: accuracy, 95.473%, AUC, 0.99531). The original Pneumonia Classification Dataset in Kaggle is split into three sub-sets, training, validation and test sets randomly at ratios of 70%, 10% and 20%. The model's performance in pneumonia detection shows that the proposed VGG-based model could effectively classify normal and abnormal X-rays in practice, hence reducing the burden of radiologists.

**Keywords:** X-ray images; pneumonia detection; convolutional network; image enhance



check for updates

**Citation:** Zhang, D.; Ren, F.; Li, Y.; Na, L.; Ma, Y. Pneumonia Detection from Chest X-ray Images Based on Convolutional Neural Network. *Electronics* **2021**, *10*, 1512. <https://doi.org/10.3390/electronics10131512>

Academic Editors: Andrzej Czyżewski and Piotr Szczuko

Received: 27 May 2021  
Accepted: 17 June 2021  
Published: 23 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pneumonia is a lung parenchyma inflammation often caused by pathogenic microorganisms, factors of physical and chemical, immunologic injury and other pharmaceuticals. There are several popular pneumonia classification methods: (1) pneumonia is classified as infectious and non-infectious based on different pathogeneses in which infectious pneumonia is then classified to bacteria, virus, mycoplasmas, chlamydial pneumonia, and others, while non-infectious pneumonia is classified as immune-associated pneumonia, aspiration pneumonia caused by physical and chemical factors, and radiation pneumonia. (2) Pneumonia is classified as CAP (community-acquired pneumonia), HAP (hospital-acquired pneumonia) and VAP (ventilator-associated pneumonia) based on different infections, among which CAP accounts for a larger part. Because of the different range of pathogens, HAP is easier to develop resistance to various antibiotics, making treatment more difficult.

Pneumonia kills more than 800,000 children under five per year, with around 2200 deaths every day. There are more than 1400 children infected with pneumonia per 100,000 children [1]. The Global Burden of Disease Study reported that lower respiratory tract infections, including pneumonia, were the second largest cause of death in 2013. In Europe, nearly 35% of patients in hospital are infected with pneumococcal disease and worldwide, the percentage is 27.3% [2]. In India, the latest report of John Hopkins Bloomberg School of Public Health has said that India suffers the most pneumonia deaths and in 2015, there were about 2.97 lakh pneumonia and diarrhea deaths in children aged less than five years old.

In addition, in 2015 worldwide, pneumonia was the top killer of children under five years old [3]. Moreover, the death rate of pneumonia is highly related to age, and the prevalence of pneumonia increases dramatically with age, especially in people older than 65 [4]. The large number of child deaths by pneumonia alarms scientists worldwide to propose more effective and acute methods to detect pneumonia. With technology developing, more and more measures are developed, in which radiology-based methods are most popular and useful. Diagnostic radiological techniques for pulmonary disease include chest X-ray imaging, computed tomography (CT), and magnetic resonance imaging (MRI), among which chest X-ray imaging is most effective and economical as it is more available and portable in hospital and has lower exposures of dose radioactivity for patients.

However, even for very professional and experienced doctors, the diagnosis of pneumonia through X-ray images is still a tremendous task because X-ray images have similar region information for different diseases, such as lung cancer. Therefore, it is very time-consuming and energy-consuming to diagnose pneumonia through traditional methods and impossible to diagnose whether a patient suffers pneumonia through a standardized process. Hence, in this study, we propose a Convolutional Neural Network to diagnose pneumonia through X-ray images automatically and obtain results of accuracy 96.07% and Area Under Curve (AUC) 0.9911.

The rest of this paper is organized as follows. Section 2 describes related literature reviews of medical image processing methods. Section 3 describes a brief process of Convolutional Neural Networks (CNN) architecture in recent several years. Section 3 illustrated a summary of background of this machine learning and deep learning. Section 4 primarily illustrates the data utilized in this study, our proposed methods, and training procedure. Section 5 provides the experiments and results. Section 6 describes the conclusion of this study.

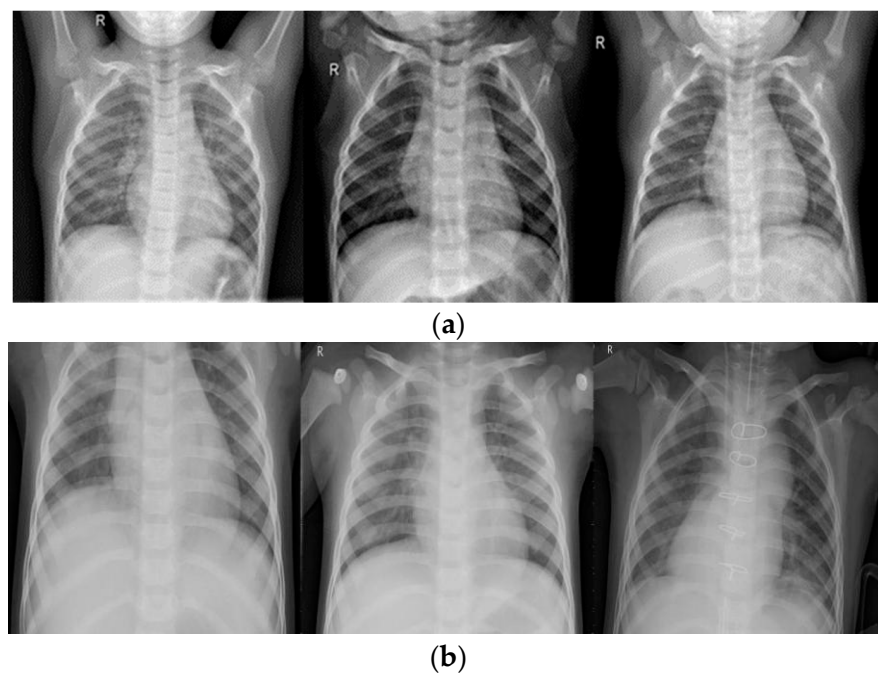
## 2. Related Work

Several methods have been introduced to describe a brief process in pneumonia detection using chest X-ray images in recent years, especially some deep learning methods. Deep Learning has been successfully applied to improve the performance of computer-aided diagnosis technology (CAD), especially in the field of medical imaging [5], image segmentation [6,7] and image reconstruction [8,9]. In 2017, Rajpurkar et al. [10] proposed a classical deep learning network named DenseNet-121 [11], which was a 121-layer CNN model to accelerate the diagnosis for pneumonia. In contrast to experienced doctors, the framework obtained a higher F1 score. Besides, in order to alleviate the effect of imbalanced classes, the team introduced Weighted Binary Cross-Entropy loss, whose difference between the Binary Cross Entropy loss was the different weights of imbalanced classes according to the number of each class.

However, the proposed loss did account for the different training difficulty levels of classes. In order to solve the problem of poor generalization ability caused by over-fitting and the problem of spatial sparseness caused by ordinary convolution operation, residual connection network [12] and dilated convolution [13] were used by Liang et al. [14] in the backbone network model. The final recall rate and F1 score of their model reached 96.7% and 92.7%, respectively. The CNN model proposed by Jain et al. [15] combined with transfer learning that effectively used the image features learned in large dataset, sped up the training procedure of the model and made it more difficult to fall into local minimum points. In addition, two models were proposed to train. Moreover, the dataset used by Jain et al. was from the world-famous organization and competition site named Kaggle, which contains numerous competitions and attracts passionate competitors to achieve a higher rank. The dataset is split into three subsets: the training subset to train the model, the validation subset to adjust parameters of model, and the test subset to verify the generalization ability of model.

Verma et al. [16] adopted several data pre-processing procedures and data augmentation methods, like a random rotation of images and a random translation of the image in horizontal and vertical, which enlarged the dataset and enhanced the representation ability of their CNN model. Finally, their model obtained an extremely outstanding accuracy. Ayan et al. [17] adopted transfer learning and fine-tuning to train two classical CNN models, Xception-Net and VGG16-Net, to classify images containing pneumonia. The authors [18] proposed four efficient CNN models, which were two pre-trained models, ResNet152V2 and MobileNetV2, a CNN architecture, and a Long Short-Term Memory (LSTM) network. In addition, they compared different parameters trained by each model. Their four models achieved results of more than 91% in accuracy, recall, F1-score, precision, and AUC.

As shown in Figure 1, invalid areas in chest X-ray images make little difference to experienced radiologists or to any proposed model and would decrease classification accuracy. In order to relieve the effect of invalid areas, attention-based models were widely adopted. Li et al. [19] proposed an improved Squeeze-and-Excitation Network (SENet) architecture to locate the pneumonia area in images. Guo et al. [20] developed a model that adaptively assigned one attention score for each dense connected layer and proposed an abnormal-aware attention module to make the network weight the learned low-level features into high-level features according to the importance of features. Furthermore, they introduced a novel angular contrastive loss to reduce the loss of intra-class and enlarge the loss of inter-class. Their method achieved 89.41% accuracy in WCE images.



**Figure 1.** Examples from the dataset. (a) normal cases, (b) pneumonia cases.

In order to make a summary about the task of pneumonia classification, Baltruschat et al. [21] compared the classification accuracy of currently widespread CNN models in pneumonia X-ray images by using the same hyper-parameter settings and same image pre-processing procedures. The method by Nahid et al. [22] proposed a novel CNN architecture which composed of two channels. The first channel processed the images, whose contrast was enhanced by the CLAHE method, while the second channel processed the images whose edges were enhanced by the Canny method. Then, these images were entered into a multichannel CNN model to detect if patients suffered from pneumonia. Researchers [23] developed a weak supervision approach to release a diagnosis burden of radiologists. They evaluated the model performance in a dataset of 30,000 chest X-ray images which

were collected by the Radiological Society of North America (RSNA). In addition, they compared the region of interest (ROI) predicted by their proposed architecture and the ROI ground truth bounding boxes provided by RSNA. Finally, they proposed several model architectures, Xception, ResNet-50, Inception and Ensemble (which meant that they used a weighted mean of three models above). Their single best model was Inception net, which obtained an accuracy of 0.782 and F1 score of 0.6411 on the binary classification in detecting pneumonia.

Other references in fields processed by CNN, like medical image segmentation, medical image reconstruction and so on, are also briefly described in this paper. The most widely used models in medical image segmentation are U-Net [6] and V-Net [7], which propose the idea of a fully connected neural network and stacked down-sampling layers followed by up-sampling layers. In addition, the networks directly connect the layers at the front to the layers behind to improve the ability of the model to extract features. The architecture [6], named U-net, consists of some stacked down-sampling convolution layers that adopt  $3 \times 3$  kernels followed by Rectified Linear unit (ReLU) and  $2 \times 2$  max pooling operation to obtain hidden feature maps. After down-sampling layers, there are several symmetric convolution layers the same as down-sampling layers to reconstruct original images. Their model design outperformed prior works on the ISBI (International Symposium on Biomedical Imaging) challenge and won the ISBI cell-tracking challenge 2015 by a large margin. Similarly, the model architecture [7] named V-net adapts stacked convolution down-sampling and up-sampling layers, whose difference between U-net is the bottom of model, in other words, like their names, U-net and V-net.

### 3. Background

In the past few decades, machine learning (ML) algorithms have gradually attracted researchers' attention. This type of algorithm could take full advantage of the giant computing power of calculators in images processing through given algorithms or specified steps. However, traditional ML methods in classification tasks need to manually design algorithms or manually set feature extraction layers to classify images.

In response to the above situation, LeCun et al. [24] proposed a CNN method, which could automatically extract features through continuously stacked feature layers and output the possibility of which class the input images belonged to. The shallow networks mainly focus on low-level features of the image. As the number of network layers increases, CNN model gradually extracts high-level features. Combining and analyzing these advanced features, CNN learns the differences between different images, and uses a back-propagation algorithm to update and record the learned parameters.

The essence of CNN is to filter the previous image or feature maps through a specific convolution kernel to generate the feature map of the next layer, and combine with operations such as pooling operations to reduce the feature map scale and reduce the computation. Then, a nonlinear activation function is added to the generated feature map to increase the characterization ability of the model. Common pooling operations include maximum pooling and average pooling. Maximum pooling means that the feature delivered into the pooling layer is split into a number of sub-regions and will output the maximum of each sub-region according to the strides in horizontal and vertical. The only difference between maximum pooling and average pooling is the output of the sub-region where the average pooling outputs the average of each sub-region. Common activation functions include ReLU (Rectified Linear Units) and Sigmoid.

For the classification task, through the continuous stacking of convolutional operations, pooling operations, nonlinear activation functions and other fully-connected layers, the features of the image are automatically extracted. Then possibility of whether the images processed by the model shows pneumonia is outputted by analyzing these extracted features. While making full use of inherent pixel-level information of the image, the generalization ability of model is improved. In recent decades, with the extensive development of deep learning, the most popular neural framework has been proposed,

such as AlexNet [25] and VGGNet [26]. However, when the number of network layers increases, the neural network will suffer from the problem of gradient disappearance or gradient explosion and will become fixated on the specific features of the training image, rather than learning more generative features, which makes the model's generalization ability poorer and causes over-fitting. In order to solve the problem of network depth, a residual connection structure was proposed. Since then, deeper neural networks have received extensive attention and development, and have become the cornerstone of many tasks. In this study, we also evaluated the performance of residual connection if used in our simplified CNN architecture with fewer layers.

#### 4. Materials and Methods

In this study, we explored: (1) the performance between original and contrast-enhanced images; (2) performance between different image resolutions; (3) performance between different loss functions, MSE (mean-square-error), BCE (binary-cross-entropy), FL (focal-loss) [27]. The comparison of trainable parameters in different models are listed in Table 1.

**Table 1.** Trainable Parameters.

Models	Trainable Parameters
VGG16	134,263,489
Res-50	23,628,673
Xception	20,808,425
DenseNet121	6,948,609
Our model	3,341,121
MobileNet	3,207,425

##### 4.1. Data

The proposed dataset used to evaluate model performance consists of total 5786 X-ray images and is from the Kaggle competition. In addition, the dataset is organized into three folders (train, test, val) and contains sub-folders for each image category (Pneumonia/Normal). All chest X-ray images (anterior-posterior) were obtained from patients aged one to five years old. In order to illustrate the generalization of different models, the original dataset was randomly split into three folders, provided with train validation, and tested by 70%, 10%, and 20%—as was done with the work by Baltruschat et al. [21].

All images were resized to a fixed shape, which in this study is a hyper-parameter from  $\Omega = \{50, 100, 224, 300\}$ . The training set includes a total of 4037 images, which were divided into normal and pneumonia sub-sets; the normal set contained 1093 images while the pneumonia set contained 2944 images. The validation set included a total of 579 images, which were divided into normal and pneumonia sub-sets; the normal set contained 157 images while the pneumonia set contained 422 images. The test set included a total of 1170 images; the normal set and the pneumonia set contain 316 and 854 images, respectively. Several examples of normal and pneumonia images are listed in Figure 1.

Chest X-ray images always have the characteristics of limited contrast on account of low-exposure dose to patients, for the reason that chest X-ray images always contain black, white, and gray components. The lungs are located on both sides of the chest cavity and the area of the lungs could be easily missed by X-rays, which are almost black. The heart, located in the middle of the lungs, is shown as almost white since the X-rays cannot completely pass through the heart. Bones are made of protein and they are too dense, so the X-rays cannot cross through them and bones are shown almost white as well. In addition, bones have clear edges.



Other tissues like the trachea and arteries are always visually unclear. The unclear information in chest X-ray images will cause an ambiguous diagnosis, not only to radiologists, but also to the CNN models. Therefore, in this study, the contrast of the original images would be enhanced and a comparison of the performance of enhanced images between non-enhanced images is illustrated.

## 4.2. Methods

### 4.2.1. Data Pre-Processing

Image pre-processing is a very common and beneficial technique in the deep learning process and it not only could enlarge the quantity of the original dataset but also enrich the information implicit in the dataset. As previously mentioned, we utilized an effective image enhancement method named Dynamic Histogram Equalization (DHE) to improve the quality of images before they were inputted into the CNN model.

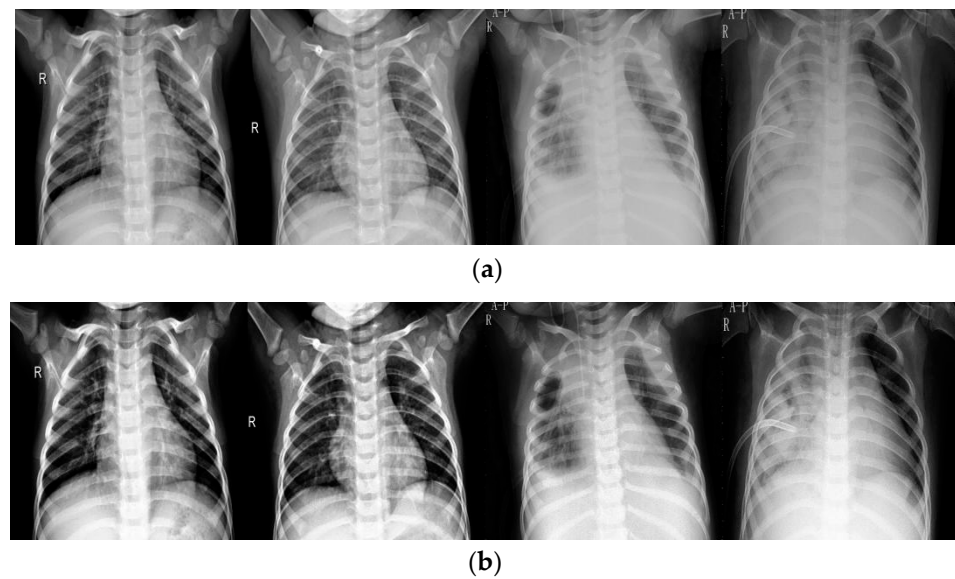
Histogram Equalization (HE), which denotes mapping from the initial narrow pixel levels to a wider extent and improves image enhancement, has been widely used in image processing. The HE technique means to convert the gray levels of an image by using cumulative effort function globally, yet always brings about the problem that elaboration information in images is damaged, leading to awful image quality. This popular image contrast enhancement method could enhance image contrast effectively in many aspects, like MRI, X-rays, CT and so on.

Scientists [28] aiming to classify brain MRI images accurately proposed a novel enhancement method called Average Intensity Replacement-Adaptive Histogram Equalization (AIR-AHE). The proposed model consisted of several existing methods to improve enhancement performance, such as contrast enhancement, window-sliding neighborhood operation and so on. Their experiments demonstrated the proposed combined enhancement methods could obtain better results compared to the traditional HE method. Another MRI brain image enhancement method [29] proposed a novel framework named particle swarm optimized texture-based histogram equalization (PSOTHE), which combs the multi-objective fitness function to obtain better enhancement performance in brain MRI images. In their study, different objective functions consisting of edge information, entropy, contrast, and the PSRN of the image had the same weight to enhance the images.

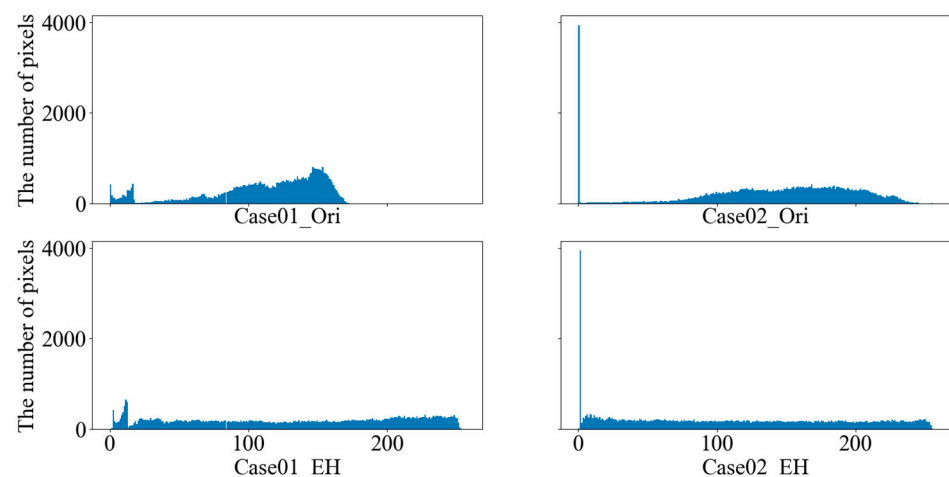
In this study, we follow Wadud et al. [30], using their adaptive DHE in order to pre-process pneumonia chest X-ray images before inputting them into the proposed CNN model. The DHE technique eliminates the possibility of the low histogram components being compressed, which may cause the information of the image to wash out. Moreover, DHE is free from any severe side effects such as checkerboard effect, over-enhancing noises, etc.

Figure 2 provides several examples between the original and enhanced images. Figure 2a represents examples from original data set, while Figure 2b represents examples from the enhanced data set. As shown, the upper images always contain gray and vague areas while lower images display almost white and black areas, showing that the enhancement technique plays a significant role in image contrast. In this study, the information implicit in original images, like bones, lung areas, and other tissues, is revealed by the DHE technique.

Figure 3 displays the histograms of original and enhanced images. The upper row depicts histograms of several original images, while the lower row depicts histograms of corresponding enhanced images by the DHE technique. As shown, the scopes of image pixel levels are extended, which means the contrast of images is enhanced effectively.



**Figure 2.** Comparison of original and enhanced images. (a) Examples of original images; (b) examples of enhanced images.



**Figure 3.** Histogram comparison between original and enhanced images. Ori means original images and EH means images enhanced by DHE. As shown, the scopes of image pixel levels are extended, which means the contrast of images is enhanced effectively.

Other techniques used throughout this paper are tabulated in Table 2. In our experiment, the image size was fixed to  $\text{SHAPE} \times \text{SHAPE}$  where SHAPE is a hyper-parameter, and in this study, we explored the performance of different shapes and sizes, and the pixel scale of image was resized to  $[0, 1]$  by dividing the maximum of image pixel. Subsequently, the channel of images was set to  $[\text{SHAPE}, \text{SHAPE}, 1]$ , as the information was already improved by DHE and the image pixel was scaled to  $[0, 1]$ . That is, there was no need to extend the shape of images to  $[\text{SHAPE}, \text{SHAPE}, 3]$  as used in other methods. In addition, several data augmentation methods were applied to boost the robustness of the proposed model. The rotation range denotes random image rotation during the training procedure in the range  $[0, 30]$ . The zoom range was utilized to randomly rescale the image size by 0.2 percent. Horizontal and vertical flip was used to randomly flip the image horizontally. Also, the width and height shift ranges are used to shift the image horizontally or vertically. These operations could intensely mitigate the negative effect of different image angles, different image heights, and different image widths.

**Table 2.** Data pre-processing techniques used in this study.

Methods	Setting
Resize	224 × 224
Normalization	[(0, 255) → (0, 1)]
Rotation Range	[0, 30]
Zoom Range	0.2
Weight_Shift_Range	0.1
Height_Shift_Range	0.1
Horizontal_Flip	True
Vertical_Flip	True

Also, the performance of different loss functions, MSE and Focal Loss, is discussed in the following section.

#### 4.2.2. Proposed Network

In this study, we designed a VGG-based CNN model to extract the features of chest X-ray images and use those features to detect if a patient suffers from pneumonia.

Let  $x_i$  be the input or the  $i$ th hidden layer feature map,  $1 \leq i \leq N$ ,  $N$  represents all layers of the model and in our architecture, set  $N = 6$ , meaning our model contains six hidden layers.  $C_i(\cdot)$ ,  $1 \leq i \leq 4$ , represents convolution operation whose kernel size and strides are  $3 \times 3$  and  $1 \times 1$ , respectively.  $W_i$ ,  $1 \leq i \leq N$ , represents the weight matrix of the  $i$ th layer;  $b_i$  represents bias value of the  $i$ th layer;

$$\psi_i = \begin{cases} \text{ReLU}, & 1 \leq i \leq 5, \\ \text{Sigmoid}, & i = 6 \end{cases} \quad (1)$$

where Sigmoid and ReLU are activation functions.  $D_i(\cdot)$ ,  $2 \leq i \leq 5$  represents the drop operation, which aims to randomly set the value of hidden layers or neuron to zero and improve the performance of model.  $M_i(\cdot)$ ,  $1 \leq i \leq 4$  represents the max-pooling operation, aiming to reduce the computation of the model and improve its robustness;  $FC_i(\cdot)$ ,  $5 \leq i \leq 6$  represents the fully-connected operation. So the  $i + 1$ th output of model is:

$$x_{i+1} = M_i(D_i(\psi_i(C_i(W_i x_i + b_i)))) \quad (2)$$

The last layer outputs final decision  $y \in \{0, 1\}$ , indicating the absence or presence of disease, respectively.

#### 4.3. Training Process

##### 4.3.1. Classification Evaluation Metrics

In this subsection, several evaluation metrics, accuracy, precision, recall, F1 score and so on, are described. According to the outputs of model, four indices, True Positive, True Negative, False Positive, False Negative, are used to analyze and identify the performance of model. The True Positive means that the chest X-ray images, which suffer from pneumonia, are signed as pneumonia as well by the model. The True Negative means if the chest X-ray images do not show pneumonia as well as the model predicts.

The remaining matrices have a similar definition.

The four metrics are given as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$



$$\text{recall} = \frac{\text{TP}}{\text{FP} + \text{FN}} \quad (5)$$

$$\text{F1} = 2 \times \frac{\text{precision}}{\text{precision} + \text{recall}} \quad (6)$$

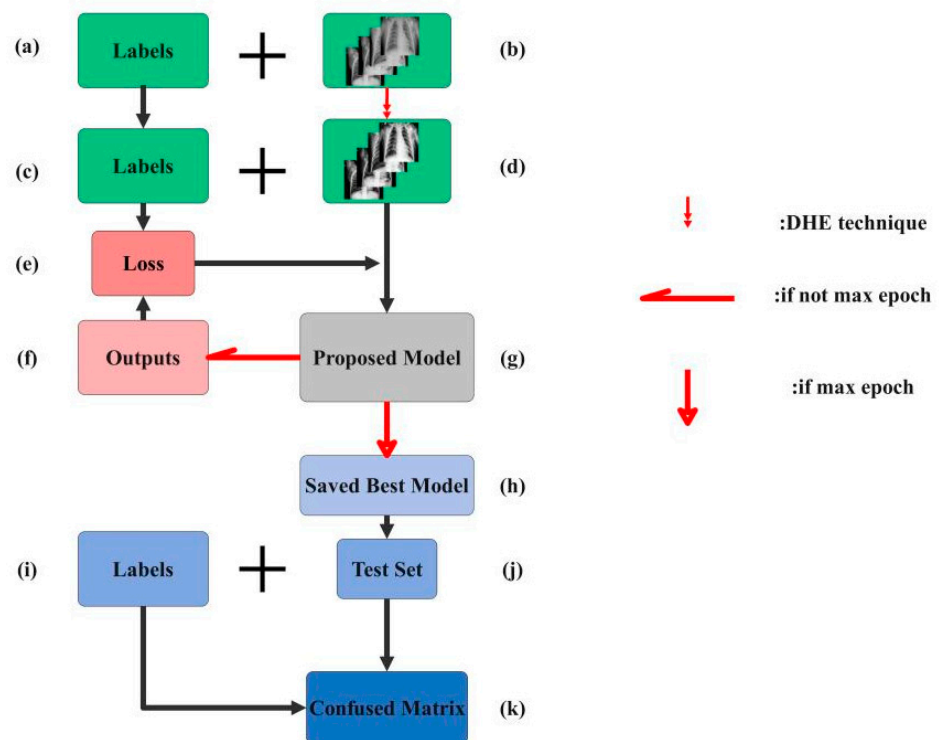
TP, TN, FP and FN mentioned by the above formulas are True Positive, True Negative, False Positive, and False Negative, respectively. Among the four metrics, the precision rate was always used to estimate how much the number of images that are truly pneumonia accounted for in the total number examples, which are classified as positive for pneumonia. That is, the pneumonia images must be identified in practical clinical diagnoses and hence, the precision rate is especially important. In most cases, the higher the precision rate gets, the lower the recall rate is. Thus, F1 score rate is widely considered as a proper criterion. In addition, the Receiver Operating Characteristic (ROC) and AUC are calculated to compare the performance of different models. The whole training procedure will be displayed in this section. First, the hyper-parameters are fixed and shown in Table 3. In our architecture, we optimize the model by Adam optimizer, an algorithm for stochastic optimization; the learning rate and the learning rate decay per epoch were set to 0.001 and 0.0001. Then, the hidden layer activation function and last classification activation function are set to ReLU and Sigmoid, respectively.

**Table 3.** Hyper-parameters of model.

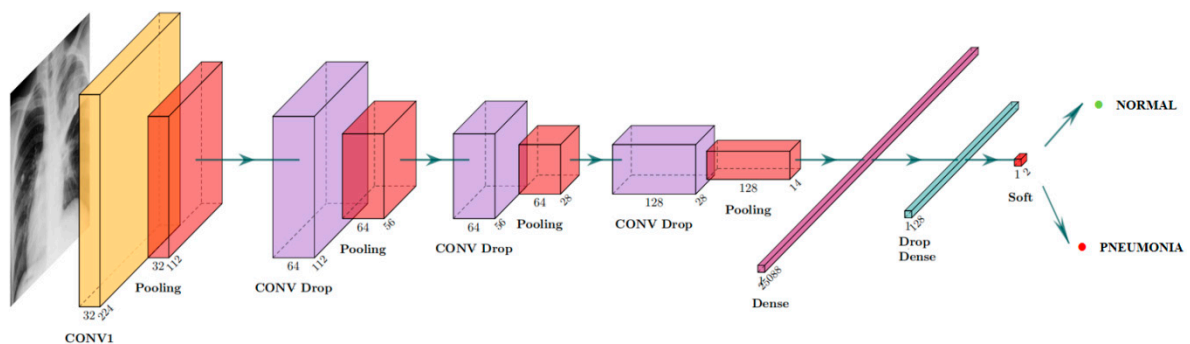
Parameters	Value
Optimizer	Adam
Learning Rate	0.001
Learning Rate Decay Per Epoch	0.0001
Batch Size	16
Hidden Layer Activation Function	ReLU
Classification Activation Function	Sigmoid

#### 4.3.2. Proposed CNN Model

Figure 4 illustrates the architecture of our proposed model that has been applied for the detection of whether the input image shows pneumonia. Figure 5 displays our model that contains a total of six layers, where we employed  $3 \times 3$  kernel convolution layers whose strides are  $1 \times 1$  and the activation function is ReLU. After each convolution layer, a  $2 \times 2$  strides kernel operation was employed as a max-pooling operation to retain the maximum of each sub-region, which is split according to strides. Besides, we set several drop layers to randomly fit weights to zero, aiming to improve the model performance. Then two densely fully-connected layers followed by Sigmoid function are utilized to take full advantage of the features extracted through previous layers, outputting the possibility of patients suffering from pneumonia or not. As illustrated above, the input channel is  $224 \times 224 \times 1$  and the output size is  $y \in \{0, 1\}$ , where 0 denotes that the image does not show pneumonia, while 1 denotes that the image shows pneumonia.



**Figure 4.** The workflow of our proposed Deep Learning architecture. (a,c) The labels of the training subset where 0 denotes the images that do not show pneumonia, while 1 denotes the images with pneumonia; (b) the original train subset and (d) the enhanced train subset processed by the DHE technique. (e) The loss used in this study; (g) is the DL model proposed; (f) is the outputs of the model where 0 denotes images without pneumonia while 1 denotes images showing pneumonia; (h) is the best model saved after training; (j) is the original test subset; (i) is the labels of the test subset where 0 denotes the images without pneumonia, while 1 denotes the images showing pneumonia; (k) is the confusion matrix to evaluate the performance of the model.



**Figure 5.** Details of proposed DL model.

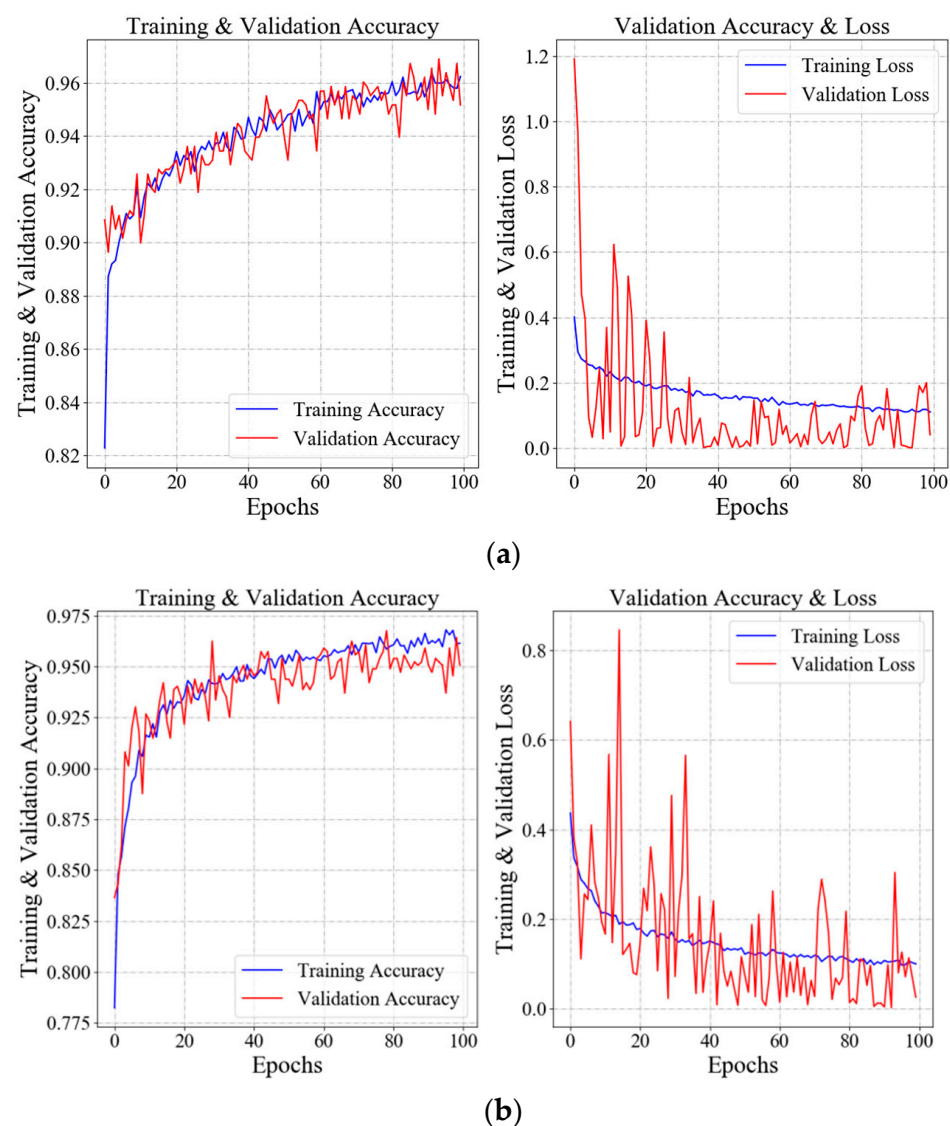
#### 4.3.3. t-SNE Visualization

The t-SNE (t-distributed Stochastic Neighbor Embedding) visualization method is also displayed in this paper. This method aims to visualize the distribution of high dimensions and assists us in analyzing the connection among the samples.

## 5. Experiments and Results

### 5.1. Experiments

Figure 6 shows the training and validation accuracy as well as the training and validation loss obtained during the training procedure for 100 epochs with our proposed model and 50 epochs with other model for the reason that the number of parameters of the other model are times than ours. The training accuracy and validation accuracy all exceeded 95%, also the training loss and validation loss were below 0.025, which means that the model we proposed has the potential power to acquire excellent performance. Furthermore, the training accuracy and training loss shows an apparent decreasing tendency during training. As a result, the best model was saved after training the last epoch. In our experiments, the loss tendency shows an instability of oscillatory probably according to the mini batch size limited by our GPU computation source.



**Figure 6.** Training procedure. (a) Accuracy and loss in training process of enhanced model with our proposed model and BCE loss function using enhanced images; (b) Accuracy and loss in training process of the original model with our proposed model and BCE loss using original images.

## 5.2. Results

### 5.2.1. Comparison of Different Models and Different Shapes

Table 4 illustrates a summary of performance of different models including our proposed model with and without residual framework, VGG, ResNet50, MobileNet, Inception and DenceNet121 with input shape 224 and loss function BCE. It can be seen that the performance of models pre-processed by contrast enhancement technology exceeds original non-enhancement models. In addition, the table shows that the best results accuracy 0.9607 and precision 0.9441 are obtained by our proposed model. Also, it can be seen that model performance pre-processed by contrast enhancement technology obtained accuracy above 0.94 except DenceNet121, which is 0.9342 and model performance with the original non-enhancement models, obtained by the best accuracy of 0.953 by our proposed model, inflecting the effect of enhancement methods in CNN models. Yet, the enhanced model accuracy of VGG and Res is below the original non-enhanced models. This is possibly caused by that the number of model parameters affects the performance of different models. In other words, the model with fewer parameters is more likely affected by enhancement methods; for example, our proposed model and MobileNet with fewer parameters all show a trend that model results of enhanced methods exceed non-enhanced methods. Another trend shown in Table 4 is that the best results, accuracy, precision, recall, F1 and AUC, are all obtained by models pre-processed by enhanced methods. The best accuracy and precision are obtained by our proposed model. The best recall rate is obtained by DenseNet121, which is 1, and the best F1 and AUC scores are obtained by Inception and DenseNet121, respectively.

**Table 4.** Summary of performance of different models with input shape 224 and loss function BCE.

		Accuracy	Precision	Recall	F1 Score	AUC
Our model	original	0.9538	0.8764	0.9652	0.9187	0.9910
	enhanced	0.9607	0.9441	0.9082	0.9258	0.9911
Our-model-Res	original	0.9453	0.8481	0.9715	0.9056	0.9877
	enhanced	0.9598	0.8899	0.9715	0.9289	0.9917
VGG-16	original	0.9479	0.8512	0.9778	0.9102	0.993
	enhanced	0.9436	0.8511	0.9589	0.9018	0.9893
ResNet50	original	0.9496	0.9132	0.8987	0.9059	0.9825
	enhanced	0.9427	0.8374	0.9778	0.9022	0.9888
MobileNet	original	0.9453	0.8706	0.93670	0.9024	0.9839
	enhanced	0.9547	0.8663	0.9842	0.9215	0.9953
Inceptionv3	original	0.9470	0.8567	0.9652	0.9077	0.9921
	enhanced	0.9589	0.8895	0.9684	0.92723	0.9924
DenseNet121	original	0.9137	0.7694	0.9715	0.8587	0.9845
	enhanced	0.9342	0.8041	1.0000	0.8914	0.9957
[14]	original	0.9050	0.8910	0.9670	0.9270	0.9530
[15]-Model 1	Original	0.8526	0.7500	0.9400	0.8900	–
[15]-Model 2	Original	0.9231	0.8700	0.9800	0.9400	–
[32]-Architecture 1	Original	0.9359	–	–	–	–
[32]-Architecture 2	Original	0.9263	–	–	–	–
[32]-Architecture 3	Original	0.9231	–	–	–	–
[32]-Architecture 4	Original	0.9054	–	–	–	–
[32]-Architecture 5	Original	0.9022	–	–	–	–

Table 5 illustrates a summary of model performance of different input shapes with our proposed model and loss function BCE. The table also shows the trend that the model performance pre-processed by enhanced methods exceeds non-enhanced methods. The best model results are also obtained by enhanced models. The best accuracy and precision are obtained by our proposed model with input shape 224. The best recall rate is obtained by our model with shape 300 and the best F1 score and AUC are obtained by our model with input shape 50. Also, the comparison between different loss functions is listed in

Table 6, which showing the same trend as Table 4; Table 5, which inflects the effect of contrast enhancement methods.

**Table 5.** Summary of performance of different input shapes with our proposed model and loss function BCE.

		Accuracy	Precision	Recall	F1 Score	AUC
50 × 50	original	0.9547	0.8997	0.9367	0.9178	0.9886
	enhanced	0.9589	0.8851	0.9747	0.9277	0.9948
100 × 100	original	0.9325	0.8143	0.9715	0.8860	0.9903
	enhanced	0.9529	0.8718	0.9684	0.9175	0.9926
224 × 224	original	0.9538	0.8764	0.9652	0.9187	0.9910
	enhanced	0.9607	0.9441	0.9082	0.9258	0.9911
300 × 300	original	0.9291	0.8074	0.9684	0.8806	0.9875
	enhanced	0.9556	0.8729	0.9778	0.9224	0.9888

**Table 6.** Summary of performance of different loss functions with our proposed model and input shape 224.

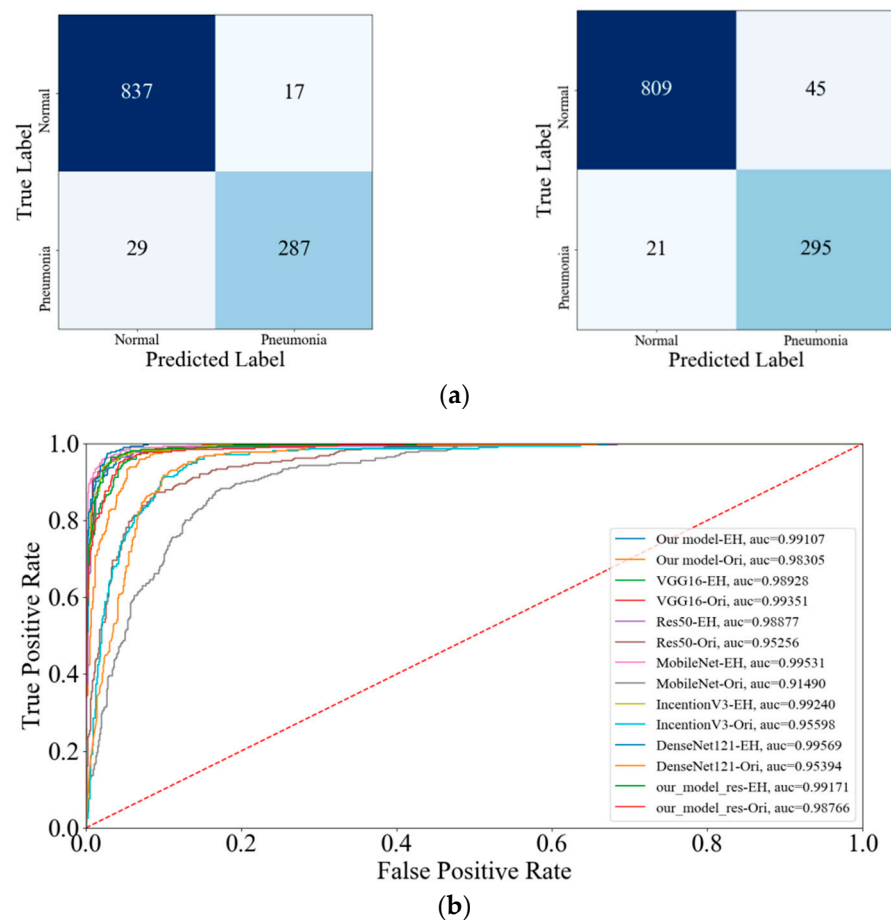
			Accuracy	Precision	Recall	F1 Score	AUC	
MSE	original		0.9564	0.8909	0.9557	0.9221	0.9872	
	enhanced		0.9556	0.8771	0.9715	0.9219	0.9929	
BCE	original		0.9538	0.8764	0.9652	0.9187	0.9910	
	enhanced		0.9607	0.9441	0.9082	0.9258	0.9911	
$\gamma = 1$	$\alpha = 0.65$	original	0.9479	0.8653	0.9557	0.9083	0.9884	
		enhanced	0.9573	0.9236	0.9177	0.9206	0.9899	
	$\alpha = 0.75$	original	0.9368	0.8253	0.9715	0.8924	0.9900	
		enhanced	0.9487	0.8657	0.9589	0.9099	0.9851	
	$\alpha = 0.85$	original	0.9043	0.7488	0.9715	0.8457	0.9869	
		enhanced	0.9145	0.7621	0.9937	0.8626	0.9932	
Focal loss	$\gamma = 2$	$\alpha = 0.65$	original	0.9393	0.8375	0.9620	0.8954	0.9903
		enhanced	0.9573	0.8800	0.9747	0.9249	0.9929	
	$\alpha = 0.75$	original	0.9308	0.9898	0.8167	0.9589	0.8821	
		enhanced	0.9538	0.8639	0.9842	0.9201	0.9936	
	$\alpha = 0.85$	original	0.8974	0.7311	0.9810	0.8378	0.9879	
		enhanced	0.9179	0.7709	0.9905	0.8670	0.9908	
	$\alpha = 0.65$	original	0.9111	0.7636	0.9715	0.8551	0.9847	
		enhanced	0.9436	0.8397	0.9778	0.9035	0.9919	
	$\gamma = 3$	$\alpha = 0.75$	original	0.9256	0.7989	0.9684	0.8755	0.9843
			enhanced	0.9256	0.7913	0.9842	0.8773	0.9904
		$\alpha = 0.85$	original	0.9299	0.8095	0.9684	0.8818	0.9900
			enhanced	0.9333	0.8099	0.9842	0.8886	0.9921

### 5.2.2. Confusion Matrix and ROC

The predicted probabilities of CNN models by stacked convolution layers combined with pooling layers and ReLU activation or Sigmoid function were calculated by Receiver Operating Characteristic Curve (ROC curve), which is shown in Figure 7. It shows that among the total of 316 pneumonia images, only 29 samples were misclassified and 17 samples from true normal images, which are 854, were predicted as pneumonia. A total of 46 images among 1170 images from the test sub-set were classified incorrectly.

The difference in discriminative capacity between each individual risk score obtained by each CNN model was evaluated by the DeLong test [31], where  $p < 0.05$  was statistically significant. In this paper, the diagnosis efficacy of ROC by our proposed model was higher than other CNN models in test subsets and the difference was statistically significant ( $p = 0.0061 < 0.05$ ).





**Figure 7.** Confusion matrix and ROC curve of model. (a) The confusion matrix of model; (b) the ROC curve and AUC value of different models where Ori denotes the model using original images and EH denotes the model using enhanced images.

### 5.2.3. Other Evaluations

The TP, TN, FP and FN were also calculated. Liang et al. [14] reported a classification accuracy rate of 90.5%. The study described by Jain et al. [15] proposed two CNN architectures, which obtained 85.26% and 92.31% accuracy rates, respectively. Although the proposed Model 2 by Jain et al. [15] obtained a 98% recall rate, which means the FN obtained is lower than our model, Model 2 obtained a higher FP, which is 42 compared to our 29. Ayan et al. [17] used two well-known CNN networks, Xception-Net and VGG16-Net, combined with transfer learning and fine-tuning to diagnose pneumonia. Varalakshmi et al. [32] proposed five CNN model architectures to classify whether the images suffer pneumonia or not.

These methods evaluated the model performance in the original test subset, which contained 624 images. Other methods changed the original data set split to training, test and validation subsets by a fixed ratio. For example, Nahid et al. [22] took a different approach to split original samples into new sample sets by 70% and 30%, and finally obtained excellent results. Stephen et al. [33] rearranged the entire data set into train and validation sets only and obtained a validation accuracy of 93.73%. Rajpurkar et al. [10] presented the biggest available X-ray image set named ChestX-ray8, which contains 112,120 frontal chest X-ray images. Ma et al. [34] proposed a novel multi-attention CNN model for disease diagnosis and location on the ChestX-ray8 dataset. Xu et al. [35] proposed a new model named CXNet-m1, which was thinner and more powerful than fine-tuning, and obtained outstanding results.

Figure 8 gives the t-SNE of the test set. First, there is no clear boundary of original non-enhanced images between the normal and pneumonia subsets (PNE-Ori and NORMAL-Ori

respectively in Figure 8), which would obscure the detection of human radiologists or other automatic models. Second, enhanced by DHE, the distribution of normal and pneumonia subsets (PNE-EH and NORMAL-EH respectively in Figure 8) is almost independent, showing that the contrast-enhanced method provides effective results in medical images classification tasks.



**Figure 8.** t-SNE of the test set where Ori denotes model using original images and EH denotes the model using enhanced images.

## 6. Conclusions

This study describes a CNN-based model aiming to diagnose pneumonia on a chest X-ray image set. The contributions in this paper are listed as follows. First, we utilized the Dynamic Histogram Equalization (DHE) technique to enhance the image contrast. This technique has the potential to enhance image contrast without washing out appearance or causing problems like checkerboard effects. Then, we designed a simple VGG-based CNN model to extract the features from original images or previous feature maps, which contained only six layers combining ReLU activation function, drop operation, and max-pooling layers. The results of the obtained accuracy rate of 96.07% and precision rate of 94.41%, shows that our proposed model performs well in comparison to state-of-the-art CNN model architectures. To illustrate the performance of our proposed model, several comparisons of different input shapes and loss functions were provided.

In the future, we will continue the research to explore more accurate classification architectures to diagnose two types of pneumonia, viruses, and bacteria. According to the description discussed above, the CNN-based model is a promising method to diagnose the disease through X-rays.

**Author Contributions:** Data analysis, validation and writing—original draft preparation, D.Z.; Study design, writing—review and editing, project administration and funding acquisition, F.R.; Writing—review and methodology, Y.L.; Medical data interpretation, L.N.; Literature search, Y.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China Grant No. 61807029, and Natural Science Foundation of Hebei Province Grant No. F2019203427, A2020203021.

**Data Availability Statement:** The dataset processed in this article is from the public Mendeley database [36] (<https://doi.org/10.17632/rscbjbr9sj.2> (accessed on 11 May 2021)) as well as Kaggle at <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> (accessed on 11 May 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. UNICEF. Available online: <https://data.unicef.org/topic/child-health/pneumonia/> (accessed on 22 March 2018).
2. Prina, E.; Ranzani, O.T.; Torres, A. Community-acquired pneumonia. *Lancet* **2015**, *386*, 1097–1108. [[CrossRef](#)]
3. Tilve, A.; Nayak, S.; Vernekar, S.; Turi, D.; Shetgaonkar, P.R.; Aswale, S. Pneumonia Detection Using Deep Learning Ap-proaches. In Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020; pp. 1–8.

4. Periselneris, N.J.; Brown, S.J.; José, R.J. Pneumonia. Available online: [https://www.medicinejournal.co.uk/article/S1357-3039\(20\)30049-9/fulltext](https://www.medicinejournal.co.uk/article/S1357-3039(20)30049-9/fulltext) (accessed on 23 April 2021).
5. Le, W.T.; Maleki, F.; Romero, F.P.; Forghani, R.; Kadoury, S. Overview of machine learning: Part 2. *Neuroimaging Clin. N. Am.* **2020**, *30*, 417–431. [[CrossRef](#)] [[PubMed](#)]
6. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
7. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
8. Jeelani, H.; Martin, J.; Vasquez, F.; Salerno, M.; Weller, D. Image quality affects deep learning reconstruction of MRI. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 357–360.
9. Chlemper, J.S.; Caballero, J.; Hajnal, J.; Price, A.N.; Rueckert, D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* **2017**, *37*, 491–503. [[CrossRef](#)] [[PubMed](#)]
10. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.
11. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
13. Yu, F.; Koltun, V.; Funkhouser, T. Dilated Residual Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 25–30 June 2017; pp. 636–644.
14. Liang, G.; Zheng, L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput. Methods Programs Biomed.* **2020**, *187*, 104964. [[CrossRef](#)] [[PubMed](#)]
15. Jain, R.; Nagrath, P.; Kataria, G.; Kaushik, V.S.; Hemanth, D.J. Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. *Measurement* **2020**, *165*. [[CrossRef](#)]
16. Verma, D.; Bose, C.; Tufchi, N.; Pant, K.; Tripathi, V.; Thapliyal, A. An efficient framework for identification of Tuberculosis and Pneumonia in chest X-ray images using Neural Network. *Procedia Comput. Sci.* **2020**, *171*, 217–224. [[CrossRef](#)]
17. Ayan, E.; Unver, H.M. Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 24–26 April 2019; pp. 1–5.
18. Elshennawy, N.M.; Ibrahim, D.M. Deep-pneumonia framework using deep learning models based on chest X-ray images. *Diagnostics* **2020**, *10*, 649. [[CrossRef](#)] [[PubMed](#)]
19. Li, B.; Kang, G.; Cheng, K.; Zhang, N. Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-Rays. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4851–4854.
20. Guo, X.; Yuan, Y. Triple ANet: Adaptive Abnormal-aware Attention Network for WCE Image Classification. In Proceedings of the 19th International Conference on Application of Concurrency to System Design (ACSD 2019), Aachen, Germany, 23–28 June 2019; pp. 293–301.
21. Baltruschat, I.M.; Nickisch, H.; Grass, M.; Knopp, T.; Saalbach, A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. *Sci. Rep.* **2019**, *9*, 1–10. [[CrossRef](#)] [[PubMed](#)]
22. Nahid, A.-A.; Sikder, N.; Bairagi, A.K.; Razzaque, A.; Masud, M.; Kouzani, A.Z.; Mahmud, M.A.P. A Novel Method to Identify Pneumonia through Analyzing Chest Radiographs Employing a Multichannel Convolutional Neural Network. *Sensors* **2020**, *20*, 3482. [[CrossRef](#)] [[PubMed](#)]
23. Odaibo, D.; Zhang, Z.; Skidmore, F.; Tanik, M. Detection of visual signals for pneumonia in chest radiographs using weak supervision. *2019 SoutheastCon* **2019**, 1–5. [[CrossRef](#)]
24. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [[CrossRef](#)]
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
27. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
28. Isa, I.S.; Sulaiman, S.N.; Mustapha, M.; Karim, N.K.A. Automatic contrast enhancement of brain MR images using Average Intensity Replacement based on Adaptive Histogram Equalization (AIR-AHE). *Biocybern. Biomed. Eng.* **2017**, *37*, 24–34. [[CrossRef](#)]
29. Acharya, U.K.; Kumar, S. Particle swarm optimized texture based histogram equalization (PSOTHE) for MRI brain image enhancement. *Optik* **2020**, *224*, 165760. [[CrossRef](#)]

30. Wadud, M.A.A.; Kabir, M.H.; Dewan, M.A.A.; Chae, O. A Dynamic Histogram Equalization for Image Contrast Enhancement. In Proceedings of the 2007 Digest of Technical Papers International Conference on Consumer Electronics, Las Vegas, NV, USA, 10–14 January 2007; pp. 1–2.
31. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **1988**, *44*, 837–845. [[CrossRef](#)] [[PubMed](#)]
32. Varalakshmi, P.; Yuvaraj, G.; Dhanasekaran, K.; Samannan, K. Diminishing fall-out and miss-rate in the classification of lung diseases using deep learning techniques. In Proceedings of the 2018 Tenth International Conference on Advanced Computing (ICoAC), Chennai, India, 13–15 December 2018; pp. 373–376.
33. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep learning approach to pneumonia classification in healthcare. *J. Healthc. Eng.* **2019**, *2019*, 1–7. [[CrossRef](#)] [[PubMed](#)]
34. Ma, Y.; Zhou, Q.; Chen, X.; Lu, H.; Zhao, Y. Multi-attention Network for Thoracic Disease Classification and Localization. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1378–1382.
35. Xu, S.; Wu, H.; Bie, R. CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning. *IEEE Access* **2019**, *7*, 4466–4477. [[CrossRef](#)]
36. Kermany, D.; Zhang, K.; Goldbaum, M. Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images. Available online: <https://data.mendeley.com/datasets/rscbjbr9sj/2> (accessed on 11 May 2021).