

## A Classification–Detection Approach of COVID-19 Based on Chest X-ray and CT by Using Keras Pre-Trained Deep Learning Models

Xing Deng<sup>1,2</sup>, Haijian Shao<sup>1,2,\*</sup>, Liang Shi<sup>3</sup>, Xia Wang<sup>4,5</sup> and Tongling Xie<sup>6</sup>

<sup>1</sup>School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, 212003, China

<sup>2</sup>School of Automation, Key Laboratory of Measurement and Control for CSE, Ministry of Education, Southeast University, Nanjing, 210096, China

<sup>3</sup>School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, 212003, China

<sup>4</sup>School of Information Science and Technology, Nantong University, Nantong, 226019, China

<sup>5</sup>Department of Electrical and Computer Engineering, National University of Singapore, 117583, Singapore

<sup>6</sup>The People's Hospital of RUGAO, Nantong, 226500, China

\*Corresponding Author: Haijian Shao. Email: jsj\_shj@just.edu.cn

Received: 05 June 2020; Accepted: 07 September 2020

**Abstract:** The Coronavirus Disease 2019 (COVID-19) is wreaking havoc around the world, bring out that the enormous pressure on national health and medical staff systems. One of the most effective and critical steps in the fight against COVID-19, is to examine the patient's lungs based on the Chest X-ray and CT generated by radiation imaging. In this paper, five keras-related deep learning models: ResNet50, InceptionResNetV2, Xception, transfer learning and pre-trained VGGNet16 is applied to formulate an classification–detection approaches of COVID-19. Two benchmark methods SVM (Support Vector Machine), CNN (Convolutional Neural Networks) are provided to compare with the classification–detection approaches based on the performance indicators, i.e., precision, recall, F1 scores, confusion matrix, classification accuracy and three types of AUC (Area Under Curve). The highest classification accuracy derived by classification–detection based on 5857 Chest X-rays and 767 Chest CTs are respectively 84% and 75%, which shows that the keras-related deep learning approaches facilitate accurate and effective COVID-19-assisted detection.

**Keywords:** COVID-19 detection; deep learning; transfer learning; pre-trained models

### Notation

<b>AI:</b>	Artificial Intelligence;
<b>AUC:</b>	Area Under Curve;
<b>AUC_LOG:</b>	ROC Curves for Logistic Regression Classification;
<b>AUC_NB:</b>	ROC Curves for Naive Bayes classification;
<b>AUC_SVM:</b>	ROC Curves for SVM classification;
<b>ATC:</b>	Average Time cost in Seconds;



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

<b>CNN:</b>	Conventional Neural Networks;
<b>COVID-19:</b>	Coronavirus Disease 2019;
<b>DCNN:</b>	Deep Convolutional Neural Network;
<b>FN:</b>	False Negative;
<b>FP:</b>	False Positive;
<b>MSE:</b>	Mean Squared Error;
<b>PTVGG16:</b>	Pre-trained VGG16;
<b>TLVGG16:</b>	Transfer learning VGG16;
<b>TN:</b>	True Negative;
<b>TP:</b>	True Positive;
<b>RL:</b>	Reinforcement learning;
<b>SARS-CoV-2:</b>	Severe Acute Respiratory Syndrome Coronavirus 2;
<b>SVM:</b>	Support Vector Machine;
<b>SSIM:</b>	Structural Similarity Index;
<b>VGG16:</b>	VGG with 16-weights layers;

## 1 Introduction

Using the chest X-ray for radiation imaging to examine the patient's lungs is one of the most effective and critical steps in the fight against COVID-19 [1–7]. Many deep learning-based artificial intelligence (AI) systems have been proposed, and the results show that the chest X-ray and CT images benefit the detection of COVID-19 infection patients, and can effectively improve the accuracy of detection. Bernheim et al. [8] analyzed the COVID-19 infection on imaging related to chest CT of 121 symptomatic patients, and figured out that about 65% patients still has a normal CT, and later comes with more clinical symptoms, such as consolidation, bilateral as well as linear opacities. Fang et al. [9] investigated the sensitivity of chest CT as well as viral nucleic acid assay to compare the results derived by baseline methods [10]. The experiments indicated that sensitivity of CT and RT-PCR for COVID-19 infection is 98% and 7%, respectively. Hellewell et al. [11] developed a stochastic transmission model to systematically analyze the quantify the effectiveness of contact tracing and isolation of cases, and provided an effective strategy to control the SARS-CoV-2-like pathogen. Shi et al. [12] discussed the descriptive research for the 81 patients with COVID-19 pneumonia by using the radiological findings, and figured out that the COVID-19 pneumonia can be effectively detected based on the chest CT imaging abnormalities in asymptomatic patients. Convolutional neural network (CNN) is a kind of feed-forward neural network with multi-level deep structure, the neurons of which is non-full connected, and share the network's weights among some neurons at the same layer. Therefore, CNN has a significant advantage in processing high-dimensional data and automatically extracting features. Therefore, deep learning is also increasingly and widely used in medically assisted analysis, such as the CNN is applied to split the cartilage in the knee based on MRI [13]. In general, CNN requires a large number of samples for training and structural adjustment in order to form a model with strong characteristic analysis capabilities. Reinforcement learning (RL) is an important learning method, the main processing purpose of which is to achieve goal optimization through learning strategies [14]. The significant advantage of the RL approach is that it can receive learning information and updating model parameters, without any training data in advance, only by receiving feedback on actions from the external environment. RL has been widely used in image analysis, financial trading system and planetary vehicle path planning [15–17]. Improving CNN's generalization ability with unsupervised competitive learning or RL, can effectively reduce the over-reliance on large samples and improve model's classification accuracy.

Deep Convolutional Neural Network (DCNN) such as VGG16 (VGG with 16-weights layers) as well as DenseNet121, with transfer learning methods were applied in [18] to enhance the generalization ability of the proposed model and perform the better classification accuracy based on the pediatric Chest X-ray dataset. Experiments based on the occlusion test was proposed to detect the appropriate area and visualize model's output. Three classification issues, such as normal, pneumonia-bacterial and pneumonia-viral based on publically available COVID-19 chest X-ray medical images were implemented through the CNN with transfer learning [19]. The precision, recall and classification accuracy for COVID-19 dataset were 89.6%, 93.0% as well as 98.2%, respectively, which shown the better performance compared with the results that reported in the literatures. Zhang et al. [20] proposed an approach based on confidence-aware anomaly detection (CAAD) to implement the binary classification. Experiments based on the 5,977 Non-COVID-19 viral pneumonia samples and 18,774 healthy controls cases etc. were provided to shown that viral pneumonia usually exhibits significant different visual appearances on the Chest X-ray images. Baltruschat et al. [21] proposed a novel insight of the deep learning approaches related ResNet-50 to classify the ChestX-ray14 dataset, as aforementioned above, the transfer learning without fine-tuning methods were jointed into the new X-ray network. The ROC statistics of the ResNet-50 with extended architecture, indicating that the proposed deep learning model can achieve the best overall results compared with different approaches. Bhandary et al. [22] developed a modified AlexNet (MAN) network with deep-learning architecture to evaluate the human lung abnormality based on the Chest CT images. The extracted features were effectively selected by the PCA to construct the principal features dataset and then the experiments compared with the *state-of-the-art* methods such as VGG16, VGG19 and ResNet50 shown that the proposed approaches can achieve the classification accuracy of 97.25%. Moreover, GPU will benefit the rapid convergence of the CNN model when a large amount of structured data requires floating-point operation. The deep CNN model can get more detailed feature than the simple one. Especially when the training samples are difficult to stimulate all the modes of the system due to the limited samples, resulting that the generalization capability of the CNN model is difficult to be guaranteed. The RL and pre-training methods are utilized to train deep CNN, i.e., VGG16 model by transferring the knowledge of the simple CNN network to optimize the deep network architecture, and to achieve high-precision model recognition and detection of Pneumonia, which undoubtedly benefits the early diagnosis of COVID-19 pneumonia.

The rest of this paper is organized as follows. In Section 2, we will introduce the experimental analysis and model performance evaluation, and provide the fundamental analysis according to the experimental results for the detection of the COVID-19. Finally, Section 3 will conclude the paper with discussions and provide the further research.

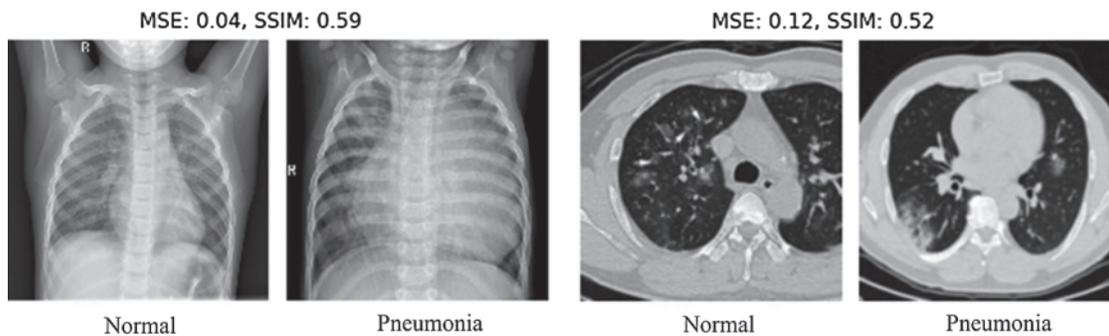
## 2 Proposed Classification–Detection Approach

In routine medical examinations, usually only the professional experience of medical imaging doctors can be qualified to accurately determine the difference between normal samples and infected samples based on the Chest X-ray and CT dataset. Because the difference in image visual display is difficult to evaluate, the performance indicator used to evaluate the images' similarity, such as statistical histogram, Mean Squared Error (MSE) as well as the Structural Similarity Index (SSIM) [23], are applied to arbitrarily selected two images from the dataset and estimate the similarity by checking that if their distribution are identical or near identical. MSE and SSIM used to calculate the quantitative results are provided by Eqs. (1) and (2), respectively.

$$Err_{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [P(i,j) - Q(i,j)]^2 \quad (1)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

where  $P(i,j)$  and  $Q(i,j)$  are both the pixel intensities corresponding to the location  $(i,j)$  of two arbitrary images,  $\mu_x$  (or  $\mu_y$ ) and  $\sigma_x^2$  (or  $\sigma_y^2$ ) are the mean and variance of the pixels intensities, respectively. MSE is a perceived differential measurement that can be used to evaluate two arbitrary images with the same size, but which is sensitive to outliers because the distance between the difference pixels intensity will be increased. In other words, it still doesn't mean that the two images are dramatically similar even if the MSE is small enough. SSIM as the performance indicator that focuses more on the similarity assessment of structural information, which can be used to effectively reflect the properties of the structure of the object between the two images, and calculate the differences in terms of brightness, contrast and structure to effectively evaluate their similarity. The quantitative results and the image visual comparison are shown in Fig. 1. Both normal Chest X-ray and CT, and the corresponding histogram of RGB pixel intensity etc., are shown in Fig. 2.



**Figure 1:** MSE and SSIM between the normal and pneumonia samples

The maximum of the RGB pixels intensity histogram of the normal Chest X-ray and CT are derived when pixels intensity equal to 0.6 and 0.2, respectively. The approximate dispersion of the former and latter are correspondingly distributed between  $[0.1, 0.9]$  and  $[0.2, 1.0]$ . Taking into account the similarity of the Chest X-ray and CT, the MSE and SSIM of the former are higher than the latter 0.08 and 0.07, respectively. This shows that there are significant differences in the feature distribution of these two types of samples.

The processing diagram of this paper is shown in Fig. 3. The construction process of the classification–detection model is divided into three parts: Firstly, the corresponding Chest X-ray and CT are divided into training samples and testing samples, and then used as the inputs in features learning; Secondly, five keras-related deep learning models, i.e., ResNet50, InceptionResNetV2, Xception, transfer learning and pre-trained VGGNet16 are provided to formulate the classification–detection approaches of COVID-19; Finally, the performance indicators, i.e., precision, recall, F1 scores, confusion matrix and AUC, are used to evaluate the classification–detection

approaches, and compare two benchmark methods such as SVM and CNN, to demonstrate the main performance of the proposed approaches.

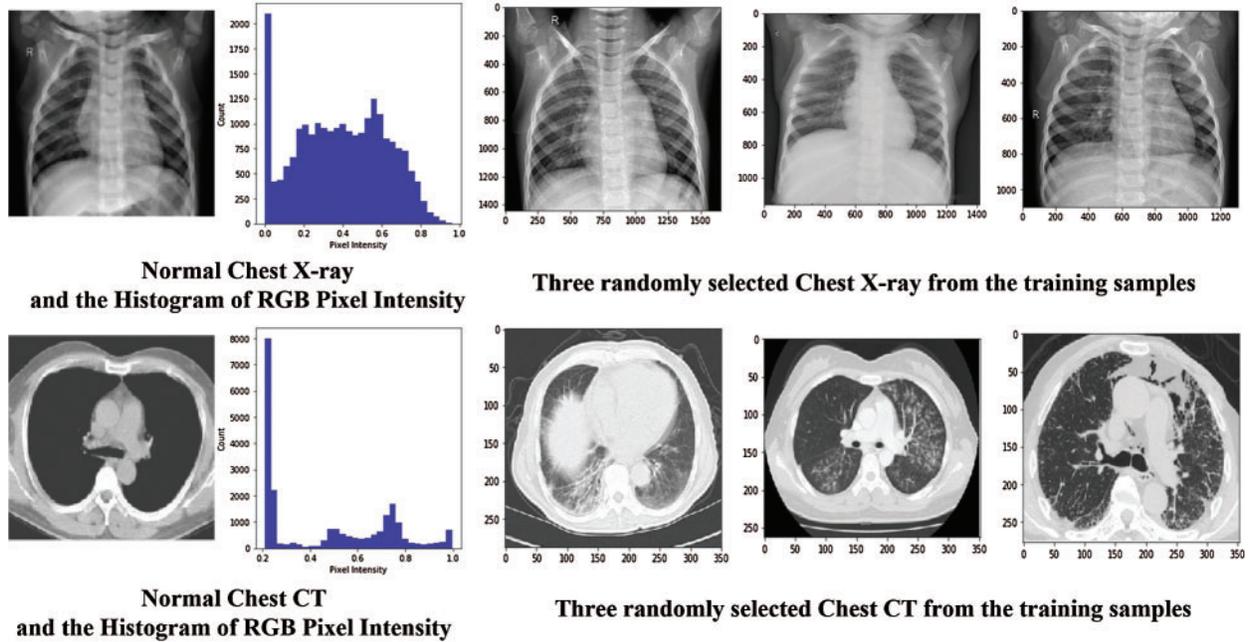


Figure 2: Normal chest X-ray, CT and three corresponding randomly selected training samples

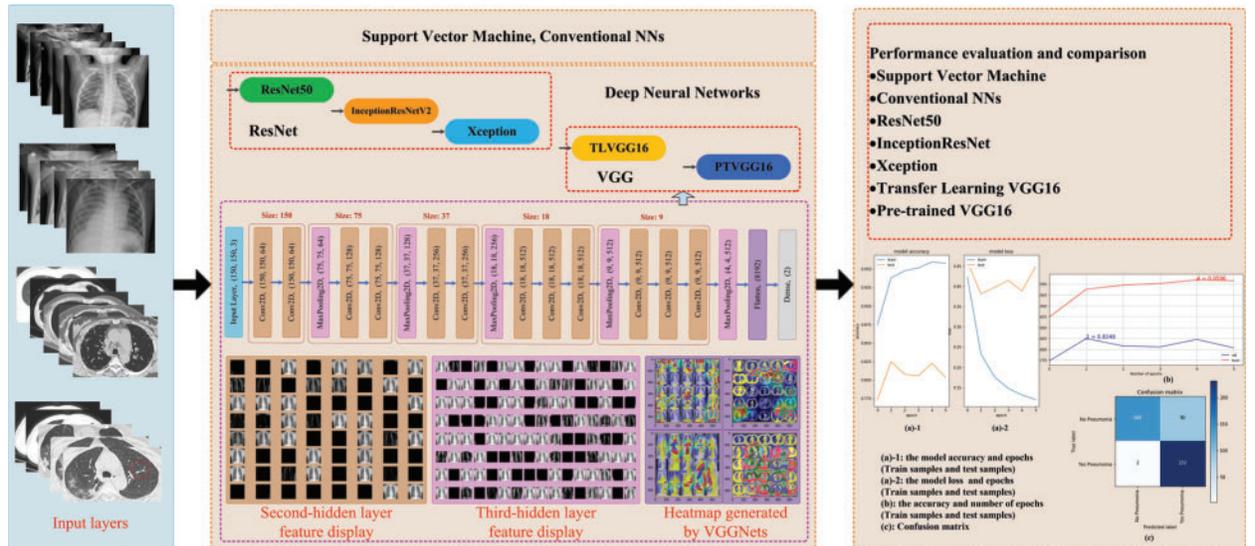


Figure 3: Processing diagram of this paper

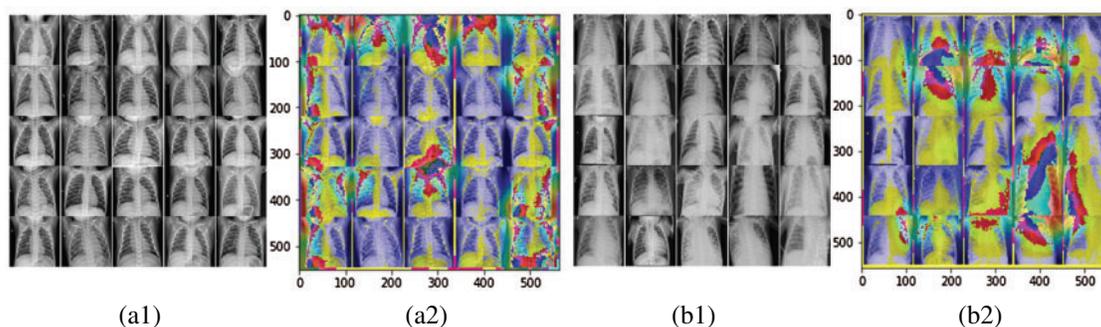
## 2.1 Experiments Setup and Design

The data collected from the dataset: Chest X-ray [24–26] and Chest CT [27–29], are precisely classified as two groups: Normal & Pneumonia. Obviously, this is essentially a typical binary classification issues to judge if the people are infected or not. The training and testing dataset are divided in Tab. 1 to evaluate the performance of the proposed approach.

**Table 1:** Sample division

Dataset	Chest X-ray	Processing speed (it/s)	Chest CT	Processing speed (it/s)
Xtrain	1349 (Normal)	4.48	300 (Normal)	81.52
ytrain	3884 (Pneumonia)	9.07	300 (Pneumonia)	41.81
Xtest	234 (Normal)	2.47	97 (Normal)	70.59
ytest	390 (Pneumonia)	8.65	70 (Pneumonia)	45.01

The explanation of the abbreviations in Tab. 1 is given as follows, Xtrain and Xtest: represents the training sample and testing samples related to images of normal persons. ytrain and ytest: Are the training sample and testing samples related to images of persons infected by Pneumonia.

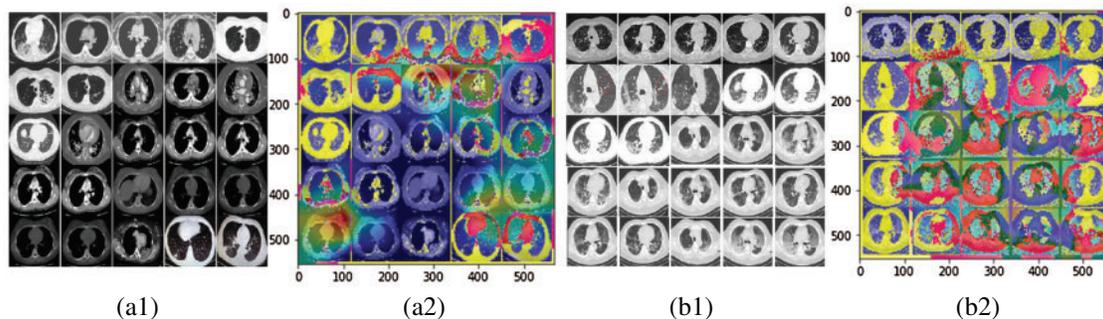


**Figure 4:** The 25 normal and pneumonia chest X-ray images. (a1) (25 normal images). (a2) (Visualizing heatmaps of a1). (b1) (25 Pneumonia images) (b2) (Visualizing heatmaps of b1)

## 2.2 Architecture of the VGGNet

All programming experiments are based on Python 3.6 under OS win10. The hardware used for experiments are configured as the i7-9750H CPU, 32G RAM, and NVIDIA P620 4G GPU. VGG16 is proposed by Visual Geometry Group from Oxford. Compared with the AlexNet, VGG utilized the successive  $3 \times 3$  convolution cores instead of larger convolution nucleus in AlexNet ( $11 \times 1$ ,  $7 \times 7$ ,  $5 \times 5$ ). For a given sensory field (the local size of the input image associated with the output), VGG using a small convolutionary that accumulates is preferable to a large convolutionary, because multi-layer nonlinear layers increase the CNN's depth to ensure learning more complex patterns at a lower cost and less parameters. In addition, 25 normal and pneumonia Chest X-ray (and CT) images and the corresponding visualization heatmap are respectively shown in Figs. 4 and 5, respectively. Essentially, because the distribution of these

two types of data (positive and negative samples related to normal and pneumonia symptoms) is uneven, such as training sample size is too much larger than the test sample size. The model performance may not be effectively evaluated and reflected, especially for the unbalanced class distribution if only the classification accuracy is provided. In other words, it may be insufficient to evaluate the model performance only based on model accuracy. To this end, more performance indicators, such as Precision, Recall, F1-score and Support, are provided to comprehensively evaluate the performance of the classification–detection approaches.



**Figure 5:** The 25 normal and pneumonia chest CT images. (a1) (25 normal images). (a2) (Visualizing heatmaps of a1). (b1) (25 Pneumonia images) (b2) (Visualizing heatmaps of b1)

### 2.3 Experimental Comparison and Analysis

In order to further verify the performance of model classification detection, SVM, two hidden layers of CNN (i.e., two layers of convolution layer and maximum pooled layer), ResNet50, InceptionResNetV2, Xception, transfer learning VGG16 and pre-trained VGG16 model, are respectively utilized to evaluate the final classification detection based on the positive and negative samples. The parameter gamma and the punitive parameters of the utilized Radial Basis Function kernel in SVM are 0.001 and default value, respectively. Similarly, based on the excellence classification performance of the deep learning, the Keras framework is used to build a CNN network classification model. The positive and negative samples as the inputs are normalized to  $150 \times 150 \times 1$ , and the two convolution layer and the maximum pooled layer are then set, and the output classified result is calculated through two layers of full-connection layer. Stochastic gradient descent algorithm is then applied to optimize the architecture of CNN model and improve feature learning ability, and the learning rate of the outlined five models i.e., ResNet50, Inception-ResNetV2, Xception, TLVGG and PTVGG is 0.0001. Performance evaluation and comparison of the SVM, CNN, ResNet50, InceptionResNetV2, Xception, TLVGG16 (Transfer learning VGG16) and PTVGG16 (Pre-trained VGG16) are given in [Tabs. 2](#) and [3](#), respectively. The computational resources consumption of the deep learning approaches are given in [Tab. 4](#). where ATC: Average time cost in seconds per step in each learning epoch. The number of the training (or validation) samples of the X-ray and CT are 2698 (468) and 600 (140), respectively. The experimental results show that the classification accuracy calculated by SVM for pneumonia is 0.77 (Chest X-ray) and 0.63 (Chest CT), respectively. More precisely, the classification accuracy related to the positive samples is significantly higher than that of negative samples by 0.20 (Chest X-ray) and 0.17 (Chest CT), respectively. This indicates that the SVM has a stronger ability to identify normal lungs, but it is still not qualified to identify patients with pneumonia. For Chest X-ray images, the significant increase in training and testing sample size will effectively improve the classification

accuracy although increasing training time. The aforementioned methods facilitates the accurate identification and detection of pneumonia.

**Table 2:** Performance evaluation and comparison (chest X-ray)

Samples	Methods	Accuracy	Precision	Recall	F1-score	Support
Normal	SVM	0.77	0.94	0.41	NA	NA
	CNN	0.63	0.79	0.40	NA	NA
	ResNet50	0.84	0.97	0.71	<b>0.82</b>	234
	InceptionResNetV2	0.82	0.97	0.67	0.79	234
	Xception	0.71	0.69	<b>0.77</b>	0.73	234
	TLVGG16	0.85	<b>0.99</b>	0.62	0.76	234
	PTVGG16	0.80	0.98	0.65	0.78	234
Pneumonia	SVM	0.77	0.74	0.98	NA	NA
	CNN	0.63	0.62	1.00	NA	NA
	ResNet50	0.84	0.77	0.98	<b>0.86</b>	234
	InceptionResNetV2	0.82	<b>0.75</b>	0.98	0.85	234
	Xception	0.71	0.74	0.66	0.70	234
	TLVGG16	0.85	0.72	<b>0.99</b>	0.83	234
	PTVGG16	0.80	0.74	<b>0.99</b>	0.84	234

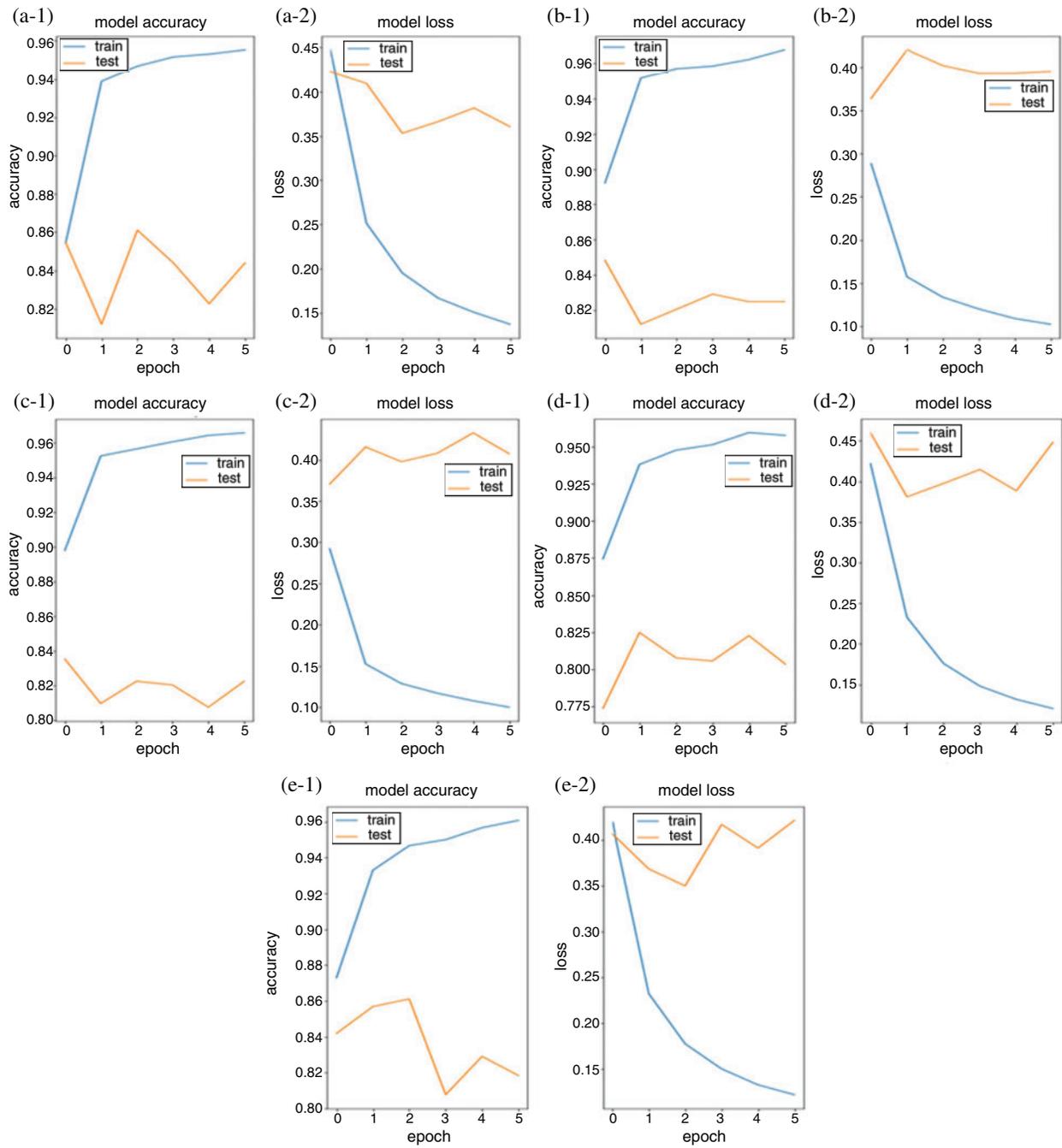
**Table 3:** Performance evaluation and comparison (chest CT)

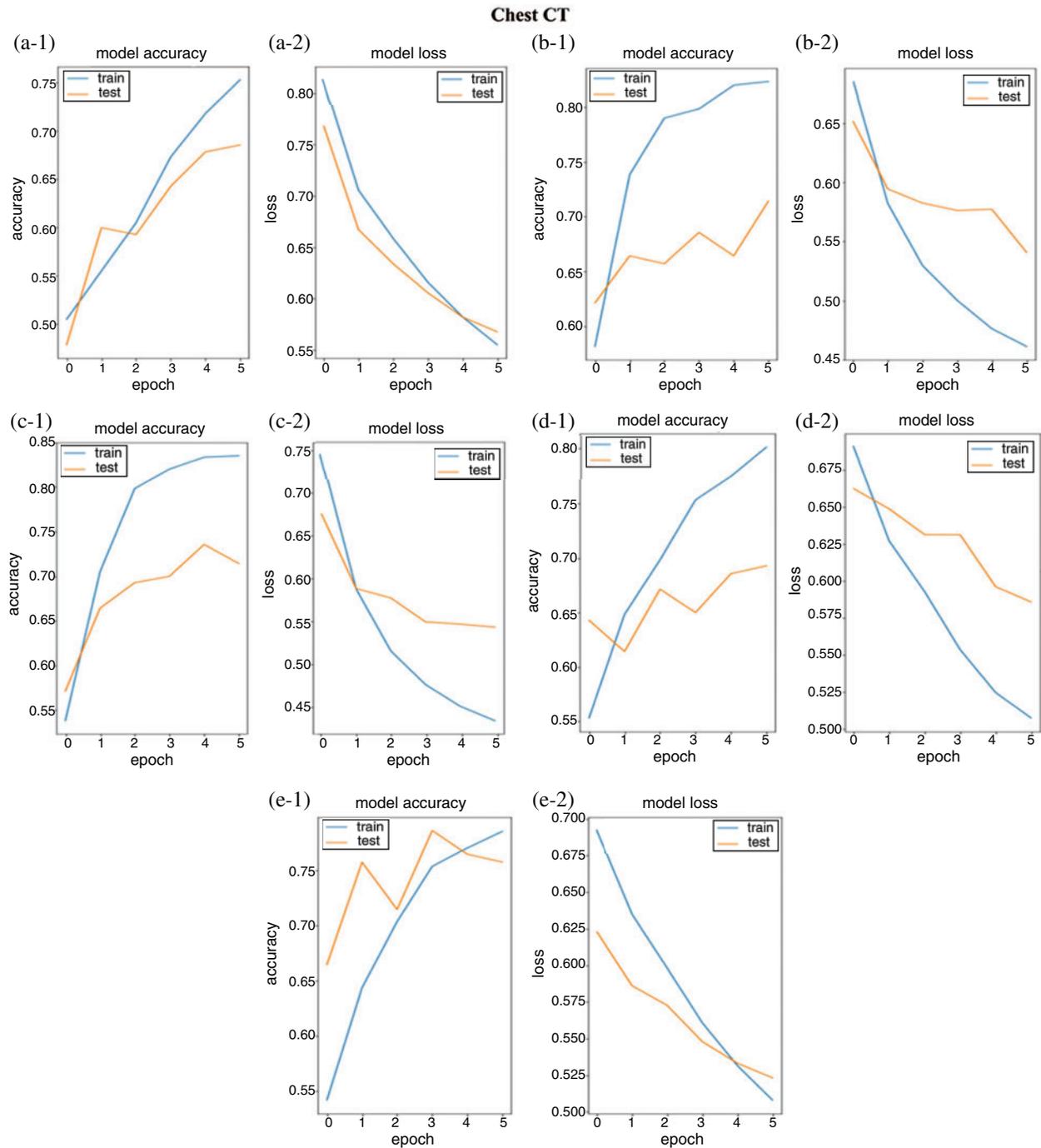
Samples	Methods	Accuracy	Precision	Recall	F1-score	Support
Normal	SVM	0.63	<b>0.81</b>	0.47	NA	NA
	CNN	0.56	0.72	0.40	NA	NA
	ResNet50	0.75	0.72	0.81	<b>0.77</b>	70
	InceptionResNetV2	0.71	0.73	0.69	0.71	70
	Xception	0.71	0.69	0.77	0.73	70
	TLVGG16	0.69	0.71	0.66	0.68	70
	PTVGG16	0.75	0.72	<b>0.83</b>	<b>0.77</b>	70
Pneumonia	SVM	0.63	0.54	<b>0.84</b>	NA	NA
	CNN	0.56	0.49	0.79	NA	NA
	ResNet50	0.75	0.79	0.69	0.73	70
	InceptionResNetV2	0.71	0.70	0.74	0.70	70
	Xception	0.71	0.74	0.66	0.70	70
	TLVGG16	0.69	0.68	0.73	0.70	70
	PTVGG16	0.75	<b>0.80</b>	0.69	<b>0.74</b>	70

**Table 4:** The comparison of computational efficiency (five deep learning models)

Methods	ResNet50	InceptionResNetV2	Xception	TLVGG16	PTVGG16
ATC (2698-X-ray)	172.064	170.230	171.064	171.063	164.728
ATC (600-CT)	178.130	181.803	178.464	137.730	137.896

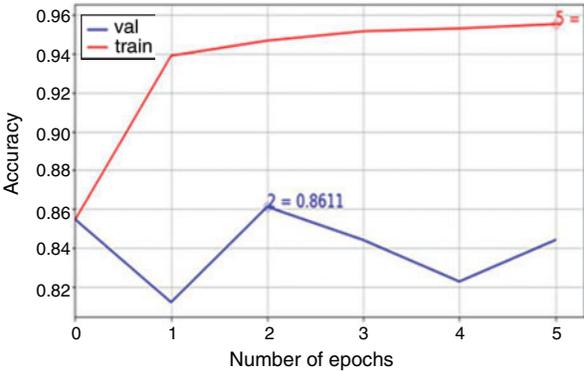
**Chest X-ray**



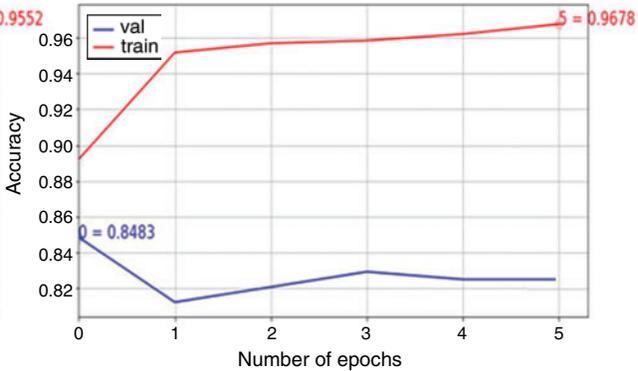


**Figure 6:** The accuracy and cost of the five deep learning models. (a-1, a-2) Model accuracy, model cost of ResNet50. (b-1, b-2) Model accuracy, model cost of InceptionResNetV2. (c-1, c-2) Model accuracy, model cost of Xception. (d-1, d-2) Model accuracy, model cost of TLVGG16. (e-1, e-2) Model accuracy, model cost of PTVGG16

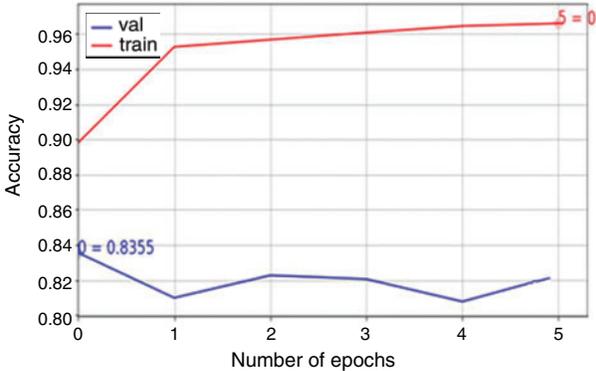
Chest X-ray



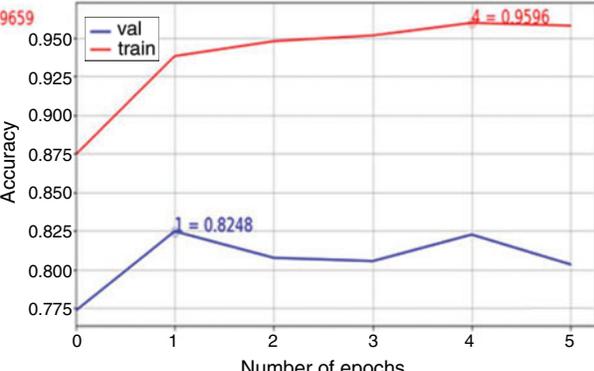
(a)



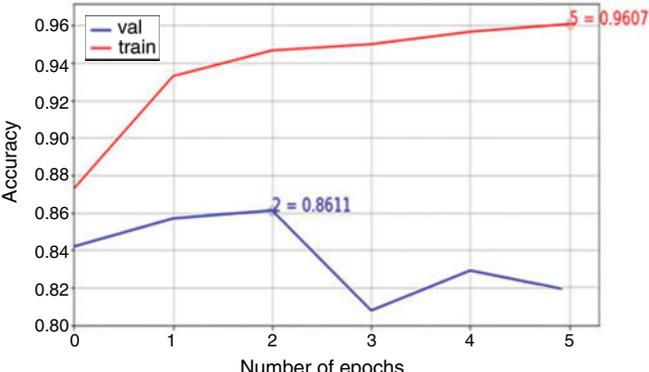
(b)



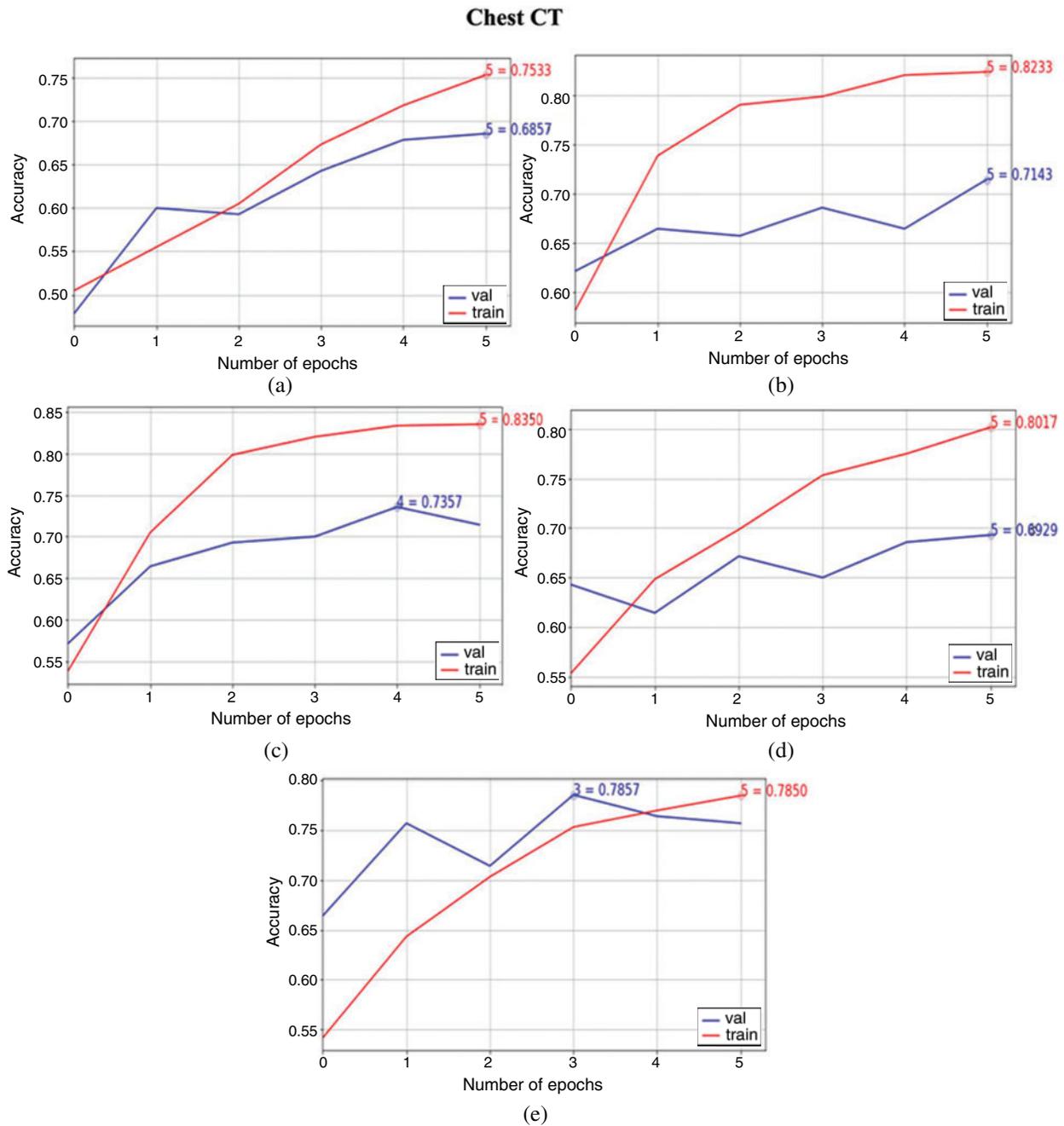
(c)



(d)

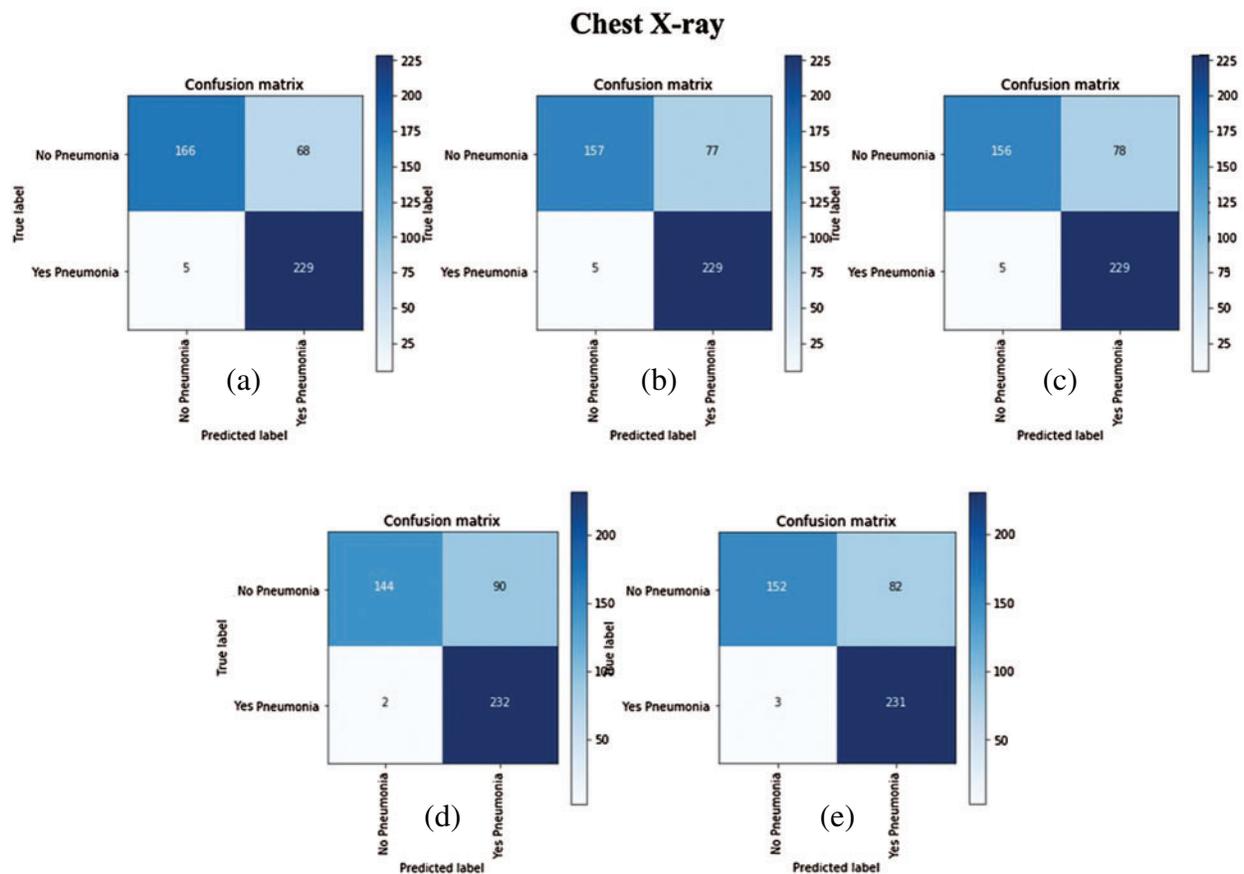


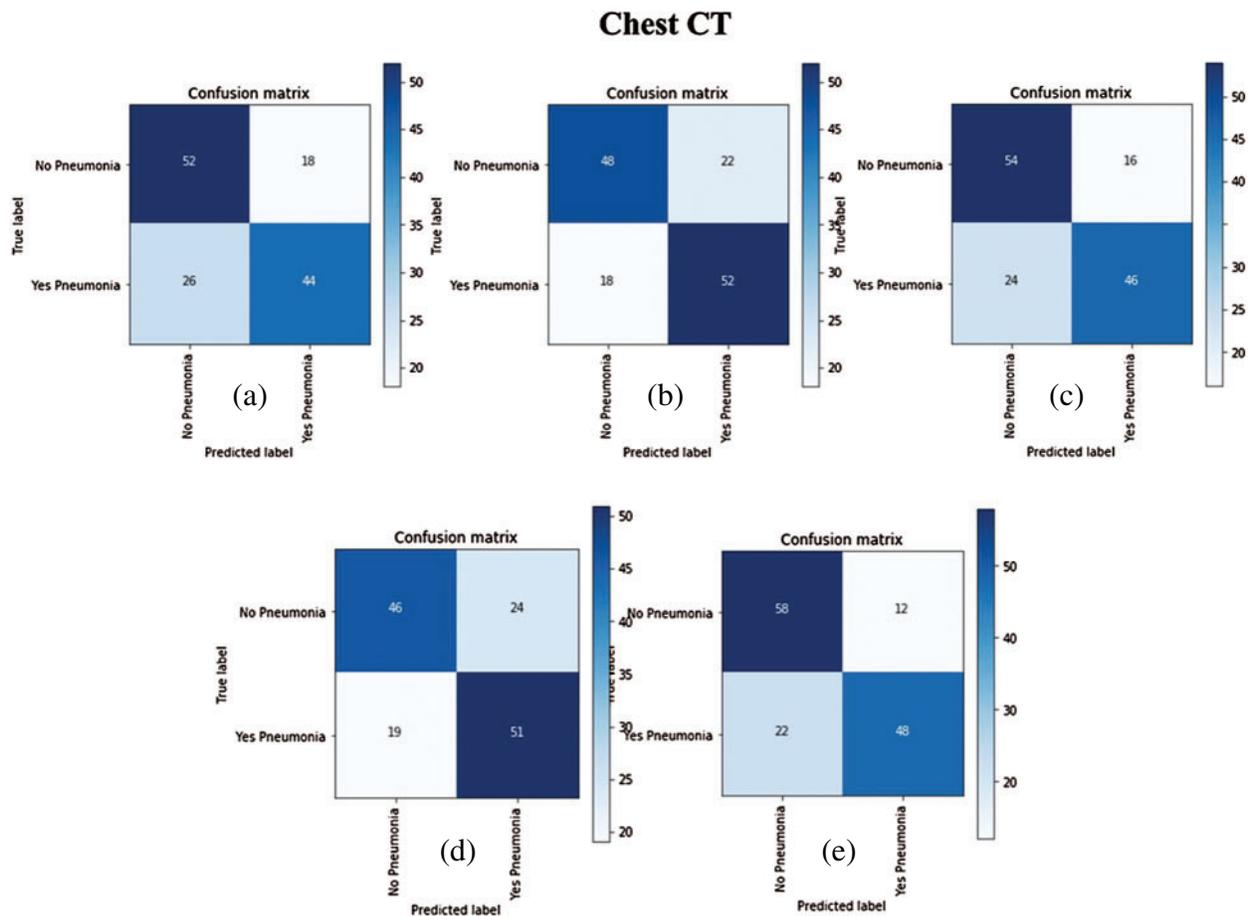
(e)



**Figure 7:** The accuracy of the five deep learning models. (a) Train and validation accuracy of ResNet50. (b) Train and validation accuracy of InceptionResNetV2. (c) Train and validation accuracy of Xception. (d) Train and validation accuracy of TLVGG16. (e) Train and validation accuracy of PTVGG16

The accuracy curve (and loss function curve) of the five outlined keras-related deep learning models are successively marked as a-1-e-1 (a-2-e-2) in Fig. 6. Moreover, the model training and the curve related to the verification curve and the number of iteration steps are also sequentially shown in a-e of the Figs. 7, while the corresponding confusion metrics are provided in a-e of the Fig. 8. In addition, the final model performance is reported by the Tabs. 5-9. The AUC (Area Under Curve) is defined as the area enclosed with the axes under the ROC (Receiver Operating Characteristic) curve, the value of which is between the interval (0.5, 1). More precisely, the closer the AUC is to 1.0, the higher the credibility of the detection methods are. Especially, the method will show meanness in the real application if the AUC equals to 0.5. In order to verify the performance of the utilized methods in this paper, AUC\_SVM, AUC\_NB and AUC\_LOG related to the ResNet50 X-ray (CT), InceptionResNetV2 X-ray (CT), Xception X-ray (CT), TLVGG16 X-ray (CT) and PTVGG16 X-ray (CT) are listed in Tab. 10. Where AUC\_NB and AUC\_LOG represent the ROC Curves for naive Bayes and logistic regression classification, respectively. SVM produces lower ROC values compared with naive Bayes and logistic regression, and the ROC curve for naive Bayes is generally equal to logistic regression, which indicates better performance than the SVM classifier method.





**Figure 8:** Confusion matrices of the five deep learning models. (a) Confusion matrix of ResNet50. (b) Confusion matrix of InceptionResNetV2. (c) Confusion matrix of Xception. (d) Confusion matrix of TLVGG16. (e) Confusion matrix of PTVGG16

**Table 5:** Performance of the ResNet50 based on chest X-ray (CT)

Methods	Precision	Recall	F1 score	Support
Normal	0.97 (0.67)	0.71 (0.74)	0.82 (0.70)	234 (70)
Pneumonia	0.77 (0.71)	0.98 (0.63)	0.86 (0.67)	234 (70)
Accuracy	NA (NA)	NA (NA)	0.84 (0.69)	468 (140)
Macro avg	0.87 (0.69)	0.84 (0.69)	0.84 (0.68)	468 (140)
Weighted avg	0.87 (0.69)	0.84 (0.69)	0.84 (0.68)	468 (140)
Keras CNN-accuracy	<b>0.84</b> (0.69)			

#### 2.4 Performance Analysis of the Proposed Approaches

Precision usually refers to the ratio of the number of correctly predicted positive samples to the total number of predicted positive samples, that is,  $Precision = TP / (TP + FP)$ . A higher value

means a lower false positive rate. The highest precision of patients with normal and pneumonia infections are: 0.99 (TLVGG16, Chest X-ray, Normal), 0.75 (InceptionResNetV2, Chest X-ray, Pneumonia), 0.81 (SVM, Chest CT, Normal), 0.80 (PTVGG16, Chest CT, Pneumonia), and generally almost all of the outlined five deep learning models are higher than 0.7 on both of the Chest X-ray and Chest CT. This shows that the performance of the outlined five deep learning model in this paper is excellent.

**Table 6:** Performance of the InceptionResNetV2 based on chest X-ray (CT)

Methods	Precision	Recall	F1 score	Support
Normal	0.97 (0.73)	0.67 (0.69)	0.79 (0.71)	234 (70)
Pneumonia	0.75 (0.70)	0.98 (0.74)	0.85 (0.72)	234 (70)
Accuracy	NA (NA)	NA (NA)	0.82 (0.71)	468 (140)
Macro avg	0.86 (0.71)	0.82 (0.71)	0.82 (0.71)	468 (140)
Weighted avg	0.86 (0.71)	0.82 (0.71)	0.82 (0.71)	468 (140)
Keras CNN-accuracy	0.82 (0.71)			

**Table 7:** Performance of the Xception based on chest X-ray (CT)

Methods	Precision	Recall	F1 score	Support
Normal	0.97 (0.69)	0.67 (0.77)	0.79 (0.73)	234 (70)
Pneumonia	0.75 (0.74)	0.98 (0.66)	0.85 (0.70)	234 (70)
Accuracy	NA (NA)	NA (NA)	0.82 (0.71)	468 (140)
Macro avg	0.86 (0.72)	0.82 (0.71)	0.82 (0.71)	468 (140)
Weighted avg	0.86 (0.72)	0.82 (0.71)	0.82 (0.71)	468 (140)
Keras CNN-accuracy	0.82 (0.71)			

**Table 8:** Performance of the TLVGG16 based on chest X-ray (CT)

Methods	Precision	Recall	F1score	Support
Normal	0.99 (0.71)	0.62 (0.66)	0.76 (0.68)	234 (70)
Pneumonia	0.72 (0.68)	0.99 (0.73)	0.83 (0.70)	234 (70)
Accuracy	NA (NA)	NA (NA)	0.80 (0.69)	468 (140)
Macro avg	0.85 (0.69)	0.80 (0.69)	0.80 (0.69)	468 (140)
Weighted avg	0.85 (0.69)	0.80 (0.69)	0.80 (0.69)	468 (140)
Keras CNN-accuracy	0.80 (0.69)			

Recall is usually also called sensitivity, which is defined as  $Recall = TP / (TP + FN)$ , and can be used to represent the ratio of the number of correctly predicted samples to the number of samples in all actual categories. That is, the number of the positive samples are correctly predicted. In other word, the deep learning model usually has better sensitivity if its value is higher than 0.5. The highest Recall of patients with normal and pneumonia infections are: 0.77 (Xception, Chest X-ray, Normal), 0.99 (TLVGG16 and PTVGG16, Chest X-ray, Pneumonia), 0.83 (PTVGG16, Chest CT, Normal), 0.84 (SVM, Chest CT, Pneumonia), and generally almost all of the outlined

five deep learning models is higher than 0.6 on both of the Chest X-ray and Chest CT. This shows that the overall performance of the model is good.

**Table 9:** Performance of the PTVGG16 based on chest X-ray (CT)

Methods	Precision	Recall	F1score	Support
Normal	0.98 (0.72)	0.65 (0.83)	0.78 (0.77)	234 (70)
Pneumonia	0.74 (0.80)	0.99 (0.69)	0.84 (0.74)	234 (70)
Accuracy	NA (NA)	NA (NA)	0.82 (0.76)	468 (140)
Macro avg	0.86 (0.76)	0.82 (0.76)	0.81 (0.76)	468 (140)
Weighted avg	0.86 (0.76)	0.82 (0.76)	0.81 (0.76)	468 (140)
Keras CNN-accuracy	0.82 ( <b>0.75</b> )			

**Table 10:** AUC of the SVM, NB and LOG

Methods	AUC_SVM	AUC_NB	AUC_LOG
ResNet50 X-ray (CT)	0.92 (0.56)	0.92 (0.76)	0.92 (0.76)
InceptionResNetV2 X-ray (CT)	0.90 (0.52)	0.90 (0.62)	0.90 (0.62)
Xception X-ray (CT)	0.92 (0.52)	0.92 (0.59)	0.92 (0.59)
TLVGG16 X-ray (CT)	0.85 (0.50)	0.85 (0.69)	0.85 (0.69)
PTVGG16 X-ray (CT)	0.86 (0.50)	0.86 (0.73)	0.86 (0.73)

$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision)$ , which is the weighted average of Precision and Recall. Especially, when the data presents an uneven class distribution, the P and R indicators sometimes are inconsistent, at this time, F-Measure is intuitively (also known as F-Score) and more helpful to measure the cost of false positives and false negatives. The highest Recall of patients with normal and pneumonia infections are: 0.82 (ResNet50, Chest X-ray, Normal), 0.86 (ResNet50, Chest X-ray, Pneumonia), 0.77 (ResNet50 and PTVGG16, Chest CT, Normal), 0.74 (PTVGG16, Chest CT, Pneumonia), and generally almost all of the outlined five deep learning models is higher than 0.7 on both of the Chest X-ray and Chest CT. This shows that the performance of the model is generally superior.

The confusion matrix is mainly used to evaluate the performance of the classification model according to a given test sample, which includes four parts: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), where TP and FP are usually related to the number of the correct positive samples and false positive samples, respectively. TN and FN respectively indicate the number of the correct negative samples and false negative samples. In other word, the less the false positives and false negatives are, the better the model's classification performance is. Based on the discussion, deep transfer learning approaches is suitable for the rapid and accurate detection of pneumonia in theoretical analysis.

### 3 Conclusions

The main purpose of this paper is to formulate the keras-related deep learning approach for classification-detection of COVID-19. Model architecture design, structure selection and model summary are illustrated in detail to illustrate the main processing diagram of the proposed

approaches. The experimental results obtained by five keras-related deep learning approaches are provided and applied to compare with the benchmark methods according to the performance indicators. The experimental evaluation indicates that the deep learning approaches benefits the accurate auxiliary detection of COVID-19. In our further work, the clinical dataset analysis in combination with the classification–detection approach will be considered to further improve the performance.

**Acknowledgement:** The authors wish to express their appreciation to Editorial board and four anonymous reviewers for their helpful suggestions which greatly improved the presentation of this paper.

**Funding Statement:** This project is supported by National Natural Science Foundation of China (NSFC) (Nos. 61902158, 61806087), Graduate student innovation program for academic degrees in general university in Jiangsu Province (No. KYZZ16-0337). The authors would like to thank the Editor and anonymous reviewers for their valuable comments and helpful suggestions to improve the quality of this manuscript, and thank Paul Mooney (Developer Advocate at Kaggle, Colorado, USA) for the programming support.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, 121, 103795. DOI 10.1016/j.combiomed.2020.103795.
2. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M. et al. (2020). A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *MedRxiv*. DOI 10.1101/2020.02.14.20023028.
3. Apostolopoulos, I. D., Mpesiana, T. A. (2020). COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 1(2), 635–640. DOI 10.1007/s13246-020-00865-4.
4. Sethy, P. K., Behera, S. K. (2020). Detection of coronavirus disease (COVID-19) based on deep features. Preprints, 2020030300.
5. Chughtai, A., Malik, A. (2020). Is coronavirus disease (COVID-19) case fatality ratio underestimated? *Global Biosecurity*, 1(3).
6. Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q. et al. (2020). Coronavirus disease 2019 (COVID-19): A perspective from china. *Radiology*, 296(2), E15–E25. DOI 10.1148/radiol.202000490.
7. World Health Organization (2020). *Coronavirus disease 2019 (COVID-19): Situation report*, 72. World Health Organization.
8. Bernheim, A., Mei, X., Huang, M., Yang, Y., Fayad, Z. A. et al. (2020). Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection. *Radiology*, 295(3), DOI 10.1148/radiol.202000463.
9. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L. et al. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, 296(2), E115–E117. DOI 10.1148/radiol.202000432.
10. Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F. et al. (2020). Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: Relationship to negative RT-PCR testing. *Radiology*, 296(2).
11. Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I. et al. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health*, 8(4), e488–e496. DOI 10.1016/S2214-109X(20)30074-7.

12. Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O. et al. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study. *Lancet Infectious Diseases*, 20(4), 425–434.
13. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E. et al. (2013). Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Medical Image Computing and Computer-Assisted Intervention*, 16, 246–253. DOI 10.1007/978-3-642-40763-5\_31.
14. Xu, D., Xu, W., Tang, Z., Liu, F. (2014). Exploiting visual saliency and bag-of-words for road sign recognition. *IEICE Transactions on Information and Systems*, 97(9), 2473–2482. DOI 10.1587/transinf.2013EDP7448.
15. Jiao, Y., Li, Z., Huang, S., Yang, X., Liu, B. et al. (2018). Three-dimensional attention-based deep ranking model for video highlight detection. *IEEE Transactions on Multimedia*, 20(10), 2693–2705. DOI 10.1109/TMM.2018.2815998.
16. Shi, J., Sun, Z., Bai, S. (2018). 3D reconstruction framework via combining one 3D scanner and multiple stereo trackers. *Visual Computer*, 34(3), 377–389. DOI 10.1007/s00371-016-1339-4.
17. Shao, H., Deng, X. (2018). Adaboosting neural network for short-term wind speed forecasting based on seasonal characteristics analysis and lag space estimation. *Computer Modeling in Engineering & Sciences*, 114(3), 277–293.
18. Behzadi-Khormouji, H., Rostami, H., Salehi, S., Derakhshande-Rishehri, T., Masoumi, M. et al. (2020). Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. *Computer Methods and Programs in Biomedicine*, 185, 105162. DOI 10.1016/j.cmpb.2019.105162.
19. Khan, A. I., Shah, J. L., Bhat, M. M. (2020). Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Computer Methods and Programs in Biomedicine*, 105581.
20. Zhang, J., Xie, Y., Li, Y., Shen, C., Xia, Y. (2020). COVID-19 screening on chest X-ray images using deep learning based anomaly detection. arXiv preprint arXiv: 2003.12338.
21. Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., Saalbach, A. (2019). Comparison of deep learning approaches for multi-label chest X-ray classification. *Scientific Reports*, 9(1), 1–10. DOI 10.1038/s41598-019-42294-8.
22. Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C. et al. (2020). Deep-learning framework to detect lung abnormality—A study with chest X-ray and lung CT scan images. *Pattern Recognition Letters*, 129, 271–278. DOI 10.1016/j.patrec.2019.11.013.
23. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. DOI 10.1109/TIP.2003.819861.
24. Kermany, D., Zhang, K., Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data*, 2.
25. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H. et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122–1131. DOI 10.1016/j.cell.2018.02.010.
26. Killeen, B. D., Wu, J. Y., Shah, K., Zapaishchykova, A., Nikutta, P. et al. (2020). A county-level dataset for informing the United States’ response to COVID-19. <https://arxiv.org/pdf/2004.00756.pdf>.
27. Cohen, J. P., Dao, L., Morrison, P., Roth, K., Bengio, Y. et al. (2020). Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *Cureus*, 12(7), e9448. DOI 10.7759/cureus.9448.
28. Cohen, J. P., Morrison, P., Dao, L. (2020). COVID-19 image data collection. arXiv 2003.11597. <https://github.com/ieee8023/covid-chestX-ray-dataset>.
29. He, X., Yang, X., Zhang, S., Zhao, J., Zhang, Y. et al. (2020). Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. *MedRxiv*. DOI 10.1101/2020.04.13.20063941.