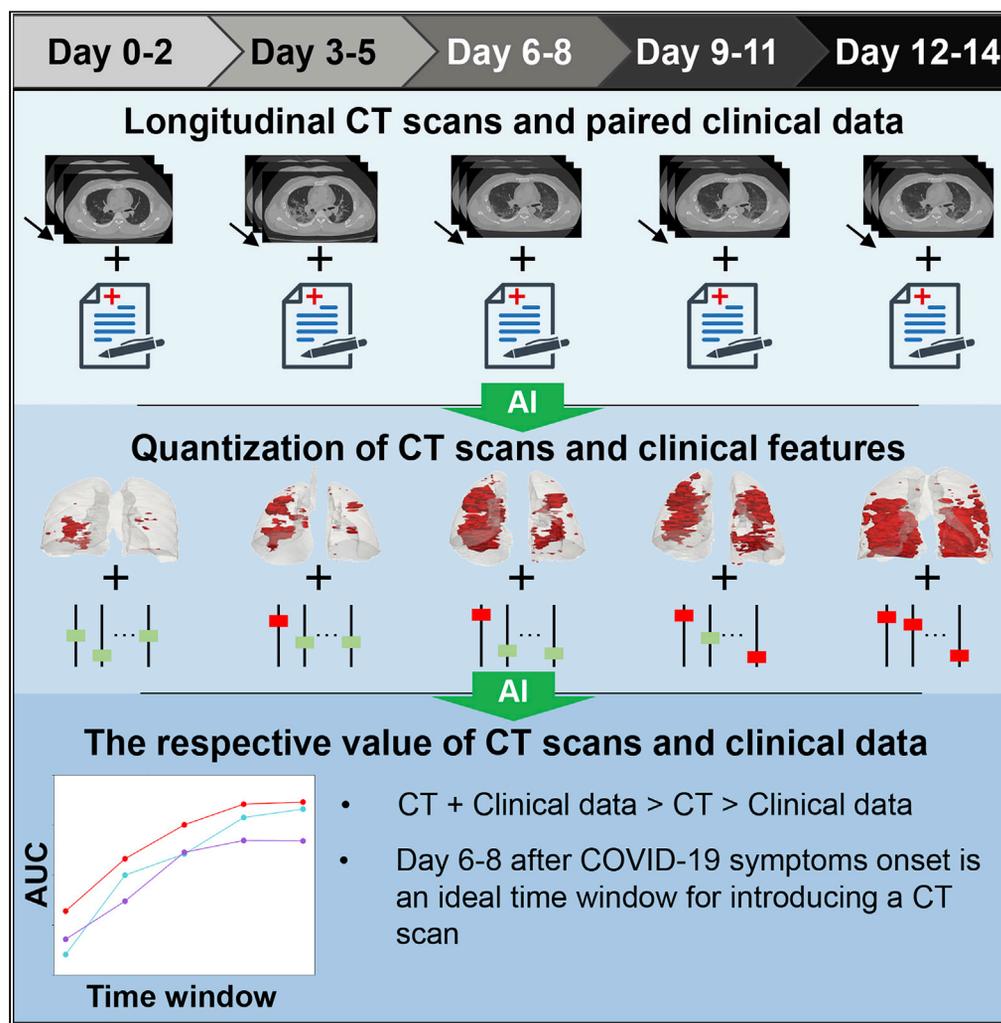


Article

The value of longitudinal clinical data and paired CT scans in predicting the deterioration of COVID-19 revealed by an artificial intelligence system



Xiaoyang Han, Ziqi Yu, Yaoyao Zhuo, ..., Fei Shan, Tingying Peng, Xiao-Yong Zhang

shanfei@shphc.org.cn (F.S.)
tingying.peng@tum.de (T.P.)
xiaoyong_zhang@fudan.edu.cn (X.-Y.Z.)

Highlights

COVID-19 patients with 341 longitudinal CT scans and paired clinical data included

A new AI model for the prediction of COVID-19 progression was developed

CT scans show significant add-on value over clinical data for the prediction

Day 6–8 after the onset of COVID-19 symptoms is an ideal time window for a CT scan



Article

The value of longitudinal clinical data and paired CT scans in predicting the deterioration of COVID-19 revealed by an artificial intelligence system

Xiaoyang Han,^{1,2,11} Ziqi Yu,^{1,2,11} Yaoyao Zhuo,^{3,4,11} Botao Zhao,^{1,2} Yan Ren,⁵ Lorenz Lamm,^{6,7} Xiangyang Xue,⁸ Jianfeng Feng,^{1,2} Carsten Marr,⁶ Fei Shan,^{4,*} Tingying Peng,^{7,*} and Xiao-Yong Zhang^{1,2,9,10,*}

SUMMARY

The respective value of clinical data and CT examinations in predicting COVID-19 progression is unclear, because the CT scans and clinical data previously used are not synchronized in time. To address this issue, we collected 119 COVID-19 patients with 341 longitudinal CT scans and paired clinical data, and we developed an AI system for the prediction of COVID-19 deterioration. By combining features extracted from CT and clinical data with our system, we can predict whether a patient will develop severe symptoms during hospitalization. Complementary to clinical data, CT examinations show significant add-on values for the prediction of COVID-19 progression in the early stage of COVID-19, especially in the 6th to 8th day after the symptom onset, indicating that this is the ideal time window for the introduction of CT examinations. We release our AI system to provide clinicians with additional assistance to optimize CT usage in the clinical workflow.

INTRODUCTION

During COVID-19 pandemic, fast and accurate prediction of disease progression in the early stage is crucial for clinical decision-making and optimal medical resource allocation. Chest computed tomography (CT) plays an important role in evaluating COVID-19 patients by showing specific image features such as ground-glass opacification and consolidation (Wong et al., 2020). Based on CT images or clinical data or both artificial intelligence (AI)-based methods have shown advantages in disease diagnosis and progression prediction (Shi et al., 2020; Wong et al., 2020; Roberts et al., 2021; Zhu et al., 2021). However, most of these AI studies focused on CT or clinical data acquired at a single time point during COVID-19 progression, which limits the prediction accuracy. To address this issue, the evaluation of both longitudinal CT images and clinical data is desirable, because it can help to better understand dynamical changes of pneumonia lesions during COVID-19 deterioration as well as the progression process following therapy, thus resulting in more accurate prognostication of patient outcomes (Feng et al., 2020; Huang et al., 2021).

Longitudinal measurements can provide temporal information that comprehensively reflects dynamic changes of COVID-19 deterioration. So far, only a few studies explored temporal information from longitudinal CT or clinical measurements using deep learning (DL)-based methods (Fang et al., 2021; Kim et al., 2021; Pu et al., 2021; Wang et al., 2021a; Shamout et al., 2021). Zhou et al., (Zhou et al., 2021) used clinical features only to predict COVID-19 severity. Pu et al. (Pu et al., 2021) and Kim et al. (Kim et al., 2021) registered longitudinal CT scans and visualized the difference in diseased vs. healthy areas between scans and found that the CT-based biomarkers may be used to monitor the development of the infection process in COVID-19. Another work (Fang et al., 2021) extracted features directly from sequential CT scans without an extra registration step and is therefore more practical using a recurrent neural network to predict the malignant progression of COVID-19. Shamout, et al., (Shamout et al., 2021) provided visually intuitive saliency maps to help clinicians interpret the model predictions of risk (Heagerty and Zheng, 2005) of deterioration over different time horizons, ranging from 24 h to 96 h with combination of Chest X-rays images and routinely collected clinical variables at only one time point. Wang Robin et al. (Wang et al., 2021a) developed an AI system to predict future deterioration to clinical illness in COVID-19 patients using chest CT and clinical data and proved that the combination of CT and clinical data improves the performance of disease progression prediction.

¹Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China

²Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Shanghai 200433, China

³Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai 200032, China

⁴Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China

⁵Department of Radiology, Huashan Hospital, Fudan University, Shanghai 200433, China

⁶Institute of AI for Health, Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

⁷Helmholtz AI, Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

⁸Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China

⁹MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

¹⁰Lead contact

¹¹These authors contributed equally

*Correspondence: shanfei@shphc.org.cn (F.S.), tingying.peng@tum.de (T.P.), xiaoyong_zhang@fudan.edu.cn (X.-Y.Z.)

<https://doi.org/10.1016/j.isci.2022.104227>



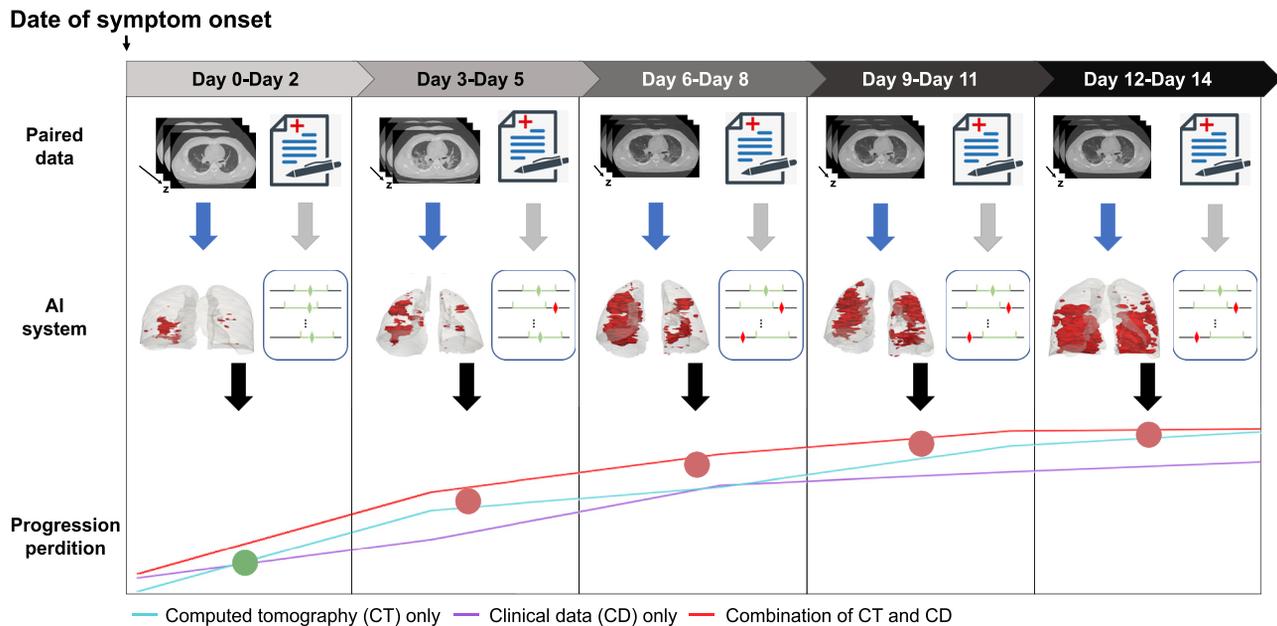


Figure 1. The workflow of our study

CT images and paired clinical data were collected at the same time point. Based on these data, we have developed an AI system to derive a pneumonia progression score by volumetric segmentation of lung (transparent) and COVID-19 lesions (red) and to combine with clinical features to predict whether the patient will progress to a severe status, and thus to reveal the respective value of clinical data and parallel CT scans for comprehensive prediction of COVID-19 progression.

Although the aforementioned studies provided important information for the prediction of COVID-19 progression, the respective values of clinical data and parallel CT scans remain unclear, because longitudinal time points of previously used clinical data and CT scans were limited and not paired. To address this issue based on well-designed longitudinal-paired CT scans and clinical data, we developed a new AI model to predict whether the patient will progress to a severe status and thus to reveal the respective value of clinical data and parallel CT scans for comprehensive prediction of COVID-19 progression reflecting its dynamic deterioration changes (Figure 1).

RESULTS

Characteristics of the study cohort

The 119 patients with 341 longitudinal CT scans and clinical data collected from Shanghai Public Health Clinical Center were divided into two patients' groups (Figure 2): 90 non-progression patients with 273 longitudinal CT scans who were discharged from the hospital in common subtype and 29 progression patients with 68 longitudinal CT scans who progressed into severe or critically ill or died. We summarized the characteristics of clinical data including patient demographics, medical history, presenting signs and symptoms, and laboratory tests of COVID-19 patients (Table 1). The median age of progression patients was higher than that of non-progression patients (63 vs. 55 years, $p < 0.01$). The median number of days from symptom onset to hospital admission was four days (Range: 0–20 days).

Automatic quantization of CT images with BCL-Net

To automatically quantify lung CT images, we designed a deep learning model, termed BCL-Net (Figure 3A). BCL-Net uses a share-weighted convolution operator in the encoding and decoding paths of 2D U-Net to extract intra-slice features. In addition, a bidirectional convolutional LSTM (BC-LSTM) module is applied to integrate cross-plane context. Based on the features extracted from CT images and clinical data, we developed an AI system to predict the progression of COVID-19 using an ensemble learning module (Figure 3B).

Because accurate pneumonia lesion segmentation is the key step for disease progression prediction, we evaluated lesion segmentation performance of BCL-Net along with four state-of-the-art methods on

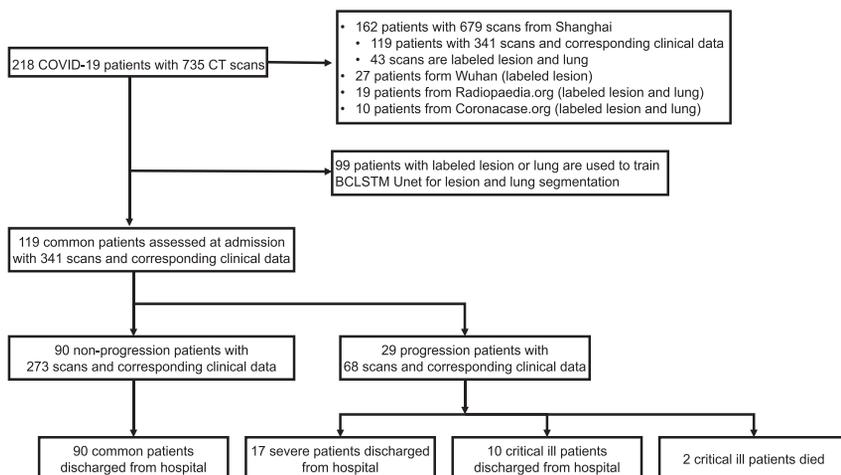


Figure 2. Flowchart of patient selection

A total of 119 with 341 CT scans out of 218 patients were selected for disease progression prediction. See also [Figure S1](#).

four datasets ([Table 2](#)) which are summarized in [Table 3](#). Annotations of lung and COVID-19 lesions of the Coronacases dataset and parts of the Radiopedia dataset are publicly available ([Ma et al., 2021](#)), which we used to train our model and to evaluate it with a five-fold cross-validation. We annotated the remaining data by two experienced radiologists and used it only for testing.

In all four datasets, BCL-Net achieved better DC of 0.840 (95% CI 0.813–0.867) than other methods for lesion segmentation ([Figure 4A and 4B](#); [Table 3](#)), suggesting that BCL-Net is robust for variations in lesion size, CT image intensities, and slice thickness. By contrast, the performance of a full 3D convolution deteriorates with increasing slice thickness ([Figure S2](#)) and could completely fail when the information of slice thickness is missing and the interpolation along the z-axis is not properly done ([Figure 4B](#), nnUNet3D). Moreover, compared to other methods, BCL-Net is three times faster than the second fastest method, Sensor3D, and hundreds of times faster than nnUNet ([Figure 4C](#); [Table 1](#)). This is an important advantage, as local hospitals generally face enormous time pressure when a COVID-19 wave hits a region.

The CT features quantified from lung and pneumonia lesion segmentation are summarized in [Table S2](#). The z-position of progression patients is significantly higher than non-progression patients (53.70 vs 47.05, $p < 0.001$), consistent with previous studies ([Yu et al., 2020](#)).

Prediction of COVID-19 progression using CT and clinical data

To accurately predict the progress of COVID-19 based on a combination of features extracted from CT scans and paired clinical data, we implemented an ensemble learning method to assess whether the patient will develop a severe state ([Figure 5](#)) and achieved a prediction with an AUC of 0.900 (95% CI 0.891–0.909) ([Figure 5A](#)) in a five-fold cross-validation. If we use only the CT data or clinical data, lower AUCs are achieved of only 0.857 (95% CI 0.846–0.868) and 0.844 (95% CI 0.833–0.856) for CT scan-only model and clinical data-only model, respectively. Practically, we may use the prediction of disease progression as guidance for hospital triage to distribute the limited hospital beds to patients who are predicted to have a severe disease progression. For interpretation, the 10 most features that contribute to the prediction were listed ([Figure 5B](#)). In addition to CT-related features and clinical data, age was found to be a key factor leading to different disease progression patterns, consistent with previous studies ([Huang et al., 2020](#); [Wang et al., 2020](#)). The detailed comparison of performance on disease progression prediction is summarized in [Table 4](#).

We visualized the progression of pneumonia in individual patients by plotting the PLV and the average trajectory of three clinical data (including CK, NLR, and LDH) over time ([Figures 5C and 5E](#)) with normal range in orange dotted lines ([Zhou et al., 2021](#)). The progression of pneumonia lesions of a progression patient and a non-progression patient are visualized in [Figure 5D](#). Although individual differences are large,

Table 1. The patient characteristics included in the COVID-19 progression prediction data set

Characteristic	All patients (n = 119, s = 341)	Severe group (n = 29, s = 68)	Non-severe group (n = 90, s = 273)	p values
Age, year	56 [23-84]	63 [30-84]	55 [23-79]	0.0028
Sex, n (%)				0.0977
Male	60 (50.4)	19 (65.5)	41 (45.6)	
Female	59 (49.6)	10 (34.5)	49 (54.4)	
Days from symptom onset to hospital admission, d	4 [0-20]	4 [2-20]	4 [0-20]	0.0916
Comorbidities, n (%)				
Diabetes	9 (7.6)	1 (3.4)	8 (8.9)	0.5756
Hypertension	34 (28.6)	10 (34.5)	24 (26.7)	0.5660
CHD	13 (10.9)	6 (20.7)	7 (7.8)	0.1104
Cerebrovascular disease	1 (0.8)	0 (0.0)	1 (1.1)	0.5488
Hepatitis B	2 (1.7)	0 (0.0)	2 (2.2)	0.9833
Chronic renal disease	1 (0.8)	0 (0.0)	1 (1.1)	0.5488
COPD	3 (2.5)	1 (3.4)	2 (2.2)	0.7529
Total comorbidities number, n (comorbidities per patient)	63 (0.53)	18 (0.62)	45 (0.50)	0.1948
Signs and symptoms, n (%)				
Fever	76 (63.9)	23 (79.3)	53 (58.9)	0.0770
Cough	45 (37.8)	16 (55.2)	29 (32.2)	0.0459
Productive cough	26 (21.8)	10 (34.5)	16 (17.8)	0.1021
Chest tightness	9 (7.6)	1 (3.4)	8 (8.9)	0.5756
Dyspnea	11 (9.2)	5 (17.2)	6 (6.7)	0.1798
Baseline laboratory examinations				
Neutrophil count, $\times 10^9/L$	3.05 [0.70–16.59]	3.97 [1.40–16.59]	2.92 [0.70–8.43]	1.4e-6
Lymphocyte count, $\times 10^9/L$	1.20 [0.30–5.48]	0.82 [0.33–5.48]	1.33 [0.30–4.00]	1.8e-8
NLR	2.38 [0.58–41.47]	3.96 [0.67–41.47]	2.10 [0.58–25.58]	1.0e-12
Lactate dehydrogenase, U/L	249.00 [22.00–911.00]	351.00 [190.00–911.00]	236.00 [22.00–552.00]	4.0e-15
Direct bilirubin, $\mu\text{mol/L}$	4.50 [1.40–24.50]	5.00 [1.70–21.70]	4.30 [1.40–24.50]	0.0058
Creatine kinase, U/L	67.00 [19.00–2925.00]	152.00 [31.00–2925.00]	61.00 [19.00–1398.00]	8.4e-11

Clinical data including patient demographics, medical history, signs and symptoms, and laboratory tests are collected from the Hospital Information System (HIS) at the same time as the CT scan time. The time point of the onset of symptoms and the start time of the severe/critical phase are also recorded for further selection of available CT scans. See also [Table S2](#).

patients who progress to a severe state usually show a sharp increase in the PLV, CK, and LDH in the first week of hospitalization, peaking at around 7–10 days, and then slowly recover in the following days. In contrast, patients who showed only mild symptoms during hospitalization had consistently lower PLV, CK, NLR, and LDH.

Longitudinal comparison of time-dependent disease progression prediction

Furthermore, to reveal the respective values of clinical data and paired CT scans in predicting the deterioration of COVID-19, we divided the data into six groups longitudinally every three days from the onset of COVID-19 symptoms and analyzed them using our AI model. As shown in [Figure 6A](#), the progression prediction model achieves time-dependent AUCs of 0.688, 0.833, 0.901, 0.942, and 0.946 with combination of CT and clinical data over three-day time points in 14 days from symptom onset. The time-dependent AUC of the time group Day 15- reaching 1.000 trained and evaluated on only ten paired CT scans and clinical data (see [Figure S3](#)) was excluded from the discussion for the limitation of data scale. As a comparison, AUCs of the disease progression prediction model are 0.656, 0.800, 0.842, 0.915, and 0.944 using CT

Table 2. A short summary of our multicentre datasets used to develop and evaluate our image segmentation model, BCL-Net

Dataset	Patients	Scans	Slices	Labeled scans	Labeled slices		
					Lesion	Lung	Slice thickness (mm)
Coronacase.org	10	10	2581	10	2581	2581	1.0–1.5
Radiopaedia.org	19	19	1342	19	1342	1342	4.0–6.0
Wuhan	27	27	3805	27	3805	N/A	3.0
Shanghai	162	679	293,349	43	2717	10,007	1.0–6.0
Total	218	735	301,077	99	10,445	13,930	1.0–6.0

Four datasets, Coronacases, Radiopedia, Wuhan and Shanghai were collected independently. Coronacases and Radiopedia were uploaded by individual community users and publicly available. Wuhan and Shanghai were collected from Wuhan Tongji hospital and Shanghai Public Health Clinical Center. Among the datasets, over 10,000 slices labeled by two radiologists were used to train and test BCL-Net and other competitive segmentation models. Moreover, the Shanghai dataset contains longitudinal scans of the same patients which reflect the progression of corona-triggered pneumonia over time.

only. Using only clinical data, the AUCs of the model are 0.672, 0.748, 0.846, 0.869, and 0.890. To more appropriately describe the performance of COVID-19 progression prediction models on the dataset with imbalanced positive and negative samples (273 vs. 68) over three-day time groups, we compared

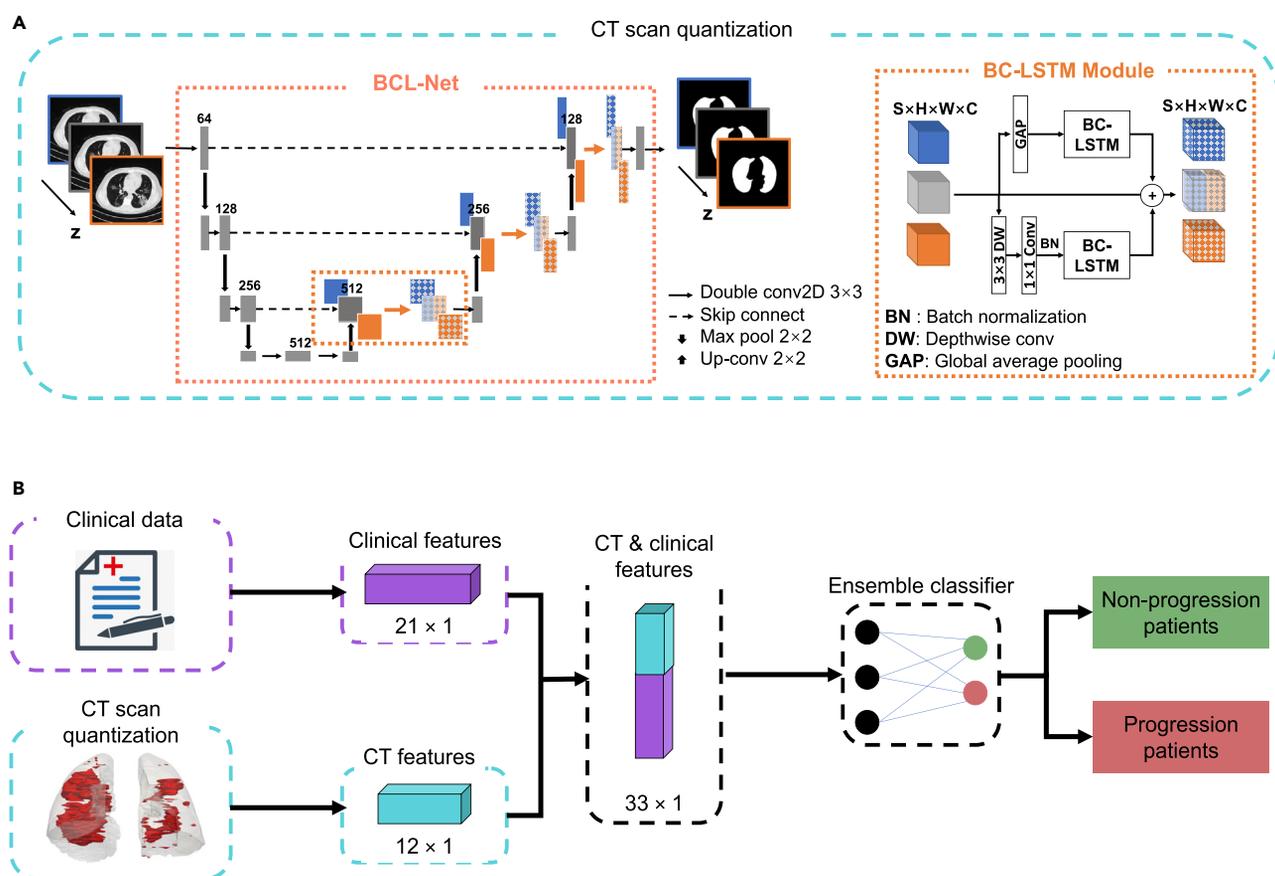


Figure 3. The overview of our proposed AI system

(A) CT features were quantified using our BCL-Net, which combines a 2D U-net to process intra-slice spatial information with an LSTM to leverage inter-slice context. BCL-Net uses the share-weighted 2D convolution in both encoding and decoding paths, avoiding computational expensive 3D convolution. Instead, it uses a BC-LSTM Module (right) to integrate inter-slice context from multiple CT slices. In addition, a dual path attention mechanism including a global average pooling and a depthwise 2D convolution followed by 1×1 convolution is applied to reduce the computation cost.

(B) We implemented an ensemble learning method to predict the disease progression using the paired CT scans and clinical data. 21 clinical measurements were combined with 12 lung and lesion volume CT features to predict the disease progression.

See also Table S1.

Table 3. Comparison of different network architectures and segmentation performances

Method	Multi-slice fusion	Parameter (M)	Inference speed (ms/slice)	Dice coefficient (95% CI)		Additional preprocessing
				Lesion	Lung	
DeepLabV3+	N/A	41.3	85	0.727 (0.650~0.804)	0.978 (0.974~0.983)	N/A
Sensor3D	ConvLSTM	18.7	50	0.653 (0.545~0.760)	0.968 (0.964~0.973)	N/A
nnUNet2D	N/A	17.8	3000	0.780 (0.725~0.835)	0.964 (0.951~0.977)	Interpolation
nnUNet3D	3D convolution kernel	30.4	3200	0.816 (0.785~0.847)	0.911 (0.857~0.966)	Interpolation
BCL-Net (ours)	BC-LSTM Module	19.5	16	0.840 (0.813~0.867)	0.982 (0.979~0.985)	N/A

In addition to BCL-Net, we tested four state-of-the-art segmentation methods, including DeepLabV3+, Sensor3D, nnUNet2D, and nnUNet3D. Among all these methods, DeepLabV3+ and nnUNet2D are purely 2D segmentation methods; nnUNet3D is a full 3D volumetric segmentation method with 3D convolution kernels; both Sensor3D and BCL-Net are hybrid 2D-3D methods, whereas Sensor3D fuses 2D segmentation with convolutional LSTM (ConvLSTM). Compared to other methods, BCL-Net achieves the highest dice score for both lung and lesion segmentation as well as the fastest inference speed. Segmentation performances are presented as dice scores along with 95% CI. Bold numbers indicate the best performance in terms of the corresponding metrics. See also [Figure S2](#).

the changes in F1 scores ([Figure 6B](#)). The progression prediction model achieved time-dependent F1 scores of 0.286, 0.563, 0.783, 0.857, and 0.889 with combination of CT and clinical data over 3-day time groups. With only features extracted from CT, the prediction model achieved F1 scores of 0.286, 0.424, 0.694, 0.815, and 0.889; with clinical data alone, the model achieved 0.250, 0.400, 0.638, 0.815, and 0.750. Time-dependent AUCs and F1 scores with different time group combinations are detailed in [Figure S3](#) and [Table S3](#).

Note that in the very early stage (Day 0–Day 2), the performance of disease progression prediction models is limited with either CT or clinical data alone or even with both. From day 3 to day 14 after the onset of symptoms, compared to using clinical data alone, the introduction of CT examinations significantly increases the prediction accuracy ($p < 0.05$). The add-on value of CT over clinical data is maximized during day 6 to day 8 (AUC improves from 0.833 to 0.901, F1-score from 0.563 to 0.783). These results suggest that the ideal time window for a CT scan is between day 6 and day 8 after the COVID-19 symptom onset.

DISCUSSION

In the current work, based on well-designed longitudinal clinical data and paired CT scans, we developed a new AI system to reveal the respective value of clinical data and parallel CT scans for comprehensive prediction of COVID-19 progression.

The CT examination and clinical data are complementary for the prediction of COVID-19 progression

CT scans and clinical data capture different characteristics of COVID-19 patients, but the complementarities between them have not been fully leveraged. Several machine learning studies have addressed this issue and are attempting to assess the risk of critical illness for COVID-19 patients during the hospital admission ([Liang et al., 2020](#); [Vaid et al., 2020](#); [Wu et al., 2020](#)). Although these studies are very promising to identify patients at high risk, they usually rely on CT scans ([Kim et al., 2021](#); [Wang et al., 2021b](#)) or clinical measurements ([Yan et al., 2020](#); [Zhou et al., 2021](#)) alone, which can lead to incomplete clinical observation. By contrast, as shown in [Figure 7](#), our correlation analysis between CT features and clinical measurements revealed there is a strong correlation between LDH and PLV ($r = 0.68$, $p < 0.01$) and PCV ($r = 0.65$, $p < 0.01$). In addition, the univariate and bivariate distributions and the correlation of the features are shown in [Figure S4](#). Based on the combination of chest CT scans and clinical data, our AI system reaches an AUC score of 0.900, which has been proven outperforming CT scan or clinical data alone. Therefore, our tool is complementary to existing AI tools for COVID-19 triage and prognosis.

The longitudinal measurements provide possibilities for precision medicine

Early prediction of disease deterioration is crucial because it allows for timely and effective treatment, which can potentially improve outcomes and reduce mortality for critically ill COVID-19 patients ([Li et al., 2020](#)). DL methods have the potential to help identify patients at risk for progression to critical illness in an early stage of disease. In contrast to published work, with longitudinal measurements, we found that

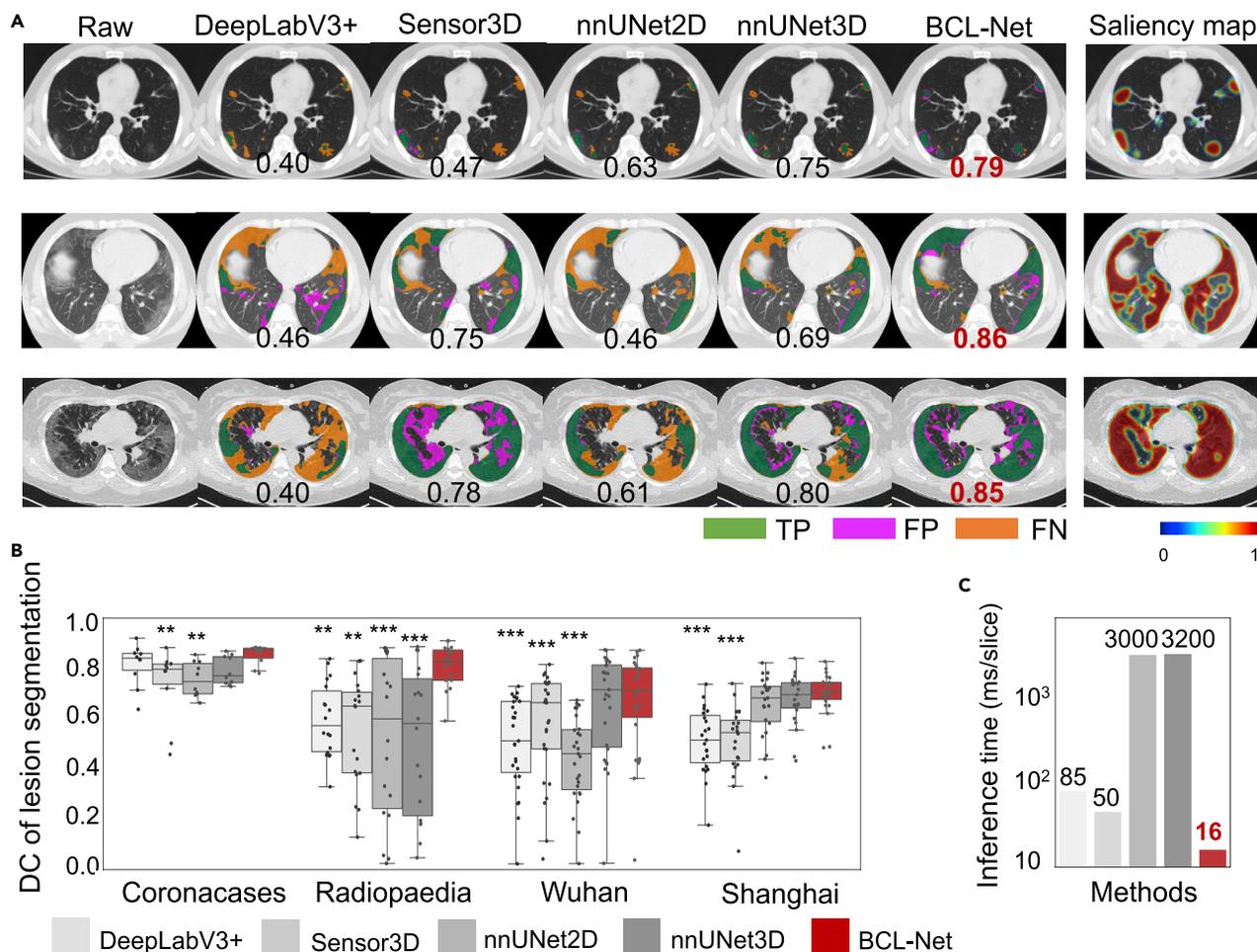


Figure 4. BCL-Net outperforms state-of-the-art methods in terms of accuracy and inference time for lung lesion segmentation

(A) Representative CT images of three COVID-19 patients in mild (top), severe (middle), and critical (bottom) stages and corresponding segmentation results of five methods. BCL-Net achieves highest dice coefficient (DC) in all three cases (red) and robustly segments small lesions in the mild stage, which are often missed by other methods (see false negative (FN) regions marked by orange). The saliency maps highlight the most prominent regions where BCL-Net makes decisions.

(B) BCL-Net achieves significantly higher DC than other methods in all three datasets (** $p < 0.01$, *** $p < 0.001$, Friedmann test with adjusted significance level).

(C) A particular highlight of BCL-Net is its fast speed. With an average inference time of 16 ms/slice, it takes less than 5 seconds to segment a conventional CT scan with about 300 slices in clinics. In comparison, nnUNet3D, because of its computationally expensive interpolation preprocessing step, needs almost 20 min to process the same CT image.

See also [Figure S2](#).

the performance of the predictive model improves as the disease progresses and reaches considerable accuracy after five days. In the very early stage (the first three days from the date of symptom onset), the performance of disease progression prediction is limited with either CT scans or clinical data. The performance of the model using combined features is greatly improved in three to eight days compared to using clinical data alone. Given the prediction accuracy of the combined approach reaches a satisfying F1-score of around 0.80, the most recommended time point for CT scan is between day 5 and day 8 from the date of symptom onset.

The stability and robustness of the model is critical for clinical application

DL-based methods often encounter performance degradation in the multicenter study, mainly because of the large data distribution discrepancy between different cohorts. As shown in [Table 3](#), BCL-Net achieves better Dice scores than other methods for both lung and lesion segmentation in all four cohorts from

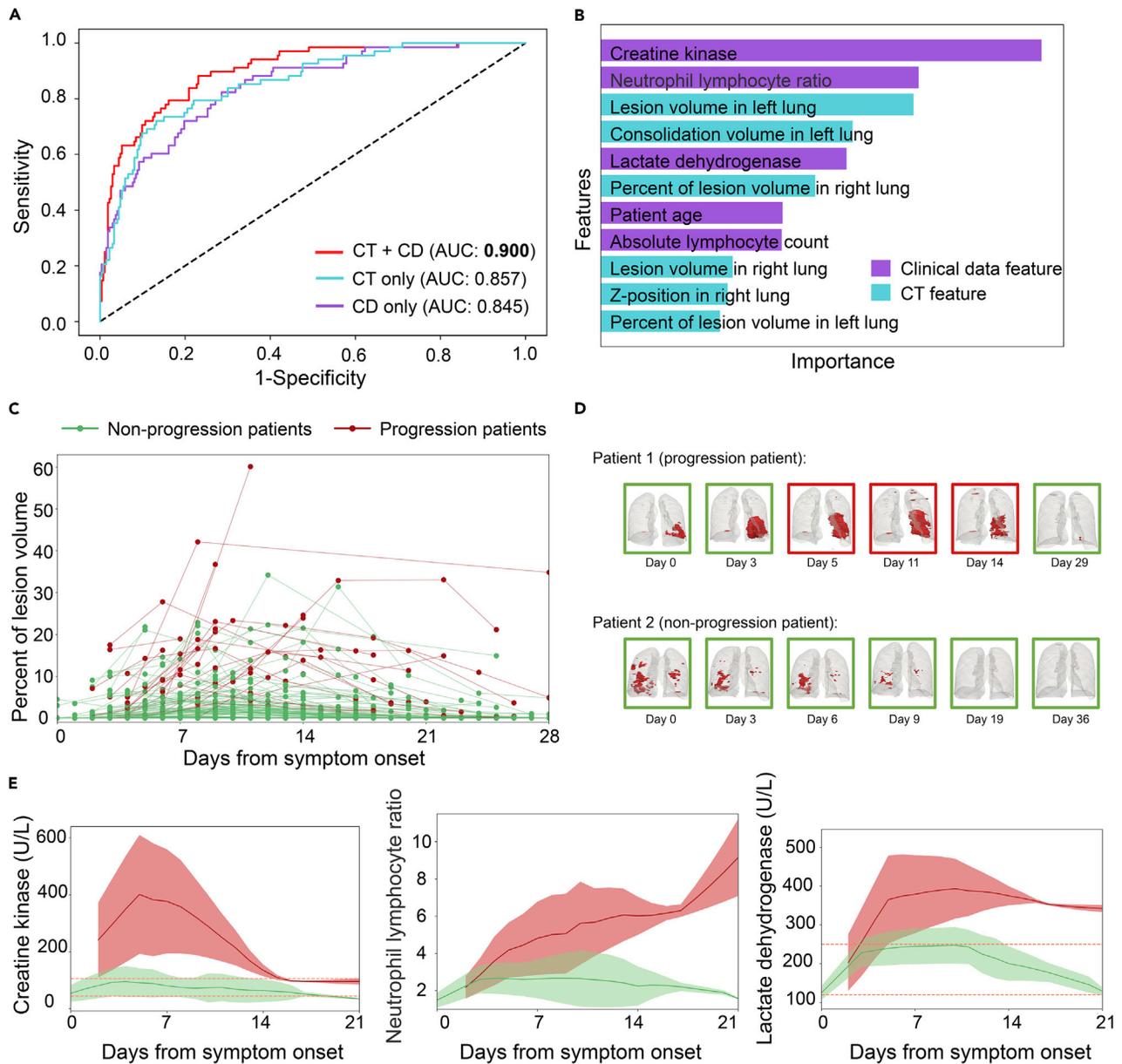


Figure 5. Prediction of COVID-19 progression using CT scan and clinical data at the same time point

(A) Using features extracted from CT scans and clinical data, we can predict whether a patient will develop severe symptoms during hospitalization with an AUC.

(B) Top 10 important features selected by our classifier to distinguish progression patients from non-progression ones.

(C) Temporal trajectories of percent lesion volume (PLV) of individual patients illustrate high variability in the COVID-triggered pneumonia progression. Progression patients tend to have larger lesion volume as compared to non-progression ones. Note, all patients showed mild symptoms when admitted to hospital.

(D) 3D visualization of the pneumonia progression for a representative progression patient and a non-progression patient.

(E) Average trajectories of CK, NLR, and LDH for non-progression vs. progression patients (shaded area represents 25–75% percentiles). The orange dotted lines show the normal ranges of these clinical data. Progression patients generally show a more acute disease progression than non-progression patients in the early stage.

CT, computed tomography; CD, clinical data; CK, Creatine kinase; NLR, Neutrophil lymphocyte ratio; LDH, Lactate dehydrogenase.

multiple centers, suggesting that BCL-Net is robust for variations in lesion size, CT image intensities, and slice thickness. We also found that our new volumetric segmentation algorithm, BCL-Net, increases the prediction accuracy by 6–8% over the current state-of-the-art method, DeepLabV3+ (Zhang et al., 2020).

Table 4. Comparison of performance of machine learning methods and ensemble learning methods using CT data and clinical data

Method	CT data	Clinical data	AUC	F1 score	Accuracy	Sensitivity	Specificity
SVM		✓	0.772 (0.758~0.786)	0.317 (0.282~0.341)	0.798 (0.788~0.808)	0.235 (0.209~0.261)	0.938 (0.931~0.945)
	✓		0.827 (0.813~0.840)	0.447 (0.416~0.477)	0.833 (0.823~0.842)	0.338 (0.314~0.362)	0.956 (0.950~0.962)
	✓	✓	0.853 (0.843~0.864)	0.584 (0.555~0.613)	0.862 (0.854~0.870)	0.485 (0.458~0.513)	0.956 (0.950~0.962)
MLP		✓	0.740 (0.720~0.760)	0.477 (0.452~0.502)	0.801 (0.791~0.810)	0.456 (0.428~0.484)	0.886 (0.878~0.895)
	✓		0.807 (0.795~0.819)	0.543 (0.519~0.567)	0.812 (0.803~0.822)	0.559 (0.533~0.585)	0.875 (0.866~0.885)
	✓	✓	0.817 (0.803~0.830)	0.533 (0.509~0.557)	0.795 (0.786~0.804)	0.588 (0.559~0.617)	0.846 (0.837~0.855)
GB		✓	0.815 (0.802~0.829)	0.534 (0.509~0.559)	0.821 (0.813~0.829)	0.515 (0.490~0.539)	0.897 (0.889~0.906)
	✓		0.858 (0.847~0.869)	0.594 (0.571~0.616)	0.848 (0.838~0.857)	0.559 (0.530~0.588)	0.919 (0.911~0.927)
	✓	✓	0.840 (0.828~0.852)	0.592 (0.569~0.615)	0.850 (0.841~0.860)	0.544 (0.519~0.588)	0.927 (0.919~0.934)
XGBoost		✓	0.840 (0.827~0.853)	0.554 (0.533~0.575)	0.806 (0.798~0.815)	0.603 (0.578~0.628)	0.857 (0.847~0.867)
	✓		0.841 (0.830~0.852)	0.556 (0.533~0.580)	0.804 (0.794~0.813)	0.618 (0.591~0.644)	0.850 (0.842~0.858)
	✓	✓	0.868 (0.858~0.879)	0.653 (0.632~0.674)	0.850 (0.842~0.859)	0.706 (0.680~0.732)	0.886 (0.878~0.895)
Ours		✓	0.844 (0.833~0.856)	0.588 (0.564~0.613)	0.836 (0.827~0.844)	0.588 (0.556~0.621)	0.897 (0.888~0.907)
	✓		0.857 (0.846~0.868)	0.611 (0.588~0.633)	0.850 (0.842~0.859)	0.588 (0.563~0.614)	0.916 (0.908~0.923)
	✓	✓	0.900 (0.891~0.909)	0.691 (0.667~0.714)	0.874 (0.866~0.882)	0.706 (0.681~0.730)	0.916 (0.908~0.923)

We compared our ensemble learning method with four independently trained and evaluated machine learning methods including support vector machine (SVM), multilayer perceptron (MLP), gradient boost (GB), and extreme gradient boosting (XGBoost). The numbers in Bold indicate the best performance. All metrics are presented with 95% CI.

Yet, DeepLabV3 is a pure 2D segmentation algorithm that neglects the rich inter-slice contexts in CT scans. By contrast, BCL-Net is a hybrid 2D-3D segmentation network that combines classical 2D intra-slice feature extraction and a bidirectional ConvLSTM network for inter-slice feature learning, yielding an improved performance. Sensor3D fuses 2D segmentation with ConvLSTM, whereas BCL-Net uses a dual path attention mechanism to reduce the computation cost. On the other hand, compared to full 3D volumetric segmentation such as 3D U-Net, BCL-Net is several orders of magnitude faster, lower in memory consumption, and robust with respect to thick-slice CT as well as thin-slice CT (Figure S2). Hence, it is more generalizable in clinical settings where different CT scanners and computing resources are used.

Limitations of the study

One limitation of this study is that the sample size is not large. Unlike the acquisition of routine clinical data and because of the radiation in CT examinations, unnecessary multiple CT scans are not only harmful to health but also consume excessive medical resources. Therefore, for the same patient, it is often impractical to obtain longitudinal multi-time CT data that exactly matches the clinical data. Fortunately, in this study, we have collected the paired CT and clinical data. We have taken advantage of these data to predict disease progression after the onset of COVID-19 symptoms. The results indicate that the addition of CT examinations between day 6 and day 8 after the onset of symptoms may be ideal for predicting COVID-19 progress. In future work, we will collect more CT data at multiple time points that are paired with clinical data and accurately analyze the predictive value of CT examinations for the prediction of COVID-19 progression and evaluate its necessity in tracking clinical treatment outcome.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Patients' cohort
- METHOD DETAILS

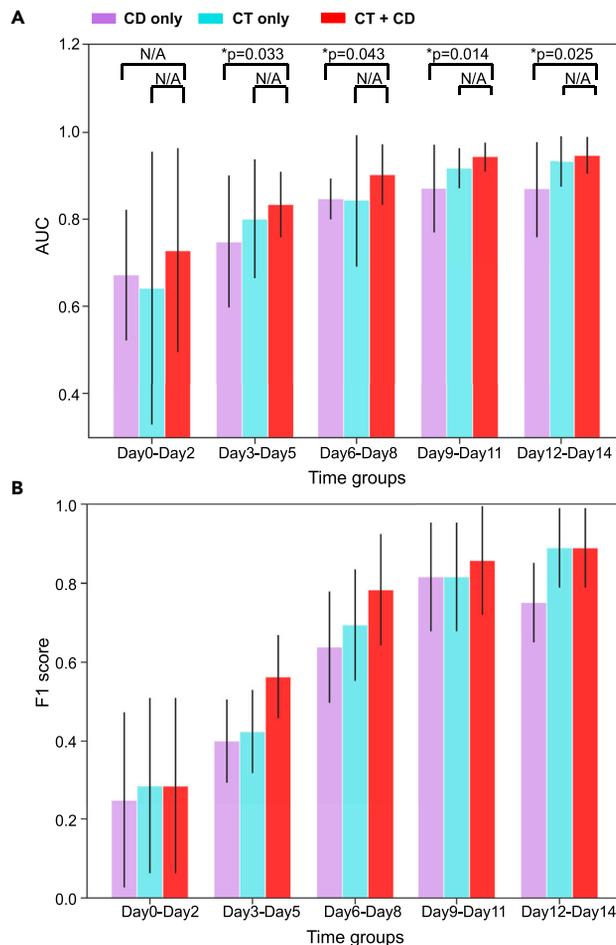


Figure 6. The time-dependent AUCs and F1 scores of the COVID-19 progression prediction model

Data are represented as mean \pm STD.

(A) The time-dependent AUCs of the disease progression prediction model using CT or clinical data or both.

(B) The F1 scores of the disease progression prediction model using CT or clinical data or both. Permutation tests for AUC and F1 score were performed on comparison of combination of CT scan and clinical data vs. CT scan only and combination of CT scan and clinical data vs. clinical data only.

CT, computed tomography; CD, clinical data.

See also [Figure S3](#) and [Table S3](#).

- Overview of our proposed AI system
- Quantization of lung and pneumonia lesion using BCL-Net
- Prediction of disease progression using ensemble learning
- Longitudinal studies for precision medicine
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104227>.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (92043301), Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), ZJLab, Shanghai Center for Brain Science and Brain-Inspired Technology, the Office of Global Partnerships (Key Projects Development Fund) at Fudan University, C.M. and funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (No. 866411).

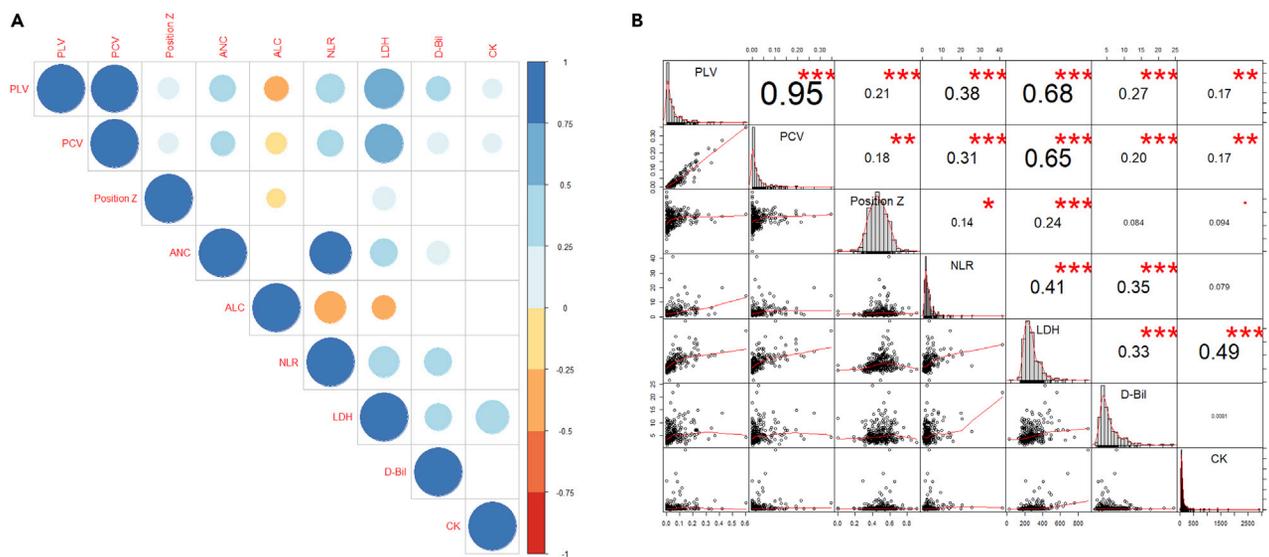


Figure 7. Correlation analysis between CT features and clinical measurements

(A) Correlation diagram heat map of CT features and clinical measurements. Dots are present when $p < 0.01$ and color represent either a positive (blue) or negative (orange) correlation coefficient.

(B) Detailed correlogram of features with significant correlation coefficients ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

PLV, percent lesion volume; PCV, percent consolidation volume; Z-position, related center of the lesion in the z axis; NLR, neutrophil lymphocyte; LDH, lactate dehydrogenase; D-BIL, direct bilirubin; CK, creatine kinase.

See also [Figure S4](#).

AUTHOR CONTRIBUTIONS

X-Y.Z., T.P., and F.S. conceived and designed the study. Y.Z., F.S., and J.F. provided the data. X.H., Z.Y., T.P., and X-Y. Z. analyzed and interpreted the data. X.H., Z.Y., and L.L. verified the underlying data. X.H. and Z.Y. drafted the manuscript, whereas X-Y.Z. and T.P. critically revised the manuscript. C.M., X.X., B.Z., and Y.R. edited the manuscript. All authors approved the final submission of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 17, 2021

Revised: March 10, 2022

Accepted: April 5, 2022

Published: May 20, 2022

REFERENCES

Released by National Health Commission & National Administration of Traditional Chinese Medicine on March 3 (2020). Diagnosis and treatment protocol for novel Coronavirus pneumonia (trial version 7). *Chin. Med. J.* 133, 1087–1095.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Comput. Vis.* 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery (KDD '16))*, pp. 785–794.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49.

Falk, T., Mai, D., Bensch, R., Cicek, O., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jackel, Z., Seiwald, K., et al. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70. <https://doi.org/10.1038/s41592-018-0261-2>.

Fang, C., Bai, S., Chen, Q., Zhou, Y., Xia, L., Qin, L., Gong, S., Xie, X., Zhou, C., Tu, D., et al. (2021). Deep learning for predicting COVID-19 malig-

nant progression. *Med. image Anal.* 72, 102096. <https://doi.org/10.1016/j.media.2021.102096>.

Feng, X., Ding, X., and Zhang, F. (2020). Dynamic evolution of lung abnormalities evaluated by quantitative CT techniques in patients with COVID-19 infection. *Epidemiol. Infect.* 148, e136. <https://doi.org/10.1017/s0950268820001508>.

Heagerty, P.J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* 61, 92–105. <https://doi.org/10.1111/j.0006-341x.2005.030814.x>.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395,

497–506. [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5).

Huang, Y., Li, Z., Guo, H., Han, D., Yuan, F., Xie, Y., Li, Z., Zhang, J., Wang, P., Yang, Y., et al. (2021). Dynamic changes in chest CT findings of patients with coronavirus disease 2019 (COVID-19) in different disease stages: a multicenter study. *Ann. Palliat. Med.* **10**, 572–583. <https://doi.org/10.21037/apm-20-2484>.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., and Maier-Hein, K.H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211. <https://doi.org/10.1038/s41592-020-01008-z>.

Kim, S.T., Goli, L., Paschali, M., Khakzar, A., Keicher, M., Czempliel, T., Burian, E., Braren, R., Navab, N., and Wendler, T. (2021). Longitudinal quantitative assessment of COVID-19 infection progression from chest CTs. Preprint at arXiv, 273–282. [eess.IV]. https://doi.org/10.1007/978-3-030-87234-2_26. <http://arxiv.org/abs/2103.07240>.

Li, L., Li, R., Wu, Z., Yang, X., Zhao, M., Liu, J., and Chen, D. (2020). Therapeutic strategies for critically ill patients with COVID-19. *Ann. Intensive Care* **10**, 45. <https://doi.org/10.1186/s13613-020-00661-z>.

Liang, W., Yao, J., Chen, A., Lv, Q., Zanin, M., Liu, J., Wong, S., Li, Y., Lu, J., Liang, H., et al. (2020). Early triage of critically ill COVID-19 patients using deep learning. *Nat. Commun.* **11**, 3543. <https://doi.org/10.1038/s41467-020-17280-8>.

Liu, F., Zhang, Q., Huang, C., Shi, C., Wang, L., Shi, N., Fang, C., Shan, F., Mei, X., Shi, J., et al. (2020). CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* **10**, 5613–5622. <https://doi.org/10.7150/thno.45985>.

Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., et al. (2021). Toward data-efficient learning: a benchmark for COVID-19 CT lung and infection segmentation. *Med. Phys.* **48**, 1197–1210. <https://doi.org/10.1002/mp.14676>.

Novikov, A.A., Major, D., Wimmer, M., Lenis, D., and Buhler, K. (2019). Deep sequential segmentation of organs in volumetric medical scans. *IEEE Trans. Med. Imaging* **38**, 1207–1215. <https://doi.org/10.1109/tmi.2018.2881678>.

Poyiadji, N., Cormier, P., Patel, P.Y., Hadied, M.O., Bhargava, P., Khanna, K., Nadig, J., Keimig, T., Spizarny, D., Reeser, N., et al. (2020). Acute pulmonary embolism and COVID-19. *Radiology* **297**, E335–E338.

Pu, J., Leader, J.K., Bandos, A., Ke, S., Wang, J., Shi, J., Du, P., Guo, Y., Wenzel, S.E., Fuhrman, C.R., et al. (2021). Automated quantification of COVID-19 severity and progression using chest CT images. *Eur. Radiol.* **31**, 436–446. <https://doi.org/10.1007/s00330-020-07156-2>.

Roberts, M., AIX-COVNET, Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A.I., Etmann, C., McCague, C., Beer, L., et al. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Machine Intelligence* **3**, 199–217. <https://doi.org/10.1038/s42256-021-00307-0>.

Shamout, F.E., Shen, Y., Wu, N., Kaku, A., Park, J., Makino, T., Jastrzebski, S., Witowski, J., Wang, D., Zhang, B., et al. (2021). An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. *NPJ Digital Med.* **4**, 80. <https://doi.org/10.1038/s41746-021-00453-0>.

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., and Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15. <https://doi.org/10.1109/RBME.2020.2987975>.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., and Woo, W.C. (2015). Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems* **28**, C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, and R. Garnett, eds. (Curran Associates, Inc), pp. 802–810.

Vaid, A., Somani, S., Russak, A.J., De Freitas, J.K., Chaudhry, F.F., Paranjpe, I., Johnson, K.W., Lee, S.J., Miotto, R., Richter, F., et al. (2020). Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York city: model development and validation. *J. Med. Internet Res.* **22**, e24018. <https://doi.org/10.2196/24018>.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in wuhan, China. *JAMA* **323**, 1061–1069. <https://doi.org/10.1001/jama.2020.1585>.

Wang, R., Jiao, Z., Yang, L., Choi, J.W., Xiong, Z., Halsey, K., Tran, T.M., Pan, I., Collins, S.A., Feng, X., et al. (2021a). Artificial intelligence for prediction of COVID-19 progression using CT imaging and clinical data. *Eur. Radiol.* **32**, 205–212. <https://doi.org/10.1007/s00330-021-08049-8>.

Wang, X., Jiang, L., Li, L., Xu, M., Deng, X., Dai, L., Xu, X., Li, T., Guo, Y., Wang, Z., et al. (2021b). Joint learning of 3D lesion segmentation and classification for explainable COVID-19 diagnosis. *IEEE Trans. Med. Imaging* **40**, 2463–2476. <https://doi.org/10.1109/tmi.2021.3079709>.

Wong, H.Y.F., Lam, H.Y.S., Fong, A.H.T., Leung, S.T., Chin, T.W.Y., Lo, C.S.Y., Lui, M.M.S., Lee, J.C.Y., Chiu, K.W.H., Chung, T.W.H., et al. (2020). Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology* **296**, E72–E78. <https://doi.org/10.1148/radiol.2020201160>.

Wu, G., Yang, P., Xie, Y., Woodruff, H.C., Rao, X., Guiot, J., Frix, A.N., Louis, R., Moutschen, M., Li, J., et al. (2020). Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicenter study. *Eur. Respir. J.* **56**, 2001104. <https://doi.org/10.1183/13993003.01104-2020>.

Yan, L., Zhang, H.T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., et al. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nat. Machine Intelligence* **2**, 283–288. <https://doi.org/10.1038/s42256-020-0180-7>.

Yu, Q., Wang, Y., Huang, S., Liu, S., Zhou, Z., Zhang, S., Zhao, Z., Yu, Y., Yang, Y., and Ju, S. (2020). Multicenter cohort study demonstrates more consolidation in upper lungs on initial CT increases the risk of adverse clinical outcome in COVID-19 patients. *Theranostics* **10**, 5641–5648. <https://doi.org/10.7150/thno.46465>.

Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., Zha, Y., Liang, W., Wang, C., Wang, K., et al. (2020). Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **182**, 1360. <https://doi.org/10.1016/j.cell.2020.08.029>.

Zhou, K., Sun, Y., Li, L., Zang, Z., Wang, J., Li, J., Liang, J., Zhang, F., Zhang, Q., Ge, W., et al. (2021). Eleven routine clinical features predict COVID-19 severity uncovered by machine learning of longitudinal measurements. *Comput. Struct. Biotechnol. J.* **19**, 3640–3649. <https://doi.org/10.1016/j.csbj.2021.06.022>.

Zhu, X., Song, B., Shi, F., Chen, Y., Hu, R., Gan, J., Zhang, W., Li, M., Wang, L., Gao, Y., et al. (2021). Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan. *Med. Image Anal.* **67**, 101824. <https://doi.org/10.1016/j.media.2020.101824>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Radiopaedia.org	https://radiopaedia.org/	N/A
Coronacases	https://coronacases.org/	N/A
Wuhan	This paper (not shared)	N/A
Shanghai	This paper (not shared)	N/A
Software and algorithms		
Python	https://www.python.org/	Version 3.7
Tensorflow	https://www.tensorflow.org/	Version 1.15.4
Keras	https://keras.io/	Version 2.24
Scikit-learn	https://scikit-learn.org/	Version 0.21.3
Scikit-image	https://scikit-image.org/	Version 0.14
XGBoost	https://github.com/dmlc/xgboost	Version 1.1.0
SimpleITK	https://simpleitk.org/	Version 2.0
Scipy	https://scipy.org/	Version 1.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Xiao-Yong Zhang (xiaoyong_zhang@fudan.edu.cn).

Materials availability

This study did not generate any new unique materials.

Data and code availability

- Partial data reported in this paper will be shared by the [lead contact](#) upon request. The longitudinal data reported in this study cannot be deposited in a public repository because patients do not want their data to be made public.
- All original code has been deposited at <https://robin970822.github.io/DABC-Net-for-COVID-19/> and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients' cohort

This longitudinal cohort study was approved by the Ethics Committee of Shanghai Public Health Clinical Center (YJ-2020-S035-01). [Figure 2](#) illustrates the flowchart of the patient selection. As shown in [Table 2](#), a total number of 218 COVID-19 patients with 735 CT scans were collected from January 21, 2020 to April 29, 2020. From public datasets available online, we obtained 27 CT scans with lesion labels from Wuhan, 19 CT scans with lesion and lung labels from [Radiopaedia.org](#) and 10 CT scans with lesion and lung labels from [Coronacase.org](#). From Shanghai Public Health Clinical Center, we obtained 679 CT scans from 162 patients including 43 CT scans with lesion and lung labels and 341 longitudinally measured CT scans and corresponding clinical data from 119 patients. The CT scans from Shanghai Public Health Clinical Center were obtained on dedicated CT scanners (Hitachi, Philips and UIH) with the following parameters: slice thickness 1-6 mm, slice gap 0 mm. All CT scans with lesion and lung labels were included to segment the lesion and lung volume on a deep learning method. 119 patients with 341 longitudinally measured

CT scans and paired clinical data from Shanghai Public Health Clinical Center were enrolled to build an AI system to predict the disease progression and investigate the respective value of clinical data and CT scans.

The longitudinal clinical data including patient demographics, medical history, presenting signs and symptoms and laboratory examinations were all collected from the hospital information system (HIS). The laboratory examinations including absolute neutrophil count (ANC), absolute lymphocyte count (ALC), neutrophil lymphocyte ratio (NLR), lactate dehydrogenase (LDH), direct bilirubin (D-BIL), and creatine kinase (CK) were measured and matched with CT scans. The time points of symptoms onset and the beginning of the severe/critical stage were recorded for the time alignment.

According to the guidelines of national diagnosis and treatment protocols for COVID-19 (the seventh version) ([Released by National Health Commission & National Administration of Traditional Chinese Medicine on March 3, 2020](#), 2020), patients with COVID-19 can be divided into four subtypes: mild, common, severe, and critically ill. The progress of the severe illness with COVID-19 is usually rapid and there is no clear separation between the severe illness and the critical illness ([Li et al., 2020](#)). Therefore, patients deteriorating into these two subtypes were combined to be the progression patients. The 119 patients who were common subtypes at admission were divided into two groups, that is, 90 non-progression patients with 273 paired CT scans and clinical data who were discharged from hospital in the common subtype, and 29 progression patients with 68 paired CT scans and clinical data who deteriorated into severe or critically ill subtype or died.

METHOD DETAILS

Overview of our proposed AI system

An AI system ([Figure 3](#)) was used to predict the disease progression with paired clinical data and CT scans at each time point for each patient. CT features are obtained by BCL-Net. Both clinical features (patient demographics, medical history, signs and symptoms, and laboratory tests) and CT features are combined and fed into the AI system for the prediction.

Quantization of lung and pneumonia lesion using BCL-Net

As a hybrid 2D-3D network, BCL-Net combines a U-shaped network (UNet) ([Çiçek et al., 2016](#); [Falk et al., 2019](#)) with shared-weighted encoder and decoder to process in-plane context and a BC-LSTM Module that uses bidirectional convolutional LSTM to integrate cross-plane context ([Figure 3B](#)).

BCL-Net uses a share-weighted convolution operator in the encoding and decoding paths of 2D U-Net to extract intra-slice features. More specifically, each convolution and transposed convolution block consists of two 3×3 shared convolution filters followed by a 2×2 max pooling layer and ReLU function. These parameters are shared among all input slices. In order to process the cross-plane context along the z-axis and retain the original high-resolution xy in-plane information, a convolutional long-short term memory (C-LSTM) ([Shi et al., 2015](#)) is used to integrate in-plane features extracted by 2D U-Net. Unlike temporal sequential data (e.g., video clips), where information flows in only one forward direction, structural CT scans have two orientations that need to be considered. Hence, a bidirectional C-LSTM module (BC-LSTM Module) is used to model both forward and backward information.

To reduce computational cost, we propose a dual path attention mechanism including a global average pooling (GAP) to generate channel-wise maps and a channel-dimension squeeze procedure with a depth-wise convolution followed by 1×1 convolution to generate spatial-wise maps. BC-LSTM only fuses the information distilled by the channel-wise maps and the spatial-wise maps and reduces the computation cost.

We trained two BCL-Nets, respectively for lung and pneumonia lesion segmentation. We further multiplied the output of lesion BCL-Net with the corresponding output of lung BCL-Net to remove possible false positive lesions outside the lung organ. Consolidations were found more common in patients >50 years old ([Poyiadji et al., 2020](#)) and could be a warning sign of severe progression. Therefore, we further outlined the consolidation region thresholding at 0.5 on normalized intensity within the lesion region ([Liu et al., 2020](#)). Additionally, we calculated the weighted volume from the inner product of the lesion and the intensity, determining the center of the lesion in the z-axis ([Yu et al., 2020](#)). In all, 12 AI-derived CT features including lung volume (LuV), lesion volume (LeV), consolidation volume (CV), percentage of lesion (PLV)

and consolidation volume (PCV), and the related center of the lesion in the z-axis (z-position) in left and right lung were obtained.

We compared BCL-Net with four state-of-the-art medical segmentation methods, namely DeepLabV3+ (Chen et al., 2018), Sensor3D (Novikov et al., 2019), nnUNet2D & nnUNet3D (Isensee et al., 2021). Table 3 summarizes the individual features of all five methods including BCL-Net in terms of model architecture, parameter size and inference speed.

Prediction of disease progression using ensemble learning

As illustrated in Figure 3A, we implemented an ensemble learning method to predict the disease progression using a total of 33 features including 12 AI-derived features extracted from CT scans and paired 21 clinical measurements collected from HIS.

The ensemble learning method was implemented in Python scikit-learn library with base learners consisting of support vector machine (SVM), k-nearest neighbors (KNN), naive Bayes (NB), multilayer perceptron (MLP), random forest (RF), gradient boost (GB), logistic regression (LR), adaptive boosting (Adaboost), and extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016). The averaged output of base learners was calculated as the final output for prediction of disease prediction. Table S1 summarizes the implementation details of base learners. We compared our ensemble learning method with four independently trained and evaluated machine learning methods including SVM, MLP, GB, and XGBoost in a five-fold cross-validation. In addition to prediction, we also highlighted important features that mostly contribute to our prediction using the average feature importances from base learners including RF, GB, Adaboost, and XGBoost.

Longitudinal studies for precision medicine

Given the fact that patients from Shanghai Public Health Clinical Center were under daily medical observation and scanned for CT about every 3 days (Figure S1), we divided 341 paired CT and clinical data in six three-day time groups, that is 27, 74, 79, 73, 78, and 10 paired CT and clinical data in group Day0-Day2 (within two days from symptom onset), group Day3-Day5 (3-5 days), group Day6-Day8 (6-8 days), group Day9-Day11 (9-11 days), group Day12-Day14 (12-14 days), and group Day15- (more than 15 days from symptom onset), separately. We compare the performance changes over the different three-day time groups of the model trained with a combination of CT and clinical data at the same time-point, the model with only CT and the model with only clinical data.

Using our proposed AI system (Figure 3), the respective value of longitudinal clinical data and paired CT scans in predicting the deterioration of COVID-19 was evaluated with AUC (area under the receiver operating characteristic curve, ROC) and F1 score in a five-fold cross-validation in each time group except group Day0-2 and Day15- in a three-fold cross-validation owing to the insufficiency of progression patients.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of features extracted from CT scans and clinical characteristics is performed in the Python statsmodel library. Continuous variables are represented by median and range, and a comparison between non-progression vs progression patients is performed using the Student's t test for normally distributed variables and the Mann-Whitney U statistics for non-normally distributed variables. Categorical variables are represented as numbers with percentages, and a comparison between patient groups is performed by chi-square. The performances of lung and pneumonia lesion segmentation are evaluated in dice coefficient (DC) and compared with four state-of-the-art in the Friedmann test. The performances of disease progression prediction are evaluated in AUCs and F1 scores and compared in the permutation test. All the metrics of the AI system are presented with 95% confidence interval (CI).