

## 1 A collaborative online AI engine for CT-based COVID-19 diagnosis

2

3 Yongchao Xu<sup>1,2#</sup>, Liya Ma<sup>1#</sup>, Fan Yang<sup>3#</sup>, Yanyan Chen<sup>4#</sup>, Ke Ma<sup>2</sup>, Jiehua Yang<sup>2</sup>, Xian Yang<sup>2</sup>, Yaobing  
4 Chen<sup>5</sup>, Chang Shu<sup>2</sup>, Ziwei Fan<sup>2</sup>, Jiefeng Gan<sup>2</sup>, Xinyu Zou<sup>2</sup>, Renhao Huang<sup>2</sup>, Changzheng Zhang<sup>6</sup>,  
5 Xiaowu Liu<sup>6</sup>, Dandan Tu<sup>6</sup>, Chuou Xu<sup>1</sup>, Wenqing Zhang<sup>2</sup>, Dehua Yang<sup>7</sup>, Ming-Wei Wang<sup>7</sup>, Xi Wang<sup>8</sup>,  
6 Xiaoliang Xie<sup>8</sup>, Hongxiang Leng<sup>9</sup>, Nagaraj Holalkere<sup>10</sup>, Neil J. Halin<sup>10</sup>, Ihab Roushdy Kamel<sup>11</sup>, Jia Wu<sup>12</sup>,  
7 Xuehua Peng<sup>13</sup>, Xiang Wang<sup>14</sup>, Jianbo Shao<sup>13</sup>, Pattanasak Mongkolwat<sup>15</sup>, Jianjun Zhang<sup>16,17</sup>, Daniel L.  
8 Rubin<sup>18</sup>, Guoping Wang<sup>5</sup>, Chuangsheng Zheng<sup>3\*</sup>, Zhen Li<sup>1\*</sup>, Xiang Bai<sup>2\*</sup>, Tian Xia<sup>2,5\*</sup>

9 <sup>1</sup>Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and  
10 Technology, Wuhan 430030, China.

11 <sup>2</sup>School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan  
12 430074, China.

13 <sup>3</sup>Department of Radiology, Union Hospital of Tongji Medical College, Huazhong University of Science and  
14 Technology, Wuhan 430022, China.

15 <sup>4</sup>Department of Information Management, Tongji Hospital, Huazhong University of Science and Technology,  
16 Wuhan 430030, China.

17 <sup>5</sup>Institute of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology,  
18 Wuhan 430030, China.

19 <sup>6</sup>HUST-HW Joint Innovation Lab, Wuhan 430074, China.

20 <sup>7</sup>The National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences,  
21 Shanghai 201203, China.

22 <sup>8</sup>CalmCar Vision System Ltd., Suzhou, China.

23 <sup>9</sup>SAIC Advanced Technology Department, SAIC, Shanghai, China.

24 <sup>10</sup>CardioVascular and Interventional Radiology, Radiology for Quality and Operations, The CardioVascular Center  
25 at Tufts Medical Center, Radiology, Tufts University School of Medicine.

26 <sup>11</sup>Russell H Morgan Department of Radiology & Radiologic Science, Johns Hopkins hospital, Johns Hopkins  
27 Medicine Institute, 600 N Wolfe St, Baltimore, MD 21205 USA.

28 <sup>12</sup>Department of Radiation Oncology, Stanford University School of Medicine, 1070 Arastradero Rd, Palo Alto,  
29 CA94304.

30 <sup>13</sup>Department of Radiology, Wuhan Children's Hospital, Wuhan, China.

31 <sup>14</sup>Department of Radiology, Wuhan Central Hospital, Wuhan, China.

32 <sup>15</sup>Faculty of Information and Communication Technology, Mahidol University, Thailand.

33 <sup>16</sup>Thoracic/Head and Neck Medical Oncology, <sup>17</sup>Translational Molecular Pathology, The University of Texas MD  
34 Anderson Cancer Center, Houston, Texas 77030, USA.

35 <sup>18</sup>Department of Biomedical Data Science, Radiology and Medicine, Stanford University, USA.

36

37 <sup>#</sup>These authors contributed equally to this work.

38 \*Correspondence should be addressed to T.X. ([tianxia@hust.edu.cn](mailto:tianxia@hust.edu.cn)), X.B. ([xbai@hust.edu.cn](mailto:xbai@hust.edu.cn)),

39 Z.L. ([zhenli@hust.edu.cn](mailto:zhenli@hust.edu.cn)), or C.Z. ([hqzcsx@sina.com](mailto:hqzcsx@sina.com)).

40

## 41 Abstract

42 Artificial intelligence can potentially provide a substantial role in streamlining chest computed  
43 tomography (CT) diagnosis of COVID-19 patients. However, several critical hurdles have  
44 impeded the development of robust AI model, which include deficiency, isolation, and  
45 heterogeneity of CT data generated from diverse institutions. These bring about lack of  
46 generalization of AI model and therefore prevent it from applications in clinical practices. To  
47 overcome this, we proposed a federated learning-based Unified CT-COVID AI Diagnostic  
48 Initiative (UCADI, <http://www.ai-ct-covid.team/>), a decentralized architecture where the AI  
49 model is distributed to and executed at each host institution with the data sources or client ends  
50 for training and inferencing without sharing individual patient data. Specifically, we firstly  
51 developed an initial AI CT model based on data collected from three Tongji hospitals in Wuhan.  
52 After model evaluation, we found that the initial model can identify COVID from Tongji CT test  
53 data at near radiologist-level (97.5% sensitivity) but performed worse when it was tested on  
54 COVID cases from Wuhan Union Hospital (72% sensitivity), indicating a lack of model  
55 generalization. Next, we used the publicly available UCADI framework to build a federated  
56 model which integrated COVID CT cases from the Tongji hospitals and Wuhan Union hospital  
57 (WU) without transferring the WU data. The federated model not only performed similarly on  
58 Tongji test data but improved the detection sensitivity (98%) on WU test cases. The UCADI  
59 framework will allow participants worldwide to use and contribute to the model, to deliver a  
60 real-world, globally built and validated clinic CT-COVID AI tool. This effort directly supports  
61 the United Nations Sustainable Development Goals' number 3, Good Health and Well-Being,  
62 and allows sharing and transferring of knowledge to fight this devastating disease around the  
63 world.

## 64      **Introduction**

65      COVID-19 has become a global pandemic. RT-PCR was adopted as the main diagnostic  
66      modality to detect viral nucleotide in specimens from patients with suspected COVID-19  
67      infection and remained as the gold standard for active disease confirmation. However, due to the  
68      greatly variable disease course in different patients, the detection sensitivity is only 60%-71%<sup>1-3</sup>  
69      leading to considerable false negative results. These symptomatic COVID 19 patients and  
70      asymptomatic carriers with false negative RT-PCR results pose a significant public threat to the  
71      community as they may be contagious. As such, clinicians and researchers have made  
72      tremendous efforts searching for alternative and/or complementary modalities to improve the  
73      diagnostic accuracy for COVID-19.

74      COVID-19 patients present with certain unique radiological features on chest computed  
75      tomography (CT) scans including ground glass opacity, interlobular septal thickening,  
76      consolidation etc., that have been used to differentiate COVID-19 from other bacterial or viral  
77      pneumonia or healthy individuals<sup>4-7</sup>. CT has been utilized for diagnosis of COVID-19 in some  
78      countries and regions with reportedly sensitivity of 56-98%<sup>2,3</sup>. However, these radiologic  
79      features are not specifically tied to COVID-19 pneumonia and the diagnostic accuracy heavily  
80      depending on radiologists' experience. Particularly, insufficient empirical understanding of the  
81      radiological morphology characteristic of this unknown pneumonia resulted in inconsistent  
82      sensitivity and specificity by varying radiologists in identifying and assessing COVID-19. A  
83      recent study has reported substantial differences in the specificity in differentiation of COVID-19  
84      from other viral pneumonia by different radiologists<sup>8</sup>. Meanwhile, CT-based diagnostic  
85      approaches have led to substantial challenges as many suspected cases will eventually need

86 laboratory confirmation. Therefore, there is an imperative demand for an accurate and specific  
87 intelligent automatic method to help to address the clinical deficiency in current CT approaches.

88 Successful development of an automatic method depends on a tremendous amount of imaging  
89 data with high quality clinical annotation for training an artificial intelligence (AI) model. We  
90 confronted several challenges for developing a robust and universal AI tool for precise COVID-  
91 19 diagnosis: 1) data deficiency. Our high-quality CT data sets were only a small sampling of the  
92 full infected cohorts and therefore it is unlikely we captured the full set radiological features. 2)  
93 data isolation, Data derived across multiple centers was difficult to transfer for training due to  
94 security, privacy, and data size concerns. and 3) data heterogeneity. Datasets were generated by  
95 different scanner machines which introduces an additional layer of complexity to the training  
96 because every vendor provides some unique capabilities. Furthermore, it is unknown whether  
97 COVID-19 patients in diverse geographic locations, ethnic groups, or demographics show  
98 similar or distinct CT image patterns. All of these may contribute to a lack of generalization for  
99 an AI model, which a serious issue for a global AI clinical solution.

100 To solve this problem, we propose here a Unified CT-COVID AI Diagnostic Initiative (UCADI)  
101 to deliver an AI-based CT diagnostic tool. We base our developmental philosophy on the  
102 concept of federated learning, which enables machine learning engineers and medical data  
103 scientists to work seamlessly and collectively with decentralized CT data without sharing  
104 individual patient data, and therefore every participating institution can contribute to AI training  
105 results of CT-COVID studies to a continuously-evolved and improved central AI model and help  
106 to provide people worldwide an effective AI model for precise CT-COVID diagnosis (Fig.1).

107

108 **Results**

109 **Building AI model using pooled data**

110 We firstly gathered a dataset of 5732 CT images from 1276 individuals collected from multiple  
111 centers of Tongji Hospital including Tongji Hospital Main Campus (3457 CT images from 800  
112 studies), Tongji Optical Valley Hospital (882 CT images from 227 studies), and Tongji Sino-  
113 French New City Hospital (1393 CT images from 241 studies) (Table 1 for patient information ).

114 Among these patients, 432 patients had COVID-19 pneumonia confirmed by RT-PCR; 76  
115 patients had other viral pneumonia including 7 cases with respiratory syncytial virus (RSV), 13  
116 with EB virus, 16 with cytomegalovirus, 3 with influenza A, 1 with parainfluenza virus and 36  
117 with mixed virus pneumonia that were confirmed PCR or antibodies against corresponding  
118 viruses; 350 patients had bacterial pneumonia confirmed CT scan and bacterial culture. The  
119 remaining 418 individuals having clinical symptoms of respiratory system were healthy  
120 individuals who had normal chest CT scans. Based on the dataset, we developed an initial deep  
121 learning model by using convolutional neural networks (CNN) (detailed in Methods).

122 Next, we validated the predictive performance of the CNN through a classification task: four-  
123 class pneumonia partition—four featured clinical diagnoses in determining suspected cases of  
124 COVID-19. This task aimed at distinguishing COVID-19 (Fig. 3. i) from three types of non-  
125 COVID-19 (Fig. 3. ii) including other viral pneumonia, bacterial pneumonia, and healthy cases  
126 (d, e, and f in Fig. 3). We selected 20% of 1036 CT cases in training and validation set for 5-fold  
127 cross-validation. The CNN demonstrated the validation result that achieved overall sensitivity of  
128 77.2% and specificity of 91.9%.

129 We further tested the previously trained CNN by conducting a comparative study of same task  
130 between the CNN and expert radiologists using previously separated test set (detailed in  
131 Methods). Six qualified radiologists (ZL [18 years' experience], LYM [9 years' experience],  
132 YZL [9 years' experience], COX [8 years' experience], HLM [4 years' experience], GC [4  
133 years' experience]) from department of radiology, Tongji Hospital (Main campus), Wuhan,  
134 China were asked to make diagnosis as one of above 4 classes based on CT study. In this task,  
135 the CNN achieved a sensitivity of 97.5% and specificity of 89.4% in differentiating COVID-19  
136 from three types of non-COVID-19 cases whereas six radiologists obtained the average 79% in  
137 sensitivity (87.5%, 90%, 55%, 80%, 68%, 93%, respectively, and 90% for the maximal voting  
138 value among six radiologists), and 90% in specificity (92%, 97%, 89%, 95%, 88%, 79%,  
139 respectively, and 95.6% for the maximal voting value) (Fig 4). In the Tongji dataset, the CNN  
140 shows performance approaching that of expert radiologists. To examine the reliability of the  
141 model, we performed class activation mapping (CAM) analysis for raw CT images in both  
142 validation and test datasets<sup>9</sup> and visualized the featured image regions which lead to  
143 classification decision. As shown in Figure 3. iii, the heatmap generated by CAM mostly  
144 characterized local lesions suggesting the model learned radiologic features rather than simply  
145 overfitting the dataset.

146 To comprehensively evaluate the comparisons of two tasks, we visualized the correlation of  
147 sensitivity and specificity via receiver operating characteristic (ROC) curve to calculate the area  
148 under the curve (AUC) for representing the CNN's classification performance. As a result, the  
149 AUC of the CNN attained 0.98, 0.88, 0.91, 0.98 in specifically identifying COVID-19 pneumonia,  
150 other viral pneumonia, bacterial pneumonia, and healthy tissue from 4 classes, and 0.92, 0.92,  
151 0.95 in assessing three ordinal severities of COVID-19. Fig. 4 illustrates the ROC curve of the

152 CNN and sensitivity-specificity points displaying radiologists' diagnosis. Importantly, the CNN  
153 performed comparable sensitivity-specificity to all six radiologists in differentiating COVID-19  
154 from non-COVID-19 cases (Fig. 4a). Meanwhile, the CNN also performed equivalent  
155 sensitivity-specificity in comparison with average radiologists in the assessment of three  
156 severities (e, f, g in Fig. 4). However, the CNN revealed insufficient capability in determining  
157 other viral pneumonia (Fig. 4b), bacterial pneumonia (Fig. 4c), and healthy case (Fig. 4d).

158 To test the generalization of the initial model that was trained exclusively on data from Tongji  
159 hospitals, we evaluated the predictive performance using CT data from 100 confirmed COVID-  
160 19 cases generated at Wuhan Union hospital. The accuracy of the model was only 72%,  
161 compared with a 97% sensitivity using reserved testing data from Tongji hospitals. This  
162 demonstrated a lack of generalization for the initial model.

### 163 **The global online AI diagnostic engine enabled with federated learning**

164 To overcome the hurdle, we proposed a federated learning framework to facilitate UCADI, a  
165 global joint effort to generate an AI based on large scale date and integration of diverse ethnic  
166 patient groups. In the traditional AI approach, sensitive user data from different sources are  
167 gathered and transferred to a central hub where models are trained and generated. The federated  
168 learning proposed by Google<sup>10</sup>, in contrast, is a decentralized architecture where the AI model is  
169 distributed to and executed at each host institution with the data sources or client ends for  
170 training and inferencing. The local copies of the AI model on the host institution eliminate  
171 network latencies and costs incurred due to sharing large size of data with the central server.  
172 Most importantly, the strategy privacy preserved by design enables medical centers collaborating  
173 on the development of models, but without need of directly sharing sensitive clinical data with  
174 each other.

175 We implemented the federated learning framework at <http://www.ai-ct-covid.team/> where we  
176 deployed the initial model to provide 1) online diagnostic interface allowing people easily query  
177 the model with patient CT images and 2) AI development federated learning interface(detailed in  
178 Methods). UCADI stakeholders can download the code and train a new model based on the  
179 initial model. Once the new model had been trained locally for several iterations, if UCADI  
180 participants share their updated version of the model, the framework will encrypt the model  
181 parameters based on Learning with Errors (LWE)-based encryption<sup>11</sup> and transfer them back to  
182 the centralized server via a customized server protocol. Participants' datasets will keep within  
183 their own secure infrastructure. The central server would then combine the contributions from all  
184 of the UCADI participants. The updated model parameters would then be shared with all  
185 participants, which enables continuation of local training. The framework is highly flexible,  
186 allowing hospitals join or leave the UCADI initiative at any moments, because it is not tied to  
187 any specific data cohorts.

188 With the framework, we deployed two experiments to validate federated learning concept on the  
189 CT COVID data. Firstly, we trained three models for each of three Tongji hospital datasets, and  
190 then transferred the datasets to three physically independent computer servers, respectively, and  
191 trained a Tongji federated model in a simulation mode (detailed in Methods). As shown in Figure  
192 4. e-h, the federated model performed close to the centralized-trained initial model and better  
193 than Tongji Main Campus model for predicting COVID-19, bacterial pneumonia and healthy  
194 case (the comparison not applied to models of Tongji Sino-French Hospital and Tongji Optics  
195 Valley because they lack of other viral pneumonia data). It shows the effectiveness of federated  
196 model. In the second experiment, we trained a federated model in real mode based on three  
197 Tongji hospital datasets (432 COVID-19 cases) and 407 confirmed COVID-19 cases from

198 Wuhan Union hospital. We tested the federated model performance on predicting the same 100  
199 confirmed Wuhan Union COVID-19 cases which we used to test the initial model previously.  
200 The result, 98% sensitivity, was improved compared to the initial model (72% sensitivity) which  
201 was centralized trained only based on data from three Tongji hospitals.

202 **Discussion**

203 COVID-19 is a global pandemic. Over 2 million people have been infected, tens of thousands  
204 hospitalized, and nearly 200,000 have died worldwide as of April 23<sup>rd</sup>, 2020. There are borders  
205 between countries. But only real border in this war is the border between human being and virus.  
206 We need a global joint effort to fight the virus. The first challenge we have confronted in this  
207 war is to deliver is deliver people precise and effective diagnosis. In this study, we introduce a  
208 globally collaborative AI initiative framework, UCADI, to assist radiologists, streamline, and  
209 accelerate CT-based diagnosis. Firstly, we developed an initial CNN model that achieved a  
210 performance comparable to expert radiologist in classifying pneumonia to identify COVID-19,  
211 and additionally assessing the severity of identified COVID-19. Furthermore, we developed a  
212 federated learning framework, based on which hospitals worldwide can join UCADI to jointly  
213 train an AI-CT model for COVID-19 diagnosis. With CT data from multiple Wuhan hospitals,  
214 we confirmed the effectiveness of this the federated learning approach. We have shared the  
215 initial model and the federated learning programmatic API source code  
216 (<https://github.com/HUST-EIC-AI-LAB/>) and encourage hospitals worldwide join UCADI to  
217 form an international collaboration to fight the virus with a globally trained AI application. It is  
218 worth noting that there is still need for improvement in the technical implementation in the  
219 framework: 1) The number of local training iterations before global parameter updating. The  
220 number of local training iterations has a direct influence on the training efficiency, effectiveness,

221 and model performance. Currently, different clients in UCADI framework train with their private  
222 data for one epoch before sending the parameter gradients to the global server. We will construct  
223 more detailed experiments about this hyper-parameter to explore the best trade-off between  
224 model performance and communication cost. 2) Private information leakage from gradients.  
225 Reconstruction of input data from the parameter gradients is possible for realistic deep  
226 architectures, and an encryption-decryption module is needed in the federated learning  
227 framework. We have adopted an additively homomorphic encryption scheme in our COVID  
228 diagnosis framework. The parameter gradients sent to the global server are encrypted while the  
229 secret key is kept confidential from the global server, which guarantees the privacy security of  
230 our framework. 3) Non-IID and unbalanced data distribution. The training data available is  
231 typically based on the patients in the hospital, and any particular hospital's local dataset will not  
232 be representative of the entire distribution. Therefore, it requires a dynamic aggregation method  
233 that aggregates different parameter gradients via dynamic weighted averaging. Hence, it can  
234 decrease the influence of non-IID and unbalanced data.

235 **Methods**

236 **CT data collecting and processing**

237 This study was approved by the Ethics Committee Tongji Hospital, Tongji Medical College of  
238 Huazhong University of Science and Technology to access this dataset for research purpose.  
239 Here we list the three major scanners used to obtain CT scans: GE Medical  
240 System/LightSpeed16, SOMATOM Definition AS+, and GE Medical Systems/Discovery 750  
241 HD. The scanning protocols of slice thicknesses and reconstruction kernel were 1.25mm and  
242 adaptive statistical iterative reconstruction (60%) for two GE scanners whilst 1mm and sinogram  
243 affirmed iterative reconstruction for the Siemens scanner. The high-quality CT image data from

244 the 432 COVID-19 patients were scanned, enrolled, selected and annotated in this study since  
245 January 7, 2020 while other image data were retrospectively collected from CT databases of the  
246 three Tongji Hospitals. In addition, we collected an independent cohort including 507 COVID-19  
247 pneumonia CT cases confirmed by chest CT from Union Hospital, Wuhan, China. The cohort  
248 was used for testing the performance of initial model and the multi-hospital model using  
249 federated learning framework.

250 We conducted image processing of the raw CT image data to reduce computing burdens. We  
251 utilized a sampling method to select 5 subsets of CT slices from all sequential images of one CT  
252 case using random starting positions and scalable sampling intervals on transverse view to  
253 picture the infected lung regions. All 5 processed subsets were separately fed to the CNN to  
254 obtain average predictive probabilities, which can effectively include impacts of different levels  
255 of lung from all CT slices. To further improve computing efficiency, we resized each slice from  
256 512 to 128 pixel regarding its width and height and rescaled the lung windows of CT to a range  
257 from -1200 to 600 and normalized them via the Z-score means before feeding the CNN.

## 258 **Building AI model using pooled data**

259 The dataset was split out into the training and validation set with 1036 cases (80% for training,  
260 20% for validation), and independent test set with 240 cases consisting of 80 COVID-19 studies  
261 (28 from Main Campus Hospital, 30 Sino-French New City Hospital, 20 Optical Valley  
262 Hospital), 20 with other viral pneumonia (19 from Main Campus Hospital, 1 Sino-French New  
263 City Hospital), 60 with bacterial pneumonia (50 from Main Campus Hospital, 8 Sino-French  
264 New City Hospital, 2 Optical Valley Hospital), and 80 healthy cases (58 Main Campus Hospital,  
265 10 Sino-French New City Hospital, 12 Optical Valley Hospital). We particularly considered the  
266 balanced data distribution of 4 classes in test set. We initially trained a four-class CNN (Fig. 2)

267 based on 3D-Densenet<sup>12</sup>, a densely connected convolutional network, which performed  
268 remarkable advantages in classifying CT images. We customized its architecture to contain 14  
269 3D-convolution layers distributed in 6 dense blocks and 2 transmit blocks (Fig. 2b indicating the  
270 architecture and data flow). The CNN took 16 resized 128-x128-pixel CT image sequences as  
271 input of each CT case, and generated a predicted pneumonia type with maximum probability as  
272 output across thousands of attached computing neurons. We defined the loss function as the  
273 weighted cross entropy between predicted probability and the true labels. Fine-tuned parameters  
274 of the network via back-propagation were optimized using batch size of 16, learning rate of 0.01,  
275 weight decay of 0.0001, momentum of 0.9, and epsilon of 0.00001. We conducted the training  
276 process utilizing a workstation equipped with 2 Tesla V100 GPUs, costing 6 hours to finish the  
277 task.

278 **Building AI model using federated learning**

279 Data preparation:

280 In experiment I, we trained with data collected from multiple centers of Tongji Hospital  
281 including Tongji Hospital Main Campus, Tongji Optical Valley Hospital, and Tongji Sino-  
282 French New City Hospital. We assigned each hospital to a federated client and place their local  
283 data on three different physical machines. In experiment II, besides data collected from above  
284 three hospitals, we added Wuhan Union Hospital as a new participant,

285 Federated model setup:

286 For all experiments, we used the same architecture (3D-Densenet) with data-centralized training  
287 and the same set of local training hyperparameters for all clients with SGD optimizer: batch size  
288 of 35, learning rate of 0.01, momentum of 0.9 and weight decay of 5e-4. In experiment I, we set

289 the number of federated rounds to 200 with one local epoch per federated round. A local epoch  
290 means each client train with its local data once before sending information to central  
291 server(cloud). We conducted the training process utilizing a workstation equipped with 3 Tesla  
292 V100 GPUs, costing 16 hours to finish. In experiment II, we set the number of federated rounds  
293 to 30 with one local epoch per federated round and start training with the global model coming  
294 from experiment I. For all experiments, we use the same evaluation metric with data-centralized  
295 training to check that our procedures are working properly. (In experiment II, we need to train 5  
296 rounds before the model achieving the same performance with data-centralized training on test  
297 data from Wuhan Union Hospital).

298 Model aggregation:

299 The server distributes a global model and receives synchronized weight updates ( $\Delta W_k^t$ ) from all  
300 clients at each federated round. Due to each client train with one epoch per federated round, so  
301 we just average all the weight updates from the client with equal weight and update the global  
302 model.

303 Privacy-preserving setup:

304 We use a variant of additively homomorphic encryption to achieve privacy-preserving, which  
305 called Learning with Errors (LWE)-based encryption. The encryption method allows us to leak  
306 no information of participants to the honest-but-curious parameter (cloud) server.

307 **Data Availability** All relevant data used for developing the initial model and federated models  
308 during the current study are not publicly available.

309

310 **Model Availability**

311 The online application of AI model is publicly available at <http://www.ai-ct-covid.team/>.

312 The initial model or offline APP is publicly available upon request at [tianxia@hust.edu](mailto:tianxia@hust.edu) or

313 [xbai@hust.edu.cn](mailto:xbai@hust.edu.cn) or through website <http://www.ai-ct-covid.team/>.

314

315 **Federated Learning Framework Availability.** The source code can be accessed at  
316 <https://github.com/HUST-EIC-AI-LAB/>.

317

## 318 References

- 319 1. Ai, T., *et al.* Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-  
320 19) in China: a report of 1014 cases. *Radiology*, 200642 (2020).
- 321 2. Fang, Y., *et al.* Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*,  
322 200432 (2020).
- 323 3. Kanne, J.P., Little, B.P., Chung, J.H., Elicker, B.M. & Ketar, L.H. Essentials for radiologists on  
324 COVID-19: an update—radiology scientific expert panel. (Radiological Society of North  
325 America, 2020).
- 326 4. Chung, M., *et al.* CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* **295**,  
327 202-207 (2020).
- 328 5. Kanne, J.P. Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan,  
329 China: key points for the radiologist. (Radiological Society of North America, 2020).
- 330 6. Shi, H., *et al.* Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan,  
331 China: a descriptive study. *The Lancet Infectious Diseases* (2020).
- 332 7. Vaseghi, G., *et al.* Clinical characterization and chest CT findings in laboratory-confirmed  
333 COVID-19: a systematic review and meta-analysis. *medRxiv* (2020).

- 334 8. Bai, H.X., *et al.* Performance of radiologists in differentiating COVID-19 from viral pneumonia  
335 on chest CT. *Radiology*, 200823 (2020).
- 336 9. Yao, T., Pan, Y., Li, Y., Qiu, Z. & Mei, T. Boosting image captioning with attributes. in  
337 *Proceedings of the IEEE International Conference on Computer Vision* 4894-4902 (2017).
- 338 10. McMahan, H.B., Moore, E., Ramage, D. & Hampson, S. Communication-efficient learning of  
339 deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).
- 340 11. Aono, Y., Hayashi, T., Wang, L. & Moriai, S. Privacy-preserving deep learning via additively  
341 homomorphic encryption. *IEEE Transactions on Information Forensics and Security* **13**, 1333-  
342 1345 (2017).
- 343 12. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K.Q. Densely connected convolutional  
344 networks. in *Proceedings of the IEEE conference on computer vision and pattern recognition*  
345 4700-4708 (2017).
- 346

347 **Acknowledgements**

348 This study was supported by HUST COVID-19 Rapid Response Call (No. 2020kfyXGYJ031,  
349 No. 2020kfyXGYJ093, No. 2020kfyXGYJ094) and the National Natural Science Foundation of  
350 China (61703171 and 81771801). This work was also supported in part by a grant from the  
351 National Cancer Institute, National Institutes of Health, U01CA242879, and Thammasat  
352 University Research fund under the NRCT, Contract No. 25/2561, for the project of “Digital  
353 platform for sustainable digital economy development”, based on the RUN Digital Cluster  
354 collaboration scheme.

355 **Author contributions**

356 T.X., X.B., Z.L., and C.Z, conceived the work. Y.X., L.M., F.Y., K.M., J.Y., X.Y, C.S., Z.F.,  
357 J.G., X.Z., R.H., C.Z., X. L., D.T., C.X., W.Z., D.Y., M.W., N.H., N.J.H., I.R.K., X.P., X.W.,

358 J.B. designed and developed the models and analyses; Y.X., K.M., D.L.R., J.Z., and T.X.  
359 interpreted results; and K.M., J.W., P.M., D.L.R., J.Z., Z.L., and T.X. wrote the paper.

360 **Competing interests**

361 The authors declare no competing interests.

362

363 **Tables**

	<b>Male</b>	<b>Female</b>	<b>0-20 years</b>	<b>20-40 years</b>	<b>40-60 years</b>	<b>60-80 years</b>	<b>&gt;80 years</b>
<b>Patient Number</b>	617	659	40	444	421	340	31

364 **Table 1** | Patient information of 1276 studies collected from Tongji Hospital regarding gender  
365 and age distribution.  
366

367

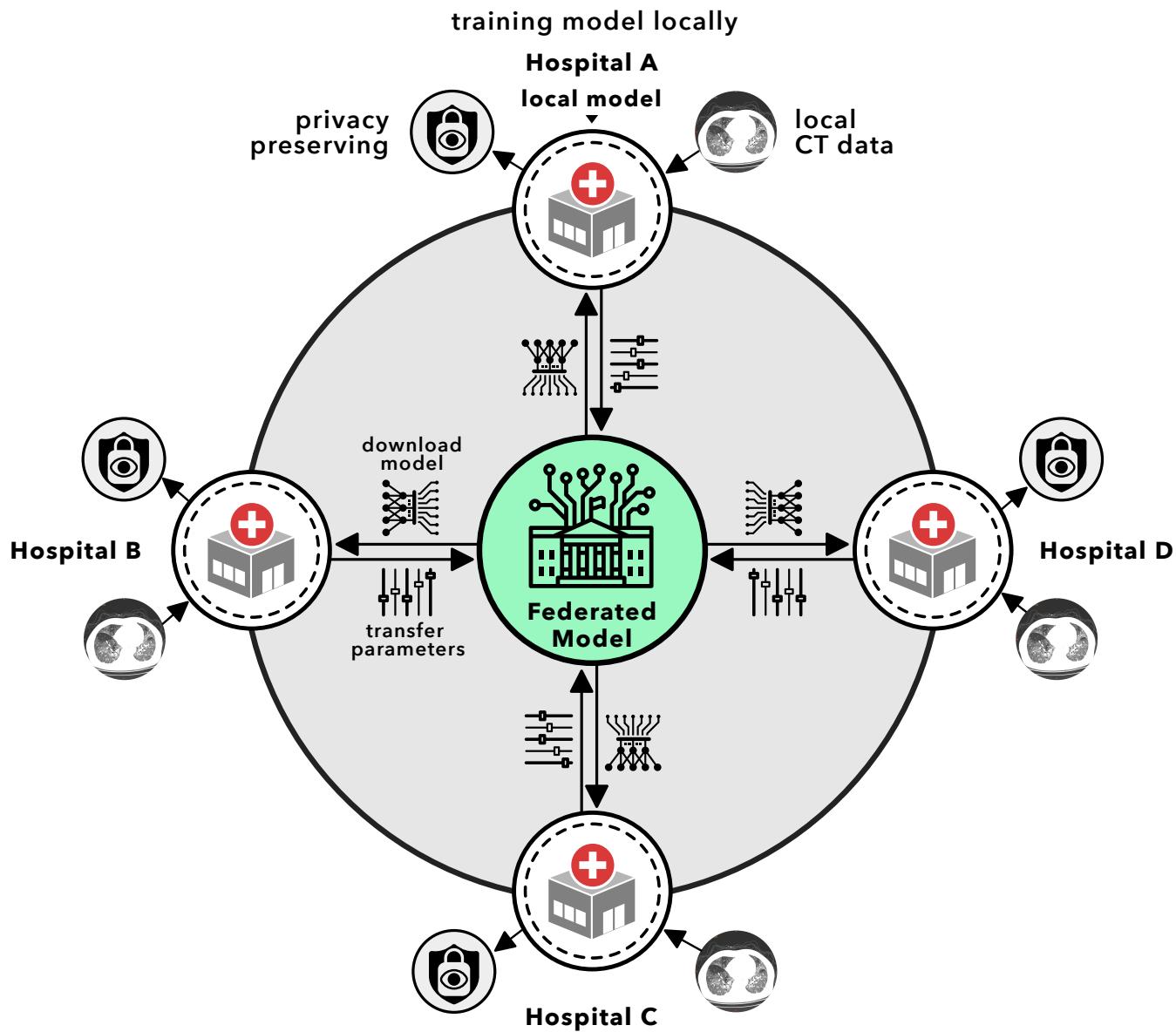
368

369

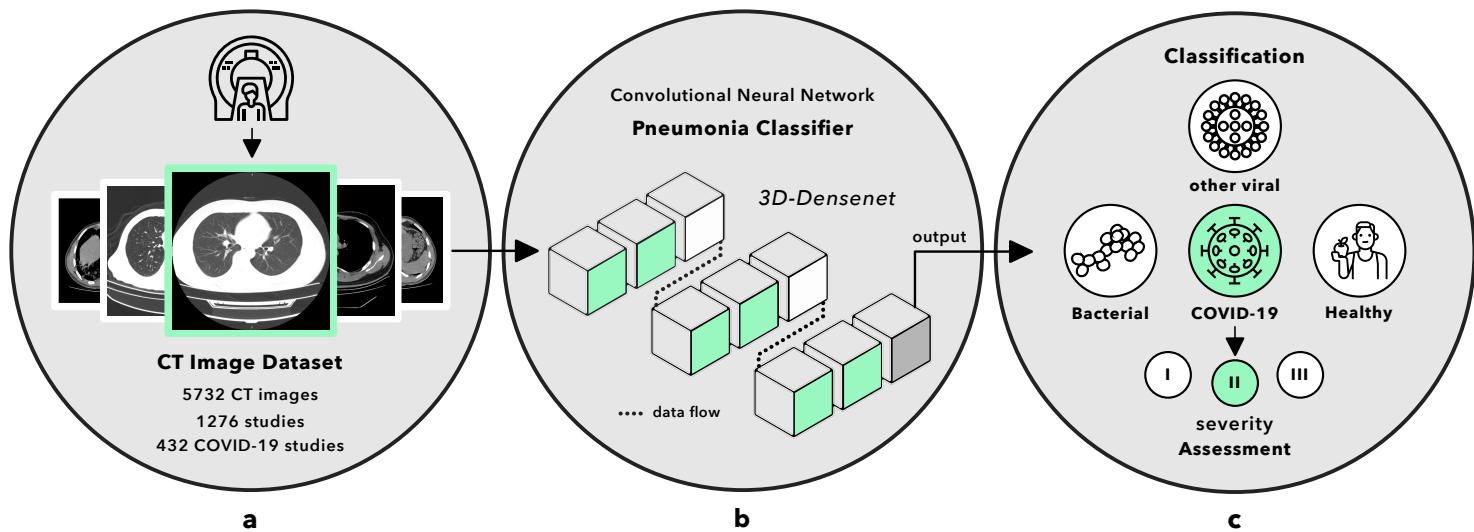
370

371

372



**Figure 1 | The conceptual architecture of UCADI on the basis of federated learning.** UCADI stakeholders firstly download the code and train a new model locally based on the initial model, and secondly transfer the encrypted model parameters back to the federated model. The central server combines the contributions shared from all of the UCADI participants.

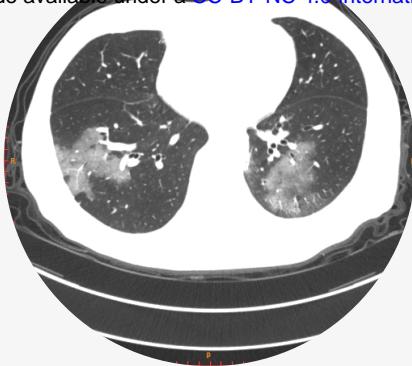


**Figure 2 | Data and strategy.** **a**, number of CT studies and total images. **b**, the CNN was developed based on 3D-Densenet, consisting of 6 dense blocks in green, 2 transmit blocks in white and an output layer in gray. Pre-processed 128-x-128-pixel CT images of one case were fed to the network across 14 3D-convolution layers and a number of functions embedded in 3D blocks, finally received the predicted classification result. **c**, the CNN classified CT case into 4 types and further assessed the severity into I or II or III if the case was predicted as COVID-19.

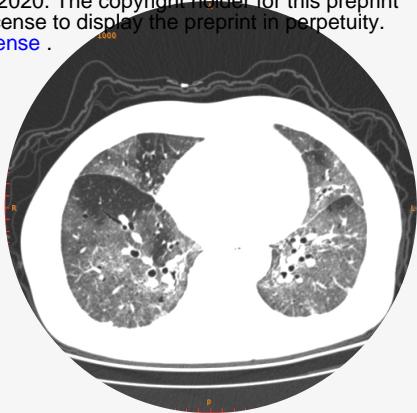
i. medRxiv preprint doi: <https://doi.org/10.1101/2020.05.10.20096072>; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.



a  
COVID-19 pneumonia



b  
COVID-19 pneumonia



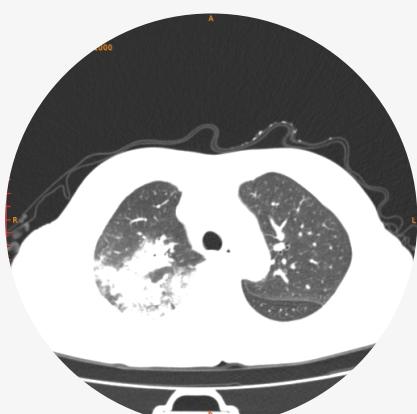
c  
COVID-19 pneumonia



d  
healthy



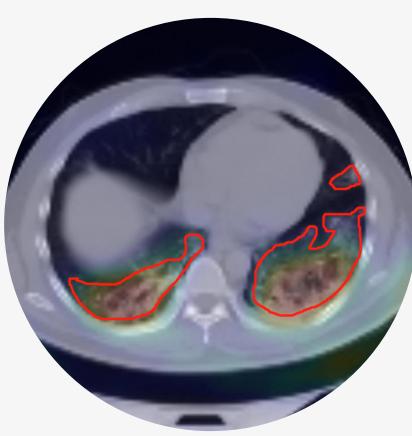
e  
other viral pneumonia



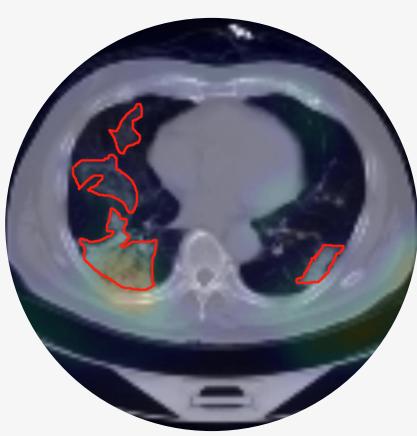
f  
bacterial pneumonia



g

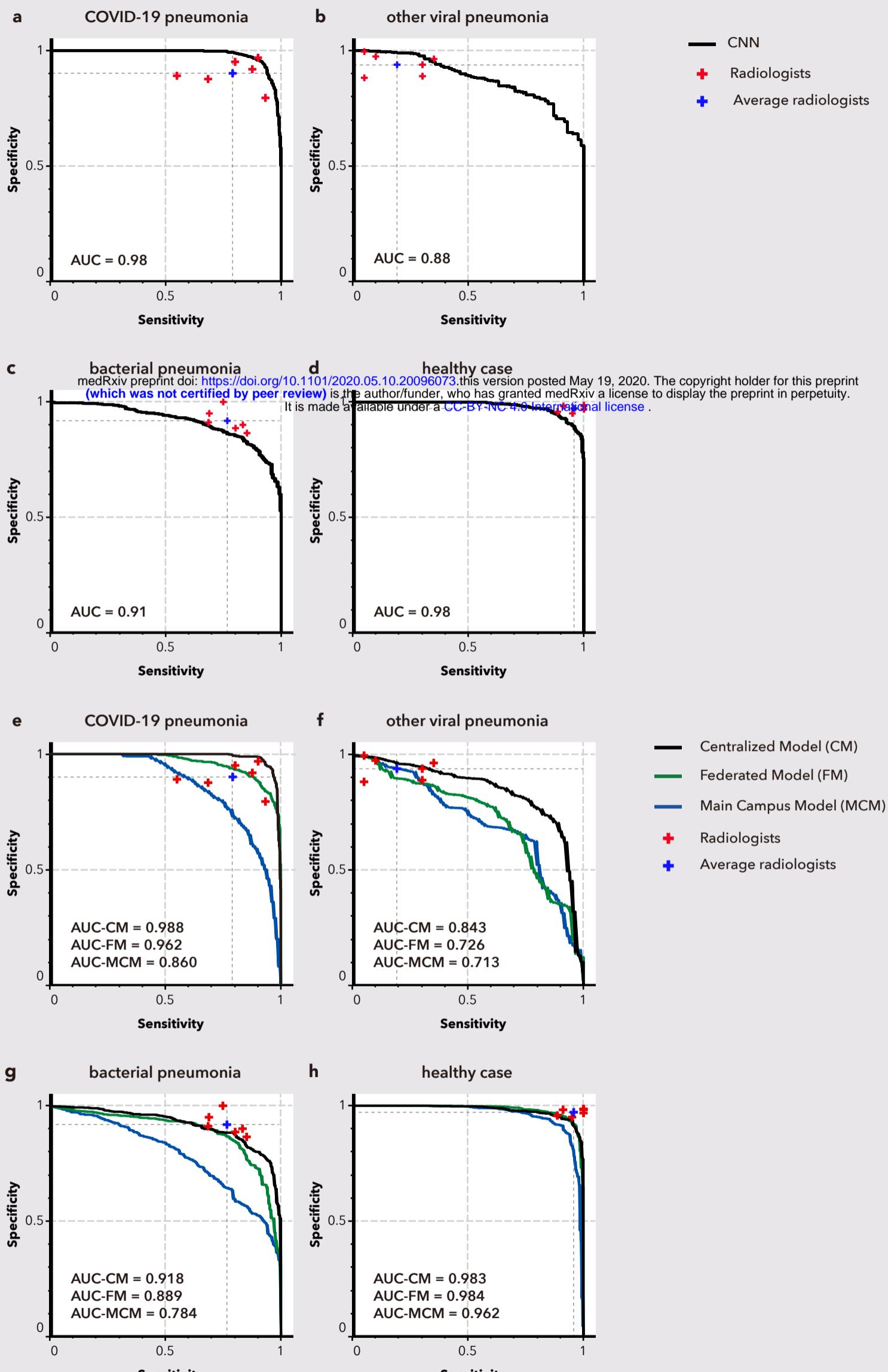


h



i

**Figure 3 | CT images. i and ii, the taxonomy of pneumonia and featured CT image for per-class. iii, the heatmap generated by GradCAM and local lesions annotated by the radiologist. i, COVID-19 pneumonia. a, b, c represent the CT images of COVID-19 defined by radiological features. ii, non-COVID-19 cases. d, e, f respectively displays the CT image of healthy case, other viral pneumonia, and bacterial pneumonia. iii, CAM visualized the image areas which lead to classification decision. The radiologist, LYM [9 years' experience], from Department of Radiology, Tongji Hospital circumscribed the local lesions with the red curved masks. g-h, patients with COVID-19 pneumonia.**



**Figure 4 | Pneumonia classification performance of CNN models and radiologists.** This figure illustrates the comparative analysis between the CNN and radiologists by correlating the ROC curve of CNN and sensitivity-specificity points of six invited radiologists for two conducted classification test tasks. **a-d**, per-class evaluation for three types of pneumonia and healthy case. The curve in black represents the performance of the CNN. Cross marks in red separately represent the performance of six radiologists and the blue mark annotates the average capability. **e-h**, comparative evaluation of centralized-trained initial model, federated model, and Tongji Main Campus model on four per-class classification tasks.