



Accurately Discriminating COVID-19 from Viral and Bacterial Pneumonia According to CT Images Via Deep Learning

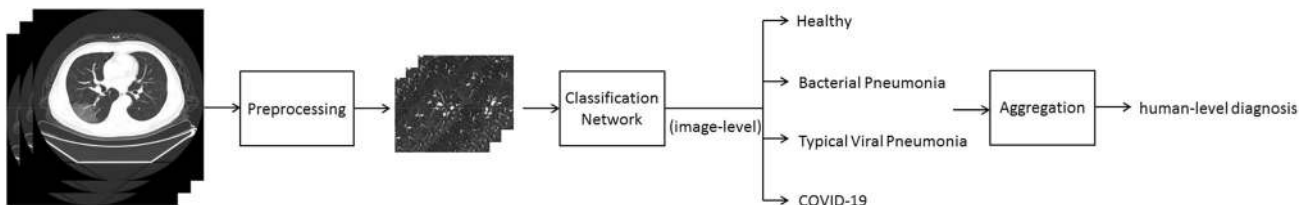
Fudan Zheng¹ · Liang Li² · Xiang Zhang³ · Ying Song⁴ · Ziwang Huang¹ · Yutian Chong⁵ · Zhiguang Chen^{1,6} · Huiling Zhu⁷ · Jiahao Wu⁸ · Weifeng Chen⁹ · Yutong Lu^{1,6} · Yuedong Yang^{1,6} · Yunfei Zha² · Huiying Zhao³ · Jun Shen³

Received: 19 August 2020 / Revised: 22 January 2021 / Accepted: 1 February 2021 / Published online: 27 February 2021
© International Association of Scientists in the Interdisciplinary Areas 2021

Abstract

Computed tomography (CT) is one of the most efficient diagnostic methods for rapid diagnosis of the widespread COVID-19. However, reading CT films brings a lot of concentration and time for doctors. Therefore, it is necessary to develop an automatic CT image diagnosis system to assist doctors in diagnosis. Previous studies devoted to COVID-19 in the past months focused mostly on discriminating COVID-19 infected patients from healthy persons and/or bacterial pneumonia patients, and have ignored typical viral pneumonia since it is hard to collect samples for viral pneumonia that is less frequent in adults. In addition, it is much more challenging to discriminate COVID-19 from typical viral pneumonia as COVID-19 is also a kind of virus. In this study, we have collected CT images of 262, 100, 219, and 78 persons for COVID-19, bacterial pneumonia, typical viral pneumonia, and healthy controls, respectively. To the best of our knowledge, this was the first study of quaternary classification to include also typical viral pneumonia. To effectively capture the subtle differences in CT images, we have constructed a new model by combining the ResNet50 backbone with SE blocks that was recently developed for fine image analysis. Our model was shown to outperform commonly used baseline models, achieving an overall accuracy of 0.94 with AUC of 0.96, recall of 0.94, precision of 0.95, and F1-score of 0.94. The model is available in <https://github.com/Zhengfudan/COVID-19-Diagnosis-and-Pneumonia-Classification>.

Graphic Abstract



Keywords Diagnosis of COVID-19 · Pneumonia classifying · CT image · Deep learning network

Fudan Zheng, Liang Li and Xiang Zhang contributed equally.

- ✉ Yuedong Yang
yangyd25@mail.sysu.edu.cn
- ✉ Yunfei Zha
zhayunfei999@126.com
- ✉ Huiying Zhao
zhaohy8@mail.sysu.edu.cn
- ✉ Jun Shen
shenjun@mail.sysu.edu.cn

Extended author information available on the last page of the article

1 Introduction

The coronavirus disease emerged in late 2019, which was named COVID-19 by the World Health Organization, has become a pandemic and poses a serious threat to international health. The disease is caused by SARS-CoV-2 [1], which can be transmitted from person to person, and the number of infected persons has increased dramatically [2, 3]. Up to August 14, 2020, more than 20 million cases have been reported in more than 216 countries and territories, resulting in more than 751 thousand deaths [4]. Therefore,

a computer aided CT diagnosis system is urgently needed to assist doctors in identifying suspected cases.

In order to detect COVID-19 infected patients and to prevent community infection brought by missed patients, persons are recommended to perform COVID-19 screening if they have fever, cough, flu-like symptoms, or close contact with a COVID-19 infected patient. CT detection is becoming an important tool in detecting infected patients because of its quickness and low false negative rate [5–7]. In addition, CT films can intuitively show the patient's lung details, including locations and characters (ground-glass opacities, consolidation shadows, fibrosis, etc. [5–7]) of the lesions or inflammations. However, the large number of CT images put large burdens on doctors to read them. As Michael J Ryan, the executive director of WHO's emergency program said on May 13, COVID-19 might become another endemic virus in our communities and might never go away [8], it is critical to develop a system that can not only diagnose COVID-19 when it is in intensive outbreak, but also distinguish COVID-19 from routine examinations when the outbreak is under control. Therefore, it is urgent to develop computer aided CT diagnosis systems to assist doctors in identifying suspected cases.

Recently, there have been many AI-assisted methods on COVID-19 and pneumonia classification on CT images or chest X-ray images. As summarized in a recent review [9], there have been mainly two types of classification, i.e., classification of COVID-19 from non COVID-19 and classification of COVID-19 from other pneumonia. For example, Song et al. [10] proposed a CT diagnosis system based on deep learning models to distinguish patients with COVID-19 from bacterial pneumonia patients and healthy persons. The model achieved accuracy of 86.0% and 94.0% for distinguishing COVID-19 from bacterial pneumonia, and for diagnosing COVID-19 infected patients from healthy persons, respectively. Xu et al. [11] proposed a classification system to identify COVID-19 patients, Influenza-A patients and healthy persons, which achieved an accuracy of 86.7%. Li et al. [12] used ResNet50 to discriminate COVID-19 from non-pneumonia or community-acquired pneumonia, acquiring a sensitivity of 90%. Then, Chen et al. [13], Zheng et al. [14], Jin et al. [15], Wang et al. [16], Shi et al. [17], Rasheed et al. [18], Zhang et al. [19], Ouyang et al. [20], Han et al. [21], Kang et al. [22], Apostolopoulos et al. [23] and Jaiswal et al. [24] also aimed to separate COVID-19 infected patients from nonCOVID-19 subjects and other pneumonia. However, all of these works have ignored typical viral pneumonia that is infected by typical virus, which is also the most challenging as COVID-19 is also a kind of virus.

In recent years, many deep learning methods have been used for processing medical images. For example, the ResNet50 [25] was commonly used as the backbone

network because the pre-trained network could capture the subtle features in CT images without introducing computational complexity and performance degradation. VGG [26] is another network commonly used to extract key features, but it has a large number of parameters and high computational complexity. DenseNet [27] has much more parameters than ResNet50, and is not so flexible to be assembled and combined with other networks. On the other hand, the CT data contain many image slices, and each slice could provide both associated and individual information. The recently developed SE block [28] provided a framework to selectively emphasize useful information and suppress useless information through network training.

In this study, we have collected CT images of 262, 100, 219, and 78 persons for COVID-19, bacterial pneumonia, typical viral pneumonia, and healthy controls, respectively. To effectively capture the subtle differences in CT images, we have constructed a new model by combining the ResNet50 backbone with SE blocks for quaternary classification of patients infected with COVID-19, bacterial pneumonia, typical viral pneumonia, and healthy persons in CT images. To the best of our knowledge, this was the first work to distinguish these four types of cases all at once by CT images. Our model achieved an overall accuracy of 0.94 with AUC of 0.96, recall of 0.94, precision of 0.95, and F1-score of 0.94, indicating that it can accurately discriminate COVID-19 from bacterial and typical viral pneumonia and healthy persons.

2 Materials and Methods

2.1 Data Acquisition

The CT images were provided by Sun Yat-sen Memorial Hospital and Renmin Hospital of Wuhan University, with totally 52973 slices of 659 persons. The CT images from Sun Yat-sen Memorial Hospital were obtained by two scanners: Somatom Sensation 64-slice spiral scanner of Siemens and Discovery CT 750 HD of GE, with the scanning parameters as follows: effective tube current of 200–250 mA; tube voltage of 120 kV; matrix of 512×512 ; FOV of 500 mm; thickness of 5.0 mm; slice spacing of 5.0 mm; reconstruction thickness of 1.0 mm; and reconstruction slice spacing of 1.0 mm. The scanning body position is the supine position. All patients underwent plain scanning, ranging from the tip of the lung to the entire area of the bottom of the lung, including the chest wall and axilla on both sides. The CT images provided by Renmin Hospital of Wuhan University were acquired by Optima 680, a 64-section scanner of GE, without using contrast materials. The scanning parameters were as follows: automatic tube current; tube voltage of 120 kV; matrix of 512×512 ; detector of 35 mm; rotation time of

0.35 second; section thickness of 5.0 mm; slice spacing of 5.0 mm; reconstruction thickness of 0.625 mm; collimation of 0.75 mm; pitch of 1–1.2; and inspiration breath holding. The images were obtained at the lung window with window width of 1000–1500 HU and window level of -700 HU, and mediastinal window with window width of 350 HU and window level of 35–40 HU.

2.2 Data Preprocessing

As shown in Fig. 1, we extracted the lung region in each slice using the following algorithm: (a) converting the image into a binary image with a density threshold of -600 HU to obtain a mask of interest; (b) removing the connected regions that are in contact with the edges of the image as these are affected by radiations from CT devices; (c) keeping the two largest areas as two lungs; (d) performing a morphological erosion with a disk of radius 2 pixels to shrink bright regions and to enlarge dark regions; (e) performing binary morphological closing to remove the small dark spots and to connect small the bright cracks, and filling small holes inside the detected lungs; (f) superimposing the binary mask on the input image, and detecting the smallest effective rectangle surrounding the lungs. Then, the image was filled with 10 translational

and rotational copies of the lungs on the background to avoid the interference of different lung contours on model training (Fig. 1g). Finally, the preprocessed images were resized into 512×512 , and sent into the subsequent processing with 3 slices as a group.

Since the CT scanners used to capture the CT images were set at a slice spacing of 5.0 mm, and the adjacent images were highly similar. We found the inclusion of all images didn't increase the performance in our task (Results not shown), and selected at most 30 image slices for each person to speed up model training and predictions. Specifically, we selected slices with the following approach: For patients with fewer than 10 slices, retained all slices. For patients with fewer than 30 slices, one slice was selected for every two. For patients with more than 30 slices, slices were selected by the step of slices number divided by 30. With this treatment, the number of slices per patient will not exceed 30, which can accelerate the computations. On the other hand, due to the strong correlation between contiguous slices, the selection of slices at certain step intervals will not result in too much information loss. Finally, we compiled a dataset of 659 persons with 5363 slices. Details of the number of patients and slices provided by the two hospitals after preprocessing are shown in Table 1.

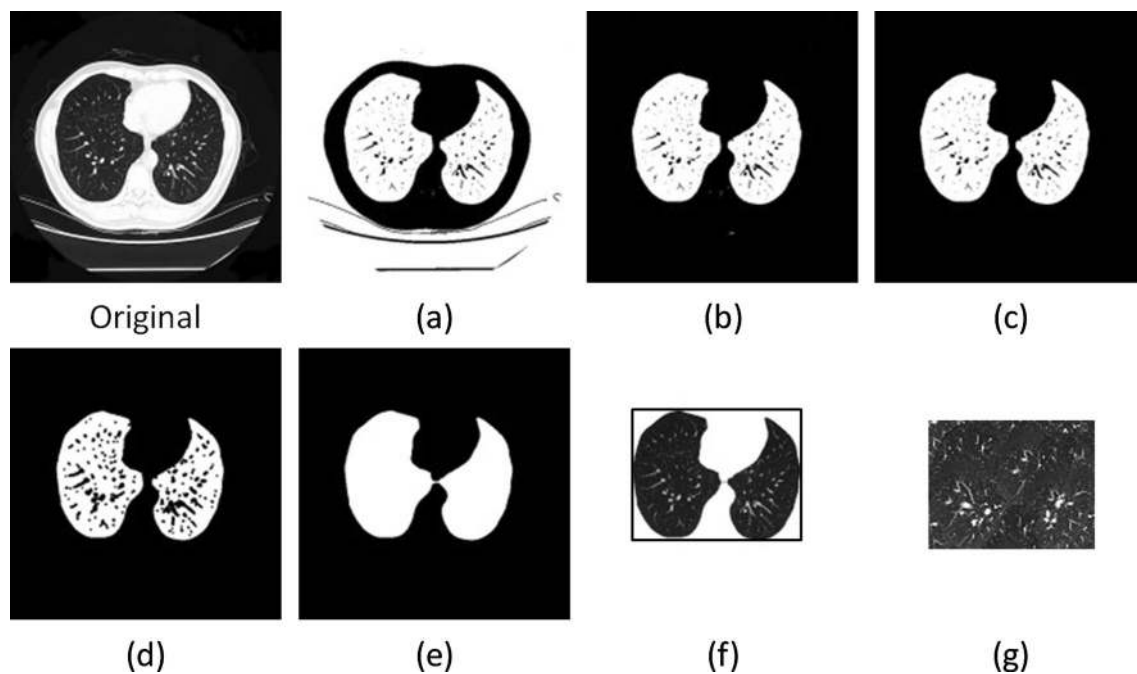


Fig. 1 The workflow of data preprocessing: **a** converting the image into a binary image with a density threshold of -600 HU, **b** removing the connected regions that are in contact with the edges of the image, **c** keeping the two largest areas, **d** performing morphological erosion, **e** performing binary morphological closing and filling

the small holes inside the binary mask of lungs, **f** superimposing the binary mask on the input image and detecting the smallest effective rectangle surrounding the lungs, **g** filling the image with 10 translational and rotational copies of the lungs on the background

Table 1 The number of persons and CT slices provided by the two hospitals after preprocessing

	COVID-19		Healthy people		Bacterial pneumonia		Typical viral pneumonia		Total	
	Patients	Slices	Persons	Slices	Patients	Slices	Patients	Slices	Persons	Slices
Sun Yat-sen Memorial Hospital	–	–	22	295	22	113	61	647	105	1055
Renmin Hospital of Wuhan University	262	2119	56	287	78	391	158	1511	554	4308
Total	262	2119	78	582	100	504	219	2158	659	5363

2.3 Data Augmentation

In total, we had CT slices from four groups of persons including patients of COVID-19, typical viral pneumonia, bacterial pneumonia, and healthy persons. However, the number of CT slices in these four categories varies greatly, and such data imbalance will affect the performance of the classification model. Moreover, since our model is based on deep learning, more samples are needed to learn image features. Therefore, we adopt the following three data enhancement methods: horizontal flipping, random translation of 0–8 pixels in four directions (up, down, left and right), and the combination of the previous two ways. The augmentation was performed at person level. Considering the number of existing slices in each category, we augmented 2 times, 8 times, 8 times and 2 times on the four groups of patients respectively, and ended up with a relatively close slices number, namely, the slices of COVID-19 infected patients, healthy people, bacterial pneumonia patients, and typical viral pneumonia patients were 4238, 4656, 4032 and 4316, respectively.

2.4 Neural Network Architecture

To accurately classify a person by his/her CT images, we developed a new framework based on deep learning neural networks. As shown in Fig. 2, the CT images were first preprocessed according to the above preprocessing steps, before they were input to the classification network

to predict the types for each image. Then, the image-level predictions of all images of each person were aggregated to provide human-level diagnosis. In this study, we simply averaged the predicted image-level probabilities of all image slices of a person by category, and chose the category with the highest score as the diagnosis result for the person.

Classification Neural Network

As illustrated in Fig. 3, we used ResNet50 [25] as the backbone network, and integrated the network with SE blocks as described in the SENet [28]. The ResNet50 was selected because we need a deep network to extract the hidden features in CT images that are more challenging than natural images. The SE blocks could make full use of the information between slices of CT images and between channels of feature maps by selectively emphasizing important information and suppressing the less important ones.

Concretely, for each building block of ResNet50, a channel squeeze and excitation operation was added for every three convolution layers (1×1 conv, 3×3 conv, 1×1 conv). In the SE block, the generated feature maps from ResNet blocks, $X \in R^{H \times W \times C}$ with $H \times W$ as the spatial dimension and C as the number of channel, were converted through a channel squeeze and excitation operation to $X' \in R^{H \times W \times C}$.

For the squeezing step $F_{CS}(\cdot)$, a simple global average pooling was used to shrink X through its spatial dimension $H \times W$, such that the c th channel of X was calculated by:

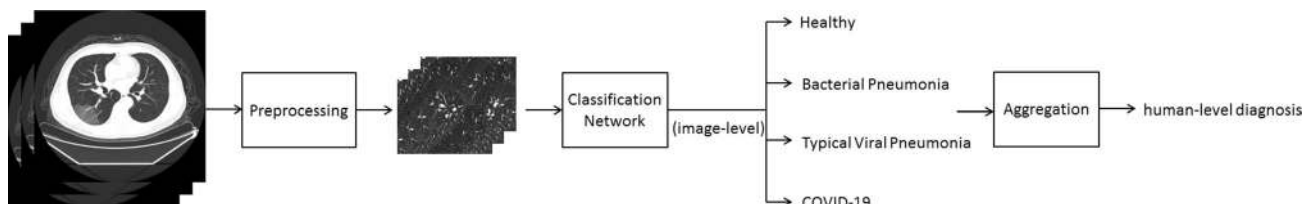


Fig. 2 The pipeline of the proposed system. The CT images were first preprocessed, and then sent to the classification network to make predictions in image level. Then, the image-level predictions of all images of each person were aggregated to provide human-level diagnosis

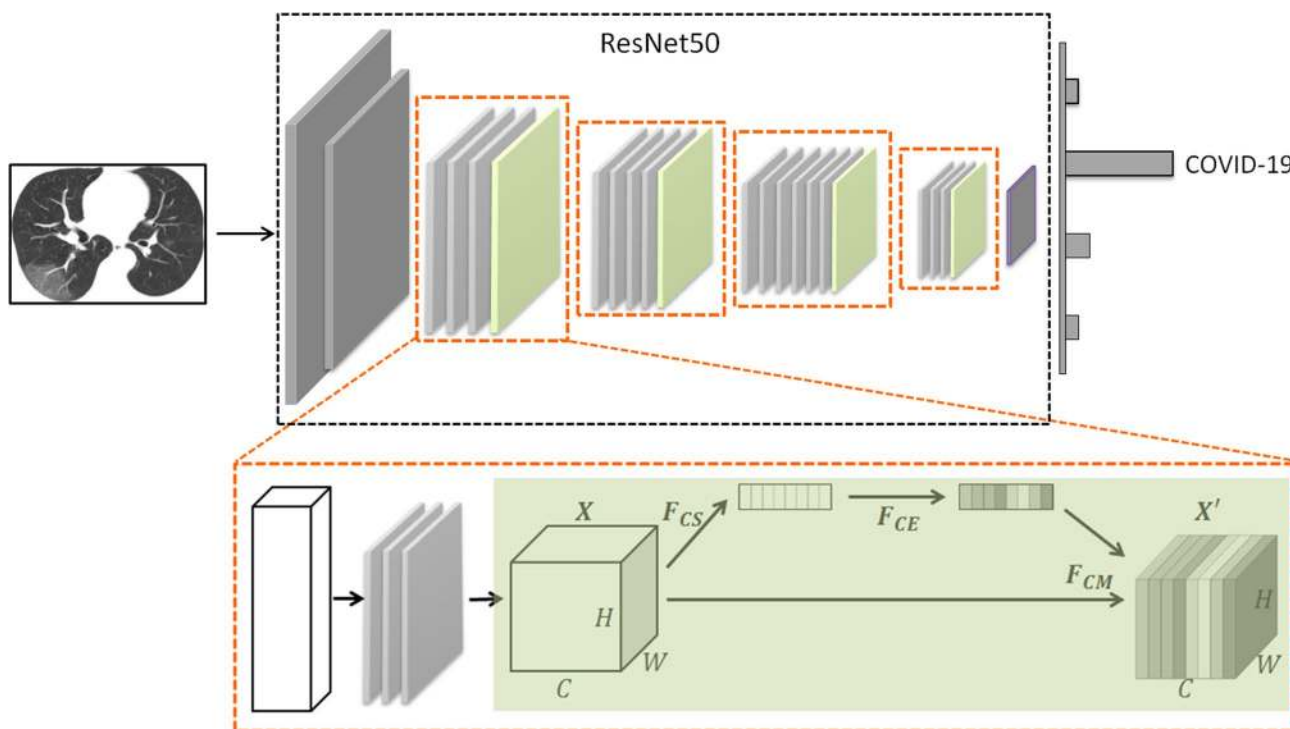


Fig. 3 The architecture of the classification model. Channel SE blocks were introduced to emphasize important channels and suppress less important ones

$$S_c = F_{CS}(X_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j). \tag{1}$$

Then, the excitation step $F_{CE}(\cdot)$ was performed with two linear transformations to the squeezed information S . The network could automatically learn the most important channels so as to endue these channels with higher attentions. The $F_{CE}(\cdot)$ operation was as follows:

$$E = F_{CE}(S, W) = \sigma(W_2 \delta(W_1 S + b_1) + b_2) \tag{2}$$

where σ and δ were the ReLU and Sigmoid functions, respectively, W_1 , W_2 , b_1 , and b_2 were weights and bias to learn. The value of each channel in E represented the importance of the channel learned by the network, which would be attached to the corresponding channel to obtain new features for the channel by:

$$X'_c = F_{CM}(E_c, X_c) \tag{3}$$

where $F_{CM}(\cdot)$ represented channel-wise multiplication.

After the above channel squeeze and excitation operations, a new feature map $X' = [X'_1, X'_2, \dots, X'_C]$ was generated, which emphasized informative channel features.

At the end of the network, a fully connected layer was used for the multi-class prediction by minimizing the cross-entropy loss.

2.5 Training Configurations and Implementation Details

Our method was implemented in Pytorch framework [29]. All experiments were conducted on a container equipped with 28 Intel Xeon Gold 6132 CPUs working at 2.6 GHz and 16 NVIDIA TESLA V100 SXM2 with 16 GB of memory. In the training stage of our method, we trained the deep networks end to end through back-propagation and Adam Optimizer [30] with an initial learning rate of $1e-5$. The model was trained for 100 epochs, which was sufficient for convergence, and the epochs with best validation performance were selected for test. The training batch size was set as 64, and the parameters were initialized by normalization [31].

2.6 Dataset Split Strategy

Our system could perform auxiliary diagnoses in both image and human levels. It is obvious that human-level results are more meaningful than image-level results for medical diagnoses. Therefore, we split the training, validation and test sets by person, so that images of one person are always in the same set.

As our data came from 2 different hospitals, and each hospital used different equipments for CT examination, so the CT slices collected were various in pixel size, spatial

resolution, layer thickness, and layer distance. The differences between these devices might interfere with the training and inference of the models. To avoid learning the differences between devices, we randomly extracted data only from the Renmin Hospital of Wuhan University to form the test set, and utilized the remaining as the training data. The number of persons and CT slices in the training set, validation set and test set for the quaternary classification task of all the four types of persons are shown in Table 2.

2.7 Metrics

The performance was evaluated by the following 5 metrics. The AUC (area under the receiver operating characteristics curve) of a classifier represents the probability that the positive instances of the prediction rank ahead of the negative ones [32]. Obviously, a classifier with a larger AUC works better. Recall, precision, F1-score and accuracy are defined as:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (5)$$

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

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (7)$$

where TP, FP, TN, and FN are the numbers of true positive, false positive, true negative, and false negative, respectively.

Table 2 The number of persons and CT slices in the training set, validation set and test set for the quaternary classification task of all the four types of persons

	COVID-19		Healthy people		Bacterial pneumonia		Typical viral pneumonia	
	Patients	Slices	Persons	Slices	Patients	Slices	Patients	Slices
Training set	205	1667	42	394	60	306	167	1713
Validation set	27	205	6	57	10	64	22	215
Test set	30	247	30	131	30	134	30	230
Total	262	2119	78	582	100	504	219	2158

Table 3 Performance of our classification model in identifying four different types of persons all at once

	AUC	Recall	Precision	F1-score	Accuracy
Macro average performance	0.96	0.94	0.95	0.94	0.94
Healthy	0.999	0.97	1	0.98	–
COVID-19	0.93	0.97	0.85	0.91	–
Bacterial pneumonia	0.97	1	0.97	0.98	–
Typical viral pneumonia	0.95	0.83	0.96	0.89	–

2.8 To Identify Four Different Types of Persons from Each Other

3 Results

We evaluated the performance of our classification model from the following aspects: (1) the ability of the model to identify four different types of persons from each other; (2) ablation study; (3) comparison with other models. Note that all the results were with data augmentation, except the comparison experiments in ablation study.

Table 3 exhibits the performance of our model to identify four different types of persons all at once. As shown in Table 3, our model achieved a macro average performance with AUC of 0.96, recall of 0.94, precision of 0.95, and F1-score of 0.94. The overall accuracy is 0.94. When considering each type, the separation of healthy persons has the highest AUC that is close to 100%. This is as expected because the other three types are different kinds of pneumonia and there are clear differences between the imaging features of healthy CT images and those of pneumonia CT images. The discriminations of bacterial and typical viral pneumonia achieved AUCs of 0.97 and 0.95, respectively. Though COVID-19 is the most difficult to discriminate, it achieved an AUC of 0.93 and a high recall of 0.97. It is worth mentioning that high recall is very important for such a COVID-19 diagnosis system because a higher recall means that fewer COVID-19 infected patients will be missed, which can greatly prevent further infection by missed diagnoses.

The receiver-operating characteristic curve and confusion matrix of our classification model in identifying four

different types of persons all at once are shown in Figs. 4a and 5a respectively. As can be seen in the confusion matrix in Fig. 5a, the model mistakenly identified some typical

viral pneumonia patients as COVID-19 infected patients, resulting in a slightly lower recall of typical viral pneumonia (as shown in Table 3). We visualized some CT images

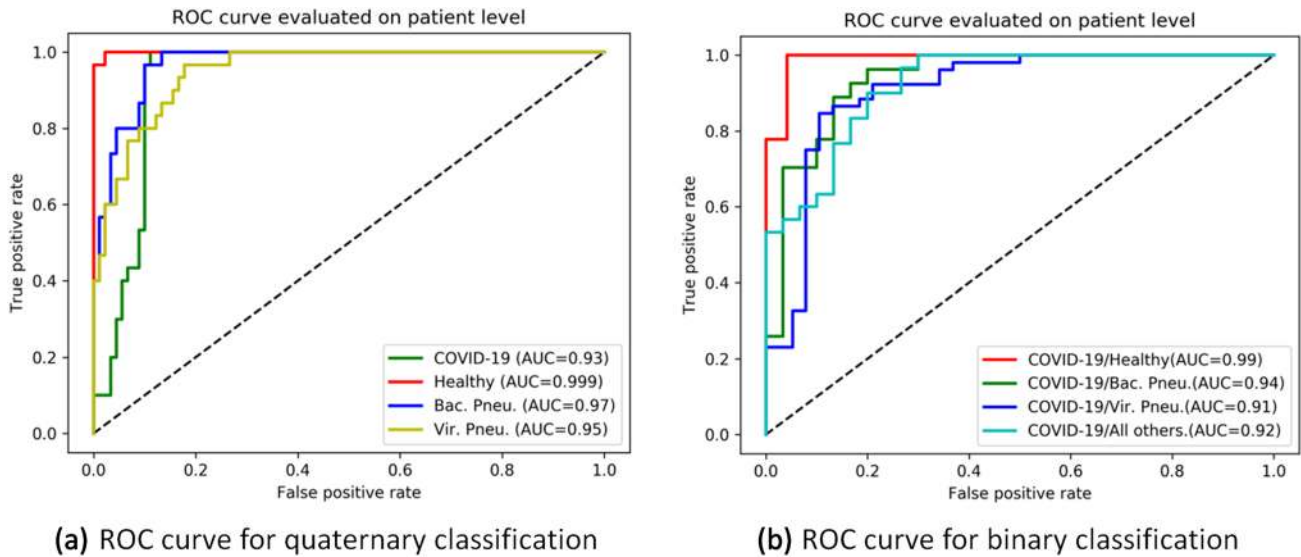


Fig. 4 The receiver-operating characteristic curves of our classification model. **a** Is for the quaternary classification in identifying four different types of persons all at once and **b** is for the binary classification in identifying COVID-19 from the other types

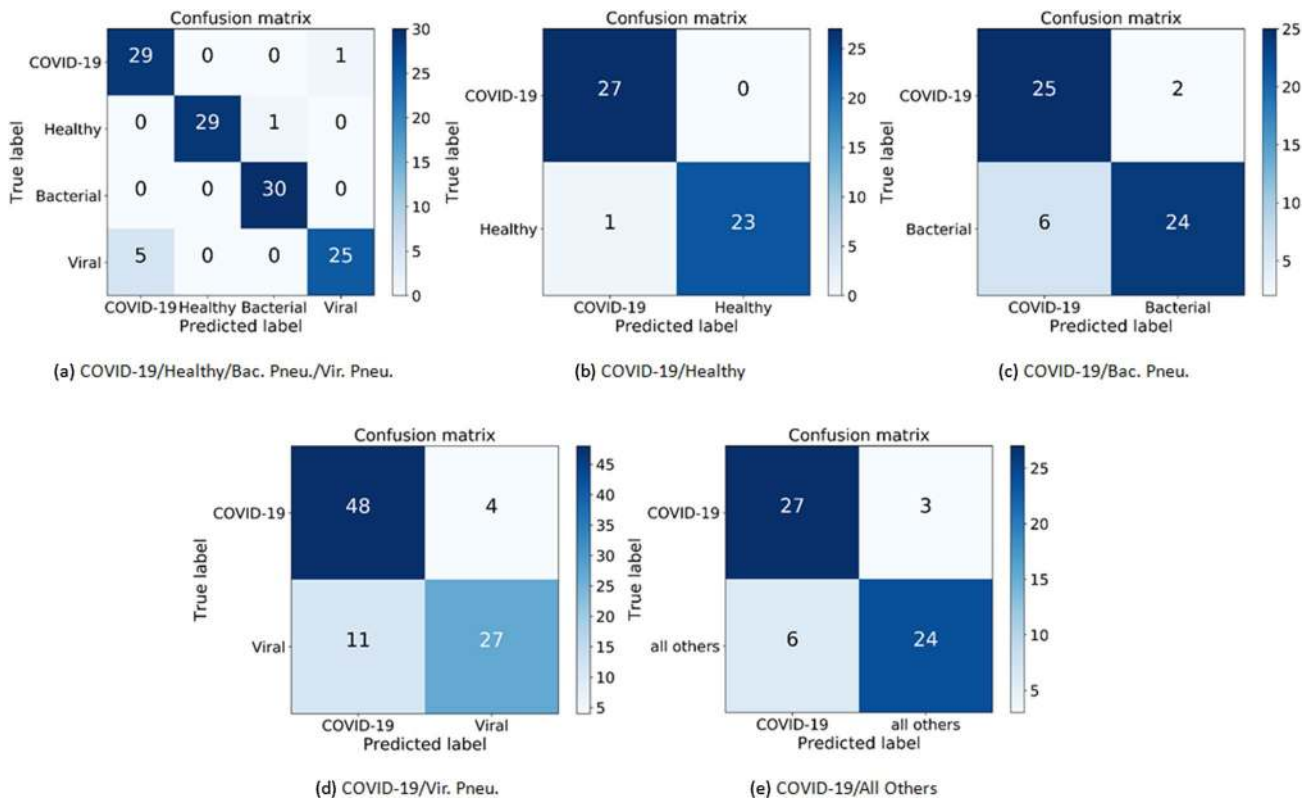


Fig. 5 Confusion matrix of our classification model. **a** Is for the quaternary classification in identifying four different types of persons all at once, **b–d** are for the binary classification in identifying COVID-19

from the other types respectively, and **e** is for the binary classification in identifying COVID-19 from all the others

of COVID-19 infected patients and typical viral pneumonia patients in Fig. 6 to figure out the reasons. As Fig. 6 shows, the correctly predicted COVID-19 images in Row (a) had very distinct imaging characteristics of COVID-19, which were very different comparing to the correctly predicted typical viral pneumonia images in Row (c). However, for the images of typical viral pneumonia that were incorrectly predicted as COVID-19, the images were indeed similar to those of COVID-19, especially in a single slice. Therefore, in future study, to improve the prediction performance, a whole CT image will be taken as input to extract 3D features. Moreover, some slices contained small lung areas, which also affected the learning of intrapulmonary characteristics. In practical clinical applications, the majority of lesions are found in the middle portion of the CT volume, so the anterior and posterior slices of the CT volume containing small lung areas can be removed to make the model focus on the learning of intrapulmonary features.

We further illustrated the feature maps of the CT images of the four different types of persons extracted by our classification model to explore the overall feature learning and representation capabilities of the network. As shown in Fig. 7, the areas where the lesions were located showed

a higher response, demonstrating that our model was able to learn the underlying characteristics of CT images of the three different types of patients.

Since the diagnosis framework may not be faced with so many types of data at the same time in daily routine examination, we removed the images of healthy persons, bacterial pneumonia patients and typical viral pneumonia patients respectively, and conducted a series of binary classification experiments to see if it still performed well. The experiments included: (1) to diagnose whether a person is healthy or with COVID-19; (2) to distinguish COVID-19 from bacterial pneumonia; (3) to distinguish COVID-19 from typical viral pneumonia; (4) to distinguish COVID-19 infected patients from all the other persons. In the above four binary classification tasks, the goal of our classification model was to detect COVID-19 infected patients. Once the predicted probability exceeded a certain threshold (threshold = 0.5 in this paper), the prediction was considered positive, otherwise it was considered negative. Table 4 shows the performance of our model in the binary classification tasks. It can be seen that the model still performed very well in distinguishing COVID-19 infected patients from healthy persons. This might due to the fact that comparing to COVID-19 infected

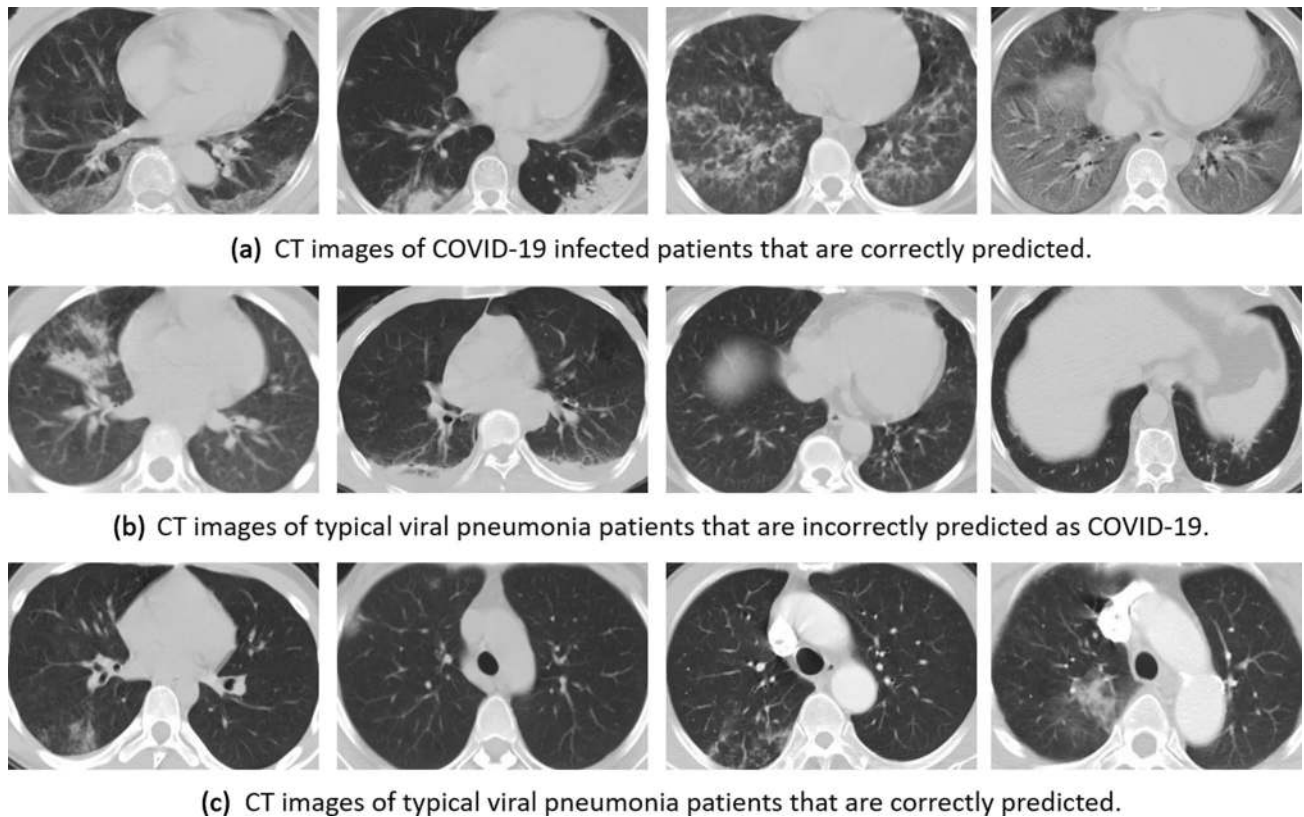


Fig. 6 CT images of COVID-19 and typical viral pneumonia patients. Row **a** are CT images of COVID-19 infected patients that are correctly predicted. Row **b** are CT images of typical viral pneumonia

patients that are incorrectly predicted as COVID-19. Row **c** are CT images of typical viral pneumonia patients that are correctly predicted

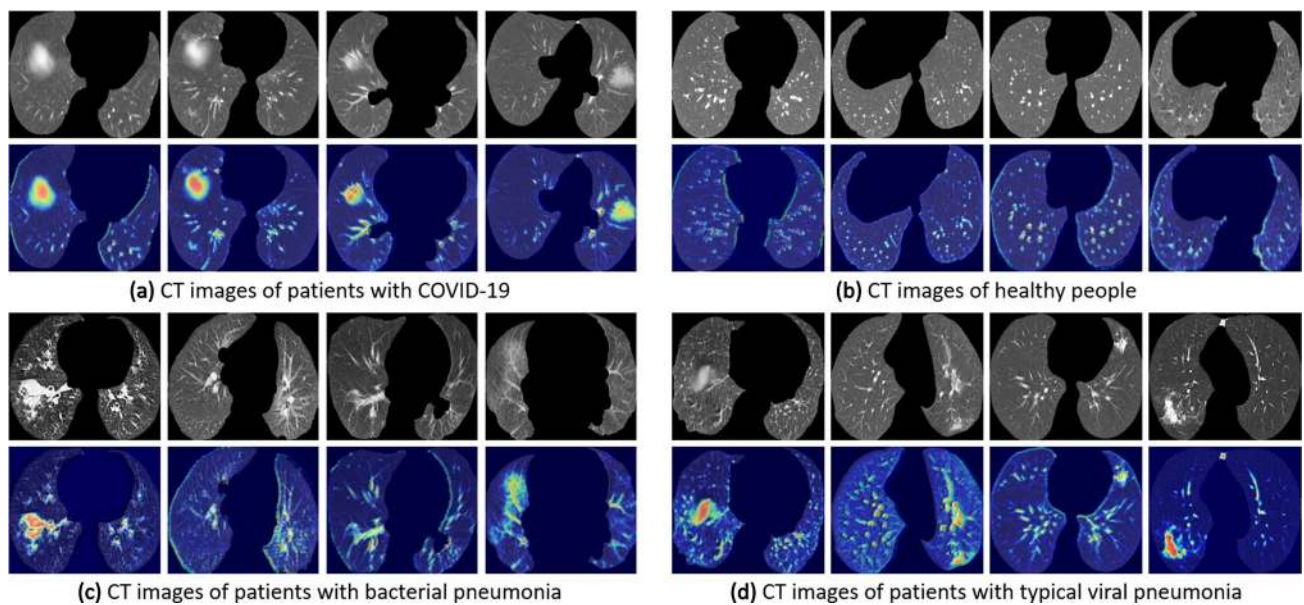


Fig. 7 The feature maps of the CT images of four types of persons extracted by our classification model

Table 4 The performance of our model in the binary classification tasks

Task	AUC	Recall	Precision	F1-score	Accuracy
COVID-19 or healthy?	0.99	1	0.96	0.98	0.98
COVID-19 or bacterial pneumonia?	0.94	0.93	0.81	0.86	0.86
COVID-19 or typical viral pneumonia?	0.91	0.92	0.81	0.86	0.83
COVID-19 or all others (healthy, bacterial pneumonia, typical viral pneumonia)?	0.92	0.90	0.82	0.86	0.85

patients, the lung parenchyma in the CT images of healthy persons was very clean and clear, without any lesions, which was very easy to distinguish. However, the results of other binary classification tasks were not as good as those of the quaternary classification tasks in Table 3. That was because in the quaternary classification, the model was fed with more diverse data, which enabled it to acquire stronger discrimination through learning. Therefore, in clinical application, it is better to train the network on more diverse data, e.g., the above four types of data, and then make auxiliary diagnosis according to the needs of daily examination. The receiver-operating characteristic curves and confusion matrixes are shown in Figs. 4b and 5b–e, respectively. As Fig. 5d shows, 11 typical viral pneumonia patients were wrongly diagnosed as COVID-19 infected patients, which was understandable. As visualized in Fig. 6 above, CT images of typical viral pneumonia are indeed similar to CT images of COVID-19, which may easily lead to misdiagnosis. In future studies, we will distinguish these two types of pneumonia more based on the combination of their pathological characteristics and CT image features. We also conducted a series of experiments of triple classification of COVID-19/Healthy/Bacterial

Pneumonia, COVID-19/Healthy/Typical Viral Pneumonia and COVID-19/Bacterial Pneumonia/Typical Viral Pneumonia, which are described in the supplementary material.

3.1 Ablation Study

Since the four groups of CT images we obtained are in different sample sizes, to avoid the performance loss caused by sample size unbalance and to prevent overfitting caused by insufficient samples, we adopted three data augmentation ways: horizontal flipping, random translation of 0–8 pixels in four directions of up, down, left and right, and the combination of the previous two. We conducted experiments to explore the effects of these three data augmentation ways. As can be seen in Row (1)–(4) and Row (7) in Table 5, the performance of the model was better when horizontal flipping, translation, or a combination of the two was used alone [Row (2), (3), (4)] than when no data enhancement was used at all [Row (1)], and the performance of the model was best when all three data enhancements were used [Row

Table 5 The performance of ablation study in quaternary classification. Row (1)–(4) are the performance of different data augmentation methods. Row (5) is the performance of our model without the SE

	Flipping	Translation	Flipping + translation	SE	Aggregation	AUC	Recall	Precision	F1-score	Accuracy
(1)				✓	✓	0.88	0.73	0.78	0.73	0.73
(2)	✓			✓	✓	0.90	0.77	0.81	0.77	0.79
(3)		✓		✓	✓	0.94	0.83	0.84	0.82	0.82
(4)			✓	✓	✓	0.95	0.87	0.92	0.88	0.87
(5)	✓	✓	✓		✓	0.94	0.80	0.82	0.79	0.80
(6)	✓	✓	✓	✓		0.92	0.84	0.82	0.83	0.84
(7)	✓	✓	✓	✓	✓	0.96	0.94	0.95	0.94	0.94

(7)], which meant that all the three data augmentation ways we used were effective.

Moreover, we conducted experiments to inquiry the impact of the SE blocks we integrated into the backbone network and the effect of aggregation. By comparing Row (5) and Row (7) in Table 5, we can conclude that the SE blocks do work. After the integration of the SE blocks, the model had a great improvement in the metrics of recall, precision, F1-score and Accuracy. The main reason, in our view, was that the importance of CT Slices of a patient varied, as did the importance of the various channels of the feature maps after the feature extraction network. With the introduction of SE Blocks, it was helpful to discover the more important slices and feature map channels, thereby directing the network to learn the more important features.

In addition, by comparing Row (6) and Row (7) in Table 5, it can be found that the aggregation of image-level results into human-level results was not only more in line with the actual diagnostic needs, but also significantly improved the diagnostic performance. This was because, in image-level prediction, there were no lesions in some slices of the patient, which would lead to slight deviation in the image-level prediction. By aggregating image-level results into human-level results, such deviation could be alleviated.

3.2 Comparison with Other Models

The model was compared with other existing deep learning models, i.e., DenseNet, VGG, and ResNet. We conducted all experiments using the same data split strategy and training configuration. The results in Table 6 show that our model outperformed other models. Consistent with the previous ablation study in Table 5, our network exceeded ResNet in every metric. On the one hand, it was due to the strong learning ability of the backbone network and the alleviation of the performance degradation of the deep network caused by the residual layer; on the other hand, it was mainly due to the addition of SE module, which ensured that the features of the multi-channel feature maps could be fully learned.

blocks. Row (6) is the performance of our model in image level without aggregation into human level. Row (7) is the performance of our model

Table 6 Performance of our classification model comparing with other existing models

Model	AUC	Recall	Precision	F1-score	Accuracy
DenseNet	0.90	0.79	0.78	0.78	0.79
VGG	0.93	0.80	0.81	0.79	0.80
ResNet	0.94	0.80	0.82	0.79	0.80
OURS	0.96	0.94	0.95	0.94	0.94

4 Discussion

Currently, identification of COVID-19 infected patients from bacterial pneumonia patients and typical viral pneumonia patients is important for taking accurate treatments for COVID-19. As indicated by many previous studies [10, 11], the CT images of typical viral pneumonia patients and bacterial pneumonia patients are similar to that of COVID-19 infected patients. Especially, these images all have shadow and ground-glass opacity. Accurately distinguishing them in short time is critical for doctors to diagnose immediately. To increase the accuracy of diagnosis and reduce the burdens of doctors in reading CT images, it is important to develop a computer-based approach to classify the pneumonia types according to CT images. However, most of the current models are constructed to classify the COVID-19 and healthy controls or bacterial pneumonia, and have ignored typical viral pneumonia. For example, Xu et al. [11] distinguished COVID-19 patients, Influenza-A patients and healthy persons using a deep learning model. Li et al. [12] used ResNet50 to discriminate COVID-19 from non-pneumonia or community-acquired pneumonia. Song et al. [10] proposed a deep CT diagnosis system to detect COVID-19 infected patients from healthy persons and bacterial pneumonia patients. Since COVID-19 is also a type of viral pneumonia and its imaging features are similar to those of typical viral

pneumonia, it is of great significance to assist doctors in distinguishing COVID-19 from typical viral pneumonia.

In this study, we integrated ResNet and SE blocks to develop a model to distinguish COVID-19 infected patients, healthy persons, bacterial pneumonia patients and typical viral pneumonia patients all at once. This model was different from previous methods in several aspects. First, it took multiple slices as input to take full advantage of the contextual information between slices. Then, it focused on the relationship between multiple slices, which was unique to medical images, and used a SE module to learn the different importance of multiple slices and multiple channels of the feature maps. Most importantly, it was trained on data of COVID-19, healthy persons, bacterial pneumonia, and typical viral pneumonia, which enabled the model to identify more types of persons and pneumonia than the previous models. Because of the properties of this model, it is accurate in distinguishing the pneumonia types. Moreover, comparison with other models showed that our model achieved higher AUC, Recall, Precision, F1-score and Accuracy. Thus, this model has the potential to become a daily used tool for doctors to classify pneumonia patients especially when the COVID-19 may become a long term existing virus. Another advantage of this model is that it can diagnose quickly. For a slice of CT, the model can give an image-level diagnosis in just 20 milliseconds.

On the other hand, this model can be further improved in many aspects. First, the current model adopted 2D CNN. Although multiple slices were used to retain the context information of channels, 2D CNN was inferior to 3D CNN in learning such volume information as CT images. Therefore, in subsequent study, we consider using 3D CNN to learn the information of the entire CT volume. Second, in a complete CT volume, the anterior and posterior slices contain very small areas of lung parenchyma, and they can provide little diagnostic information. Therefore, these slices can be removed in subsequent study to prevent the network from learning irrelevant information, so as to improve the efficiency of diagnosis. Third, as experimental results show, it is more difficult to distinguish typical viral pneumonia from COVID-19. One possible reason is that the model based on deep learning needs a great deal of samples for training, but currently there are not enough samples. Therefore, it is considered to collect more samples of COVID-19 and typical viral pneumonia in subsequent studies, so that the model can rely on more samples for training to extract more discriminative features in CT images of the two types. Then, since CT images of typical viral pneumonia patients are very similar to that of COVID-19 infected patients, pathological characteristics of these two types of pneumonia can be used to assist in discrimination. Finally, inspired by [33] and [34], on the basis of the existing category label, we consider to increase the disease severity, such as the area ratio of the

lesion to the lung, as an additional label to perform multi-label classification.

5 Conclusion

We have developed a CT image diagnosis system via deep learning for rapid COVID-19 diagnosis by integrating ResNet with SE blocks. This model can identify COVID-19 CT from CT of healthy persons, CT of bacterial pneumonia patients and CT of typical viral pneumonia patients separately. This is the first model to distinguish between so many different types of pneumonia all at once. Experimental results indicated that our model achieves high AUC, recall and precision, which indicated the reliability of the model. The model performed better than the model using ResNet only, which indicated the effectiveness of SE in feature extraction.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12539-021-00420-z>.

Acknowledgements The work was supported in part by the National Key R&D Program of China (2018YFC1315405), National Natural Science Foundation of China (U1611261, 61772566, 81801132 and 81871332), Guangdong Frontier and Key Tech Innovation Program (2018B010109006 and 2019B020228001), Natural Science Foundation of Guangdong, China (2019A1515012207), and Introducing Innovative and Entrepreneurial Teams (2016ZT06D211).

Compliance with ethical standards


Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Lu R, Zhao X, Li J et al (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395(10224):565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
2. Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
3. Li Q, Guan X, Wu P et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 382:1199–1207. <https://doi.org/10.1056/NEJMoa2001316>
4. WHO (2020) Coronavirus disease (COVID-19) situation report—207. <https://www.who.int/docs/default-source/coronavirus/situation-reports/20200814-covid-19-sitrep-207.pdf>. Accessed 15 Aug 2020.
5. Lei J, Li J, Li X, Qi X (2020) CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology* 295(1):18. <https://doi.org/10.1148/radiol.2020200236>
6. Shi H, Han X, Zheng C (2020) Evolution of CT manifestations in a patient recovered from 2019 novel coronavirus (2019-nCoV)

- pneumonia in Wuhan, China. *Radiology* 295(1):20. <https://doi.org/10.1148/radiol.2020200269>
7. Song F, Shi N, Shan F et al (2020) Emerging 2019 novel coronavirus (2019-nCoV) Pneumonia. *Radiology* 295(1):210–217. <https://doi.org/10.1148/radiol.2020200274>
 8. COVID-19 Virtual Press Conference 13 May 2020. <https://www.who.int/docs/default-source/coronaviruse/transcripts/who-press-conference-13may2020.pdf>. Accessed 15 Aug 2020.
 9. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z et al (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Rev Biomed Eng*. <https://doi.org/10.1109/RBME.2020.2987975>
 10. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z et al (2020) Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *MedRxiv*. <https://doi.org/10.1101/2020.02.23.20026930>
 11. Xu X, Jiang X, Ma C, Du P, Li X, Lv S et al (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 6(10):1122–1129. <https://doi.org/10.1016/j.eng.2020.04.010>
 12. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B et al (2020) Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 296(2):E65–E71. <https://doi.org/10.1148/radiol.2020200905>
 13. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y et al (2020) Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Sci Rep* 10:1–11. <https://doi.org/10.1038/s41598-020-76282-0>
 14. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H et al (2020) Deep learning-based detection for COVID-19 from chest CT using Weak Label. *MedRxiv*. <https://doi.org/10.1101/2020.03.12.20027185>
 15. Jin C, Cheny W, Cao Y, Xu Z, Zhang X, Deng L et al (2020) Development and evaluation of an AI system for COVID-19 diagnosis. *MedRxiv*. <https://doi.org/10.1101/2020.03.20.20039834>
 16. Wang B, Jin S, Yan Q, Xu H, Luo C, Wei L et al (2020) AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system. *Appl Soft Comput*. <https://doi.org/10.1016/j.asoc.2020.106897>
 17. Shi F, Xia L, Shan F, Wu D, Wei Y, Yuan H et al (2020) Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. [arXiv:2003.09860v1](https://arxiv.org/abs/2003.09860v1)
 18. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F (2021) A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip Sci Comput Life Sci*. <https://doi.org/10.1007/s12539-020-00403-6>
 19. Zhang R, Guo Z, Sun Y, Lu Q, Xu Z, Yao Z et al (2020) COVID19XrayNet: a two-step transfer learning model for the COVID-19 detecting problem based on a limited number of chest X-ray images. *Interdiscip Sci Comput Life Sci* 12:555–565. <https://doi.org/10.1007/s12539-020-00393-5>
 20. Ouyang X, Huo J, Xia L, Shan F, Liu J, Mo Z et al (2020) Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans Med Imaging* 39(8):2595–2605. <https://doi.org/10.1109/TMI.2020.2995508>
 21. Han Z, Wei B, Hong Y, Li T, Cong J, Zhu X et al (2020) Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. *IEEE Trans Med Imaging* 39(8):2584–2594. <https://doi.org/10.1109/TMI.2020.2996256>
 22. Kang H, Xia L, Yan F, Wan Z, Shi F, Yuan H et al (2020) Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. *IEEE Trans Med Imaging* 39(8):2606–2614. <https://doi.org/10.1109/TMI.2020.2992546>
 23. Apostolopoulos ID, Mpesiana TA (2020) COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43:635–640. <https://doi.org/10.1007/s13246-020-00865-4>
 24. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M (2020) Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1788642>
 25. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778. <https://doi.org/10.1109/CVPR.2016.90>
 26. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations, pp 1–14. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
 27. Huang G, Liu Z, Maaten LVD, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
 28. Hu J, Shen L, Albanie S, Sun G, Wu E (2018) Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 42(8):2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
 29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G et al (2019) PyTorch: an imperative style, high-performance deep learning library. In: International conference on neural information processing systems, pp 8024–8035. <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>. Accessed 17 Aug 2020.
 30. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: International conference on learning representations. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
 31. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. *J Mach Learn Res* 9:249–256
 32. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
 33. Chu Y, Shan X, Chen T, Jiang M, Wang Y, Wang Q, Salahub DR, Xiong Y, Wei DQ (2020) DTI-MLCD: predicting drug-target interactions using multi-label learning with community detection method. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbaa205>
 34. Shan X, Wang X, Li CD, Chu Y, Zhang Y, Xiong Y, Wei DQ (2019) Prediction of CYP450 enzyme-substrate selectivity based on the network-based label space division method. *J Chem Inf Model* 59(11):4577–4586. <https://doi.org/10.1021/acs.jcim.9b00749>

Authors and Affiliations

Fudan Zheng¹  · Liang Li² · Xiang Zhang³ · Ying Song⁴ · Ziwang Huang¹ · Yutian Chong⁵ · Zhiguang Chen^{1,6} · Huiling Zhu⁷ · Jiahao Wu⁸ · Weifeng Chen⁹ · Yutong Lu^{1,6} · Yuedong Yang^{1,6} · Yunfei Zha² · Huiying Zhao³ · Jun Shen³

¹ School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China

² Department of Radiology, Renmin Hospital of Wuhan University, Wuhan 430060, China

- ³ Department of Radiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510530, China
- ⁴ School of Systems Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China
- ⁵ Department of Radiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510220, China
- ⁶ National Supercomputing Center in Guangzhou, Guangzhou 510006, China

- ⁷ College of Information Science and Technology, Jinan University, Guangzhou 510632, China
- ⁸ School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510006, China
- ⁹ School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510006, China