



Research  
Coronavirus Disease 2019—Article

## A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia



Xiaowei Xu <sup>a, #</sup>, Xiangao Jiang <sup>b, #</sup>, Chunlian Ma <sup>c, #</sup>, Peng Du <sup>d, #</sup>, Xukun Li <sup>d</sup>, Shuangzhi Lv <sup>e</sup>, Liang Yu <sup>a</sup>, Qin Ni <sup>a</sup>, Yanfei Chen <sup>a</sup>, Junwei Su <sup>a</sup>, Guanqing Lang <sup>a</sup>, Yongtao Li <sup>a</sup>, Hong Zhao <sup>a</sup>, Jun Liu <sup>a</sup>, Kaijin Xu <sup>a</sup>, Lingxiang Ruan <sup>e</sup>, Jifang Sheng <sup>a</sup>, Yunqing Qiu <sup>a</sup>, Wei Wu <sup>a, \*</sup>, Tingbo Liang <sup>f, \*</sup>, Lanjuan Li <sup>a, \*</sup>

<sup>a</sup> State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

<sup>b</sup> Department of Infectious Disease, Wenzhou Central Hospital, Wenzhou 325000, China

<sup>c</sup> Department of Infectious Disease, The First People's Hospital of Wenling, Wenling 317500, China

<sup>d</sup> Artificial Intelligence Lab, Hangzhou AiSmartVision Co., Ltd., Hangzhou 310012, China

<sup>e</sup> Department of Radiology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

<sup>f</sup> Department of Hepatobiliary and Pancreatic Surgery & Key Lab of Pancreatic Diseases Research of Zhejiang Province & The Innovation Centre for the Study of Pancreatic Diseases of Zhejiang Province & Clinical Medical Research Center of Hepatobiliary and Pancreatic Diseases in Zhejiang Province & Precision Innovation Center of the Diagnosis and Treatment of Hepatobiliary and Pancreatic Diseases of Zhejiang University, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

### ARTICLE INFO

#### Article history:

Received 23 March 2020

Revised 21 April 2020

Accepted 28 April 2020

Available online 27 June 2020

#### Keywords:

COVID-19

Location-attention classification model

Computed tomography

### ABSTRACT

The real-time reverse transcription-polymerase chain reaction (RT-PCR) detection of viral RNA from sputum or nasopharyngeal swab had a relatively low positive rate in the early stage of coronavirus disease 2019 (COVID-19). Meanwhile, the manifestations of COVID-19 as seen through computed tomography (CT) imaging show individual characteristics that differ from those of other types of viral pneumonia such as influenza-A viral pneumonia (IAVP). This study aimed to establish an early screening model to distinguish COVID-19 from IAVP and healthy cases through pulmonary CT images using deep learning techniques. A total of 618 CT samples were collected: 219 samples from 110 patients with COVID-19 (mean age 50 years; 63 (57.3%) male patients); 224 samples from 224 patients with IAVP (mean age 61 years; 156 (69.6%) male patients); and 175 samples from 175 healthy cases (mean age 39 years; 97 (55.4%) male patients). All CT samples were contributed from three COVID-19-designated hospitals in Zhejiang Province, China. First, the candidate infection regions were segmented out from the pulmonary CT image set using a 3D deep learning model. These separated images were then categorized into the COVID-19, IAVP, and irrelevant to infection (ITI) groups, together with the corresponding confidence scores, using a location-attention classification model. Finally, the infection type and overall confidence score for each CT case were calculated using the Noisy-OR Bayesian function. The experimental result of the benchmark dataset showed that the overall accuracy rate was 86.7% in terms of all the CT cases taken together. The deep learning models established in this study were effective for the early screening of COVID-19 patients and were demonstrated to be a promising supplementary diagnostic method for frontline clinical doctors.

© 2020 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

At the end of 2019, the coronavirus disease 2019 (COVID-19) was reported [1–4]. On 24 January 2020, Huang et al. [5] summarized the clinical characteristics of 41 patients with

COVID-19, indicating that the common onset symptoms were fever, cough, myalgia, or fatigue. All 41 of these patients had pneumonia, and their chest computed tomography (CT) examination showed abnormalities. Complications included acute respiratory distress syndrome, acute heart injury, and secondary infections. Thirteen (31.7%) patients were admitted to the intensive care unit (ICU), and six (14.6%) died. Chan et al. [6] at the University of Hong Kong, China found evidence of human-to-human transmission of COVID-19 for the first time.

\* Corresponding authors.

E-mail addresses: [1198042@zju.edu.cn](mailto:1198042@zju.edu.cn) (W. Wu), [liangtingbo@zju.edu.cn](mailto:liangtingbo@zju.edu.cn) (T. Liang), [ljl@zju.edu.cn](mailto:ljl@zju.edu.cn) (L. Li).

# These authors contributed equally to this work.

The *Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7)* [7] recommended the etiological confirmation of patients with COVID-19 by means of two technologies: nucleic acid testing or specific antibody testing. The accessibility of nucleic acid testing has greatly improved in the past three months, but shortcomings still exist, such as high operating requirements, a time-consuming procedure, and a relatively low positive rate [8,9]. A study from Wuhan showed that the positive rate of nucleic acid testing from oral swabs, anal swabs, and the blood of infected patients was 53.3%, 26.7%, and 40.0% respectively [10]. Furthermore, antibody detection is not appropriate for early screening, since there is a window phase for antibodies testing. Long et al. [11] reported that the median time of seroconversion for both immunoglobulin G (IgG) and IgM was about 13 d post onset.

At the same time, radiology experts noticed that the manifestations of COVID-19 cases as seen through CT imaging had their own characteristics, which differed from those of the CT imaging manifestations of other viral pneumonias, such as influenza-A viral pneumonia (IAVP), as shown in Fig. 1. Therefore, clinical doctors chose to replace nucleic acid testing with lung CT imaging as one of the early diagnostic criteria for this new type of pneumonia [12], with the aim of immediately curbing transmission.

With the rapid development of computer technology, digital image processing technology has been widely applied in the medical field, including organ segmentation and image enhancement and repair, thereby providing support for subsequent medical diagnosis [13,14]. Deep learning technologies, such as the convolutional neural network (CNN) with its strong ability of non-linear modeling, have also been applied extensively in medical image processing [15–18]. Relevant studies have been conducted on the diagnosis of pulmonary nodules [19], the classification of benign and malignant tumors [20,21], and pulmonary tuberculosis analysis and disease prediction [22–24] worldwide.

In this study, multiple CNN models were used to classify CT image datasets and calculate the infection probability of COVID-19. These findings might greatly assist in the early screening of patients with COVID-19.

## 2. Materials and methods

### 2.1. Study dataset

A total of 618 transverse-section CT samples were collected in this study, including 219 from 110 patients (mean age 50 years; 63 (57.3%) male patients) with COVID-19 from the First Affiliated Hospital, College of Medicine, Zhejiang University; Wenzhou Central Hospital; and the First People's Hospital of Wenling from 19 January to 14 February 2020. All three hospitals are designated COVID-19 hospitals in Zhejiang Province. Every COVID-19 patient was confirmed with real-time reverse transcription-polymerase chain reaction (RT-PCR) testing from sputum or nasopharyngeal swab, and cases with no image manifestations on the chest CT images were excluded. In addition, there was a gap of at least two days between CT datasets scanned from the same patient in order to ensure the diversity of samples. The remaining 399 CT samples were collected from the First Affiliated Hospital, College of Medicine, Zhejiang University as the controlled experiment group. Of these, 224 CT samples were from 224 patients (mean age 61 years; 156 (69.6%) male patients) with IAVP including H1N1, H3N2, H5N1, H7N9, and so forth; 175 CT samples (mean age 39 years; 97 (55.4%) male patients) were from healthy people. The diagnosis of IAVP was proved by the RT-PCR detection of viral RNA from sputum or nasopharyngeal swab. There were 198 (90.4%) COVID-19 and 196 (87.5%) IAVP cases in early or progres-

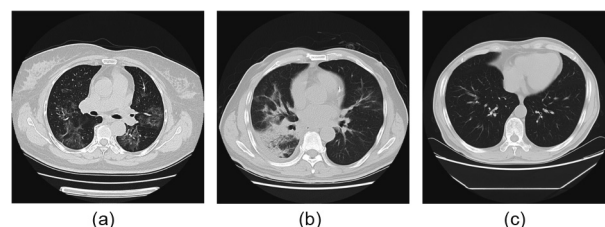


Fig. 1. Typical transverse-section CT images: (a) COVID-19; (b) IAVP; and (c) no pneumonia manifestations. Both (a) and (b) were taken within 10 d from the onset of the symptoms.

sive stages; the remaining 9.6% and 12.5% cases, respectively, were in the severe stage; no significant differences of stages between two diseases. IAVP CT samples were used because it was critical to distinguish IAVP from patients with suspected COVID-19 currently in China.

The ethics committee of the First Affiliated Hospital, College of Medicine, Zhejiang University approved this study, and the research was performed in accordance with relevant guidelines and regulations. All participants and/or their legal guardians signed an informed consent form before the study took place.

A total of 528 CT samples (85.4%) were used for training and validation sets, including 189 samples from patients with COVID-19, 194 samples from patients with IAVP, and 145 samples from healthy people. The remaining 90 CT sets (14.6%) were used as the test set, including 30 COVID-19 cases, 30 IAVP cases, and 30 healthy cases. Furthermore, the test cases of the CT set were selected from people who had not been included in the training stage.

### 2.2. Process

Fig. 2 shows the whole process of COVID-19 diagnostic report generation in this study. First, the CT images were preprocessed to extract the effective pulmonary regions. Second, a three-dimensional (3D) CNN segmentation model was used to “segment” multiple candidate image cubes. The center image together with its two neighbors of each cube was collected for further steps. Third, an image classification model was used to categorize all the image patches into three types: COVID-19, IAVP, and irrelevant to infection (ITI). Image patches from the same cube “voted” for the type and confidence score of this candidate as a whole. Finally, the overall analysis report for one CT sample was calculated using the Noisy-OR Bayesian function [25].

## 3. Theory and calculation

### 3.1. Dataset preprocessing and candidate region segmentation

The study was expedited by using the same method and models for the data preprocessing and candidate region segmentation stages as those used in a previous study on pulmonary tuberculosis [23]. The focus of the infections from pulmonary tuberculosis had multiple structures and types, including miliary, infiltrative, caseous, tuberculoma, and cavitory. The model, VNet–inception-residual network (IR)–region proposal network (RPN) [23], used a VNet [26] backbone with an IR [27] structure as the feature-extracting part, which was followed by an RPN to segment candidate regions. Although it was trained for the purpose of pulmonary tuberculosis detection, the model was verified by professional radiologists to be good enough to separate candidate patches for viral pneumonia.

The 3D segmentation process mentioned here is not a traditional pixel-level segmentation operation such as VNet or 3D

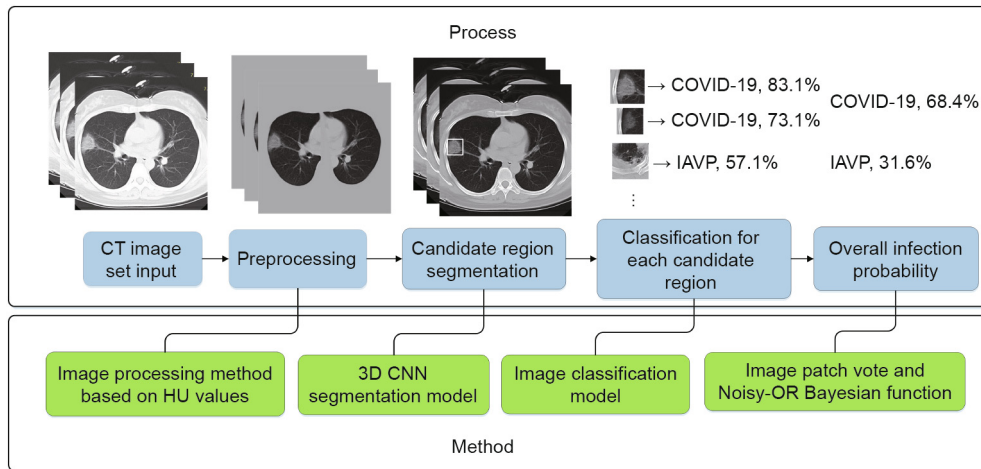


Fig. 2. Process flow chart—take one COVID-19 case as example. HU: Hounsfield unit.

u-net [28]. As the border between a healthy region and the focus of infection is usually blurred and indistinct, it was extremely difficult to label pixel-level masks for lesion regions of the pneumonia. The variable 3D structures of the lesion regions also aggravated this work. Therefore, the segmentation operation used here was more like the operations in an object-detection algorithm. An RPN structure was used to capture the region of interest with 3D bounding boxes instead of pixel-level segmented masks.

Moreover, in the study of pulmonary tuberculosis, the VNet-IR-RPN model was used for both segmentation and classification. Only the segmentation-related bounding box regression part was preserved, regardless of the classification results, because only the former task was required at this stage in this study.

### 3.2. Image data processing and augmentation

A large number of non-infection regions irrelevant to this study were also separated using the 3D CNN segmentation model, including the pulmonary fibrotic structure, calcification spots, and healthy regions that had been identified incorrectly. Therefore, the extra category of ITI was added to the COVID-19 and IAVP categories.

This study included 618 CT samples (219 COVID-19, 224 IAVP, and 175 healthy cases). Subsequently, a total of 3957 candidate cubes were generated from the 3D CNN segmentation model. Only the territory close to the middle of this cube contained maximum information about the focus of infection. Hence, only the center image together with its two neighbors of each cube was collected to represent this region for further classification steps. Next, all image patches were manually classified by two professional radiologists into two types: ITI and pneumonia. The images in the latter category were recognized automatically as COVID-19 or IAVP based on the clinical diagnosis results.

A total of 11 871 image patches were acquired from the aforementioned steps, including 2634 COVID-19, 2661 IAVP, and 6576 ITI. According to the previous dataset assignment, the training and validation sets had 528 CT samples, which were equivalent to 10 161 (85.6%) images, including 2301 COVID-19, 2244 IAVP, and 5616 ITI images. The remaining 1710 (14.4%) images were reserved for the test dataset.

The sampling possibility of the COVID-19 and IAVP cases was expanded three times to balance the specimen number of ITI cases, in order to reduce the influence of the uneven distribution of different image types on the present dataset. At the same time, generic data-expansion mechanisms, such as random clipping,

left-right flipping, up-down flipping, and mirroring operation, were performed on specimens to increase the number of training samples and prevent data overfitting.

### 3.3. Deep learning model for classification

#### 3.3.1. Location-attention classification

The work of Kanne [29] and Chung et al. [30] has shown at least three distinguishing features of COVID-19: a ground-glass appearance, peripheral distribution along with the pleura, and usually more than one independent focus of infection in a single case, as shown in Fig. 3.

The models were optimized based on these findings. The image classification model was designed to distinguish the appearance and structure of different infections. Moreover, relative distance from the edge was used as an extra weight for the model in order to learn the relative location information of the patch on the pulmonary image. Focuses of infection that were located close to the pleura were more likely to be recognized as COVID-19.

The relative distance from the edge of each patch was calculated as follows:

Step 1: Measure the minimum distance from the mask to the center of the patch (double-headed arrow in Fig. 3(c)).

Step 2: Obtain the diagonal of the minimum circumscribed rectangle of the pulmonary image (Fig. 3(d)).

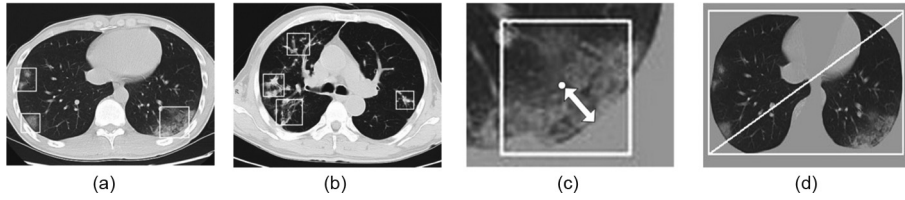
Step 3: Obtain the relative distance from the edge by dividing the distance from Step 1 by the diagonal from Step 2.

#### 3.3.2. Network structure

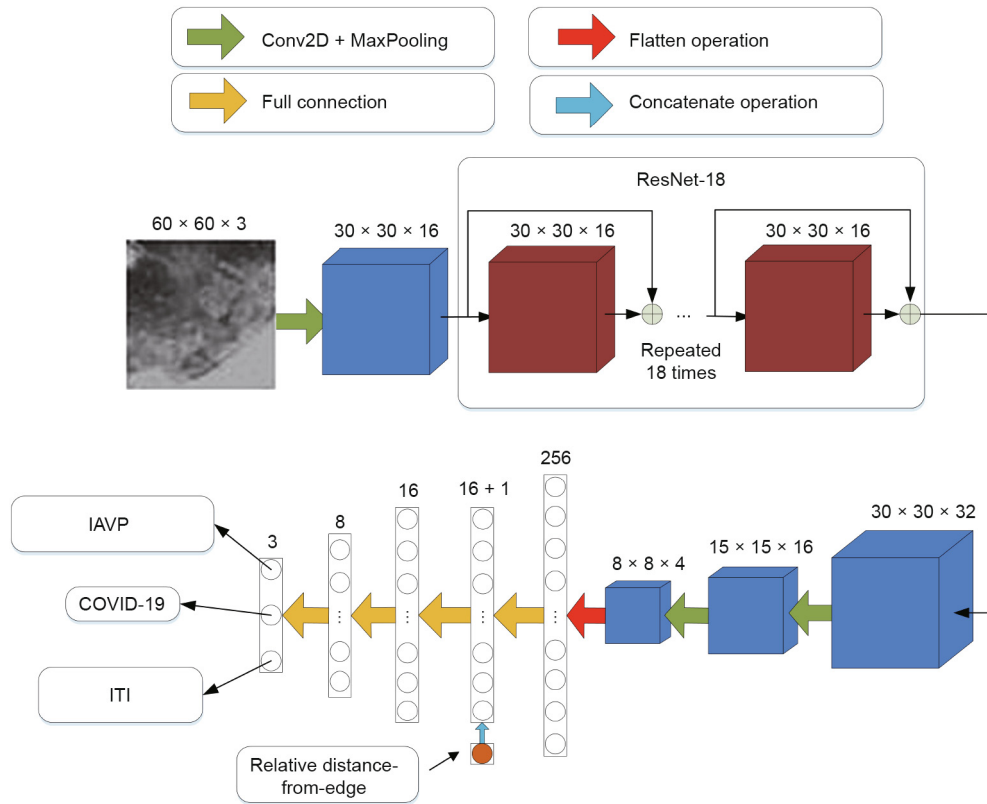
Two classification models were evaluated in this study, as shown in Fig. 4. One was a relative traditional residual network (ResNet)-based [31] model and the other was designed based on the first network structure by concatenating the location-attention mechanism in the full-connection layer to improve the overall accuracy rate. This mechanism was added to the first full-connection layer to maximize the influence of this factor on the whole network.

The classical ResNet-18 network structure was used for image feature extraction. Pooling operations were also used for the dimensional reduction of data in order to prevent overfitting and improve the problem of generalization.

The output of the convolution layer was flattened to a 256-dimensional feature vector and then converted into a 16-dimensional feature vector using a full-connection network. For the location-attention classification model, the value of the



**Fig. 3.** (a) COVID-19 image with three ground-glass focuses of infections; (b) IAVP image with four focuses of infections; (c) the minimum distance from the mask to the center of the patch (double-headed arrow); (d) diagonal of the minimum circumscribed rectangle of the pulmonary image.



**Fig. 4.** The network structure of traditional ResNet-18-based classification model (without the relative distance-from-edge mechanism). The location-attention classification model was built on the backbone of ResNet-18 by concatenating the location-attention mechanism in the full-connection layer to improve the overall accuracy rate. Conv2D: convolution 2D.

relative distance from the edge was first normalized to the same order of magnitude, and then concatenated to this full-connection network. Next, three full-connection layers were followed to output the final classification result together with the confidence score.

### 3.4. Diagnostic report

#### 3.4.1. “Voting” for each candidate region

Inspired by the theory of the Bagging prediction algorithm [32] in machine learning technology, one candidate region was represented by three image patches: the center image and its two neighbors. These three images “voted” for the whole region according to the following strategies:

- (1) If at least two images were categorized into the same type, then the image with the maximum confidence score in this type was selected.
- (2) Otherwise, the image with the maximum confidence score was picked (no type dominated).

Regions that “voted” as the ITI type were ignored in the next step.

#### 3.4.2. Noisy-OR Bayesian function to deduce the overall report

One of the remarkable features of COVID-19 is that there is more than one independent focus of infection in a single CT case [29,30]. It is reasonable for the overall probability to be much larger than 50% if a patient has two COVID-19 regions, both having a 50% probability. Accordingly, the total infection confidence score (C) for one infection type was calculated using the probability formula of the Noisy-OR Bayesian function as follows:

$$C = 1 - \prod_i (1 - C_i) \quad (1)$$

where  $C_i$  represents the confidence of the  $i$ th region.

The confidence scores of two types,  $C_{\text{COVID-19}}$  and  $C_{\text{IAVP}}$ , were deduced accordingly; this CT sample was then categorized into the corresponding group according to the dominating C value.

Moreover, the following strategies were used to export the confidence possibility of an entire CT sample to provide a reasonable reference for clinical doctors:

- (1) If both  $C_{\text{COVID-19}}$  and  $C_{\text{IAVP}}$  were equal to 0, then this CT sample belonged to the healthy case.

(2) If one of the  $C$  values was equal to 0, then the other  $C$  value was exported directly as the confidence possibility of this CT sample.

(3) Otherwise, the softmax function was used to generate two confidence scores.

$$S_j = \frac{e^{C_j}}{\sum_k e^{C_k}} \quad (2)$$

where  $j, k \in (\text{COVID-19, IAVP})$ .  $S_j$  was exported as the confidence score for each type of infection. The softmax operation normalized the sum of  $S_j$  to 100% and did not alter the judgment result for infection types. However, manual investigation should be involved from the perspective of clinical doctors, as some COVID-19 suspected regions were captured by the models even though they might not be the majority type.

## 4. Results

### 4.1. Evaluation platform

An Intel i7-8700k central processing unit (CPU) with NVIDIA graphics processing unit (GPU) GeForce GTX 1080ti was used as the testing server. The processing time largely depended on the number of image layers in one CT set. On average, it took less than 30 s for a CT set with 70 layers to go from data preprocessing to the output of the report.

### 4.2. Training process

As one of the most classical loss functions used in classification models, cross entropy was used in this study. When the epoch number of training iterations increased to more than 1000, the loss value did not obviously decrease or increase, suggesting that the models converged well to a relative optimal state without distinct overfitting. The training curves of the loss value and the accuracy rate for two classification models are shown in Fig. 5. The model with the location-attention mechanism achieved better performance on the training dataset, in comparison with the original ResNet.

### 4.3. Performance on test dataset

#### 4.3.1. Performance measurement

A confusion matrix was used, which is a table that is often used to describe the performance of a classification model on test dataset for which the true values are known. It allows the visualization of the performance of an algorithm.

The accuracy ( $A$ ) of a method determines how correct the predicted values are. Precision ( $P$ ) determines the reproducibility of the measurement, or how many of the predictions are correct. Recall ( $R$ ) indicates how many of the correct results are discovered. The  $f_1$ -score uses a combination of precision and recall to calculate a balanced average result. The following equations show how to

calculate these values, where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

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

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

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$f_1\text{-score} = \frac{2 \times P \times R}{P + R} \quad (6)$$

#### 4.3.2. Image preprocessing and segmentation

A total of 90 CT samples were randomly selected from each group (30 CT sets from COVID-19, 30 from IAVP, and 30 from healthy cases) for the test dataset. The choice of the test dataset followed the rule that any CT of this person had not been trained in the previous stage, in order to avoid having a similar CT that had been learned by the models. Moreover, the thresholds for both the image preprocessing and the segmentation were optimized to be more suitable for the current study. In the image preprocessing stage, the threshold of the Hounsfield unit (HU) value, which was used to binarize the resampled images, was raised to  $-200$  in order to maximize the filtering out of valid lung. The segmentation model VNet-IR-RPN was configured to reduce the proposal's threshold to maximize separate candidate regions, even through many normal regions could be included. We noticed that one CT case from the COVID-19 group that had no image patches was segmented as COVID-19 or IAVP, and was hence wrongly categorized as healthy case, as shown in Fig. 6. These focuses of infection were barely noticeable with the human eye, and seemed too tenuous to be captured by the segmentation model in this study.

#### 4.3.3. Classification for a single image patch

A total of 1710 image patches were acquired from 90 CT samples, including 357 COVID-19, 390 IAVP, and 963 ITI (ground truth). To determine the optimal approach, the design of each methodology was assessed using a confusion matrix. Two classification models were evaluated: with and without the location-attention mechanism, as shown in Tables 1 and 2.

The average  $f_1$ -score and the overall accuracy rate for the two models were 0.750/0.764 and 78.5%/79.4%, respectively. Furthermore, the location-attention mechanism was used to improve the respective accuracy rate of the COVID-19 and IAVP groups, and was shown to result in a remarkable improvement of 5.0% (260/273) and 1.4% (276/280). This evidence indicates that the second model with the location-attention mechanism achieved better performance. Therefore, that model was used for the rest of this study.

Moreover, as the ITI group was used to remove disturbing factors in this study, it was ignored and not counted by the

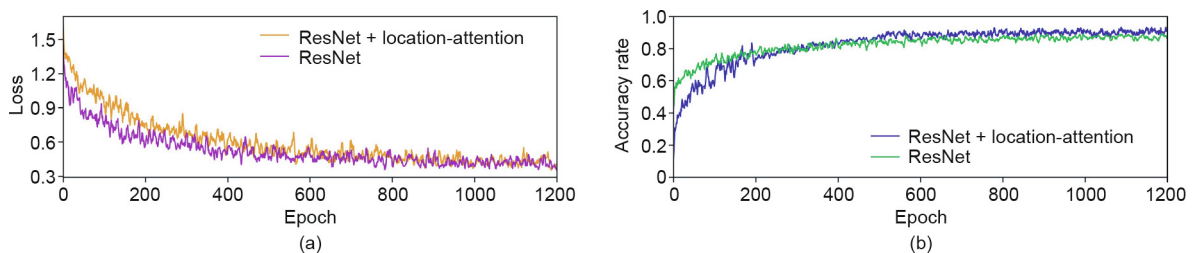
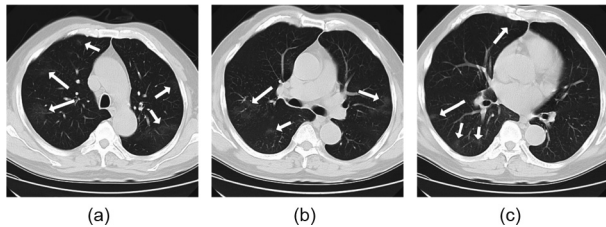


Fig. 5. Training curve of (a) loss and (b) accuracy rate for the two classification models.



**Fig. 6.** All CT images (a–c) were from a single CT case. The focuses of infections were pointed out by arrows.

**Table 1**  
Confusion matrix of two classification models for the COVID-19, IAVP, and ITI groups.

Group	COVID-19 (M <sub>1</sub> /M <sub>2</sub> )	IAVP (M <sub>1</sub> /M <sub>2</sub> )	ITI (M <sub>1</sub> /M <sub>2</sub> )
COVID-19 (M <sub>1</sub> /M <sub>2</sub> )	260/273	47/32	50/52
IAVP (M <sub>1</sub> /M <sub>2</sub> )	55/46	276/280	59/64
ITI (M <sub>1</sub> /M <sub>2</sub> )	75/77	81/82	807/804

M<sub>1</sub>: the ResNet model; M<sub>2</sub>: the ResNet model with the location-attention mechanism.

**Table 2**  
Recall, precision,  $f_1$ -score, and accuracy rate<sup>a</sup> of two classification models for the COVID-19, IAVP, and ITI groups.

Group	R	P	$f_1$ -score
COVID-19 (M <sub>1</sub> /M <sub>2</sub> )	0.728/0.765	0.667/0.689	0.696/0.725
IAVP (M <sub>1</sub> /M <sub>2</sub> )	0.708/0.718	0.683/0.711	0.695/0.714
ITI (M <sub>1</sub> /M <sub>2</sub> )	0.838/0.835	0.881/0.874	0.859/0.854

<sup>a</sup> The overall accuracy rates of M<sub>1</sub> and M<sub>2</sub> for the three groups are 78.5% and 79.4%; the accuracy rates of M<sub>1</sub> and M<sub>2</sub> for COVID-19 and IAVP groups are 71.8% and 74.0%.

Noisy-OR Bayesian function in the final step. To retain consistency in the next steps, we further compared the average  $f_1$ -score and accuracy rate for the first two groups, which were 0.720 and 74.0%, respectively.

#### 4.3.4. “Voting” for a region

Each image patch “voted” to represent the entire candidate region. A total of 570 candidate cubes were distinguished, including 119 COVID-19, 130 IAVP, and 321 ITI regions (ground truth). The confusion matrix of the voting result and the corresponding recall, precision, and  $f_1$ -score are shown in Tables 3 and 4, respectively.

The average  $f_1$ -score and overall accuracy rate for the three categories were 0.856 and 89.3%, respectively, which showed a respective improvement of 12.0% and 12.5% when compared with the previous step. As for the first two groups, the average  $f_1$ -score and accuracy rate were 0.806 and 78.3%, respectively, which showed a respective increase of 11.9% and 5.8%.

#### 4.3.5. Result of the classification of CT samples as a whole

Noisy-OR Bayesian function was used to identify the dominating infection types. Three kinds of results were exported in the final report: COVID-19, IAVP, and healthy cases. The experimental results are summarized in Tables 5 and 6.

Only the average  $f_1$ -score and the accuracy rate of the COVID-19 and IAVP groups were counted to compare with the previous results. These were 0.843 and 85.0%, respectively, which showed a promotion of 4.6% and 8.6% in this step.

A consistent improvement of the average  $f_1$ -score and accuracy rate was observed. The accuracy rate of the classification of COVID-19 and IAVP was promoted from 74.0% (single image patch) to 78.3% (image cube), and then to 85.0% (overall CT case based on

**Table 3**  
Confusion matrix for the COVID-19, IAVP, and ITI groups after voting.

Group	COVID-19	IAVP	ITI
COVID-19	97	15	7
IAVP	18	98	14
ITI	5	2	314

**Table 4**  
Recall, precision,  $f_1$ -score, and accuracy rate<sup>a</sup> for the COVID-19, IAVP, and ITI groups after voting.

Group	R	P	$f_1$ -score
COVID-19	0.815	0.808	0.811
IAVP	0.754	0.852	0.800
ITI	0.978	0.937	0.957

<sup>a</sup> The overall accuracy rate for the three groups was 89.3%; and the average accuracy rate for COVID-19 and IAVP groups was 78.3%.

**Table 5**  
Confusion matrix of the export of the Noisy-OR Bayesian function for COVID-19, IAVP, and healthy cases.

Group	COVID-19	IAVP	Healthy
COVID-19	26	3	1
IAVP	4	25	1
Healthy	2	1	27

**Table 6**  
Recall, precision,  $f_1$ -score, and accuracy rate<sup>a</sup> of the export of the Noisy-OR Bayesian function for COVID-19, IAVP, and healthy cases.

Group	R	P	$f_1$ -score
COVID-19	0.867	0.813	0.839
IAVP	0.833	0.862	0.847
Healthy	0.900	0.931	0.915

<sup>a</sup> The overall accuracy rate for the three groups is 86.7% and the average accuracy rate for COVID-19 and IAVP groups was 85.0%.

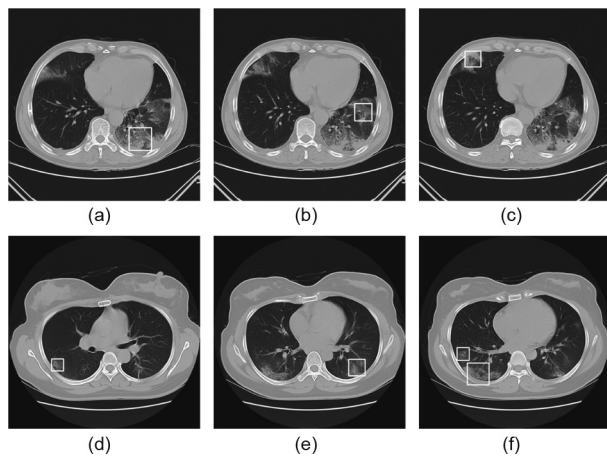
the dominating infection types). Measured by all three benchmark groups, the overall classification accuracy rate was 86.7%.

Moreover, a series of images with highlighted focuses of infection would also be exported (examples shown in Fig. 7).

## 5. Discussion

COVID-19 has caused serious public health and safety problems, and hence has become a global concern [33–35]. In the early stage of COVID-19, some patients may already have positive pulmonary imaging findings but no sputum or negative nucleic acid testing results from sputum or nasopharyngeal swabs. These patients are not diagnosed as suspected or confirmed cases. Thus, they are not isolated or treated in a timely manner, making them potential sources of infection. Meanwhile, CT examination is routinely performed on every patient with fever and respiratory symptoms in the early stage, and is repeated for dynamic observation, since it is cheap and easy to operate. Using CT images to screen patients can improve the early detection of COVID-19, and ease the pressure on laboratory nucleic acid testing.

The CT imaging of COVID-19 presents several distinct manifestations, according to previous studies [29,30,36]. These manifestations include focal ground-glass shadows mainly distributed along the pleura, multiple consolidation shadows accompanied by the “halo sign” of the surrounding ground-glass shadow, multiple consolidations of different sizes, and grid-shaped high-density shadows. An experienced radiologist can make judgments on the



**Fig. 7.** Examples of two CT cases reports with bounding boxes on the original images to highlight the focus of infections. Images (a)–(c) are from one case of IAVP. Images (d)–(f) are from one case of COVID-19. The segmented region of pulmonary was an image cube and only the center image was marked with a bounding box to facilitate the interpreting of the lesions.

possibility of COVID-19 based on his or her clinical experience; however, such judgments are easily influenced by subjective factors and individual proficiency. In comparison, deep-learning-system-based screen models reveal more specific and reliable results by digitizing and standardizing the image information. Hence, they can assist physicians in making a clinical decision more accurately.

There have been many precedents for artificial intelligence (AI)-assisted models that are now widely used in clinical practice, such as the pulmonary nodules diagnostic system. In June 2019, Ardila et al. [20] proposed a deep learning algorithm that used a patient's current and prior CT volumes to predict the risk of lung cancer. This model achieved an accuracy rate of nearly 94.4% on 6716 cases, and performed similarly on an independent clinical validation set of 1139 cases. In addition to having high accuracy rate, an AI-assisted model can do the work faster and more efficiently than a human.

In this study, deep learning technology was used to design a classification network for distinguishing COVID-19 from IAVP. In terms of the network structure, the classical ResNet was used for feature extraction. A comparison was made between models with and without an added location-attention mechanism. The experiment showed that the aforementioned mechanism could better distinguish COVID-19 cases from others. Furthermore, multiple enhancement methods were involved in our study, such as image patch vote and Noisy-OR Bayesian function, in order to determine the dominating infection types. All these efforts produced a consistent improvement in the average  $f_1$ -score and accuracy rate.

This study has some limitations. First, the manifestations of COVID-19 may have some overlap with the manifestations of other pneumonias such as IAVP, organizing pneumonia, and eosinophilic pneumonia. We only compared the CT manifestation of COVID-19 with that of IAVP. A clinical diagnosis of COVID-19 still needs to combine the patient's contact history, travel history, first symptoms, and laboratory examination. Second, the number of model samples was limited in this study. The number of training and test samples should be expanded to improve the accuracy in the future. More multi-center clinical studies should be conducted to cope with the complex clinical situation. Moreover, efforts should be made to improve the segmentation and classification model. A better exclusive model could be designed for training, the segmentation and classification accuracy of the model could be improved, and the generalization performance of this algorithm could be verified with a larger dataset.

## 6. Conclusions

In this multi-center case study, we presented a novel method that can screen CT images of COVID-19 automatically by means of deep learning technologies. Models with a location-attention mechanism can classify COVID-19, IAVP, and healthy cases with an overall accuracy rate of 86.7%, and would be a promising supplementary diagnostic method for frontline clinical doctors.

## Acknowledgements

This study was supported by the National Science and Technology Major Project of China (20182X10101-001).

## Authors' contribution

Wei Wu, Tingbo Liang, Lanjuan Li, and Xiaowei Xu initiated the project and provided clinical expertise and guidance on the study design. Xukun Li and Peng Du designed the network architecture and data/modeling infrastructure, training, testing setup, and statistical analysis. Xiaowei Xu, Wei Wu, Xukun Li, and Peng Du wrote the manuscript. Xiangao Jiang, Chunlian Ma, Shuangzhi Lv, Liang Yu, Yanfei Chen, Junwei Su, Guanqing Lang, Yongtao Li, Hong Zhao, Kaijin Xu, and Lingxiang Ruan collected the datasets and interpreted the data. Xiaowei Xu, Xiangao Jiang, Chunlian Ma, and Peng Du contributed equally to this article.

## Compliance with ethics guidelines

Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Qin Ni, Yanfei Chen, Junwei Su, Guanqing Lang, Yongtao Li, Hong Zhao, Jun Liu, Kaijin Xu, Lingxiang Ruan, Jifang Sheng, Yunqing Qiu, Wei Wu, Tingbo Liang, and Lanjuan Li declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727–33.
- [2] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020;382(13):1199–207.
- [3] Cohen J, Normile D. New SARS-like virus in China triggers alarm. *Science* 2020;367(6475):234–5.
- [4] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25(3):23–30.
- [5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497–506.
- [6] Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020;395(10223):514–23.
- [7] National Health Commission of the People's Republic of China, National Administration of Traditional Chinese Medicine. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7) [Internet]. Beijing: National Health Commission of the People's Republic of China; [cited 2020 Mar 8]. Available from: <http://www.nhc.gov.cn/jzygj/s7653p/202003/46c9294a7df4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf>. Chinese.
- [8] Loeffelholz MJ, Tang YW. Laboratory diagnosis of emerging human coronavirus infections—the state of the art. *Emerg Microbes Infect* 2020;9(1):747–56.
- [9] Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020;323(18):1843–4.
- [10] Zhang W, Du R, Li B, Zheng X, Yang X, Hu B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect* 2020;9(1):386–9.
- [11] Long Q, Deng H, Chen J, Hu J, Liu B, Liao P, et al. Antibody responses to SARS-CoV-2 in COVID-19 patients: the perspective application of serological tests in clinical practice. 2020. medRxiv:2020.03.18.20038018.

- [12] National Health Commission of the People's Republic of China, National Administration of Traditional Chinese Medicine. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 5) [Internet]. Beijing: National Health Commission of the People's Republic of China; [cited 2020 Feb 5]. Available from: <http://www.nhc.gov.cn/yzygj/s7653p/202002/d4b895337e19445f8d728fcaf1e3e13a/files/ab6bec7f93e64e7f998d802991203cd6.pdf>. Chinese.
- [13] Liu X, Guo S, Yang B, Ma S, Zhang H, Li J, et al. Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks. *J Digit Imaging* 2018;31(5):748–60.
- [14] Gharbi M, Chen J, Barron JT, Hasinoff SW, Durand F. Deep bilateral learning for real-time image enhancement. *ACM Trans Graph* 2017;36(4):118.
- [15] Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging* 2019;32(4):582–96.
- [16] Akagi M, Nakamura Y, Higaki T, Narita K, Honda Y, Zhou J, et al. Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. *Eur Radiol* 2019;29(11):6163–71.
- [17] Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, Washko GR, Rahaghi FN, Ledesma-Carbayo MJ, et al. Pulmonary artery-vein classification in CT images using deep learning. *IEEE Trans Med Imaging* 2018;37(11):2428–40.
- [18] Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys* 2019;46(2):576–89.
- [19] Huang P, Park S, Yan R, Lee J, Chu LC, Lin CT, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. *Radiology* 2018;286(1):286–95.
- [20] Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25(6):954–61.
- [21] Esteva A, Kuprel B, Novoa R A, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115–8.
- [22] Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284(2):574–82.
- [23] Wu W, Li X, Du P, Lang G, Xu M, Xu K, et al. A deep learning system that generates quantitative CT reports for diagnosing pulmonary tuberculosis. 2019. arXiv:1910.02285v1.
- [24] Li L, Huang H, Jin X. AE-CNN classification of pulmonary tuberculosis based on CT images. In: Proceedings of the 9th International Conference on Information Technology in Medicine and Education (ITME); 2018 Oct 19–21; Zhejiang, China. New York: IEEE; 2018.
- [25] Oniško A, Druzdzel M, Wasyluk H. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *Int J Approx Reason* 2001;27(2):165–82.
- [26] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. 2016. arXiv:1606.04797v1.
- [27] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. 2016. arXiv:1602.07261.
- [28] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D u-net: learning dense volumetric segmentation from sparse annotation. 2016. arXiv:1606.06650.
- [29] Kanne JP. Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist. *Radiology* 2020;295(1):16–7.
- [30] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* 2020;295(1):202–7.
- [31] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. arXiv:1512.03385.
- [32] Breiman L. Bagging predictors. *Mach Learn* 1996;24(2):123–40.
- [33] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020;395(10223):470–3.
- [34] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020;382(10):929–36.
- [35] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395(10223):507–13.
- [36] Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 2020;20(4):425–34.