**Supplementary Materials**

**CT acquisition**

All 176 chest CT scans were performed using one of the following 17 multi-detector CT scanners: Somatom Emotion 16, Somatom Sensation 64, Somatom Definition, Somatom Definition AS, Somatom Definition AS+, Somatom Definition Edge, Somatoma Definition Flash, Somatom Force and Somatom Perspective (Siemens Healthineers, Erlangen, Germany); Brilliance 64 (Philips Healthcare, Amsterdam, Netherlands); Aquilion One (Canon Medical Systems, Otawara, Japan); Optima CT520, Optima CT660, Revolution CT, Revolution EVO, and Discovery CT 750 (GE Healthcare, Milwaukee, WI, USA); NeuViz 16 (Neusoft Medical System, Shenyang, China). The CT examinations were performed with the patient in the supine position at full inspiration with or without contrast medium using the following CT parameters: a tube voltage of 80-140 kVp; a tube current of 21-247 mAs; and a volume CT dose index of 1.21-16.67 mGy; preferential reconstruction matrix of 512×512. Axial images were reconstructed with a sharp or standard reconstruction kernel at a 0.75-7.0 mm slice thickness. The thickness of the sections was 1 mm in 79 scans, 0.75 mm in one scan; 1.25 mm in 48 scans; 1.5 mm in six scans; 2 mm in one scan; 2.5 mm in six scans; 3 mm in 24 scans; 4 mm in one scan; 5 mm in five scans; and 7 mm in five scans. Resampling of pixel spacing was not done in the CT scans, and we adjusted an in-plane resolution of 512×512 when required.

**Development and distribution of the deep neural network**

The training data in the lung window setting were normalized by Hounsfield units converted to floating values (0.00 to 1.00) for deep learning input. Our 2D U-Net received an input with a size of 512×512×1 and consisted of initial convolutions, four encoders, four decoders, and a final convolution (Supplementary table 2). Except for the final convolution, which was a 1×1 convolution, every convolutional layer consisted of a 3×3 convolution followed by batch normalization (1) and the rectified linear unit (ReLu) activation function (2). For decoders, upsampling with bilinear interpolation was used, followed by concatenation to conserve information before down-sampling (Figure 2).

The Kaiming He initialization method (3) was used for weight initialization. A sigmoid function was used in the final layer, and the model was trained using the stochastic gradient descent algorithm and the binary cross entropy loss function. Training was done with a learning rate of 0.001, batch size of 8, and a total of 100 epochs. After the training was completed, the tuning dataset was used to choose the best weight, which was saved after each epoch.

**Distribution of the software**

To avoid violating the data privacy of COVID-19 patients, we decided to implement the network as downloadable software instead of a cloud system. The 2D U-Net was distributed as free standalone software (MEDIP COVID19) by a functional modification of the FDA-approved MEDIP PRO v2.0.0.0. MEDIP COVID19 v1.0.0.0 was initially released on a website (http://medicalip.com/mobile/shop/covid19.php) on 18 March 2020 and updated with the current version of v1.2.0.0 on 27 April. The software quantified pneumonia in 1 minute with the following specifications: operating system, Microsoft Windows 7 (64 bit; IBM Corporation, San Jose, CA, USA) or higher; central processing unit, Intel i5 (Intel Corporation, Santa Clara, CA, USA) or higher; random-access memory, 8GB or higher; and graphics processing unit, GeForce 1000 series or higher (NVIDIA Corporation, Santa Clara, CA, USA) with a memory of 2 gigabytes or larger supporting the Compute Unified Device Architecture with the latest drivers. The software was downloaded in 1,256 institutions from 52 countries so far (Supplementary Figure 3).

**Calculation of pneumonia weight**

The weight of pneumonia (g) was calculated as follows: The attenuation values of the lesions were converted to the density of lung tissue by adding 1,000 to the HU values of each voxel and dividing by 1,000 (4). Densities in the range from air (-1024 HU) to water (0 HU) were approximately equal to the corresponding physical densities, and the density of lung tissue was assumed to be 1.065 g/mL (5).

Weight (g) =lesion attenuation value+1000 HU)1000∗1.065 (g/mL) ∗ volume of pneumonia (cm3)



**External validation datasets**

In the Japanese dataset, two thoracic radiologists evaluated the visual severity scores of the 103 CT images in consensus. The radiologists were blinded to the patients’ clinical information. A visual CT severity scoring system was used to obtain a semi-quantitative estimate of the pulmonary involvement of the GGO, consolidation, intralobular and interlobular septal thickening, and linear opacities (6). The extent of lung lesions was scored from 0 to 5 in each lobe and summed up (0, normal; 1, 1-5% involvement; 2, 6-25% involvement; 3; 26-50% involvement; 4, 51-75%; 5, 76-100% involvement) (6). Two CT scans were excluded due to motion artifacts.

The second and third datasets were public CT datasets that comprised 100 single axial scan images of Italian COVID-19 patients (7) and nine volumetric CT scans from the Radiopaedia website (8), all in DICOM format. MedSeg provided manually segmented lesion masks by experienced radiologists in each dataset (7). The masks comprised three layers: GGO, consolidation, and pleural effusion. The GGO and consolidation masks were merged and served as reference masks. The detailed segmentation method is described on the following website: https://medium.com/@hbjenssen/covid-19-radiology-data-collection-and-preparation-for-artificial-intelligence-4ecece97bb5b. In the second dataset, as one CT image lacked a reference mask, 99 CT images were included in the analysis.

The fourth dataset was a public dataset comprising 20 NifTi files of COVID-19 CT scans, labeled by two radiologists and verified by an experienced radiologist, from China (9). Of them, 10 CT scans with preserved Hounsfield unit values were included in the analysis. The fifth dataset was obtained from de-identified “COVID data save lives” data from HM Hospitales, Spain (10) in DICOM format. The study population included 115 COVID-19 patients who were admitted between February and April and underwent a chest CT scan within seven days of the admission date.

**References**

1. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167 2015.

2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM 2017;60(6):84-90.

3. He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision2015; p. 1026-1034.

4. Coxson HO, Rogers RM, Whittall KP, et al. A quantification of the lung surface area in emphysema using computed tomography. American journal of respiratory and critical care medicine 1999;159(3):851-856.

5. Hedlund LW, Vock P, Effmann EL. Evaluating lung density by computed tomography. Seminars in Respiratory Medicine: Copyright© 1983 by Thieme Medical Publishers, Inc., 1983; p. 76-88.

6. Inui S, Fujikawa A, Jitsu M, et al. Chest CT findings in cases from the cruise ship “Diamond Princess” with coronavirus disease 2019 (COVID-19). Radiology Cardiothoracic Imaging 2020;2(2).

7. Medseg. COVID-19 CT segmentation dataset. [Medseg Web site]. Published 2020. Available at: http://medicalsegmentation.com/covid19/. Accessed June 13, 2021.

8. Bell DJ. COVID-19. [Radiopaedia Web site]. Published 2020. Available at: https://radiopaedia.org/articles/covid-19-3. Accessed June 13, 2021.

9. Zhao W, Zhong Z, Xie X, et al. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. American Journal of Roentgenology 2020;214(5):1072-1077.

10. HMhospitales. Covid Data Save Lives. [HM hospitales Web site]. Published 2020. Available at: https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version. Accessed June 13, 2021.

**Supplementary Table**

**Supplementary Table 1. Demographics and CT parameteres of the Datasets**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristics** | **Training dataset** | **Tunning dataset** | **Test**  **dataset** | **First external validation dataset (Japan)** | **Second external validation dataset (Italy)** | **Third**  **external validation dataset (Radiopaedia)** | **Fourth external validation dataset (China)** | **Fifth**  **external**  **validation**  **dataset**  **(Spain)** |
| **Number of subjects** | 146 | 10 | 20 | 101 | 99 | 9 | 10 | 115 |
| **Age (years)** | 45.9±16.5 | 41.3±9.8 | 41.2±16.3 | N/A | N/A | N/A | N/A | 67.3±15.6 |
| **Gender (M:F)** | 79:46 | 5:3 | 1:1 | N/A | N/A | N/A | N/A | 67:48 |
| **CT parameters** |  |  |  |  |  |  |  |  |
| **Tube voltage (kVp)** | 80-140 | 110-120 | 100-130 | 130 | N/A | N/A | N/A | N/A |
| **Effective mAs (mAs)** | 21-247 | 7-203 | 1-206 | 46-139 | N/A | N/A | N/A | N/A |
| **Slice thickness (mm)** | 0.75-5 | 1-5 | 1-7 | 1 | N/A | N/A | N/A | N/A |
| **CTDIvol (mGy)** | 1.20-16.67 | 3.37-15.63 | 3.52-16.36 | N/A | N/A | N/A | N/A | N/A |

N/A: not available

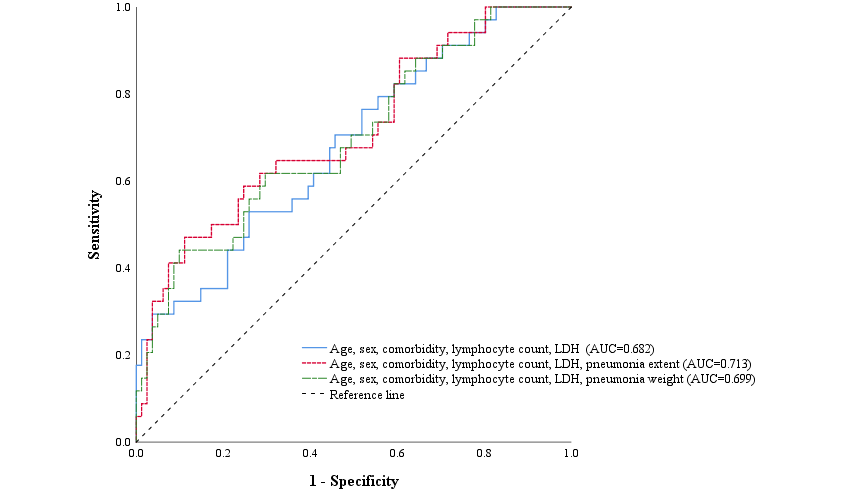
**Supplementary Table 2. Details of the 2D U-Net model and training process**

|  |  |
| --- | --- |
| **A. Model** | |
| Convolution layers (total) | 19 |
| Downsampling block | 2×2 MaxPooling layer, followed by two 3× 3 convolution layers |
| Upsampling block | Bilinear upsampling, concatenation, followed by two 3× 3 convolution layers |
| Activation function | Rectified Linear Unit (RELU) |
| Output activation | Softmax |
| **B. Hyperparameters** | |
| Learning rate | 0.001 |
| Optimizer | Stochastic Gradient Descent |
| Loss | Cross Entropy |
| Batch size | 8 |
| **C. Training** | |
| Epochs | 30 (followed by epoch selection with tuning data) |
| Software | Pytorch |
| Hardware | NVIDIA GeForce GTX 1080 Ti x2 |

**Supplementary Figures**

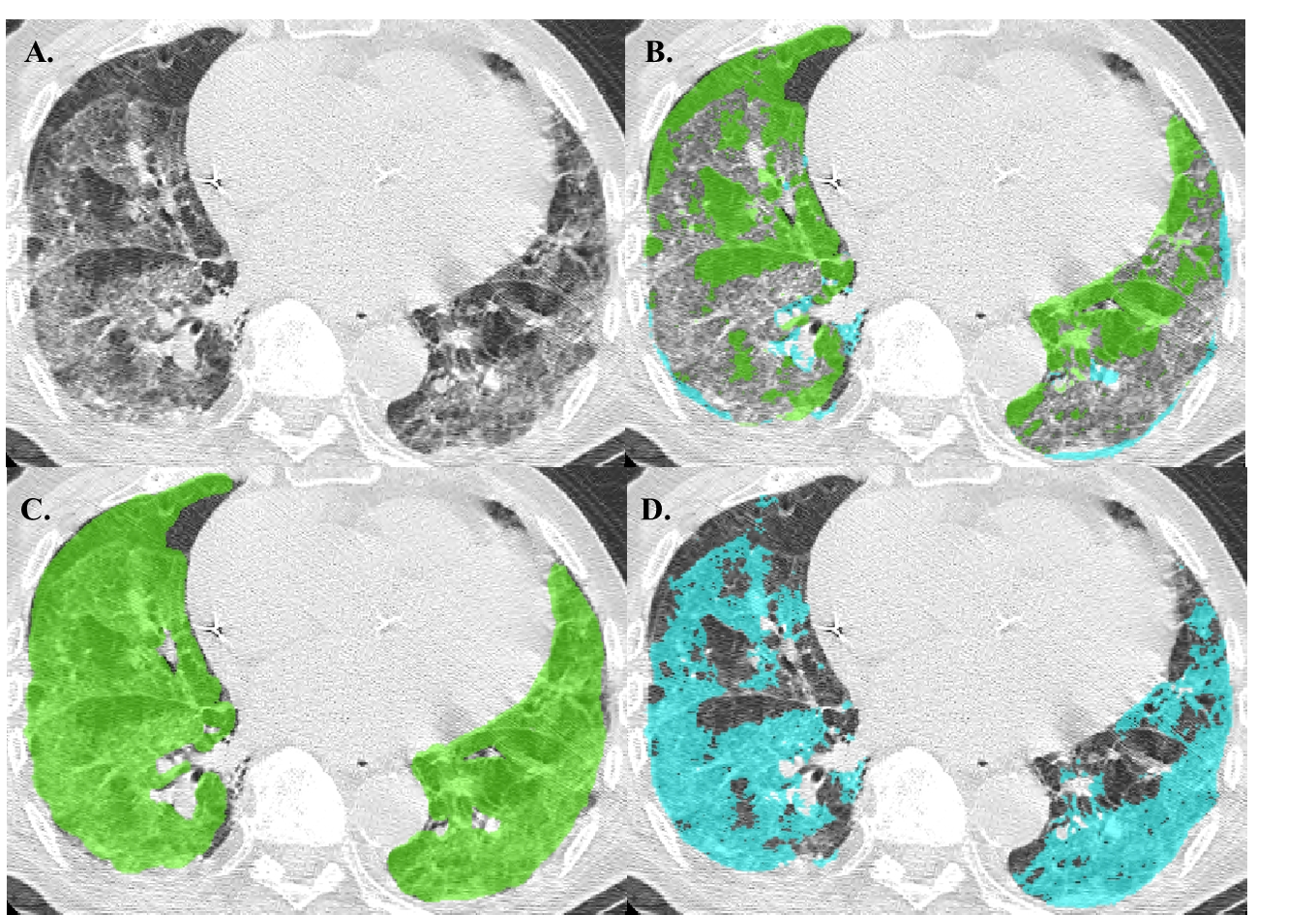
**Supplementary Figure 1. Comparison of Receiver operating characteristic (ROC) curves predicting composite outcome in the Spanish dataset.**

The area under the curve (AUC) was 0.682 for a model using clinical variables (age, sex, comorbidity, lymphocyte count, LDH). AUCs were 0.713 for the model using clinical variables and pneumonia extent and 0.699 for the model using clinical variables and pneumonia weight.

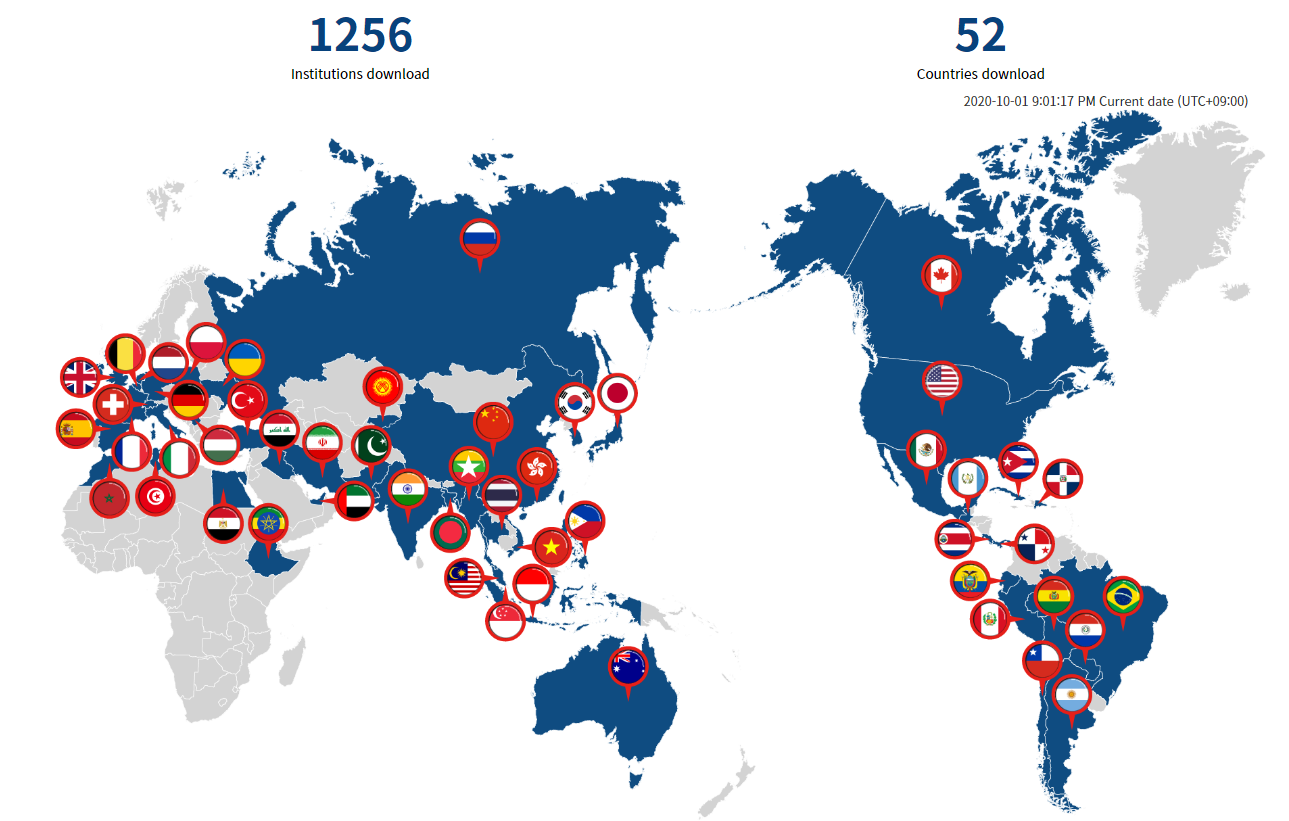


**Supplementary Figure 2. Representative images of a chest CT axial scan image of COVID-19 pneumonia in the Italian external validation dataset.**

A chest CT image shows diffuse ground-glass opacities in both lungs. Although definite ground-glass opacities were included in both the reference mask (C) and our 2D U-Net mask (D), a considerable mismatch is seen in the relatively normal lung areas in the subtracted images (B). The Dice similarity coefficient, sensitivity, and positive predictive value were 68.2%, 54.5%, and 91.2%.



**Supplementary Figure 3. Geographical distribution of institutions where the software was downloaded since March, 18 2020.**

****