



# Artificial intelligence for stepwise diagnosis and monitoring of COVID-19

Hengrui Liang<sup>1,2</sup> · Yuchen Guo<sup>3,4</sup> · Xiangru Chen<sup>5</sup> · Keng-Leong Ang<sup>2,6</sup> · Yuwei He<sup>4,7</sup> · Na Jiang<sup>8</sup> · Qiang Du<sup>5</sup> · Qingsi Zeng<sup>1,9</sup> · Ligong Lu<sup>10</sup> · Zebin Gao<sup>5</sup> · Linduo Li<sup>11</sup> · Quanzheng Li<sup>12</sup> · Fangxing Nie<sup>5</sup> · Guiguang Ding<sup>3,4,7</sup> · Gao Huang<sup>5,7</sup> · Ailan Chen<sup>1,13</sup> · Yimin Li<sup>1,14</sup> · Weijie Guan<sup>1</sup> · Ling Sang<sup>1,14</sup> · Yuanda Xu<sup>1,14</sup> · Huai Chen<sup>1,9</sup> · Zisheng Chen<sup>1</sup> · Shiyue Li<sup>1</sup> · Nuofu Zhang<sup>1</sup> · Ying Chen<sup>1</sup> · Danxia Huang<sup>1</sup> · Run Li<sup>1</sup> · Jianfu Li<sup>1,2</sup> · Bo Cheng<sup>1,2</sup> · Yi Zhao<sup>1,2</sup> · Caichen Li<sup>1,2</sup> · Shan Xiong<sup>1,2</sup> · Runchen Wang<sup>1,2</sup> · Jun Liu<sup>1,2</sup> · Wei Wang<sup>1,2</sup> · Jun Huang<sup>1,2</sup> · Fei Cui<sup>1,2</sup> · Tao Xu<sup>15</sup> · Fleming Y. M. Lure<sup>16</sup> · Meixiao Zhan<sup>10</sup> · Yuanyi Huang<sup>17</sup> · Qiang Yang<sup>18</sup> · Qionghai Dai<sup>3,4</sup> · Wenhua Liang<sup>1,2</sup> · Jianxing He<sup>1,2,19</sup> · Nanshan Zhong<sup>1</sup>

Received: 7 May 2021 / Revised: 6 July 2021 / Accepted: 7 August 2021 / Published online: 6 January 2022  
© European Society of Radiology 2021

## Abstract

**Background** Main challenges for COVID-19 include the lack of a rapid diagnostic test, a suitable tool to monitor and predict a patient's clinical course and an efficient way for data sharing among multicenters. We thus developed a novel artificial intelligence system based on deep learning (DL) and federated learning (FL) for the diagnosis, monitoring, and prediction of a patient's clinical course.

**Methods** CT imaging derived from 6 different multicenter cohorts were used for stepwise diagnostic algorithm to diagnose COVID-19, with or without clinical data. Patients with more than 3 consecutive CT images were trained for the monitoring algorithm. FL has been applied for decentralized refinement of independently built DL models.

**Results** A total of 1,552,988 CT slices from 4804 patients were used. The model can diagnose COVID-19 based on CT alone with the AUC being 0.98 (95% CI 0.97–0.99), and outperforms the radiologist's assessment. We have also successfully tested the incorporation of the DL diagnostic model with the FL framework. Its auto-segmentation analyses co-related well with those by radiologists and achieved a high Dice's coefficient of 0.77. It can produce a predictive curve of a patient's clinical course if serial CT assessments are available.

**Interpretation** The system has high consistency in diagnosing COVID-19 based on CT, with or without clinical data. Alternatively, it can be implemented on a FL platform, which would potentially encourage the data sharing in the future. It also can produce an objective predictive curve of a patient's clinical course for visualization.

## Key Points

- *CoviDet could diagnose COVID-19 based on chest CT with high consistency; this outperformed the radiologist's assessment. Its auto-segmentation analyses co-related well with those by radiologists and could potentially monitor and predict a patient's clinical course if serial CT assessments are available. It can be integrated into the federated learning framework.*
- *CoviDet can be used as an adjunct to aid clinicians with the CT diagnosis of COVID-19 and can potentially be used for disease monitoring; federated learning can potentially open opportunities for global collaboration.*

**Keywords** Coronavirus disease 2019 · AI (artificial intelligence) · Computer-assisted diagnosis

---

Hengrui Liang and Yuchen Guo are equal contributors.

Extended author information available on the last page of the article

## Abbreviations

AUC	Area under curve
CC-CCII	China Consortium of Chest CT Image Investigation
CI	Confidence interval
CLAIM	Artificial Intelligence in Medical Imaging
CNN	Convolutional neural networks
COVID-19	Coronavirus disease 2019
CT	Computed tomography
DC	Dice coefficient
DL	Deep learning
EMR	Electronic medical records
FL	Federated learning
GN	Group normalization
GPR	Gaussian process regression
GPU	Graphics processing unit
GRU	Gated recurrent unit
NPV	Negative prediction value
PCR	Polymerase chain reaction
PPV	Positive prediction value
RNN	Recurrent neural net
ROC	Receiver operating characteristic curve
SVM	Support vector machine
WS	Weight standardization

## Introduction

The current gold standard diagnostic test for coronavirus disease 2019 (COVID-19) using polymerase chain reaction (PCR) has several limitations [1]. Its accuracy also depends on the patient's clinical condition, sampling technique, and test kits used [2]. It also does not provide any information about the patient's disease severity. Chest computed tomography (CT) has emerged as a useful adjunct, but there were concerns about its accuracy as certain CT features of COVID-19 mimic other viral chest infections [3].

To overcome these challenges, we have developed the CoviDet, a deep learning (DL) system for CT analysis of COVID-19, with and without electronic medical record (EMR) data. We have also included an auto-segmentation algorithm to test its ability to perform automated monitoring and predicting of COVID-19 changes in serial CT.

To enable its future applications globally, we have tested its application within a federated learning (FL) framework. FL has been recently proposed as a promising strategy for decentralized refinement of independently built DL models [4]. As only encrypted parameters are exchanged, it avoids data protection and privacy issues associated with the sharing of sensitive personal data to the main study site. This will open up opportunities to speed up global collaboration against COVID-19 and refine the model generalizability.

## Methods

### Study design overview and image datasets

This study was approved by the ethics committee of the First Affiliated Hospital of Guangzhou Medical University (2020–94). Informed consent was waived because of its retrospective nature. We present the following article in accordance with the strengthening the reporting of Artificial Intelligence in Medical Imaging (CLAIM) guideline checklist. The CoviDet system uses five algorithms for two main functions: (1) stepwise diagnosis of COVID-19 from other chest diseases based on CT with or without clinical data and (2) auto-segmentation and monitoring/predicting of COVID-19 changes on CT. Figure 1 summarizes the data set and design for each algorithm, and the application process is summarized in Supplementary Fig. 1.

A total of 1,552,988 CT slices from 4804 patients were used for this study. They were grouped into 6 cohorts according to the data source (Supplementary Fig. 2 to Supplementary Fig. 6 and methods). Cohort 1 and 2 comprised confirmed COVID-19 cases. Cohort 3 refers to the control group (non-viral infection causes), and cohort 4 were patients with non-COVID-19 viral pneumonia. The segmentation and monitoring of clinical progress utilized CT data from COVID-19 patients with more than three consecutive repeat CT scans from a subset of cohort 2 for training and cohort 5 for validation. Cohort 6 was an independent validation cohort from the China National Center for Bioinformatics database, which is constructed from cohorts from the China Consortium of Chest CT Image Investigation (CC-CCII). All CT images in cohort 6 are classified into laboratory-confirmed COVID-19 ( $n=944$ ), common pneumonia ( $n=712$ ), and normal controls ( $n=621$ ). The stepwise diagnosis model utilizes DL #1 to distinguish between viral pneumonia (including COVID-19) from other lung conditions (pulmonary nodule, pulmonary tuberculosis, and ordinary lung scans). Transferred learning from DL #1 is then used in DL #2 to develop another model to distinguish COVID-19 from other causes of viral pneumonia (Supplementary Fig. 7 and methods) [5]. DL #3 then combines DL #2 with the support vector machine (SVM) trained using EMR data to evaluate the further diagnostic performance [6]. DL #4 is used for lesion segmentation and DL #5 is used for automated monitoring and predicting of CT changes relating to COVID-19.

Four experienced radiologists were included in the annotation process to establish the ground truth for lung tissue classification in positive cases (COVID-19 or other viral pneumonia). Two radiologists were involved in the process of creating fine boundaries with the use of Materialise's interactive medical image control system (Materialise Mimics V20.0) to highlight pathology in the same CT section to avoid the inclusion of noisy data.

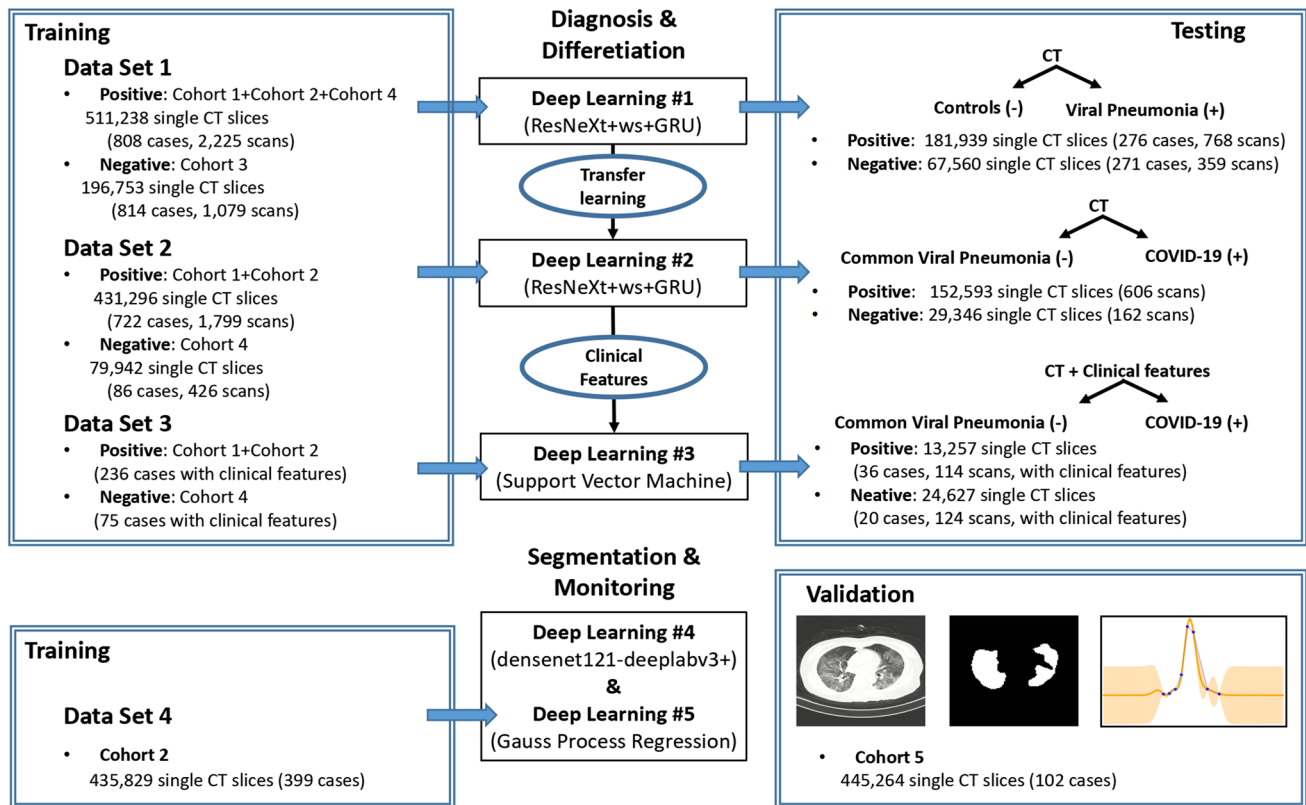


Fig. 1 Flow chart for the development and testing of the CoviDet system

### Construction and validation of the deep learning model for the stepwise COVID-19 diagnosis system

The data cleaning and pre-processing model is a small residual network (ResNet-18) achieving fast training and reasoning (Supplementary Fig. 8 and methods).

The learning models for the diagnosis system were based on the ResNeXt-50 + weight standardization (ws) + gated recurrent unit (GRU) (Supplementary Fig. 9a, Supplementary methods, and Supplementary parameters). A set of CT slides was selected for classification, and the GRU, an implementation of recurrent neural net (RNN), encodes the features selected by ResNeXt-50 to obtain the final output of the whole scan. In this approach, there is no need to specifically annotate the special location of a lesion; instead, a positive or negative label for the full scan was sufficient for model training. To solve the problem of memory usage, the ResNeXt-50 + ws model was used for feature extraction, which has been shown to provide acceptable outcomes with small batch sizes. A 2080Ti graphics processing unit (GPU) was used for training. We use ws + group normalization (gn) in our model instead of the more widely used batch normalization due to the small batch size in order to achieve better performance. As a result, we used the pre-trained

ResNeXt-50 + ws + gn structure for image feature extraction from the whole model (Supplementary Fig. 9a). We then applied a classification method based on a combination of a deep convolutional neural network (CNN) that extracts discriminative visual features from each slide automatically and an RNN that integrates features from each slide given by CNN into the features of the whole scan to obtain the classification result [7].

To include clinical features such as laboratory tests and clinical symptoms, SVM is used (Supplementary Fig. 9b, Supplementary methods, and Supplementary parameters), which is a classical classifier that functions well in many different tasks [8]. With different kernels, SVM can deal with a high-dimensional vector for classifying. In this paper, three kinds of kernels are used for classifying the clinical data for DL #3: a linear kernel, a polynomial kernel, and a radial basis function (Gaussian) kernel (RBF kernel).

### Construction and validation of the deep learning model for clinical progress prediction of COVID-19

**DL #4** The first step was to train the segmentation model to provide the segmented lesion area and calculate the lesion volume. DeeplabV3 + with Densenet121 as the backbone is used for the segmentation model deployed on the website

(Supplementary Fig. 9c, Supplementary methods, and Supplementary parameters) [9]. For better visualization, this model was reconstructed and displayed as a 3-dimensional pulmonary model. To evaluate the performance of the segmentation model, the Dice coefficient (DC) was used.

**DL #5** Patients who had at least 3 different consecutive scans after admission were used for the monitoring and predicting of the disease course. Gaussian process regression (GPR) was used to mimic the lesion curve and to predict the changes of the patients' lesion area ratio (lesion area/pulmonary area) (Supplementary Fig. 9d and methods) [10].

Using the average lesion area of the 3 known time points measured by DL #4, DL #5 was used to provide an individualized visual prediction of a patient's progress. For those with 4th and 5th time points, the agreement between the predicted values using this curve was compared with that of the actual values of these time points.

### Federated learning

To test the feasibility of using CoviDet within the FL framework, we have implemented it in a simulated FL environment. In the FL model, training was performed locally in different sites rather than in a single central site, and only encrypted parameters were shared, as depicted in Supplementary Fig. 9e and methods.

### Statistical analysis

The classification results in the diagnosis and differentiation model were evaluated based on the area under curve (AUC) of the receiver operating characteristic curve (ROC), sensitivity, specificity, accuracy (true predictions/total number of cases), positive prediction value (PPV), negative prediction value (NPV), and F1 score. The model was run 5 times to obtain a 95% confidence interval (CI) of the AUC and average sensitivity, specificity, PPV, NPV, and F1 score of the test set. Sensitivity was the ratio of positives that were correctly discriminated, and specificity was the ratio of negatives that were correctly discriminated.  $PPV = \text{true positive} / (\text{true positive} + \text{false positive})$ ,  $NPV = \text{true negative} / (\text{true negative} + \text{false negative})$ . The F1 score was a measure of the accuracy of a binary model. The Yoden index was used to choose the best cut-off value. R (version 4.0.0) was used for statistical analysis.

### Role of the funding source

The funder had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in

the study and had final responsibility for the decision to submit for publication.

## Results

### Diagnostic performance of deep learning #1: to distinguish viral pneumonia from control cases based on chest CT

DL #1 can diagnose viral pneumonia (COVID-19 and other viral pneumonia) from controls with a sensitivity of 0.98 (95% CI 0.98–0.99), a specificity of 1.00 (95% CI 0.99–1.00), an AUC of the receiver of 0.99 (95% CI 0.99–0.99) (Fig. 2a) and an f1 score of 0.98 (95% CI 0.98–0.98). The average diagnostic time was  $5.74 \pm 5.01$  s. In addition, the PPV of the algorithm was 1.00 (95% CI 1.00–1.00), the NPV was 0.97 (95% CI 0.96–0.97), and the accuracy was 0.96 (95% CI 0.96–0.97).

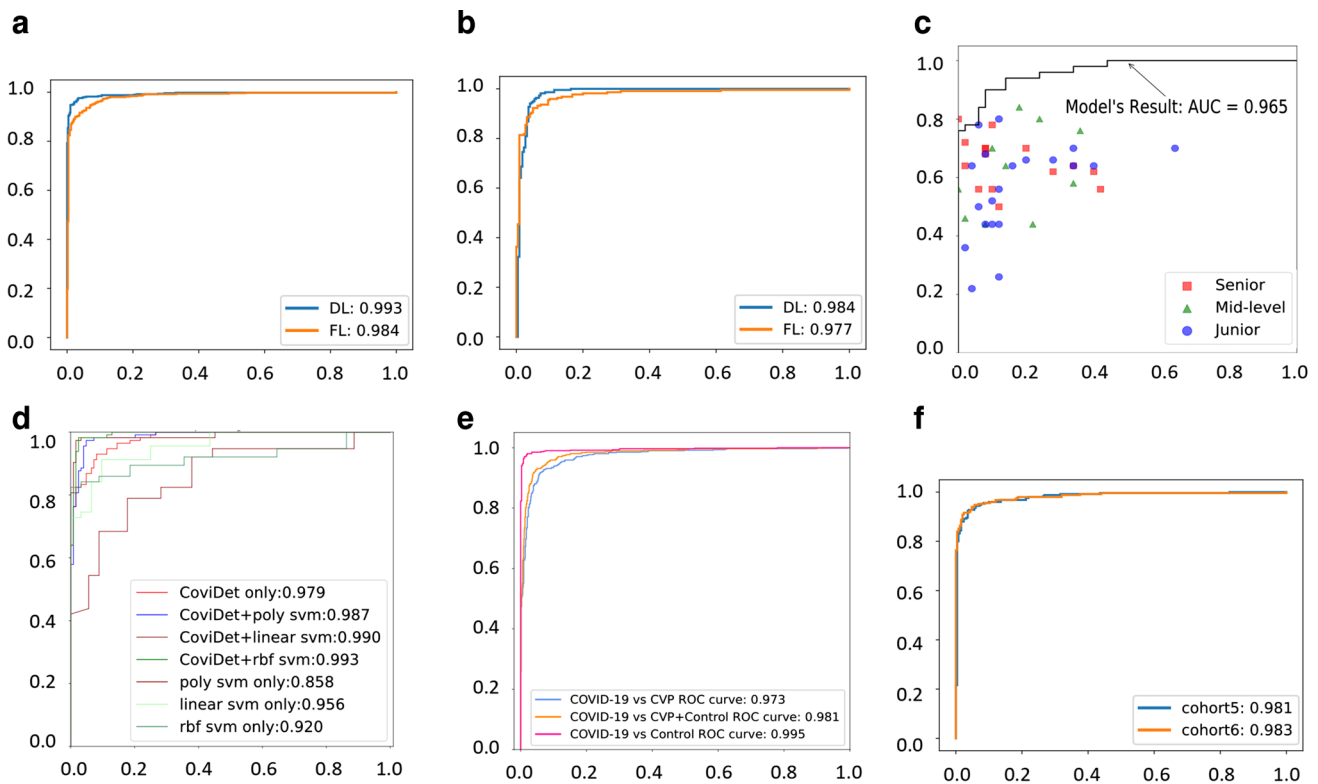
### Diagnostic performance of deep learning #2: to distinguish COVID-19 from other viral pneumonia based on chest CT

To train the DL #2 to distinguish COVID-19 from other viral pneumonia, the model parameters in DL #1 were used as initial parameters. A transfer learning strategy from DL #1 was used to improve the accuracy of DL #2 (Supplementary Fig. 7). This algorithm identifies COVID-19 from other viral pneumonia with a sensitivity of 0.98 (95% CI 0.98–0.98), a specificity of 0.95 (95% CI 0.95–0.96), and an AUC of 0.98 (95% CI 0.98–0.99) (Fig. 2b). The f1 score of this model was 0.97 (95% CI 0.97–0.98). The average diagnostic time was  $6.75 \pm 6.04$  s. The PPV was 0.91 (95% CI 0.90–0.92), and the NPV was 0.99 (95% CI 0.98–0.99); the accuracy was 0.93 (95% CI 95% CI 0.93–0.93).

### Comparison of the diagnostic performance of the deep learning #2 algorithm to those of radiologists for distinguishing COVID-19 from other viral pneumonia based on CT

Forty-five radiologists from 15 hospitals were asked to analyze 100 CT scans: 50 randomized CT scans from COVID-19 patients and 50 randomized CT scans from other viral pneumonia. Fifteen of them were senior radiologists with over 15 years' experience, 10 were middle-grade radiologists with 10 to 15 years' experience, and 20 were junior radiologists with less than 10 years' experience.

The overall sensitivity and specificity for the radiologists were 0.68 and 0.98 respectively, with an average AUC of 0.86 (95% CI 0.78–0.94). The specific diagnostic outcomes varied with the experience of the radiologist and are



**Fig. 2** Performance of the CoviDet system for diagnosing different conditions. **a** The ROC curve of DL #1 on the test set between viral pneumonia (COVID-19 and other types of viral pneumonia) and controls (pulmonary nodule, pulmonary tuberculosis, and normal lung) based on chest CT. **b** The ROC curve of DL #2 on the test set between COVID-19 and other types of viral pneumonia based on chest CT. **c** The comparison of the diagnostic performance between DL #2 and radiologists with different experience. **d** The ROC curve of DL #3 on the test set between COVID-19 and other types of viral pneumonia based on CT and clinical features from EMR (“CoviDet only” is the deep learning model CoviDet with only CT input; SVM

uses clinical features as input; “poly,” “linear,” and “rbf” are polynomial, linear, and RBF kernels for SVM, respectively; “X svm only” (X for poly, linear, or RBF) is the model with only clinical features; “CoviDet+X svm” is the model combining outputs of both CoviDet and SVM). **e** The ROC curve of the stepwise diagnosis system (stepwise DL #1 and DL #2) for COVID-19 from board population (all samples in the test group). **f** The ROC curve of the stepwise diagnosis system (stepwise DL #1 and DL #2) for COVID-19 of independent validation in cohort 5 and 6. ROC, receiver operating characteristic curve; AUC, area under curve; DL, deep learning; FL, federated learning; EMR, electronic medical record

summarized in Supplementary Table 1, with the AUC ranging from 0.81 to 0.89. To compare with the radiologist’s performance, the same 100 scans were analyzed by DL #2. This yielded a sensitivity of 0.90, a specificity of 0.92, and an average AUC of 0.97 (95% CI 0.94–1.00) (Fig. 2c), which was superior to the expert’s assessments. The average diagnosis time of radiologists (from the first sight to make the diagnosis) is  $59.91 \pm 24.07$  s, which is much longer than that of DL #2 ( $6.95 \pm 5.03$  s;  $p < 0.001$ ).

### Diagnostic performance of deep learning #2 when used in combination with clinical data to distinguish COVID-19 from other viral pneumonia (deep learning #3)

To investigate the diagnostic performance when clinical data was included, clinical features that were analyzed by SVM

were combined with DL #2 to generate DL #3. The clinical features are summarized in Supplementary Table 2 and 3. The sensitivity was 0.97 (95% CI 0.97–0.97), specificity was 0.98 (95% CI 0.98–0.98), and AUC was 0.99 (95% CI 0.99–1.00;  $p < 0.001$ ) (Fig. 2d). The f1 score of this model was 0.97 (95% CI 0.97–0.97). The average diagnostic time was  $6.75 \pm 6.04$  s. The PPV was 0.97 (95% CI 0.97–0.97), the NPV was 0.98 (95% CI 0.98–0.98), and the accuracy was 0.97 (95% CI 0.97–0.98). Meanwhile, in the radiologists’ test, the classification performance was not improved when adding with the clinical data. The sensitivity and specificity for the radiologists were 0.67 and 0.96, respectively, and the average AUC was 0.84 (95% CI 0.80–0.91), whether or not the clinical data was added in the same test data set.

## Diagnostic performance of CoviDet to distinguish COVID-19 from non-COVID-19 based on CT alone (stepwise combination of deep learning #1 and deep learning #2)

Stepwise combination of DL #1 and DL #2 was used to evaluate the detection of COVID-19 cases from non-COVID-19 cases based on CT alone, which is its most important task for clinical application. Our results showed that the CoviDet stepwise diagnostic algorithm can differentiate COVID-19 cases from all non-COVID-19 cases (other viral pneumonia and control cases), with a sensitivity of 0.93 (95% CI 0.93–0.93), a specificity of 0.95 (95% CI 0.95–0.95), an accuracy of 0.95 (95% CI 0.95–0.96), and an AUC of 0.98 (95% CI 0.97–0.99) (Fig. 2e).

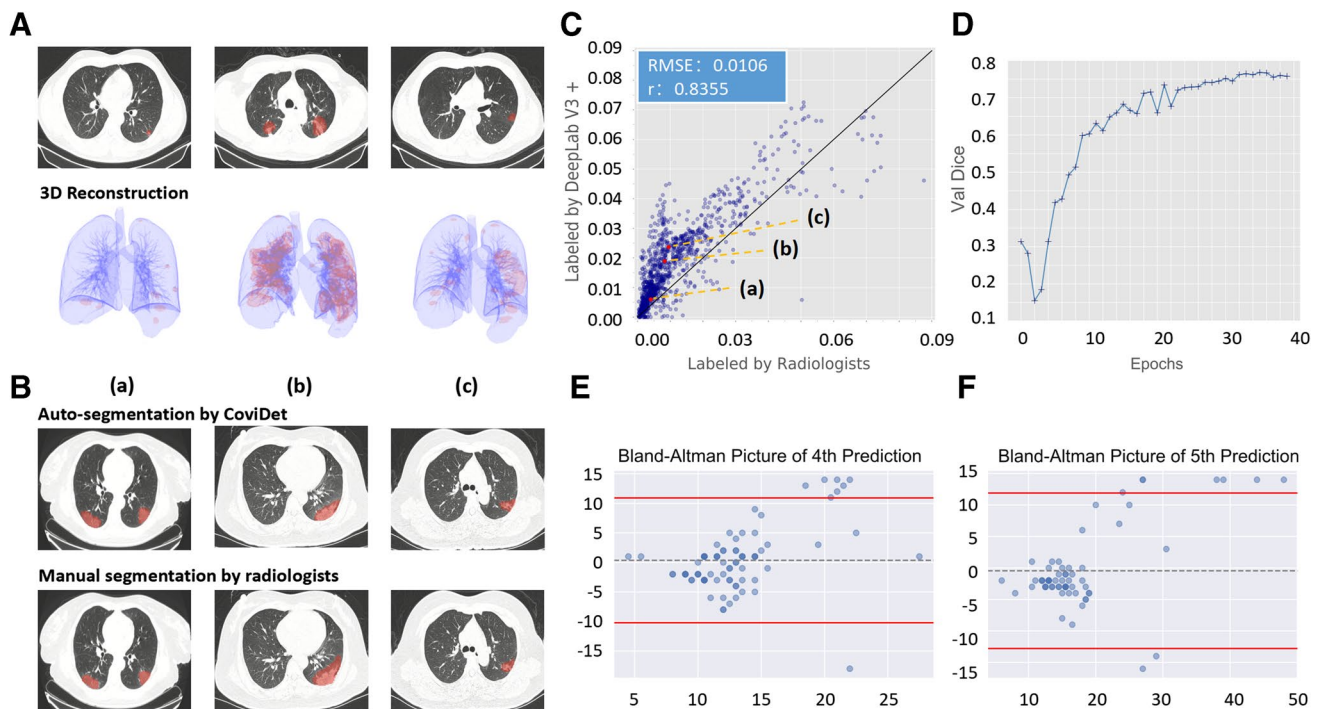
The external validation was conducted using two independent cohorts (cohort 5 and cohort 6) (Fig. 2f). The CoviDet stepwise diagnostic algorithm (the combination of DL #1 and #2) can differentiate COVID-19 cases from all non-COVID-19 cases with a sensitivity of 0.96 (95% CI 0.96–0.97), a specificity of 0.95 (95% CI 0.95–0.97), and an

AUC of 0.98 (95% CI 0.98–0.99) in cohort 5 and a sensitivity of 0.95 (95% CI 0.95–0.96), a specificity of 0.96 (95% CI 0.95–0.98), and an AUC of 0.98 (95% CI 0.98–0.99) in cohort 6.

## Diagnostic performance of CoviDet in the federated learning framework

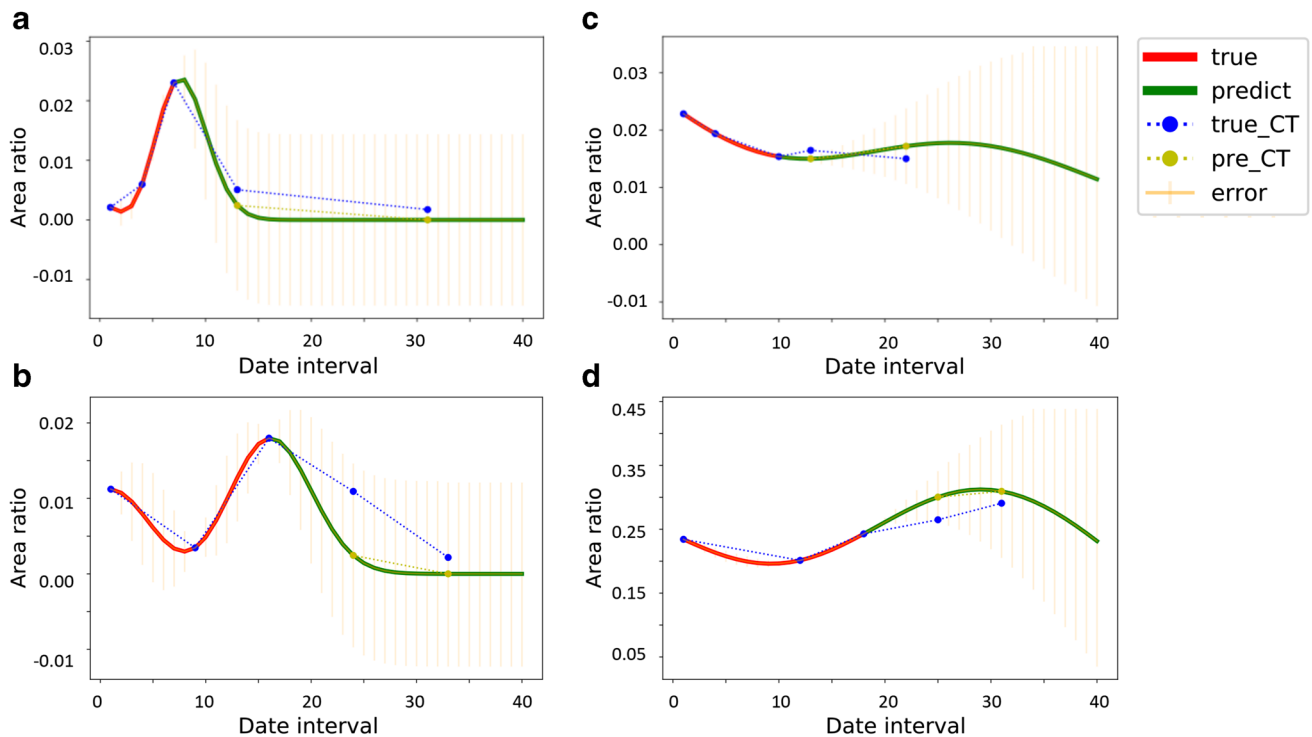
To test the feasibility of CoviDet operating within a FL framework, we have applied the DL diagnostic algorithm in a simulated FL environment. For DL #1 (Fig. 2a), the AUC of the FL model was 0.98 (95% CI 0.98–0.98), the sensitivity was 0.97 (95% CI 0.97–0.98), and the specificity was 0.94 (95% CI 0.94–0.95); for DL #2 (Fig. 2b), the AUC was 0.98 (95% CI 0.97–0.98), the sensitivity was 0.96 (95% CI 0.96–0.96), and the specificity was 0.92 (95% CI 0.92–0.93).

In FL, CoviDet can distinguish COVID-19 from non-COVID-19 cases with a sensitivity of 0.95 (95% CI 0.97–0.98), a specificity of 0.96 (95% CI 0.97–0.98), an accuracy of 0.97 (95% CI 0.97–0.98), and an AUC of 0.98 (95% CI 0.98–0.99).



**Fig. 3** The performance of the CoviDet system on segmentation. **A** The performance of DL #4 on a lesion segmentation task shown in examples at different stages; the upper pictures were two-dimensional segmentation display, and the lower pictures were three-dimensional segmentation display. **B** The performance of DL #4 on a lesion segmentation task is shown in the upper row of pictures, and the corresponding segmentation by radiologists is shown in the lower row of

pictures. **C** The consistency of CT segmentation between DL #4 and radiologists. **D** Dice on the validation data set changing with training epochs. **E** Bland–Altman plot of the predicted and actual time interval of the 3rd to 4th time point (spot indicates sample; dotted line indicates average difference; red line indicates upper and lower limits of agreement). **F** Bland–Altman plot of predicted and actual time intervals of the 3rd to 5th time point



**Fig. 4** The Gaussian process curve of dynamic lesion prediction based on repeated CT in a single patient. **a** Disease showing peak and recovery. **b** Disease showing recovery but relapsing. **c** Disease showing a plateau. **d** Disease showing deterioration

### Algorithm for auto-segmentation, disease severity assessment, and automated monitoring of COVID-19 CT changes (DL #4 and #5)

To assess the severity and monitor the CT changes associated with COVID-19 over time, an auto-segmentation model (DL #4) was constructed to outline the abnormal lesion area on CT and to calculate the lesion area/normal lung area ratio, which were displayed as a 2-dimensional and reconstructed as a 3-dimensional model (Fig. 3a). Compared with manual segmentation by radiologists (Fig. 3b, c), the auto-segmentation correlated closely with that performed by radiologists, with a low root mean square error of 0.01; with 40 epochs, the segmentation model can achieve a high Dice's coefficient of 0.78 (Fig. 3d).

Using the average lesion area/normal lung area ratio, obtained CT scans from 3 known time points measured by DL #4 and DL #5 were used to test if it was possible to provide automated monitoring and prediction of COVID-19-related CT changes. The graph plot by DL #5 provided a visual graphic trend of the CT changes. Majority in our test cohort have a radiological course similar to that reported by Pan et al [11] (Fig. 4a), where the radiological changes peaked and recovered. Some had shown signs of recovery but relapsed

(Fig. 4b); others had either plateaued (Fig. 4c) or were showing features of deterioration (Fig. 4d).

To evaluate the model's ability to predict CT changes beyond the 3rd time point, we have used those COVID-19 patients with 3 or more chest CT scans to predict the time interval taken between the 3rd and 4th or the 3rd and 5th time points, based on the average lesion area/normal lung area ratio measured at those respective time points available. This was compared with the actual known time interval. There was no statistical difference between the predicted and actual time intervals of the 3rd to the 4th time point (predicted  $5.00 \pm 5.08$  days vs. actual  $4.61 \pm 2.61$  days;  $p=0.53$ ) (Fig. 3e) and the 3rd to the 5th time point (predicted  $7.72 \pm 10.6$  days vs. actual  $9.82 \pm 5.84$  days;  $p=0.17$ ) (Fig. 3f).

### Systematic review of current work

To evaluate the novelty of our work, we conducted a systematic review based on peer-reviewed publications on the use of AI for chest CT analysis of COVID-19 (Supplementary Fig. 10 and methods). Finally, 16 studies were included (Table 1) [12–27]. All 16 published studies were based on either a 2D CNN or a 3D CNN only, with an AUC range

**Table 1** Systematic review of current work on AI COVID-19 diagnosis

Author	Region	Sample size	Additional function	Control group	AUC (95% CI)	Accuracy	Sensitivity	Specificity	Model	Original algorithm
Liang et al (this study)	China	4804	Dynamic monitoring, FL	Other types of viral pneumonia; controls	DL: 0.98 (0.97–0.99) FL: 0.98 (0.98–0.99)	DL: 95.00% FL: 97.00%	DL: 93.00% FL: 95.00%	DL: 95.00% FL: 96.00%	2D CNN + RNN	Yes
Wang et al	China	5372	Prognosis	CAP	0.88 (0.86–0.90)	80.12%	79.35%	81.16%	3D CNN	Yes
Li et al	China	3322	N/A	CAP	0.96 (0.94–0.99)	N/A	90.00%	96.00%	2D CNN + MLP	Yes
Zhang et al	China	4154	Prognosis	CAP/normal	0.97 (0.95–0.99)	90.70%	92.51%	85.92%	3D CNN	Yes
Mei et al	China	905	N/A	Non-COVID-19	0.92 (0.89–0.95)	N/A	84.30%	82.80%	2D CNN	Not
Ko et al	Korea/Italy	3993	N/A	CAP/normal	0.99	98.67%	97.39%	99.64%	2D CNN (4 models combined)	Yes
Ardakani et al	Iran	194	N/A	Non-COVID-19	0.99	99.51%	100.00%	99.02%	2D CNN (10 models)	Not
Wu et al	China	495	N/A	Non-COVID-19	0.82 (0.67–0.97)	76.00%	81.00%	61.50%	2D CNN (3 Angles)	Yes
Yang et al	China	295	N/A	Normal	0.98 (0.97–1.00)	92.00%	97.00%	87.00%	2D CNN (DenseNet)	Not
Song et al	China	201	N/A	Non-COVID-19	0.85	N/A	80.00%	75.00%	2D CNN (Big-BiGAN)	Not
Bai et al	China/USA	1186	N/A	Non-COVID-19	0.90	87.00%	89.00%	86.00%	2D CNN (EfficientNet) + MLP	Yes
Jaiswal et al	India	2492	N/A	Non-COVID-19	0.97	96.25%	N/A	96.21%	2D CNN (DenseNet) + Con-vLSTM (DADLM)	Yes
Han et al	China	329	N/A	Non-COVID-19	0.99	97.90%	N/A	N/A	2D CNN	Yes
Ouyang et al	China	1588	N/A	CAP	0.94	87.50%	86.9%	90.10%	3D CNN (ResNet)	Not
Jun Wang et al	China	4657	N/A	ILD/normal	0.97 (0.96–0.98)	93.30%	87.6%	95.50%	3D CNN (ResNet)	Not
Sun et al	China	2522	N/A	CAP	0.96	91.79%	93.05%	89.95%	2D CNN	Yes
Harmon et al	China/Japan/Italy/USA	2617	N/A	Non-COVID-19	0.95	90.80%	84.00%	95.10%	3D CNN (DenseNet)	Not

DL deep learning, FL federated learning model, N/A no information, CAP community-acquired pneumonia, CNN convolutional neural network, MLP multi-layer perceptron, AUC area under curve of receiver operating characteristic curve (ROC), ILD interstitial lung diseases



of 0.82 to 0.99. Only two studies reported additional functions of prognostic prediction. The control groups in majority of the studies were community-acquired pneumonia, non-COVID-19, or normal chest CT. No study was based on stepwise diagnostic design. There is no published peer-reviewed work on the use of FL for this field at present.

## Discussion

Our proposed CoviDet (Supplementary Fig. 1–14) is novel and different from other published reports on the use of AI for CT diagnosis of COVID-19 in several ways.

### Novel use of the 2D CNN combined with RNN for CT analysis

Unlike most reported AI frameworks of COVID-19, which relied on 3D CNN [13, 22], we have used the 2D CNN + RNN framework. This is because 3D kernels are typically small and require down-sampling of the image. This can lead to data loss and affect prediction accuracy [28]. To avoid this, we use 2D CNN to extract features from each slide. Then, GRU, an implementation of RNN, used information extracted from several adjacent slides, integrating them into a global feature vector for the whole CT scan. This also avoids the time-consuming annotation required for 2D CNN and avoids the large memory system required for 3D CNN. It is more consistent with the strategy of radiologists. All these in combination led to better performance and higher accuracy than previous works, even when CT without clinical data was used on its own.

### Novelty of the stepwise diagnosis algorithm

The rationale for using the stepwise diagnosis algorithm is to mimic the decision-making process made by clinicians during the CT assessment for COVID-19. From a technical perspective, such a stepwise approach was also easier to implement as the first step to distinguish viral pneumonia and from other pulmonary conditions is a coarse-grained classification function. In this study, the AUC of our model for this process is 0.99.

However, it is a more challenging task to distinguish COVID-19 pneumonia from other viral pneumonia on CT scans for our expert radiologists as well as for DL algorithm implementation. This is because it is a fine-grained classification task, which requires the model to learn more subtle changes. With the use of transfer learning from the first step, the model can distinguish COVID-19 from other viral pneumonia cases with the AUC as high as 0.98, outperforming experienced radiologists.

## Incorporation of federated learning

To avoid the need for transmission of any potentially sensitive data to another country or site, FL has been put forward as a way to refine DL models by implementing the learning process locally [4]. Only encrypted parameters are transferred to the coordinating site for further model refinement. We therefore have successfully tested the incorporation of our CoviDet model with the FL framework. This will open up new opportunities for global collaboration especially during this pandemic as this approach can avoid issues associated with data sharing ownership and privacy breaches. It will allow the rapid development of this model and improve its generalizability.

### Potential for automated CT monitoring of COVID-19

The other unique and useful feature of CoviDet is the incorporation of DL to provide automated assessment of disease severity and CT monitoring of COVID-19 changes. The plot of the lesion area changing over time provides an automated graphic visualization of CT changes for clinicians to monitor a patient's progress and make clinical recommendations. We had tried to use the GPR to mimic and predict the clinical course for COVID-19 patients individually. This is a practical and important property for COVID-19 lesion area prediction because the conditions of patients can vary greatly as seen in Fig. 4. Thus, it is unreasonable to handle all of them by a pre-trained prediction model.

### Study limitations

There are several limitations to this study. First, the sample size used for the monitoring of COVID-19 CT changes is modest; a large data set is required to confirm its clinical usefulness and improve its accuracy. Second, the therapeutic-related information was not involved in this model, and the clinical course might be affected by the medical care. Third, the generalizability of the FL model needs to be validated in the future.

## Conclusions

The CoviDet can accurately diagnose COVID-19 based on CT scans with or without clinical data. There is potential to further apply CoviDet for the monitoring and prediction of a patient's clinical course. Alternatively, it can be implemented on a federated learning platform.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-021-08334-6>.

**Acknowledgements** We thank Keng-Leong Ang and Lindsey Hamblin for the help in language editing. We thank the hospital staff for their efforts in collecting the information. We thank hospitals that offered the data from around the nation (Electronic supplementary material hospital list). We thank all the patients who consented to donate their data for analysis and the medical staff working on the front line.

**Funding** This study is supported by a high-level university construction project of Guangzhou Medical University (Grant No. 20182737, 201721007, 201715907, 2017160107); the Guangdong high-level hospital construction “reaching peak” plan; the Guangzhou Institute of Respiratory Health Open Project (funds provided by the China Evergrande Group)—Project No. (2020GIRHHMS01, 2020GIRHHMS09, 2020GIRHHMS10, 2020GIRHHMS25); the National Natural Science Foundation of China (No. 61571269); the National Key R&D Program of China (No. 2020AAA0105500); and the National Natural Science Foundation of China (No. 61971260).

## Declarations

**Guarantor** The guarantors are Jianxing He, Wenhua Liang and Hengrui Liang.

**Conflict of interest** The authors have declared no conflicts of interest.

**Statistics and biometry** One of the authors has significant statistical expertise.

**Informed consent** Consent was waived due to the use of anonymized retrospective data.

**Ethical approval** This study was approved by the ethics committee of the First Affiliated Hospital of Guangzhou Medical University; consent was waived due to the use of anonymized retrospective data.

**Study subjects or cohorts overlap** The cohorts have been described in the manuscript and supplementary materials.

**Methodology** • Retrospective study, diagnostic test research.


## References

- Tang YW, Schmitz JE, Persing DH, Stratton CW (2020) Laboratory diagnosis of COVID-19: current issues and challenges. *J Clin Microbiol* 58(6):e00512–20. <https://doi.org/10.1128/JCM.00512-20>
- Wang W, Xu Y, Gao R et al (2020) Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 323(18):1843–1844. <https://doi.org/10.1001/jama.2020.3786>
- Ai T, Yang Z, Hou H et al (2020) Correlation of chest CT and RT-PCR testing for Coronavirus Disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 296(2):E32–E40. <https://doi.org/10.1148/radiol.2020200642>
- Sattler F, Muller KR, Samek W (2021) Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints. *IEEE Trans Neural Netw Learn Syst* 32(8):3710–3722. <https://doi.org/10.1109/TNNLS.2020.3015958>
- Kim DH, MacKinnon T (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 73(5):439–445
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297. <https://doi.org/10.1007/BF00994018>
- Acharya J, Basu A (2020) Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE Trans Biomed Circuits Syst* 14(3):535–544
- Poovizhi S, Ganesh Babu TR (2020) An efficient skin cancer diagnostic system using Bendlet transform and support vector machine. *An Acad Bras Cienc* 92(1):e20190554
- Sabottke CF, Breaux MA, Spieler BM (2020) Estimation of age in unidentified patients via chest radiography using convolutional neural network regression. *Emerg Radiol* 27(5):463–468. <https://doi.org/10.1007/s10140-020-01782-5>
- Kim J, Lim KT, Park K, Kim Y, Cho G (2020) Uncertainty estimation of the dose rate in real-time applications using Gaussian process regression. *Sensors (Basel)* 20(10):2884. <https://doi.org/10.3390/s20102884>
- Pan F, Ye T, Sun P et al (2020) Time course of lung changes at chest CT during recovery from Coronavirus Disease 2019 (COVID-19). *Radiology* 295(3):715–721. <https://doi.org/10.1148/radiol.2020200370>
- Yang S, Jiang L, Cao Z et al (2020) Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study. *Ann Transl Med* 8(7):450
- Zhang K, Liu X, Shen J et al (2020) Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181(6):1423–1433 e1411
- Ardakani AA, Kanafi AR, Acharya UR, 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. *Comput Biol Med* 121:103795
- Song J, Wang H, Liu Y et al (2020) End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT. *Eur J Nucl Med Mol Imaging* 47(11):2516–2524. <https://doi.org/10.1007/s00259-020-04929-1>
- Wu X, Hui H, Niu M et al (2020) Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *Eur J Radiol* 128:109041
- Wang S, Zha Y, Li W et al (2020) A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 56(2):2000775. <https://doi.org/10.1183/13993003.00775-2020>
- Ko H, Chung H, Kang WS et al (2020) COVID-19 Pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation. *J Med Internet Res* 22(6):e19569
- Mei X, Lee HC, Diao KY et al (2020) Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med* 26(8):1224–1228. <https://doi.org/10.1038/s41591-020-0931-3>
- Bai HX, Wang R, Xiong Z et al (2020) Ai augmentation of radiologist performance in distinguishing covid-19 from pneumonia of other etiology on chest ct. *Radiology* 296(3):201491
- Li L, Qin L, Xu Z et al (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology* 200905
- Harmon SA, Sanford TH, Xu S et al (2020) Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* 11(1):4080
- Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M (2021) Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *J Biomol Struct Dyn* 39(15):5682–5689. <https://doi.org/10.1080/07391102.2020.1788642>
- Sun L, Mo Z, Yan F et al (2020) Adaptive feature selection guided deep forest for COVID-19 classification with chest CT. *IEEE J Biomed Health Inform* 24(10):2798–2805. <https://doi.org/10.1109/JBHI.2020.3019505>

25. Wang J, Bao Y, Wen Y et al (2020) Prior-attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Trans Med Imaging* 39(8):2572–2583
26. Ouyang X, Huo J, Xia L et al (2020) Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans Med Imaging* 39(8):2595–2605
27. Han Z, Wei B, Hong Y et al (2020) Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. *IEEE Trans Med Imaging* 39(8):2584–2594
28. You Y, Lu C, Wang W, Tang CK (2019) Relative CNN-RNN: learning relative atmospheric visibility from images. *IEEE Trans Image Process* 28(1):45–55

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Hengrui Liang<sup>1,2</sup>  · Yuchen Guo<sup>3,4</sup> · Xiangru Chen<sup>5</sup> · Keng-Leong Ang<sup>2,6</sup> · Yuwei He<sup>4,7</sup> · Na Jiang<sup>8</sup> · Qiang Du<sup>5</sup> · Qingsi Zeng<sup>1,9</sup> · Ligong Lu<sup>10</sup> · Zebin Gao<sup>5</sup> · Linduo Li<sup>11</sup> · Quanzheng Li<sup>12</sup> · Fangxing Nie<sup>5</sup> · Guiguang Ding<sup>3,4,7</sup> · Gao Huang<sup>5,7</sup> · Ailan Chen<sup>1,13</sup> · Yimin Li<sup>1,14</sup> · Weijie Guan<sup>1</sup> · Ling Sang<sup>1,14</sup> · Yuanda Xu<sup>1,14</sup> · Huai Chen<sup>1,9</sup> · Zisheng Chen<sup>1</sup> · Shiyue Li<sup>1</sup> · Nuofu Zhang<sup>1</sup> · Ying Chen<sup>1</sup> · Danxia Huang<sup>1</sup> · Run Li<sup>1</sup> · Jianfu Li<sup>1,2</sup> · Bo Cheng<sup>1,2</sup> · Yi Zhao<sup>1,2</sup> · Caichen Li<sup>1,2</sup> · Shan Xiong<sup>1,2</sup> · Runchen Wang<sup>1,2</sup> · Jun Liu<sup>1,2</sup> · Wei Wang<sup>1,2</sup> · Jun Huang<sup>1,2</sup> · Fei Cui<sup>1,2</sup> · Tao Xu<sup>15</sup> · Fleming Y. M. Lure<sup>16</sup> · Meixiao Zhan<sup>10</sup> · Yuanyi Huang<sup>17</sup> · Qiang Yang<sup>18</sup> · Qionghai Dai<sup>3,4</sup> · Wenhua Liang<sup>1,2</sup> · Jianxing He<sup>1,2,19</sup> · Nanshan Zhong<sup>1</sup>

✉ Qionghai Dai  
qhdai@tsinghua.edu.cn

✉ Wenhua Liang  
liangwh1987@163.com

✉ Jianxing He  
drjianxing.he@gmail.com

<sup>1</sup> National Clinical Research Center for Respiratory Disease, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

<sup>2</sup> Department of Thoracic Surgery, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

<sup>3</sup> Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China

<sup>4</sup> Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China

<sup>5</sup> Beijing XiaoBaiShiJi Network Technical Co., Ltd, Beijing 100084, China

<sup>6</sup> Department of Thoracic Surgery, Glenfield Hospital, Leicester LE3 9QP, UK

<sup>7</sup> School of Software, Tsinghua University, Beijing 100084, China

<sup>8</sup> Department of Gastroenterology, Wuhan Hankou Hospital, Wuhan 430000, China

<sup>9</sup> Department of Radiology, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

<sup>10</sup> Zhuhai Interventional Medical Center, Zhuhai Precision Medical Center, Zhuhai People's Hospital, Zhuhai Hospital Affiliated with Jinan University, Jinan University, Zhuhai 519000, China

<sup>11</sup> College of Engineering, Northeastern University, 360 Huntington Ave., Boston, MA 02115, USA

<sup>12</sup> Department of Radiology, Massachusetts General Hospital, White-427 55 Fruit St, Boston, MA 02114, USA

<sup>13</sup> Department of Cardiology, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

<sup>14</sup> Department of Intensive Care Unit, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

<sup>15</sup> Biomanufacturing Center Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

<sup>16</sup> College of Engineering, University of Texas, El Paso, USA

<sup>17</sup> Department of Radiology, Jingzhou Central Hospital, The Second Clinical Medical College, Yangtze University, Jingzhou, Hubei, China

<sup>18</sup> Hong Kong University of Science and Technology and weBank, Hong Kong, China

<sup>19</sup> South China Medical University, Guangzhou 510000, China