



Machine Learning-Based Prediction of COVID-19 Severity and Progression to Critical Illness Using CT Imaging and Clinical Data

Subhanik Purkayastha^{1*}, Yanhe Xiao^{2*}, Zhicheng Jiao³, Rujapa Thepumnoeysuk¹, Kasey Halsey^{1,4}, Jing Wu², Thi My Linh Tran^{1,4}, Ben Hsieh^{1,4}, Ji Whae Choi^{1,4}, Dongcui Wang², Martin Vallières⁵, Robin Wang³, Scott Collins¹, Xue Feng⁶, Michael Feldman⁷, Paul J. Zhang⁷, Michael Atalay¹, Ronnie Sebro³, Li Yang², Yong Fan³, Wei-hua Liao², Harrison X. Bai^{1,4}

¹Department of Diagnostic Imaging, Rhode Island Hospital, Providence, RI, USA; ²Department of Radiology, Xiangya Hospital, Central South University, Changsha, China; ³Department of Radiology, Center for Biomedical Image Computation and Analytics, University of Pennsylvania, Philadelphia, PA, USA; ⁴Warren Alpert Medical School at Brown University, Providence, RI, USA; ⁵Department of Computer Science, Université de Sherbrooke, Sherbrooke, Canada; ⁶Carina Medical, Lexington, KY, USA; ⁷Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA

Objective: To develop a machine learning (ML) pipeline based on radiomics to predict Coronavirus Disease 2019 (COVID-19) severity and the future deterioration to critical illness using CT and clinical variables.

Materials and Methods: Clinical data were collected from 981 patients from a multi-institutional international cohort with real-time polymerase chain reaction-confirmed COVID-19. Radiomics features were extracted from chest CT of the patients. The data of the cohort were randomly divided into training, validation, and test sets using a 7:1:2 ratio. A ML pipeline consisting of a model to predict severity and time-to-event model to predict progression to critical illness were trained on radiomics features and clinical variables. The receiver operating characteristic area under the curve (ROC-AUC), concordance index (C-index), and time-dependent ROC-AUC were calculated to determine model performance, which was compared with consensus CT severity scores obtained by visual interpretation by radiologists.

Results: Among 981 patients with confirmed COVID-19, 274 patients developed critical illness. Radiomics features and clinical variables resulted in the best performance for the prediction of disease severity with a highest test ROC-AUC of 0.76 compared with 0.70 (0.76 vs. 0.70, $p = 0.023$) for visual CT severity score and clinical variables. The progression prediction model achieved a test C-index of 0.868 when it was based on the combination of CT radiomics and clinical variables compared with 0.767 when based on CT radiomics features alone ($p < 0.001$), 0.847 when based on clinical variables alone ($p = 0.110$), and 0.860 when based on the combination of visual CT severity scores and clinical variables ($p = 0.549$). Furthermore, the model based on the combination of CT radiomics and clinical variables achieved time-dependent ROC-AUCs of 0.897, 0.933, and 0.927 for the prediction of progression risks at 3, 5 and 7 days, respectively.

Conclusion: CT radiomics features combined with clinical variables were predictive of COVID-19 severity and progression to critical illness with fairly high accuracy.

Keywords: COVID-19; Machine learning; CT; Radiomics; Severity

Received: September 6, 2020 **Revised:** January 4, 2021 **Accepted:** January 6, 2021

This study was supported by the Brown COVID-19 Research Seed Award, the Amazon Web Services Diagnostic Development Initiative, RSNA Research Scholar Grant, and National Institutes of Health/National Cancer Institute R03 grant (R03CA249554) to Dr. Harrison X. Bai, as well as NIH grants (CA223358, DK117297, MH120811, and EB022573) to Dr. Yong Fan.

*These authors contributed equally to this work.

Corresponding author: Wei-hua Liao, MD, Department of Radiology, Xiangya Hospital, Central South University, 932 Lushan S Rd, Yuelu District, Changsha 410008, China.

• E-mail: owenliao@csu.edu.cn; and

Harrison X. Bai, MD, Department of Diagnostic Imaging, Rhode Island Hospital and Warren Alpert Medical School of Brown University, Providence, 222 Richmond St, Providence, RI 02903, USA.

• E-mail: harrison_bai@brown.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Coronavirus Disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), a virus that can precipitate pneumonia, acute respiratory distress syndrome, and subsequent death [1]. In addition to pulmonary complications, which often require intubation, mechanical ventilation, and intensive care unit (ICU) level of care to treat, COVID-19 may also be associated with a host of other symptoms, including cardiovascular [2], neurological [3], hepatic, renal, olfactory, gustatory, ocular, cutaneous, and hematological manifestations [4]. Since the disease can cause multi-organ sequelae and death, COVID-19 patients who have a poor prognosis and are likely to deteriorate to critical status need to be identified promptly. In addition, it is difficult for medical systems to accommodate the high prevalence of critically ill patients with COVID-19. This disease has been detrimental to medical resource availability [5].

Early intervention has been shown to reduce mortality in COVID-19 patients [6]. When providers are aware of a patient's potential deterioration, they can promptly obtain an ICU bed, acquire a mechanical ventilator, and consider initiating experimental COVID-19 treatments [7]. Clinical data, including symptoms of fever, cough, and dyspnea, as well as laboratory findings, such as lymphopenia, elevated inflammatory markers, and atypical coagulation factor tests, have all been useful for the diagnosis and prognostic predictions of COVID-19 [8-10]. However, several of these signs and symptoms are non-specific for COVID-19 pneumonia [8].

Medical imaging, specifically chest CT, is a more specific modality for the diagnosis of COVID-19 [11], and it also has the potential to aid in predicting the severity of COVID-19 [10,12]. Patients with COVID-19 can show characteristic signs on chest CT, such as multi-focal ground-glass opacities (GGO) and consolidation with bilateral and multi-lobe involvement typically localized to the lower lung [13,14]. These findings also seem to be time-dependent [14], which can further aid disease assessment and prognosis. Artificial intelligence (AI) can recognize features and patterns that are not easily discernible to the human eye, and it has been used to improve the diagnostic and prognostic accuracy of chest CT for COVID-19 [15-21]. In this study, a machine learning (ML) pipeline was developed to predict COVID-19 disease severity and the risk of progression to critical illness within specific time

intervals using chest CT and clinical data.

MATERIALS AND METHODS

Patient Cohorts

A total of 981 patients with COVID-19 confirmed by RT-PCR and chest CT imaging suggestive of pneumonia were retrospectively identified based on data from nine hospitals in the Hunan Province in China, the Hospital of the University of Pennsylvania in Philadelphia in Philadelphia in PA, the Rhode Island Hospital in Providence in RI, and open-source data from a previously published paper [16]. The CT scans of the identified patients were directly downloaded from the hospital Picture Archiving and Communications System and reviewed by a radiologist. Publicly available chest CT images and clinical metadata of COVID-19 patients were directly downloaded from the China National Center for Bioinformatics website. A diagram illustrating patient inclusion and exclusion criteria is shown in Figure 1.

This data for the cohort were randomly divided into training, validation, and testing sets with a 7:1:2 split ratio to build the severity and progression prediction models.

Clinical Information

The patient data on demographics and co-morbidities were retrospectively collected. The patients' conditions were determined to be critical or severe if they reached any of the following endpoints: ICU admission, mechanical ventilation, or death. If not, their conditions were non-critical or non-severe. For critical or severe patients, the duration for their progression to critical events was calculated from the time of CT to the earliest time of developing one of the aforementioned critical events. A plot of the time distribution from CT and critical outcomes is shown in Supplementary Figure 1.

The patient data, including age, sex, symptoms (presence or absence of fever), white blood cell count, lymphocyte count, comorbidity status (cardiovascular disease, hypertension, chronic obstructive pulmonary disease, diabetes, chronic liver disease, chronic kidney disease, cancer, and human immunodeficiency virus), and history of exposure to the COVID-19 epicenter and/or another patient with COVID-19, were collected. These were used as the 15 clinical variables for model training. The use of mechanical ventilation, ICU care, and progression to death was also recorded. For all patients, admission and discharge

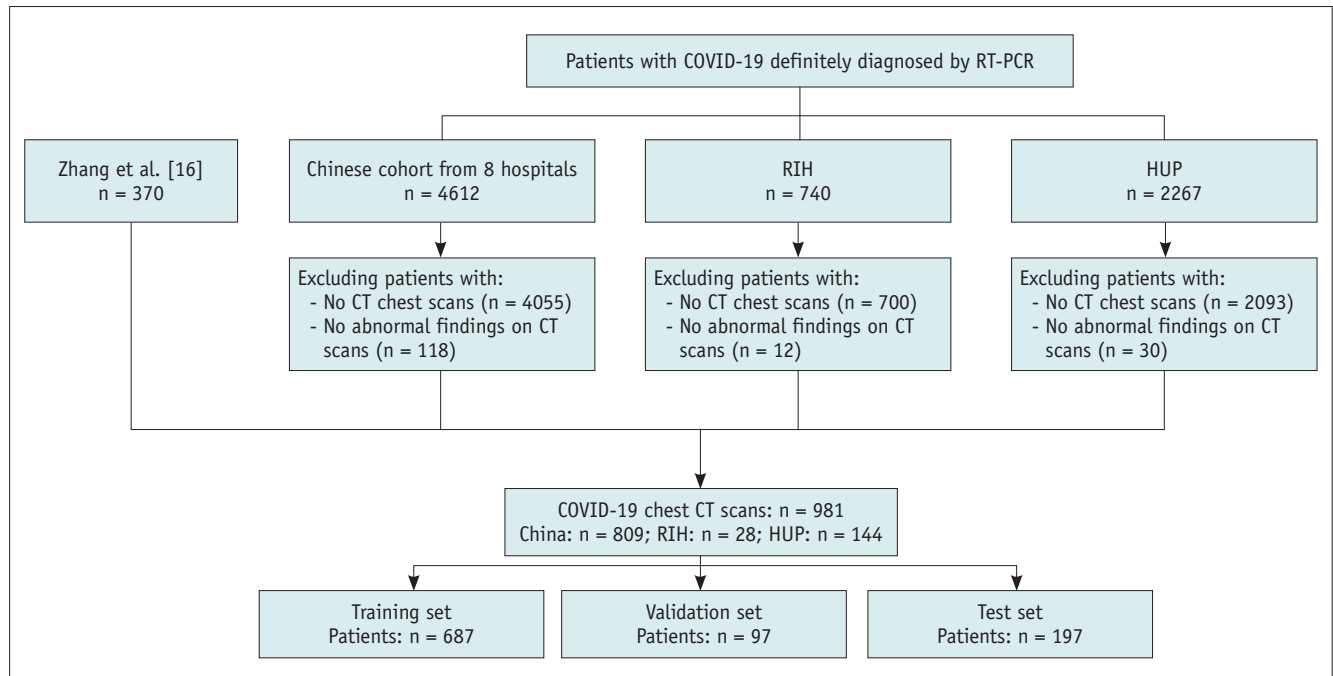


Fig. 1. Illustration of patient inclusion and exclusion. Adapted from Zhang et al. Cell 2020;181:1423-1433.e11 [16]. HUP = Hospital of the University of Pennsylvania, RIH = Rhode Island Hospital, RT-PCR = reverse transcriptase-polymerase chain reaction

times were also recorded. The missing values were imputed in groups using K-nearest neighbors (KNN) and iterative imputation methods [22,23]. A comparison of clinical data across institutions is shown in Supplementary Table 1.

Machine Learning Pipeline

First, the lung tissues and abnormalities caused by COVID-19 were automatically segmented on CT images using a deep-learning model based on a deep convolutional neural network. Second, radiomics features were extracted from CT, and an ML pipeline utilizing both image-based and clinical variables were used to predict a patient's COVID-19 severity and progression to critical illness at the time of the CT scan. An illustration of our workflow is provided in Figure 2.

Visual CT Severity Scoring

Chest CT scans were assessed using a scoring system adopted for convalescent patients after severe acute respiratory syndrome, as introduced by Chang et al. [24]. The severity scores range from 0 to 5 for each lung lobe depending on the extent of GGO (0 = no involvement, 1 = < 5%, 2 = 5–25%, 3 = 26–49%, 4 = 50–75%, 5 ≥ 75% involvement). The values for each lobe were summed to determine a final score ranging from 0 to 25. The scores were summed for each patient to represent a visual CT consensus severity score. All CT scans included in the study

were divided into two parts. Each half was assessed by two independent radiologists in consensus. They have practiced 5–10 years of thoracic radiology and have had direct clinical experience with COVID-19 chest CT scans. We chose this scoring system since it has been used in numerous studies on COVID-19. For example, a recent study used this method for the first pulmonary CT scans that were obtained at a mean of 2 ± 2 days after the onset of symptoms [25]. Another study utilized this scoring system to analyze a group of CT scans obtained within a mean of 2.2 ± 1.8 days and another group obtained within a mean of 6.6 ± 4.0 days [26]. Furthermore, the scoring system by Chang et al. [24] was successfully adapted for another chest CT severity score for assessing the severity of COVID-19 [12].

Severity Prediction

Radiomics features were extracted from the patients' CT scans. For each image space, 79 non-texture (morphology and intensity-based) and 94 texture features were extracted according to the guidelines by the Image Biomarker Standardization Initiative [27]. Each of the 94 texture features was computed 16 times using the following combinations of extraction parameters, a process known as "texture optimization" [28,29]: 1) isotropic voxels of size 2 mm and 4 mm, 2) fixed bin number (FBN) discretization algorithm with and without equalization, and 3) the number

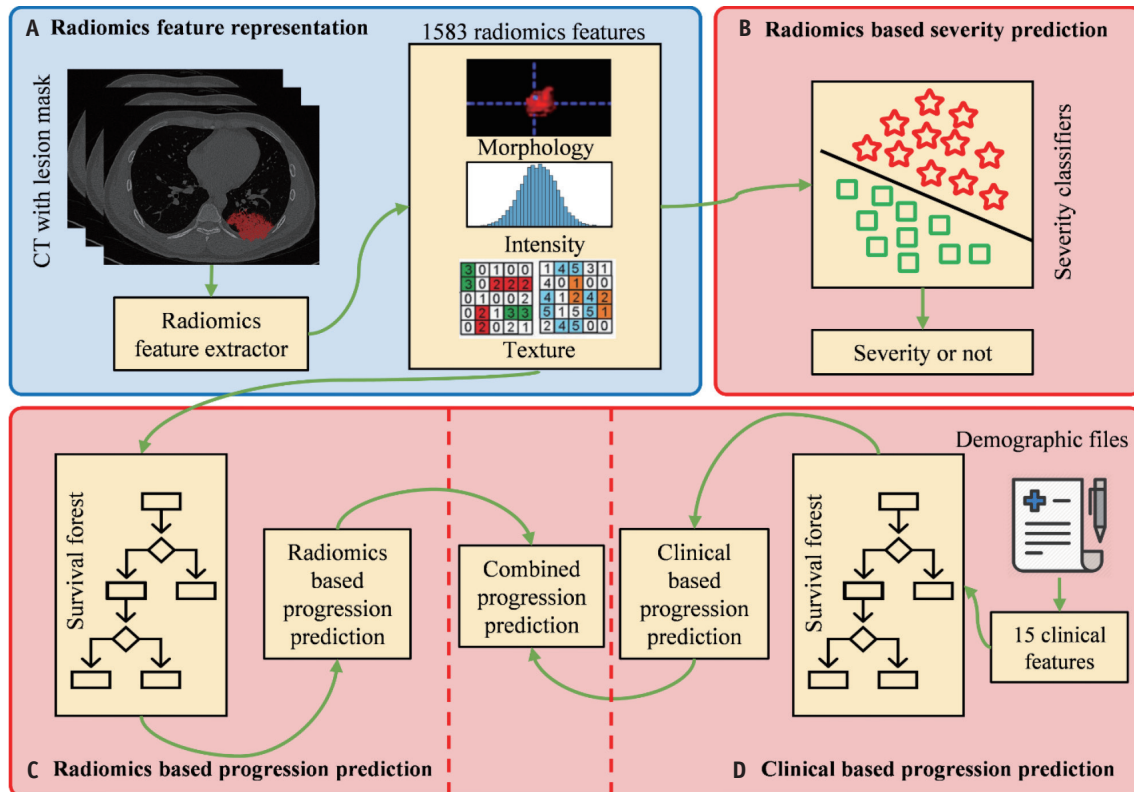


Fig. 2. Illustration of our analysis pipeline.

A. Radiomics feature representation. For each patient, 1583 radiomics features were extracted from automatically segmented lung regions. **B.** Radiomics based severity prediction. Binary classifiers were applied to classify the patients into severe or non-severe classes based on the radiomics features. **C.** Radiomics based progression prediction. A random survival forest model was optimized based on the 1583 radiomics features to assign risk scores to different subjects. **D.** Clinically based progression prediction. Fifteen clinical variables extracted from demographic recordings were input to another survival forest model to assign risk scores to different subjects. Finally, for each patient, the deep learning-based and clinical-based predictions were added with two balanced weights to obtain the combined progression risk score.

of gray levels of 8, 16, 32, and 64 for FBN. A total of (79 + 16 x 94), or 1583, radiomics features were extracted in this study.

The ML models were built using only radiomics features, only clinical variables, a combination of radiomics features and clinical variables, or a combination of visual CT severity scores and clinical variables. Feature selection and classifier optimization methods were used to build the models. To reduce the dimensionality of the datasets, the features were selected for training using five different feature selection methods. Ten ML classifiers were trained on the selected features for every combination of selection and classification. The detailed feature selection methods and classifiers used are shown in Supplementary Table 2. The classifiers were trained by decreasing the number of features from 100 to 15 selected features. Their performances were optimized on the validation set. In addition to the manually optimized ML pipelines, Tree-based Pipeline Optimization Tool (TPOT) [30], an automatic ML algorithm, was used.

TPOT automatically outputs the most optimized pipeline after being trained and validated.

Progression Prediction

Two time-to-event models were built on 1583 radiomics features and 15 clinical variables to predict progression represented by risk scores. Specifically, these progression prediction models were based on survival forests [31] that were optimized to assign risk scores to patients with different progression outcomes according to their input features (radiomics features or clinical variables). The missing values of some clinical variables were imputed by a widely used imputation method [32]. Survival forests use a collection of decision trees for predictions and the ranking of radiomics features or clinical variables by their importance for time-to-event risk prediction [32]. Both survival forests were trained and validated on the same training, validation, and test sets in a 7:1:2 split of patient data that were used for the radiomics models. The

missing values for the radiomics features were imputed using the group mean values for every parameter. The detailed parameters of our applied survival forest models are summarized in Supplementary Table 3. We selected to test our model using 3, 5, and 7 day-time points because the number of critical patients increased in approximate proportions for these intervals. The prediction based on the clinical variables combined with CT radiomics is the sum of the risk scores obtained from the clinical data-based prediction and CT radiomic feature-based prediction by a ratio of 0.52 vs. 0.48, while the prediction based on the clinical variables combined with the visual CT severity scores is the sum of the risk scores obtained from the clinical variables-based prediction and the visual CT severity score-based prediction by a ratio of 0.52 vs. 0.48.

Statistical Analysis

For the severity prediction, the following performance metrics were calculated: accuracy, sensitivity, specificity, and positive and negative predictive values with a probability of 0.50 for the operating point between the binary classifications of severe and non-severe and area under the receiver operating characteristic curve (ROC-AUC). The adjusted Wald method was used to calculate the 95% confidence intervals (CIs) [33]. The `binom_test` function in `scipy.stats` was used to statistically compare the ROC-AUC values. The C-index for the right-censored data [34] was applied to evaluate the performance of the time-to-event models for progression prediction to determine if they efficiently assigned high-risk scores to patients with poor critical outcomes and vice versa. We added the CI by bootstrapping the test set several times (10 or more) and used our optimized model to obtain a series of C-index values. Subsequently, we used these values to calculate the 95% CI. Brier scores were computed to confirm the model calibration. Time-dependent ROC-AUC was calculated from the obtained risk scores and progression information via the Kaplan-Meier method [35] to further evaluate the progression prediction performance. The 'timeROC' R package (<https://www.r-project.org/>) was used to statistically compare the time-dependent ROC-AUC values, and the 'compareC' package in R was used to statistically compare the C-index values [36,37]. The statistical tests were 2 sided. $P < 0.05$ was considered statistically significant.

Further information on our patient cohort, segmentation techniques, and code availability can be found in the Supplementary Materials section and Supplementary Figure 2.

RESULTS

Patient Characteristics

Of the 981 patients with RT-PCR-confirmed COVID-19 and chest CT, 274 developed critical illness. The median age of patients who progressed to critical illness was higher than those who did not (58 vs. 46 years, $p < 0.001$). The median duration from admission to critical illness was 0.4 days. The median durations from symptom onset to presentation, symptom onset to hospitalization, and symptom onset to CT were 4, 4, and 9 days, respectively. However, these medians are for Chinese cases from Hunan Province only (range: 0 to 30 days). The clinical characteristics of COVID-19 patients across the training, validation, and test sets and those with critical and noncritical illnesses are shown in Tables 1 and 2, respectively.

Severity Prediction Models

The chi-squared feature selection method facilitated the highest ROC-AUC score of our severity prediction model when used in tandem with KNN and Boosting classifiers on the test set. Training on the top 25 features facilitated the highest ROC-AUC for the test set.

The hand-optimized ML model combining the top 25 radiomics features and clinical variables achieved a higher ROC-AUC than that based on the visual CT severity scores and clinical variables (0.76 vs. 0.70, $p = 0.023$). The performance metrics of the top-performing models are detailed in Table 3. Heatmaps depicting the performance of the pipeline trained on different datasets are shown in Supplementary Figures 3–5. The results from the automatic ML via TPOT are shown in Supplementary Table 4. The top 25 combined radiomics and clinical variables are shown in Supplementary Table 5.

Progression Prediction Models

The combination of CT radiomics-based and clinical-based predictions achieved the highest C-index of 0.868 (95% CI: 0.830–0.907), when compared with 0.767 (95% CI: 0.706–0.828) for CT radiomics features alone ($p < 0.001$), 0.847 (95% CI: 0.803–0.892) for clinical variables alone ($p = 0.110$), and 0.860 (95% CI: 0.820–0.900) for the combination of visual CT severity scores and clinical variables ($p = 0.549$). This demonstrated success in assigning risk scores consistent with the progression outcomes of patients. The performance metrics for each model are shown in Table 4. As shown in Figure 3, the

Table 1. Comparison of Patient Characteristics Across the Training, Validation, and Test Sets

	Training Set (n = 687)	Validation Set (n = 97)	Test Set (n = 197)	P
Age, year				0.393
Median ± interquartile range	49 ± 24 (range of 0–92)	48 ± 27 (range of 0–85)	49 ± 28 (range of 0–87)	
< 20	24 (3)	4 (4)	10 (5)	
20–39	169 (25)	29 (30)	57 (29)	
40–59	298 (43)	34 (35)	70 (36)	
≥ 60	196 (29)	30 (31)	60 (30)	
Sex				0.954
Male	351 (51)	49 (51)	103 (52)	
Female	332 (48)	47 (48)	93 (47)	
Presence of fever				0.942
Fever	297 (43)	38 (39)	87 (42)	
No fever	118 (17)	15 (15)	35 (15)	
White blood cell count				0.397
Elevated	45 (7)	9 (9)	12 (6)	
Normal	370 (54)	45 (46)	108 (55)	
Lymphocyte count				0.613
Normal	182 (26)	30 (31)	55 (28)	
Decreased	254 (37)	32 (33)	74 (38)	
Comorbidities				
Cardiovascular disease	50 (7)	11 (11)	15 (8)	0.316
Hypertension	94 (14)	14 (14)	32 (16)	0.817
COPD	20 (3)	3 (3)	7 (4)	0.946
Diabetes	48 (7)	10 (10)	24 (12)	0.076
Chronic liver disease	18 (3)	2 (2)	4 (2)	0.822
Chronic kidney disease	16 (2)	2 (2)	7 (4)	0.689
Malignant tumor	13 (2)	2 (2)	2 (1)	0.628
HIV	0 (0)	0 (0)	0 (0)	1.000
Outcomes*				
Ventilator	64 (9)	9 (9)	20 (10)	0.991
Intensive care unit	76 (11)	11 (11)	25 (13)	0.924
Death	20 (3)	1 (1)	3 (2)	0.312
Unknown critical [†]	104 (15)	13 (13)	27 (14)	0.882
Discharged	235 (34)	30 (31)	70 (36)	0.833
Progression to critical event, days				0.149
Median	0.72 (range of 0–21)	0.59 (range of 0–30)	0.08 (range of 0–13)	
Day 1	111 (16)	14 (14)	38 (19)	
Day 2	16 (2)	6 (6)	3 (2)	
Day 3	8 (1)	2 (2)	2 (1)	
≥ Day 4	56 (8)	5 (5)	12 (6)	
Progression to discharge, days				0.244
Median	12 (range of 0–46)	11 (range of 0.2–31)	11.6 (range of 0–38)	
0–4	33 (5)	7 (7)	17 (9)	
5–9	89 (13)	10 (10)	25 (13)	
10–14	146 (21)	24 (25)	41 (21)	
≥ 15	150 (22)	16 (16)	33 (17)	
Epidemiologic contact				
Epicenter [‡]	129 (19)	9 (9)	30 (15)	0.031
COVID-19 patient	87 (13)	13 (13)	31 (16)	0.762

Unless specified otherwise, data are number of patients with the percentage in parentheses. *Patients with multiple critical outcomes may be counted in multiple categories, [†]For patients from public data source (Adapted from Zhang et al. Cell 2020;181:1423-1433.e11 [16]), the type of critical condition was not specified, [‡]Epidemiologic contact with epicenter includes patients who have visited Wuhan, China and New York, NY, USA. COPD = chronic obstructive pulmonary disease, HIV = human immunodeficiency virus

combination of CT radiomics and clinical variables allowed the progression prediction model to achieve time-dependent ROC-AUCs of 0.897, 0.933, and 0.927 for predicting progression risks at 3, 5 and 7 days, respectively. The

results obtained with the visual CT severity score are shown in Supplementary Figure 6. The model calibration results are shown in Supplementary Table 6.

Table 2. Clinical Characteristics of Critical and Non-Critical COVID-19 Patients

	Critical (n = 274)	Non-Critical (n = 707)	P
Age, year			< 0.001
Median ± interquartile range	57.5 ± 23.8 (range of 0 to 92)	46 ± 22.5 (range of 0 to 84)	
< 20	18 (7)	20 (3)	
20–39	29 (11)	226 (32)	
40–59	100 (36)	302 (43)	
≥ 60	127 (46)	159 (22)	
Sex			0.273
Male	148 (54)	355 (50)	
Female	124 (45)	348 (49)	
Presence of fever			< 0.001
Fever	103 (38)	319 (45)	
No fever	20 (7)	148 (21)	
White blood cell count			< 0.001
Elevated	45 (16)	21 (3)	
Normal	79 (29)	444 (63)	
Lymphocyte count			0.001
Normal	78 (28)	189 (27)	
Decreased	45 (16)	215 (30)	
Comorbidities			
Cardiovascular disease	42 (15)	34 (5)	< 0.001
Hypertension	62 (23)	78 (11)	< 0.001
COPD	15 (5)	15 (2)	< 0.001
Diabetes	36 (13)	46 (7)	< 0.001
Chronic liver disease	6 (2)	18 (3)	0.495
Chronic kidney disease	19 (7)	6 (1)	< 0.001
Malignant tumor	9 (3)	8 (1)	< 0.001
HIV	0 (0)	0 (0)	1.000
Outcomes*			
Ventilator	93 (34)	N/A	
Intensive care unit	112 (41)	N/A	
Death	24 (9)	N/A	
Unknown critical [†]	144 (53)	N/A	
Progression to critical event, days			
Median	0.3 (range of 0 to 30)	N/A	
Day 1	163 (59)	N/A	
Day 2	15 (5)	N/A	
Day 3	12 (4)	N/A	
> Day 3	73 (27)	N/A	
Epidemiologic Contact			
Epicenter [‡]	14 (5)	154 (22)	< 0.001
COVID-19 patients	26 (9)	105 (15)	0.662

Unless specified otherwise, data are number of patients with the percentage in parentheses. *Patients with multiple critical outcomes may be counted in multiple categories, [†]For patients from public data source (Adapted from Zhang et al. Cell 2020;181:1423-1433.e11 [16]), the type of critical condition was not specified, [‡]Epidemiologic contact with epicenter includes patients who have visited Wuhan, China and New York, NY, USA. COPD = chronic obstructive pulmonary disease, HIV = human immunodeficiency virus

DISCUSSION

To lower patient mortality and improve overall outcomes, COVID-19 should be detected early [6]. When medical facilities operate at maximum capacity, it becomes increasingly difficult for them to allocate high-demand

resources such as mechanical ventilators or ICU beds to patients [38]. In this study, an ML model was developed to predict COVID-19 severity and progression to critical illness using chest CT and clinical variables with good accuracy. This technology shows potential for informing prognostic decision making for COVID-19 patients, which may improve

Table 3. Performance Metrics of Our Manually Optimized ML Pipelines Predicting Severity on the Test Set Using Radiomics Features Alone, Clinical Variables Alone, Combined Radiomics and Clinical Variables, and Visual CT Severity Score and Clinical Variables

Dataset	Pipeline	AUC	Accuracy	PPV	NPV	Sensitivity	Specificity	P*
Radiomics	TSCR + KNN	0.74	0.79	0.68	0.84	0.62	0.85	0.147
	Lower 95% CI	0.72	0.77	0.66	0.83	0.60	0.82	-
	Upper 95% CI	0.75	0.81	0.70	0.86	0.65	0.87	-
Clinical	CHSQ + BY	0.70	0.68	0.61	0.78	0.73	0.67	0.023
	Lower 95% CI	0.67	0.66	0.57	0.76	0.71	0.65	-
	Upper 95% CI	0.72	0.71	0.63	0.80	0.75	0.70	-
Radiomics + clinical	CHSQ + KNN	0.76	0.80	0.69	0.87	0.62	0.87	-
	Lower 95% CI	0.73	0.77	0.65	0.85	0.59	0.85	-
	Upper 95% CI	0.79	0.82	0.72	0.89	0.65	0.89	-
Visual CT severity score + clinical	CHSQ + BST	0.70	0.77	0.60	0.79	0.56	0.85	0.023
	Lower 95% CI	0.67	0.74	0.57	0.77	0.53	0.83	-
	Upper 95% CI	0.73	0.79	0.62	0.82	0.59	0.87	-

*P value in comparison with the radiomics + clinical model AUC. AUC = area under the curve, BST = boosting, BY = bayesian, CHSQ = chi-square score, CI = confidence interval, KNN = k-nearest neighbors, NPV = negative predictive value, PPV = positive predictive value, TSCR = *t* test score

Table 4. Performance Metrics of Our Radiomics-Based, Clinical-Based, Combined Radiomics and Clinical-Based, Visual CT Severity Score, and Combined Clinical and Visual CT Severity Score-Based Progression Prediction Models

Metric	Clinical	Radiomics	Clinical + Radiomics	Visual CT Severity Score	Clinical + Visual CT Severity Score
iAUC	0.814	0.775	0.829	0.740	0.829
Standard error	0.023	0.028	0.023	0.030	0.017
Lower 95% CI	0.768	0.720	0.784	0.682	0.795
Upper 95% CI	0.859	0.829	0.873	0.799	0.863
C-index	0.847	0.767	0.868	0.742	0.860
Standard error	0.023	0.031	0.020	0.034	0.020
Lower 95% CI	0.803	0.706	0.830	0.676	0.820
Upper 95% CI	0.892	0.828	0.907	0.809	0.900
3-day ROC AUC	0.874	0.792	0.897	0.807	0.910
Standard error	0.029	0.040	0.025	0.041	0.023
Lower 95% CI	0.816	0.714	0.848	0.726	0.865
Upper 95% CI	0.931	0.870	0.947	0.888	0.955
5 day ROC AUC	0.918	0.812	0.933	0.783	0.932
Standard error	0.022	0.037	0.019	0.041	0.018
Lower 95% CI	0.875	0.739	0.896	0.702	0.896
Upper 95% CI	0.961	0.884	0.971	0.864	0.968
7-day ROC AUC	0.897	0.817	0.927	0.764	0.907
Standard error	0.025	0.036	0.020	0.041	0.025
Lower 95% CI	0.847	0.746	0.888	0.683	0.858
Upper 95% CI	0.946	0.888	0.966	0.845	0.956

AUC = area under the curve, CI = confidence interval, iAUC = incremental AUC, ROC = receiver operating characteristic

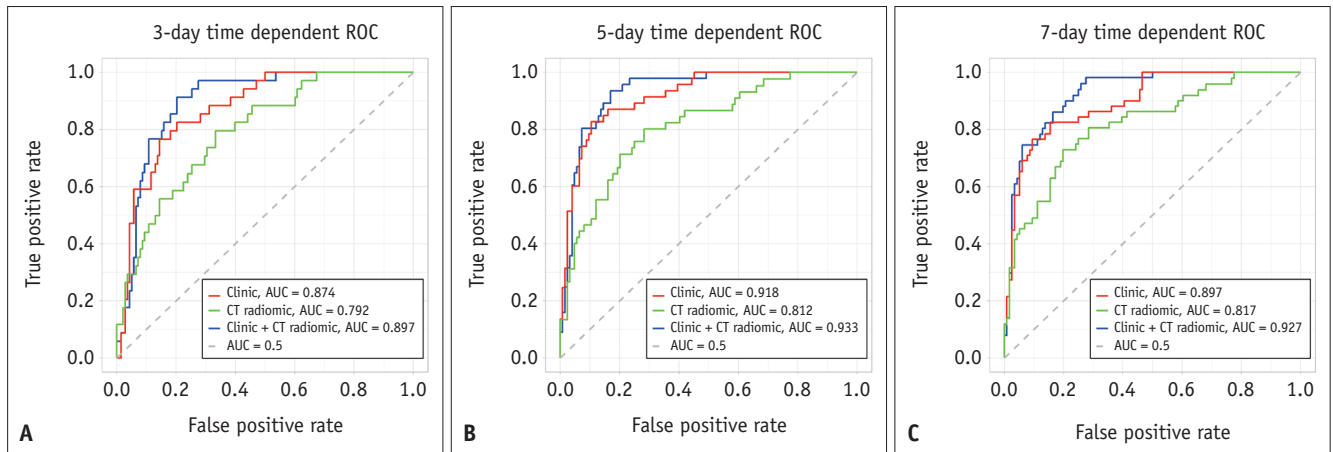


Fig. 3. Time-dependent ROC curves and AUCs for days 3, 5, and 7 for three progression models.

A-C. The results for the three models are shown: one trained on radiomics features, one trained on clinical variables, and one trained on the combination of radiomics features and clinical variables. The x-axis represents the false-positive rate and the y-axis represents the true-positive rate. AUC = area under the curve, ROC = receiver operating characteristic

patient outcomes and resource allocation. Furthermore, this study demonstrates the ability of chest CT data to marginally increase the utility of clinical information for developing severity predictions. It also shows that a model based on the combination of chest CT and clinical variables can facilitate a similar performance to that based on the combination of visual severity scores and clinical variables.

Early detection of COVID-19 enables early medical intervention, which has proven to be a major determinant for improving clinical outcomes and reducing mortality [6,39,40]. CT-based visual severity scoring by radiologists is time-consuming and costly, whereas our ML pipeline can be fully automated for segmentation and feature extraction and used to predict the severity and progression risk. The fact that the combined chest CT and clinical approach achieved similar performance to a combined visual severity score and clinical information approach indicates that ML may be used in a similar manner to expert radiologists in assigning disease progression risk scores to COVID-19 patients. This has the potential to decrease manual labor, save invaluable time, and reduce cost.

In comparison with this study, recently published studies on deep learning or radiomics-based models for assessing the prognosis of COVID-19 utilized smaller cohorts and failed to build specific time-to-critical event prediction models [41-44]. Wang et al. [44] used a deep learning model based on chest CT data to distinguish COVID-19 pneumonia from non-COVID-19 pneumonia and stratify COVID-19 patients based on the risk of developing severe disease. Although this study had a large cohort for training

the models to distinguish between COVID-19 and non-COVID-19 pneumonia, only 471 patients had follow-up for prognostic analysis, and the time-to-event analysis was based on the duration from admission to the development of a critical event, instead of the time of CT [44]. Similarly, another study by Liu et al. [45] used AI algorithms to detect the features of COVID-19 pneumonia on chest CT and predict prognosis. However, unlike the present study, which only used one image at the beginning of a patient's disease course to develop severity predictions, the study by Liu et al. [45] had to use imaging from admission and follow-up imaging on day 4 of a patient's hospital stay—when only imaging from admission was used, the accuracy of prognosis prediction was greatly decreased. This is not ideal because rapid prognostication leads to better outcomes, and several patients who present with severe disease may start deteriorating within the first four days of care before follow-up imaging can be acquired. If the precise time-to-critical-event progression window is known for a given patient, proper equipment can be obtained for their care, and the risk-benefit analysis can be more accurate.

This study has several limitations. First, this was a retrospective study with patient selection bias. Data heterogeneity may have also affected the model performance. Further, the current study defined critical outcomes as mechanical ventilation, admission to the ICU, and death, whereas other studies may have different definitions that may account for the different overall mortality rates for their cohorts. Considering the critical outcomes of mechanical ventilation, ICU admission, and

death as separate events, instead of a composite category, may have been beneficial but it requires a larger sample with sufficient statistical power. It is also worth noting that this study did not include patients without chest CT abnormalities since they did not develop severe disease. The different treatment histories of the patients may have caused bias since only outcomes were used. Laboratory results, including various compounds, such as lactate dehydrogenase, D-dimer, and direct bilirubin, which are associated with adverse outcomes in patients with COVID-19, were not available for a significant portion of our patient cohort [46,47]. Additionally, the current study did not include an external validation set. However, we ensured that the training and independent test groups were completely separate, and there was no leak of information.

In conclusion, an ML model based on radiomics features obtained from chest CT and clinical variables predicted COVID-19 severity and progression to critical events with good accuracy. The model based on the combination of chest CT data and clinical variables also showed higher performance than the model based on only clinical variables, and similar performance to the model based on the combination of the visual CT severity scores and clinical variables. Further research and development are needed to determine the practical role ML can play in COVID-19 severity predictions in the clinical setting.

Supplement

The Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2020.1104>.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Harrison X. Bai, Wei-hua Liao, Yong Fan, Li Yang, Jing Wu. Data curation: Ronnie Sebro, Michael Atalay, Paul J. Zhang, Michael Feldman, Xue Feng, Scott Collins, Dongcui Wang, Jing Wu, Thi My Linh Tran, Ben Hsieh, Kasey Halsey, Ji Whae Choi. Formal analysis: Subhanik Purkayastha, Yanhe Xiao, Rujapa Thepumnoeysuk, Zhicheng Jiao, Robin Wang. Funding acquisition: Harrison X. Bai, Yong Fan. Investigation: Harrison X. Bai, Subhanik Purkayastha, Yanhe Xiao. Methodology: Subhanik Purkayastha, Martin Vallières, Robin Wang, Ji Whae Choi, Dongcui Wang, Yanhe

Xiao. Project administration: Subhanik Purkayastha, Harrison X. Bai. Resources: Harrison X. Bai, Wei-hua Liao, Yong Fan, Li Yang, Ronnie Sebro, Michael Atalay, Paul J. Zhang, Michael Feldman, Xue Feng, Scott Collins, Martin Vallières. Software: Subhanik Purkayastha, Yanhe Xiao, Rujapa Thepumnoeysuk, Martin Vallières, Zhicheng Jiao. Supervision: Harrison X. Bai, Subhanik Purkayastha. Validation: Subhanik Purkayastha, Yanhe Xiao. Visualization: Thi My Linh Tran, Kasey Halsey, Ben Hsieh, Subhanik Purkayastha, Rujapa Thepumnoeysuk. Writing—original draft: Subhanik Purkayastha, Yanhe Xiao, Kasey Halsey. Writing—review & editing: Subhanik Purkayastha, Yanhe Xiao, Kasey Halsey, Ji Whae Choi, Harrison X. Bai.

ORCID iDs

Subhanik Purkayastha

<https://orcid.org/0000-0003-3067-2070>

Yanhe Xiao

<https://orcid.org/0000-0003-1553-0883>

Zhicheng Jiao

<https://orcid.org/0000-0002-6968-0919>

Rujapa Thepumnoeysuk

<https://orcid.org/0000-0002-1637-1768>

Kasey Halsey

<https://orcid.org/0000-0002-6746-8615>

Jing Wu

<https://orcid.org/0000-0001-9602-7387>

Thi My Linh Tran

<https://orcid.org/0000-0002-8126-1258>

Ben Hsieh

<https://orcid.org/0000-0001-5209-801X>

Ji Whae Choi

<https://orcid.org/0000-0003-1061-3102>

Dongcui Wang

<https://orcid.org/0000-0002-4435-2214>

Martin Vallières

<https://orcid.org/0000-0001-7639-8172>

Robin Wang

<https://orcid.org/0000-0001-7369-0602>

Scott Collins

<https://orcid.org/0000-0002-8117-164X>

Xue Feng

<https://orcid.org/0000-0002-2181-9889>

Michael Feldman

<https://orcid.org/0000-0002-6661-4940>

Paul J. Zhang

<https://orcid.org/0000-0001-6810-8681>

Michael Atalay

<https://orcid.org/0000-0003-3549-2634>

Ronnie Sebro

<https://orcid.org/0000-0001-7232-4416>

Li Yang

<https://orcid.org/0000-0001-5973-3653>

Yong Fan

<https://orcid.org/0000-0001-9869-4685>

Wei-hua Liao

<https://orcid.org/0000-0002-1926-3527>

Harrison X. Bai

<https://orcid.org/0000-0002-7460-8866>

REFERENCES

1. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020;8:475-481
2. Long B, Brady WJ, Koefman A, Gottlieb M. Cardiovascular complications in COVID-19. *Am J Emerg Med* 2020;38:1504-1507
3. Sheraton M, Deo N, Kashyap R, Surani S. A review of neurological complications of COVID-19. *Cureus* 2020;12:e8192
4. Lai CC, Ko WC, Lee PI, Jean SS, Hsueh PR. Extra-respiratory manifestations of COVID-19. *Int J Antimicrob Agents* 2020;56:106024
5. Aziz S, Arabi YM, Alhazzani W, Evans L, Citerio G, Fischkoff K, et al. Managing ICU surge during the COVID-19 crisis: rapid guidelines. *Intensive Care Med* 2020;46:1303-1325
6. Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. *Ann Intensive Care* 2020;10:33
7. Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. The epidemiology, diagnosis and treatment of COVID-19. *Int J Antimicrob Agents* 2020;55:105955
8. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020;324:782-793
9. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020;323:1061-1069
10. Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest Radiol* 2020;55:327-331
11. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, et al. Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. *Radiology* 2020;296:E46-E54
12. Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, et al. Chest CT severity score: an imaging tool for assessing severe COVID-19. *Radiology: Cardiothoracic Imaging* 2020;2:e200047
13. Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. *AJR Am J Roentgenol* 2020;214:1072-1077
14. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology* 2020;295:200463
15. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology* 2020;296:E156-E165
16. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 2020;181:1423-1433.e11
17. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi TN, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Med Image Anal* 2021;67:101860
18. Zheng Y, Xiao A, Yu X, Zhao Y, Lu Y, Li X, et al. Development and validation of a prognostic nomogram based on clinical and CT features for adverse outcome prediction in patients with COVID-19. *Korean J Radiol* 2020;21:1007-1017
19. Park B, Park J, Lim JK, Shin KM, Lee J, Seo H, et al. Prognostic implication of volumetric quantitative ct analysis in patients with COVID-19: a multicenter study in Daegu, Korea. *Korean J Radiol* 2020;21:1256-1264
20. Yin X, Min X, Nan Y, Feng Z, Li B, Cai W, et al. Assessment of the severity of coronavirus disease: quantitative computed tomography parameters versus semiquantitative visual score. *Korean J Radiol* 2020;21:998-1006
21. Sun D, Li X, Guo D, Wu L, Chen T, Fang Z, et al. CT quantitative analysis and its relationship with clinical features for assessing the severity of patients with COVID-19. *Korean J Radiol* 2020;21:859-868
22. He Y. Missing data analysis using multiple imputation: getting to the heart of the matter. *Circ Cardiovasc Qual Outcomes* 2010;3:98-105
23. Xiang Q, Dai X, Deng Y, He C, Wang J, Feng J, et al. Missing value imputation for microarray gene expression data using histone acetylation information. *BMC Bioinformatics* 2008;9:252
24. Chang YC, Yu CJ, Chang SC, Galvin JR, Liu HM, Hsiao CH, et al. Pulmonary sequelae in convalescent patients after severe acute respiratory syndrome: evaluation with thin-section CT. *Radiology* 2005;236:1067-1075
25. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19). *Radiology* 2020;295:715-721
26. Zhou S, Wang Y, Zhu T, Xia L. CT features of coronavirus

- disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China. *AJR Am J Roentgenol* 2020;214:1287-1294
27. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardization initiative. *arXiv preprint* 2016;arXiv:1612.07003
 28. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;60:5471-5496
 29. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep* 2017;7:10117
 30. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. *Automating biomedical data science through treebased pipeline optimization*. In: Squillero G, Burelli P, eds. *Applications of evolutionary computation. EvoApplications 2016. Lecture notes in computer science*. Cham: Springer, 2016:123-137
 31. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17:520-525
 32. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat* 2008;2:841-860
 33. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998;52:119-126
 34. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-387
 35. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005;61:92-105
 36. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* 2015;34:685-703
 37. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32:5381-5397
 38. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, et al. Triage of scarce critical care resources in COVID-19: an implementation guide for regional allocation: an expert panel report of the Task Force for Mass Critical Care and the American College of Chest Physicians. *Chest* 2020;158:212-225
 39. Million M, Lagier JC, Gautret P, Colson P, Fournier PE, Amrane S, et al. Early treatment of COVID-19 patients with hydroxychloroquine and azithromycin: a retrospective analysis of 1061 cases in Marseille, France. *Travel Med Infect Dis* 2020;35:101738
 40. Salazar E, Perez KK, Ashraf M, Chen J, Castillo B, Christensen PA, et al. Treatment of coronavirus disease 2019 (COVID-19) patients with convalescent plasma. *Am J Pathol* 2020;190:1680-1690
 41. Homayounieh F, Ebrahimian S, Babaei R, Karimi Mobin H, Zhang E, Bizzo BC, et al. CT radiomics, radiologists and clinical information in predicting outcome of patients with COVID-19 pneumonia. *Radiology: Cardiothoracic Imaging* 2020;2:e200322
 42. Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, et al. Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. *Theranostics* 2020;10:7231-7244
 43. Wei W, Hu XW, Cheng Q, Zhao YM, Ge YQ. Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics. *Eur Radiol* 2020;30:6788-6796
 44. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020;56:2000775
 45. Liu F, Zhang Q, Huang C, Shi C, Wang L, Shi N, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* 2020;10:5613-5622
 46. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;18:934-943
 47. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med* 2020;180:1081-1089