A Multimodality Machine Learning Approach to Differentiate Severe and Nonsevere COVID-19: Model Development and Validation

Monitoring Editor: Corey Basch

Reviewed by Felix Agakov

Yuanfang Chen, MSc,[#1,2] Liu Ouyang, MD,[#3] Forrest S Bao, PhD,[#4] Qian Li, MD,[#5] Lei Han, MD, PhD,[#1,6] Hengdong Zhang, MD, PhD,[1,6] Baoli Zhu, MSc,[1,6,7] Yaorong Ge, PhD,[8] Patrick Robinson, MD,[9] Ming Xu, MD, PhD,[⊠1,6,9] Jie Liu, MD, PhD,[10] and Shi Chen, PhD[9,11]

[1] Public Health Research Institute of Jiangsu Province, Nanjing, China

[2] Institute of HIV/AIDS/STI Prevention and Control, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[3] Department of Orthopaedics, Union Hospital, Huazhong University of Science and Technology, Wuhan, China

[4] Department of Computer Science, Iowa State University, Ames, IA, United States

[5] Department of Pediatrics, Affiliated Kunshan Hospital of Jiangsu University, Kunshan, China

[6] Department of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[7] School of Public health, Nanjing Medical University, Nanjing, China

[8] Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, United States

[9] Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States

[10] Department of Radiology, Union Hospital, Huazhong University of Science and Technology, Wuhan, China

[11] School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States

Ming Xu, Department of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Control and Prevention, 172 Jiangsu Road, Nanjing, 210009, China, Phone: 86 85393210, Email: sosolou@126.com.

[⊠]Corresponding author.
[#]Contributed equally.
Corresponding Author: Ming Xu sosolou@126.com

Abstract

Background

Effectively and efficiently diagnosing patients who have COVID-19 with the accurate clinical type of the disease is essential to achieve optimal outcomes for the patients as well as to reduce the risk of overloading the health care system. Currently, severe and nonsevere COVID-19 types are differentiated by only a few features, which do not comprehensively characterize the complicated pathological, physiological, and immunological responses to SARS-CoV-2 infection in the different disease types. In addition, these type-defining features may not be readily testable at the time of diagnosis.

Objective

In this study, we aimed to use a machine learning approach to understand COVID-19 more comprehensively, accurately differentiate severe and nonsevere COVID-19 clinical types based on multiple medical features, and provide reliable predictions of the clinical type of the disease.

Methods

For this study, we recruited 214 confirmed patients with nonsevere COVID-19 and 148 patients with severe COVID-19. The clinical characteristics (26 features) and laboratory test results (26 features) upon admission were acquired as two input modalities. Exploratory analyses demonstrated that these features differed substantially between two clinical types. Machine learning random forest models based on all the features in each modality as well as on the top 5 features in each modality combined were developed and validated to differentiate COVID-19 clinical types.

Results

Using clinical and laboratory results independently as input, the random forest models achieved >90% and >95% predictive accuracy, respectively. The importance scores of the input features were further evaluated, and the top 5 features from each modality were identified (age, hypertension, cardiovascular disease, gender, and diabetes for the clinical features modality, and dimerized plasmin fragment D, high sensitivity troponin I, absolute neutrophil count, interleukin 6, and lactate dehydrogenase for the laboratory testing modality, in descending order). Using these top 10 multimodal features as the only input instead of all 52 features combined, the random forest model was able to achieve 97% predictive accuracy.

Conclusions

Our findings shed light on how the human body reacts to SARS-CoV-2 infection as a unit and provide insights on effectively evaluating the disease severity of patients with COVID-19 based on more common medical features when gold standard features are not available. We suggest that clinical information can be used as an initial screening tool for self-evaluation and triage, while laboratory test results should be applied when accuracy is the priority.

Keywords: COVID-19, clinical type, multimodality, classification, machine learning, machine learning, diagnosis, prediction, reliable, decision support

Introduction

COVID-19 is a pandemic disease caused by the novel SARS-CoV-2 virus. As of January 12, 2021, COVID-19 had spread through at least 220 countries and regions, resulting in more than 88 million cases and almost 2 million deaths [1]. It has become the single most severe pandemic in the 21st century, dwarfing other coronavirus-caused epidemics, such as severe acute respiratory syndrome (SARS) in 2003 and Middle East respiratory syndrome (MERS) in 2012. COVID-19 is especially challenging to health professionals and the general population. Unlike in the preceding SARS and MERS epidemics, patients with COVID-19 can be either asymptomatic or symptomatic, and the virus has been demonstrated to be transmissible in both states to varying degrees [2-5]. In addition, the distinct clinical types of COVID-19, nonsevere and severe, require different treatment and care plans [6]. In current studies, patients with COVID-19 can be differentiated from patients who do not have the disease; however, further detection of nonsevere or severe types of COVID-19 has not been comprehensively explored. Patients with nonsevere COVID-19 can be accommodated with less intensive clinical monitoring and intervention, including treating pre-existing conditions and preventing health care–associated infections and other comorbidities [7]. In contrast, patients with severe disease require close monitoring, usually in the intensive care unit (ICU), by more clinicians [6]. Therefore, effectively and efficiently classifying clinical types of COVID-19 is essential for triage, resource optimization, and care planning for frontline clinicians and health care systems as well as for the patients [6,8].

Currently, nonsevere and severe COVID-19 types are classified based on only a few clinical features in China, including shortness of breath, O_2 saturation, and PaO_2 [9]. Because of the complexity of the pathological, physiological, and immunological response of COVID-19, these three features do not sufficiently characterize the difference between nonsevere and severe types in patients with COVID-19 [9-11]. Although shortness of breath can be self-monitored, O_2 saturation and PaO_2 cannot be accurately self-evaluated and may not be readily assessed in clinical settings, especially for socioeconomically disadvantaged patients. In addition, some patients with severe disease may not present shortness of breath initially. However, without proper medical intervention, their clinical course will worsen abruptly, often resulting in respiratory failure with high mortality [6]. Therefore, these gold standard features bear the risk of misclassification and misdiagnosis. Misclassification of COVID-19 clinical types can result in inappropriate early treatment decisions; this can place patients at risk of progression due to insufficiently aggressive supportive therapy or expose other patients to overly invasive treatment, both of which have negative clinical consequences. In addition, the three defining features may not be readily available during initial diagnosis when resources are inadequate.

It is therefore critical to provide a rapid, accurate, and efficient method to determine the severity of COVID-19 infection and identify the clinical type using alternative features. This determination will enable optimization of treatment plans for patient care and improve utilization of health care resources and staff. We suggest that additional readily available medical features, including the patient's comorbidities (eg, hypertension and diabetes) and symptoms (eg, fever and chest pain), as well as laboratory test results, can be used to develop an effective method to determine the clinical type and severity of COVID-19 [12,13]. Angiotensin-converting enzyme 2 (ACE-2) receptors, which facilitate SARS-CoV-2 infiltration, are distributed across multiple organs and systems in the human body [14]. More recent discoveries have found that in addition to the respiratory system, SARS-CoV-2 can invade digestive, reproductive, and even neural systems [15-18]. In other words, all clinical and laboratory test information of patients with COVID-19 could be consequences or risk factors of SARS-CoV-2 infection. In clinical practice to treat COVID-19, clinicians not from respiratory units or ICUs may rely only on the referenced features [9] while neglecting diverse and important clinical features of COVID-19, and they may miss critical signs leading to undesirable prognosis.

The potential power of clinical and laboratory testing features, as well as their combinations, to determine COVID-19 clinical type is currently being explored [19-24]. To use such diverse multimodality information as alternative evidence to facilitate accurate classifications, we propose a data mining and machine learning (ML) framework as an alternative to commonly used hypothesis-driven parametric models. The goal of this study is to provide reliable data-driven support for clinicians, even those who do not have comprehensive experience in diagnosing the emerging disease COVID-19. We aim to explore and contrast the distributions of clinical and laboratory testing features between nonsevere and severe COVID-19 types. We will identify key features that differ substantially between the two clinical types. Next, we will investigate whether a single modality or specific combination of features across modalities are able to provide accurate classification models via ML techniques. Specifically, we aim to identify a small and practical set of input features that can accurately differentiate COVID-19 clinical types. The insights gained from this study, as well as the developed end-to-end multimodal data analysis and ML framework, will enable us to better understand the comprehensive pathology of COVID-19, further distinguish COVID-19 from other respiratory infections, and apply the framework to other diseases with multimodal medical data in the future.

Methods

Data Source and Clinical Feature Extraction

In this study, we recruited 362 patients with COVID-19 from January to March 2020, including 148 patients presenting with severe disease and 214 patients lacking criteria for severe disease during admission, from Wuhan Union Hospital, China. The definitions of nonsevere and severe cases were mainly adopted from the official COVID-19 Diagnosis and Treatment Plan from the National Health Commission of China, and we also consulted guidelines from the American Thoracic Society [9-11]. Patients with severe COVID-19 should present any one of the following features: (1) respiratory rate >30 breaths per minute; (2) oxygen saturation <93% at rest; or (3) PaO_2/fraction of inspired oxygen <300 mm Hg (40 kPa). Each patient with COVID-19 was confirmed by two independent quantitative reverse transcriptase–polymerase chain reaction tests before being included in this study. All patients or their responsible surrogates signed

informed consent forms prior to study inclusion. The patients' symptoms were evaluated and blood samples were drawn upon admission to perform laboratory testing. No pediatric patients aged less than 18 years were included.

The patients' deidentified medical information include two major modalities of features, both of which were assessed at the time of admission. The first modality was a total of 26 pre-existing comorbidities and symptoms, referred to as "clinical features" hereinafter. These features included gender, age, hypertension, coughing, and different types of fever. A detailed description of these 26 features is provided in Table S1 in Multimedia Appendix 1. All clinical features were coded as 0-1 binary variables (age was dichotomized using 50 years as the threshold).

In addition, we collected the patients' laboratory test results. The laboratory tests were plasma, serum, or whole blood assays for commonly obtained biochemistry tests, complete blood counts with differential counts and percentages, immunologic markers, such as interleukin 6 (IL-6), dimerized plasmin fragment D (D-dimer) and high-sensitivity C-reactive protein (hsCRP). After initial screening, several features with too many missing data, such as calcitonin, were excluded. In addition, respiratory rate, oxygen saturation, and PaO_2 without supplemental oxygen were excluded because they are type-defining features according to the official National Diagnosis and Treatment Plan of China [9]. We used 26 laboratory test features in this study. Detailed descriptions and units of these features are provided in Table S2 in Multimedia Appendix 1. All these laboratory testing features were continuous features, in contrast to the binary features used in the clinical feature modality.

Patient-specific identifying information (eg, name and address of residence) was removed from the data collected for this study. This study was evaluated and approved by the IRB committee of Union Hospital, Wuhan, China (approval number: 2020-IEC-J-345).

Data Mining on Multimodal Features

Initial data mining on the multimodal COVID-19 data was conducted. The patients' clinical data were complete. Approximately 5% of the laboratory testing data were missing. Predictive mean matching (PMM) was applied to impute the missing data. To evaluate the effectiveness of PMM, we used a subset of the original data set with no data missing, randomly dropped 5% data to simulate potential data loss, re-extrapolated the data with PMM, and evaluated the root mean square error (RMSE) between the original and imputed data sets. The RMSE was less than 0.05, indicating that the extrapolation was feasible and reliable. The imputed data were then passed on to successive data mining and ML steps.

The prevalence of each clinical feature was calculated as the number of positive test results divided by the number of patients in the nonsevere and severe groups as defined by the Diagnosis and Treatment Plan [9]. The z test was applied to detect any statistically significant differences in the features between the two types. In addition, a forest plot of the odds ratios (ORs) and 95% confidence intervals of the clinical features between severe and nonsevere COVID-19 types was graphed.

For the continuous laboratory testing features, we characterized and contrasted the distribution of each feature between the two types. Because the values of most features were not normally distributed, we applied a 2-sided Kolmogorov-Smirnov test instead of the Student t test to determine whether distributions of the

feature values differed significantly between the two clinical types.

COVID-19 Clinical Type Classification via ML

Commonly used hypothesis-driven parametric models rely heavily on human decisions of how features interact with each other (eg, interaction terms in the logistic regression model), which may not reflect the underlying medical reality. In addition, these models have strict prerequisites to perform correctly, including normality of residuals, homoscedasticity, and independence of input features. Our initial exploratory analyses showed the that input features in both the clinical and laboratory testing modalities had nonnormality and high collinearity among the features. Another technical challenge to logistic regression in this study was the mixture of binary clinical and continuous laboratory testing input features.

Due to these problems, logistic regression would not be a preferred modeling approach to accurately classify and predict COVID-19 clinical types. Our exploratory analysis showed that logistic regression could only achieve average predictive accuracies of 68% and 77% on an 80-20 training-testing split using clinical and laboratory testing feature data sets, respectively (Table S3, Multimedia Appendix 1). Thus, logistic regression is less feasible in clinical settings, where high accuracy, sensitivity, and specificity are required to differentiate COVID-19 clinical types.

On the other hand, state-of-the-art ML classification models work directly with data to avoid possible human bias. In addition, ML models do not have restrictions on how input data should be distributed or related. Therefore, in this study, we determined that ML classification would be a more appropriate modeling approach to predict COVID-19 clinical type with a complicated data structure. We developed an end-to-end ML analytical framework to accurately predict the clinical type of patients with COVID-19 based on clinical and laboratory testing modality features. We built random forest (RF) classification models, as RF enables excellent interpretability of the relative importance of an input variable to provide a more comprehensive understanding of the pathobiology of COVID-19. RF is a widely used ML model based on decision theory and the decision tree approach. Due to the internal validation process with out-of-bag error measurement, RF is especially accurate and reliable. Unlike other commonly used ML models (eg, support vector machine or k-nearest neighbor), which usually require a separate cross-validation set, the RF model performs internal validation and is especially suitable when the data set is not large. In addition, RF is robust against data loss and data unbalancing (eg, there are more patients with nonsevere than severe disease in our study [25-29]). Because the major goal of this study was not to compare the performance of different ML models, we focused on RF to deliver the most accurate classification possible.

For the single modality RF model, we used 50% randomly selected data for both clinical and laboratory testing blood biochemistry features. In this step, 107 patients with nonsevere COVID-19 and 74 patients with severe COVID-19 were randomly chosen, while the other patients' information was held to build the multimodal RF model. We assigned severe cases as "positive" and nonsevere cases as "negative" in the classification. The goal of ML classification through RF was to accurately predict the patient's COVID-19 type, either positive (severe) or negative (nonsevere), based on features from different clinical modalities. In this part of the study, we first used a single modality of features, either clinical or laboratory testing, as the input. The detailed RF modeling and validation processes are provided in Multimedia Appendix 2. We trained the model with 100 independent runs; in each run, a different set of 80% of the data was randomly selected for training, while the remaining 20% of the data were held for testing only. This step was per-

formed to explore whether the RF model was robust against different input data and to assess the generalizability of the model. Hyperparameters in this RF model include using Gini impurity to determine the decision tree split, a minimum of 2 samples for tree spit, a minimum of 1 sample at any leaf node, and a total of 8 trees for the model ensemble [26]. Important ML performance metrics, including accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC) value based on the receiver operating characteristic (ROC) curve, were computed for the testing set only.

In addition, RF can evaluate the relative importance of the input variables based on their Gini importance scores [28]. We further quantified the Gini impurity importance scores of the input features in each RF run during the model development stage on the training set (80% randomly selected data). We then identified the top contributing features based on the Gini importance scores in each of the 100 runs, aggregated over 100 runs; identified the overall top contributing features; and explored the clinical relevance and interpretability of these features for COVID-19. Note that the Gini impurity importance was calculated from the RF model based on the training set only and not on the testing set. In addition, each run of the RF model was based on a completely different, randomly sampled, and independent set of 80% training data, from which the Gini importance was calculated. Therefore, this approach avoided potential issues of overfitting and inflated performance [30]. If an RF model is robust, important input features should be consistent with the different 80% portions of the data used as the training set to develop the model. The most important features to differentiate COVID-19 clinical types were also cross-checked with our results from exploratory data mining, including the prevalence of the clinical features and the distribution of the laboratory testing features.

COVID-19 Clinical Type Classification With Multimodal ML

More importantly, we explored whether and how combining features across feature modalities improved classification performance. We developed another RF model with the same hyperparameter setting that incorporated features from both modalities. The modeling process using a single modality was similar. Instead of putting all 52 features into the model, we selected only the top 5 features from each of the two modalities as new inputs to reduce the increase model feasibility in case certain features in the total 52-feature pool would not be readily available. These top features were identified from the Gini importance of the single modality RF models (highlighted in Table S1 and Table S2 in Multimedia Appendix 1). The data set for developing the multimodal RF model was a completely new data set, as in, the other 50% of the original data was based on 107 additional patients with nonsevere COVID-19 and 74 additional patients with severe COVID-19 whose data were not used in the development of the single modality RF model.

We explored whether only 10 important features from different modalities could perform sufficiently well to address the clinical challenge of differentiating COVID-19 clinical types. This study can serve as alternative and supplemental tool to the gold standard features, which may not be readily available at the time of diagnosis.

All statistical analyses and ML models were built in R 4.0.2 (R Project) and Python 3.7 with additional supporting packages. The complete codes and fully deidentified data are freely available on GitHub [31].

Results

Clinical Findings in Nonsevere and Severe COVID-19

The prevalence of clinical features in patients with nonsevere and severe COVID-19 at the time of entry into the study were calculated and compared (Figure 1). A more detailed comparison of the clinical features between the two COVID-19 clinical types is provided in Table S1 in Multimedia Appendix 1, which shows the ORs, confidence intervals, and associated P values. For patients with the two clinical types of COVID-19, the prevalence was distinct for a number of different features. Patients with severe COVID-19 were statistically much more likely to be older (aged ≥50 years, OR 13.77, 95% CI 7.33-25.86, $P<.001$) and male (OR 1.89, 95% CI 1.24-2.90, $P=.003$) and to have renal diseases (OR 8.51, 95% CI 1.86-38.99, $P<.001$), cardiovascular diseases (OR 5.61, 95% CI 2.81-11.20, $P<.001$), hypertension (OR 5.37, 95% CI 3.36-8.56, $P<.001$), diabetes (OR 4.61, 95% CI 2.53-8.38, $P<.001$), loss of appetite and taste (OR 3.20, 95% CI 1.70-6.01, $P<.001$), chills (OR 2.21, 95% CI 1.16-4.22, $P=.01$), and chest congestion (OR 1.88, 95% CI 1.22-2.89, $P=.003$) than their counterparts with nonsevere COVID-19. The only exception was sore throat, which patients with severe COVID-19 were significantly much less likely to develop (OR 0.30, 95% CI 0.14-0.61, $P<.001$). These discoveries are further demonstrated in the forest plot of the ORs and confidence intervals in Figure 2, which shows the differences between the two clinical types. Therefore, these relatively easily measured and acquired clinical features could be used to clinically evaluate the disease severity of patients with COVID-19. Our findings, especially for patients with severe COVID-19, echoed the US Centers for Disease Control and Prevention's recently updated list of symptoms of COVID-19 [32] and more recent characterizations of patients with COVID-19 in the United States [33]. Our findings showed that older male patients with COVID-19 who had cardiovascular disease, respiratory disease, renal disease, and diabetes were at much higher risk of developing serious complications of COVID-19, such as acute respiratory distress syndrome (ARDS) and even death [20,21]. In addition, we discovered that Chinese patients with renal diseases were significantly more likely to develop severe COVID-19, which has not been widely reported. Clinical evidence has shown that ACE-2 expression is associated with kidney diseases; thus, kidney disease is a potential complication of SARS-CoV-2 infection [34,35]. This finding would inform clinicians that they should also monitor kidney dysfunction, such as acute kidney injury, as a clinical sign or consequence of severe COVID-19 complications.

For the laboratory testing modality features, we compared the distributions of the continuous features between nonsevere and severe COVID-19. The results are demonstrated in Figure 3. A more detailed comparison of these 26 laboratory testing features between the two clinical types is provided in Table S2 in Multimedia Appendix 1, which shows the P values from the Kolmogorov-Smirnov tests. Based on the 2-sided Kolmogorov-Smirnov test, severe and nonsevere COVID-19 types differed significantly in most laboratory features, except for platelet (PLT), hemoglobin (HGB), CD3, and CD4. Among all laboratory features, IL-6, high-sensitivity troponin I (hsTNI), and D-dimer had the most significant differences between nonsevere and severe COVID-19 types.

In conclusion, after extensive clinical feature extraction and data mining, we obtained strong qualitative and quantitative evidence that nonsevere and severe COVID-19 types differ substantially with regard to clinical features and laboratory test results. These findings pave the way toward creating an effective ML classifier to accurately differentiate these two COVID-19 types in clinical practice.

Clinical Type Classification via ML

Comorbidity and Symptom (Clinical) Modality We first explored whether relatively simple binary features could provide accurate insights in identifying COVID-19 disease severity. The performance of this model is summarized in the upper section of Table 1. Based on 100 independent runs, the RF model reached a median of >99% and 94% accuracy for the training and testing sets, respectively (Table 1). Median is reported instead of mean value because the performance metrics were not normally distributed. The AUC was 90.2% (range 82.9%-97.6%) based on the ROC curve (Figure 4, left panel). The model performed better in detecting true positives (ie, severe clinical type) than true negatives (ie, nonsevere type). In other words, clinical features alone in the RF models were very unlikely to misclassify a severe case as a nonsevere case but had a higher likelihood of predicting a nonsevere case to be a severe case. In clinical practice, this would be a lesser concern, as a false positive (failure to detect nonsevere type) would be more tolerable than a false negative (failure to detect severe type).

Our RF model also identified the major influential features to differentiate COVID-19 types based on their contributions to the Gini importance in the training set. The top influential clinical features, in descending order, were age, gender, hypertension, diabetes, and cardiovascular diseases, in accordance with existing literature reports [36]. Other important clinical features included fatigue, chest congestion, sore throat, phlegm, and fever. Most of these findings aligned well with our parametric data mining with OR comparison (Figure 2, Table S1 in Multimedia Appendix 1) but showed much higher accuracy (94% accuracy on the testing set of the RF model compared to 68% accuracy from logistic regression). The only exception was renal disease, which was not considered to be a major differentiating factor based on its Gini importance (Table S1, Multimedia Appendix 1).

Clinically, older male patients with pre-existing comorbidities, especially hypertension, diabetes, and cardiovascular diseases, are much more vulnerable to COVID-19 and have a much higher risk of developing severe disease [19,21]. Therefore, we suggested using the comorbidity and symptom features of patients with COVID-19 as the first round of evaluation of severity with reasonable accuracy.

Laboratory Testing Modality The RF model with 26 laboratory testing features was highly effective in differentiating nonsevere and severe COVID-19. The RF model achieved >99% and >95% accuracy for the training and testing data sets, respectively. The sensitivity, specificity, and F1 scores were all >95% when using only 8 trees in the RF model (Table 1, middle section). The AUC was 97% based on the ROC curve (Figure 4, middle panel). Although this study focused on ML methods, we evaluated the model performance of non-ML logistic regression in Table S3 (Multimedia Appendix 1) as a reference point to show the improvement that state-of-the-art ML models could achieve.

The top differentiating features in the laboratory testing modality were D-dimer, hsTNI, neutrophil, IL-6, lactate dehydrogenase (LDH), and hsCRP, in descending order. The clinical interpretation of their important roles was that patients with severe COVID-19 experience more intense immune responses and hyperinflammation, such as cytokine storm syndrome, with substantially increased IL-6 [37]. Research has also shown that SARS-CoV-2 can infect many organs other than the lungs, including the heart, and induce dysfunction of these organs [38,39]. Increasing hsTNI was found to be a sign of heart tissue damage from SARS-CoV-2 infection [40]. In addition, patients with severe COVID-19 may have microthrombosis, which induces higher D-dimer levels [19,21,41-43]. Abnormal levels of neutrophils may be responsible for

cytokine storms and ARDS in patients with severe COVID-19 [13,44]. hsCRP, a biomarker of acute inflammation, cardiovascular disease, and ischemic events, was also confirmed to be a major contributing factor of COVID-19 mortality [19]. LDH is a biomarker of tissue damage and has been used to predict the clinical course of patients with COVID-19 [45]. These findings add further clinical insights to how multiple organs and systems, not just the lungs, respond to SARS-CoV-2 infection in different clinical types [14,46,47].

Multimodal Features   We further developed a multimodal RF model that incorporated both clinical and laboratory testing modalities with a completely new data set that was not used for the single modality model development. We used only the 5 most important features from the clinical and laboratory testing modalities, based on their Gini importance scores. The results showed that the top 10 of a total of 52 features from both modalities achieved almost >95% in every model performance metric, including accuracy, sensitivity, specificity, and F1 score (Table 1). The AUC was >97% as well (Figure 4, right panel). Therefore, we concluded that a two-step evaluation and triaging process would be feasible to differentiate the clinical types of patients with COVID-19 when the gold standard type-defining features were not readily available.

These findings reflect our clinical understanding that SARS-CoV-2 attacks multiple organs and systems, and the human body reacts in a unity against infection. Different features (eg, comorbidity, symptom, and laboratory testing results) complemented each other to provide a more comprehensive characterization of how the human body as a united entity, not only the respiratory system, reacted to SARS-CoV-2 infection [14]. In addition, the decent model performance supports the feasibility of multimodal data mining in detecting and differentiating patients with nonsevere COVID-19 from patients with severe disease.

Comparing the original 52 features in both modalities, which may not be all available at the same time during COVID-19 diagnosis, the top 10 most differentiating multimodal features provided a more practical input combined with the highly accurate ML model. Therefore, we concluded that our work would help effectively optimize health care operations during the pandemic and avoid overloading of the health care system [8].

Discussion

Principal Findings

This study provides a novel analytical framework that combines the power of multiple clinical features from different modalities to differentiate COVID-19 clinical types via ML techniques. Practically, it enables the delivery of a more comprehensive understanding of the pathobiology of COVID-19. It can aid the development of optimal treatment plans for individual patients, such as sending them to a mobile cabin hospital or admitting to a hospital with an ICU [7]. In addition, it will enable more effective triaging and optimization of health care system resources and personnel. This will substantially reduce the risk of overloading the health care system by admitting all patients with COVID-19 to hospital, decrease potential health care–associated infections, and improve clinical outcomes for the patients, especially during the COVID-19 pandemic [8].

In addition to accurately detecting vulnerable patients with COVID-19 who are likely to have severe disease, this study also provides insights on why these patients may have severe disease. ML models work directly with data and therefore are generally not good at providing clear interpretations. In this study, we combined the power of both hypothesis-driven and data-driven ML models. The highest-contributing comorbidities, symptoms, and biochemical features help predict and explain potential COVID-19 clinical courses and prognoses. Our research echoes recent studies that characterize and predict the clinical course, critical illness, and mortality of patients with COVID-19 [13,19,21]. In particular, another decision tree–based algorithm, extreme gradient boosting (XGBoost), showed promising performance in predicting the mortality of patients with COVID-19 [19]. RF is technically similar to XGBoost, and our results were consistent in identifying the key differentiating features, including LDH and hsCRP.

A continuous-valued risk score calculator for predicting risk of transitioning to critical-type COVID-19 (an even more severe type that requires ICU hospitalization, an invasive ventilator, or extracorporeal membrane oxygenation, and has a mortality rate as high as 50%) has been developed for patients with COVID-19 [21]. As a comparison, although our RF model predicts a 0-1 binary outcome for nonsevere and severe type disease, the internal RF modeling process through decision tree approach actually calculates an intermediate score between 0 and 1. By using a cutoff threshold, the RF model reports a final dichotomized 0-1 outcome. Therefore, our analytical framework can also be readily adjusted to provide a continuous risk score for clinical evaluation and triaging of patients with COVID-19, if needed.

Many patients with severe COVID-19 present symptoms in lungs, especially ground-glass opacity (GGO), which can be detected by biomedical imaging techniques such as computed tomography (CT). However, a major clinical challenge of COVID-19 lies in the asymptomatic patient problem, which creates far worse difficulties than other coronavirus epidemics, including the original SARS and MERS epidemics. These patients show few or no classic symptoms related to viral pneumonia, and they present no GGO; however, they are almost as capable of transmitting the virus as symptomatic patients [4-6]. We suggest that the term "asymptomatic" is used due to lack of a comprehensive characterization and understanding of this novel pathogen and the pathophysiology of the host; we also suggest that these patients are not truly "asymptomatic," as in, without any clinical symptoms or signs. Approximately 10% of the patients with mild COVID-19 in our study cohort did not show typical respiratory symptoms, including fever, coughing, and chest pain, upon admission. However, they showed other symptoms from the more comprehensive modality of 26 clinical features.

Future Work

The next step of this study is to further include a biomedical imaging modality. A technical barrier is that a CT scan is a high-dimensional feature set, while clinical and laboratory test data have relatively low dimensionality. Therefore, the CT scan, in its original form of imaging, cannot be effectively combined with other modalities. We will evaluate the feasibility of using a convolution neural network (CNN, another ML technique) first to reduce the feature space in CT scans and extract a fully connected layer in the CNN as a representation of the CT scan feature. A fully connected layer is a 1D vector and has the same dimensionality as the other two modalities. Therefore, in theory, we would be able to further combine CT scans with other clinical features and investigate the association between these features with regard to COVID-19.

COVID-19 is a complex disease in which the pathogen not only attacks the respiratory system but other organs and systems that possess ACE-2 receptors as well [14,33]. Our findings reveal the complicated pathological, physiological, and immunological responses to SARS-CoV-2 infection and shed light in understanding the complex interactions between the virus and the human body. Although our multimodal data mining and ML framework was developed with data from patients with severe and nonsevere COVID-19, we suggest that the end-to-end framework is applicable to many disease systems in which multimodal inputs are common, including demographic information, comorbidity, laboratory testing, imaging, and -omics data. Having a more holistic viewpoint and approach will enable us to understand and respond to these emerging diseases, especially the unprecedented COVID-19, more readily in the field. Another feasible analytical solution is ensembling. Each input feature modality can be used independently to train a specific model, and the final prediction of COVID-19 clinical type can be made through ensembling. We will further explore this analytical framework and transfer our insights to future clinical studies, such as differentiating healthy patients from patients with non–COVID-19 viral pneumonia, nonsevere COVID-19, and severe COVID-19.

Limitations

In this study, we recruited participants from a single hospital in Wuhan, the first epicenter of COVID-19. There will inevitably be selection bias, as the ethnic group is currently limited to Chinese participants who are mostly of Han ethnicity. It is possible that ethnicity and race and their confounding risk factors (eg, socioeconomic status, nutrition conditions, accessibility of care, and other social determinants of health) are different in various studies. Therefore, we wish to share our findings with our colleagues worldwide and determine whether different demographic backgrounds influence feature distributions between nonsevere and severe COVID-19 in patients. Some of our findings of the top contributing clinical and laboratory testing features were supported in other COVID-19 studies across different ethnic groups, while others were not [32,40]. For example, while we found male gender to be a strong influencing factor of severe COVID-19, other studies did not reach a similar conclusion [36]. The findings in this study on Chinese ethnicity could actually complement other existing studies on other ethnic groups and reveal the clinical and epidemiological complexity of this unprecedented ongoing pandemic.

Another limitation of this study is that the patients were evaluated at the time of admission; therefore, the study was a cross-sectional instead of a longitudinal study. Future studies could examine both diagnosis and prognosis and further explore how and why some patients with nonsevere COVID-19 may transition to a severe disease state and whether ML techniques are able to identify critical predictive features to undesirable prognoses such as death.

Additionally, different subtypes of SARS-CoV-2, their specific pathogenicity and virulence, and their host-pathogen interactions should be taken into consideration when conducting and comparing studies across different regions of the world. The other factors that this study did not include are behavioral and societal aspects, such as whether and how using mobile cabin hospitals to treat patients with nonsevere COVID-19 reduces the rate of transition to severe type. The COVID-19 epidemic, like all infectious disease epidemics, has individual clinical, epidemiological, behavioral and societal factors. Therefore, we will also explore cross-scale individual clinical course and population-level epidemics in future studies.

Conclusion

We trained and validated ML RF models to predict COVID-19 severity based on 26 comorbidity and symptom features and 26 laboratory testing features from a cohort of 214 patients with nonsevere COVID-19 and 148 patients with severe COVID-19. We identified the top features from both feature modalities to differentiate the clinical types, and we achieved predictive accuracies of >90%, >95%, and >99% when clinical features, laboratory test data, and the top 5 features from each modality combined were used as inputs, respectively. The results will help patients with COVID-19 self-evaluate their condition, help clinicians to evaluate disease severity and triage patients, and optimize health care resource utilization during the COVID-19 pandemic.

Acknowledgments

Abbreviations

ACE-2	angiotensin-converting enzyme 2
ARDS	acute respiratory distress syndrome
AUC	area under the curve
CT	computed tomography
D-dimer	dimerized plasmin fragment D
GGO	ground-glass opacity
hsCRP	high-sensitivity C-reactive protein
hsTNI	high-sensitivity troponin I
ICU	intensive care unit
IL-6	interleukin 6
LDH	lactate dehydrogenase
MERS	Middle East respiratory syndrome
ML	machine learning
OR	odds ratio
PMM	predictive mean matching
RF	random forest
RMSE	root mean square error
ROC	receiver operating characteristic

SARS   secure acute respiratory syndrome

XGBoost   extreme gradient boosting

Appendix

Multimedia Appendix 1

Supplementary tables S1-S3.

Multimedia Appendix 2

Supplementary information regarding the model development and complete feature distributions.

Footnotes

Contributed by

Authors' Contributions: YC designed the study. LO and JL derived and processed the data. LO, FSB, QL, LH, BZ, JL, PR, and SC interpreted the results. MX and SC performed analyses. YG and SC supervised the study. PR and SC developed the manuscript. YC, LO, and FSB contributed equally. JL (liu_jie0823@163.com), MX (sosolou@126.com), and SC (schen56@uncc.edu) serve as corresponding authors of this study equally.

Conflicts of Interest: None declared.

References

1. Weekly epidemiological update - 12 January 2021. *World Health Organization.* 2021. Jan 12, [2021-03-18]. https://www.who.int/publications/m/item/weekly-epidemiological-update---12-january-2021.

2. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' heel of current strategies to control covid-19. *N Engl J Med.* 2020 May 28;382(22):2158–2160. doi: 10.1056/nejme2009758. [PMCID: PMC7200054] [PubMed: 32329972] [CrossRef: 10.1056/nejme2009758]

3. Wölfel Roman, Corman Victor M, Guggemos Wolfgang, Seilmaier Michael, Zange Sabine, Müller Marcel A, Niemeyer Daniela, Jones Terry C, Vollmar Patrick, Rothe Camilla, Hoelscher Michael, Bleicker Tobias, Brünink Sebastian, Schneider Julia, Ehmann Rosina, Zwirglmaier Katrin, Drosten Christian, Wendtner Clemens. Virological assessment of hospitalized patients with COVID-2019. *Nature.* 2020 May;581(7809):465–469. doi: 10.1038/s41586-020-2196-x. doi: 10.1038/s41586-020-2196-x. [PubMed: 32235945] [CrossRef: 10.1038/s41586-020-2196-x] [CrossRef: 10.1038/s41586-020-2196-x]

4. Luo Yi, Trevathan Edwin, Qian Zhengmin, Li Yirong, Li Jin, Xiao Wei, Tu Ning, Zeng Zhikun, Mo Pingzheng, Xiong Yong, Ye Guangming. Asymptomatic SARS-CoV-2 Infection in Household Contacts of a Healthcare Provider, Wuhan, China. *Emerg Infect Dis.* 2020 Aug;26(8):1930–1933. doi: 10.3201/eid2608.201016. doi: 10.3201/eid2608.201016. [PMCID: PMC7392452] [PubMed:

32330112] [CrossRef: 10.3201/eid2608.201016] [CrossRef: 10.3201/eid2608.201016]

5. Ye F, Xu S, Rong Z, Xu R, Liu X, Deng P, Liu H, Xu X. Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster. *Int J Infect Dis.* 2020 May;94:133–138. doi: 10.1016/j.ijid.2020.03.042. https://linkinghub.elsevier.com/retrieve/pii/S1201-9712(20)30174-0. [PMCID: PMC7129961] [PubMed: 32247826] [CrossRef: 10.1016/j.ijid.2020.03.042]

6. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. *JAMA.* 2020 Apr 07;323(13):1239–1242. doi: 10.1001/jama.2020.2648. [PubMed: 32091533] [CrossRef: 10.1001/jama.2020.2648]

7. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Bärnighausen T, Wang C. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. *Lancet.* 2020 Apr;395(10232):1305–1314. doi: 10.1016/s0140-6736(20)30744-3. [PMCID: PMC7270591] [PubMed: 32247320] [CrossRef: 10.1016/s0140-6736(20)30744-3]

8. Truog Robert D, Mitchell Christine, Daley George Q. The toughest triage - allocating ventilators in a pandemic. *N Engl J Med.* 2020 May 21;382(21):1973–1975. doi: 10.1056/NEJMp2005689. [PubMed: 32202721] [CrossRef: 10.1056/NEJMp2005689]

9. Novel coronavirus pneumonia diagnosis and treatment plan (provisional 7th edition). Webpage in Chinese. *National Health Commission of China.* 2020. Mar 04, [2021-03-18]. http://www.nhc.gov.cn/yzygj/s7652m/202003/a31191442e29474b98bfed5579d5af95.shtml.

10. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, Cooley LA, Dean NC, Fine MJ, Flanders SA, Griffin MR, Metersky ML, Musher DM, Restrepo MI, Whitney CG. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med.* 2019 Oct 01;200(7):e45–e67. doi: 10.1164/rccm.201908-1581st. [PMCID: PMC6812437] [PubMed: 31573350] [CrossRef: 10.1164/rccm.201908-1581st]

11. Neill AM, Martin IR, Weir R, Anderson R, Chereshsky A, Epton MJ, Jackson R, Schousboe M, Frampton C, Hutton S, Chambers ST, Town GI. Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax.* 1996 Oct 01;51(10):1010–6. doi: 10.1136/thx.51.10.1010. https://thorax.bmj.com/lookup/pmidlookup?view=long&pmid=8977602. [PMCID: PMC472650] [PubMed: 8977602] [CrossRef: 10.1136/thx.51.10.1010]

12. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. *China CDC Wkly.* 2020;2(8):113–122. doi: 10.46234/ccdcw2020.032. [CrossRef: 10.46234/ccdcw2020.032]

13. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020 Feb;395(10223):497–506. doi: 10.1016/s0140-6736(20)30183-5. [PMCID: PMC7159299] [PubMed: 31986264] [CrossRef: 10.1016/s0140-6736(20)30183-5]

14. Wadman M. How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. *Science.* 2020 Apr 18; doi: 10.1126/science.abc3208. [CrossRef: 10.1126/science.abc3208]

15. Liu H, Wang R, Qu G. A general report on the systematic anatomy of COVID-19. Article in Chinese. *J Forensic Med.* 2020;36:1–3. doi: 10.12116/j.issn.1004-5619.2020.01.005. [CrossRef: 10.12116/j.issn.1004-5619.2020.01.005]

16. Cholankeril G, Podboy A, Aivaliotis VI, Tarlow B, Pham EA, Spencer SP, Kim D, Hsing A, Ahmed A. High prevalence of concurrent gastrointestinal manifestations in patients with severe acute respiratory syndrome coronavirus 2: early experience from California. *Gastroenterology.* 2020 Aug;159(2):775–777. doi: 10.1053/j.gastro.2020.04.008.

http://europepmc.org/abstract/MED/32283101. [PMCID: PMC7194555] [PubMed: 32283101] [CrossRef: 10.1053/j.gastro.2020.04.008]

17. Jing Yan, Run-Qian Li, Hao-Ran Wang, Hao-Ran Chen, Ya-Bin Liu, Yang Gao, Fei Chen. Potential influence of COVID-19/ACE2 on the female reproductive system. *Mol Hum Reprod.* 2020 Jun 01;26(6):367–373. doi: 10.1093/molehr/gaaa030. http://europepmc.org/abstract/MED/32365180. [PMCID: PMC7239105] [PubMed: 32365180] [CrossRef: 10.1093/molehr/gaaa030]

18. Moriguchi T, Harii N, Goto J, Harada D, Sugawara H, Takamino J, Ueno M, Sakata H, Kondo K, Myose N, Nakao A, Takeda M, Haro H, Inoue O, Suzuki-Inoue K, Kubokawa K, Ogihara S, Sasaki T, Kinouchi H, Kojin H, Ito M, Onishi H, Shimizu T, Sasaki Y, Enomoto N, Ishihara H, Furuya S, Yamamoto T, Shimada S. A first case of meningitis/encephalitis associated with SARS-Coronavirus-2. *Int J Infect Dis.* 2020 May;94:55–58. doi: 10.1016/j.ijid.2020.03.062. https://linkinghub.elsevier.com/retrieve/pii/S1201-9712(20)30195-8. [PMCID: PMC7195378] [PubMed: 32251791] [CrossRef: 10.1016/j.ijid.2020.03.062]

19. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, Sun C, Tang X, Jing L, Zhang M, Huang X, Xiao Y, Cao H, Chen Y, Ren T, Wang F, Xiao Y, Huang S, Tan X, Huang N, Jiao B, Cheng C, Zhang Y, Luo A, Mombaerts L, Jin J, Cao Z, Li S, Xu H, Yuan Y. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell.* 2020 May 14;2(5):283–288. doi: 10.1038/s42256-020-0180-7. doi: 10.1038/s42256-020-0180-7. [CrossRef: 10.1038/s42256-020-0180-7] [CrossRef: 10.1038/s42256-020-0180-7]

20. Liu J, Li S, Liu J, Liang B, Wang X, Wang H, Li W, Tong Q, Yi J, Zhao L, Xiong L, Guo C, Tian J, Luo J, Yao J, Pang R, Shen H, Peng C, Liu T, Zhang Q, Wu J, Xu L, Lu S, Wang B, Weng Z, Han C, Zhu H, Zhou R, Zhou H, Chen X, Ye P, Zhu B, Wang L, Zhou W, He S, He Y, Jie S, Wei P, Zhang J, Lu Y, Wang W, Zhang L, Li L, Zhou F, Wang J, Dittmer U, Lu M, Hu Y, Yang D, Zheng X. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine.* 2020 May;55:102763. doi: 10.1016/j.ebiom.2020.102763. https://linkinghub.elsevier.com/retrieve/pii/S2352-3964(20)30138-9. [PMCID: PMC7165294] [PubMed: 32361250] [CrossRef: 10.1016/j.ebiom.2020.102763]

21. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, Li Y, Guan W, Sang L, Lu J, Xu Y, Chen G, Guo H, Guo J, Chen Z, Zhao Y, Li S, Zhang N, Zhong N, He J, China Medical Treatment Expert Group for COVID-19 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 Aug 01;180(8):1081–1089. doi: 10.1001/jamainternmed.2020.2033. http://europepmc.org/abstract/MED/32396163. [PMCID: PMC7218676] [PubMed: 32396163] [CrossRef: 10.1001/jamainternmed.2020.2033]

22. Wynants L, Van Calster B, Collins Gary S, Riley Richard D, Heinze Georg, Schuit Ewoud, Bonten Marc M J, Dahly Darren L, Damen Johanna A A, Debray Thomas P A, de Jong Valentijn M T, De Vos Maarten, Dhiman Paul, Haller Maria C, Harhay Michael O, Henckaerts Liesbet, Heus Pauline, Kammer Michael, Kreuzberger Nina, Lohmann Anna, Luijken Kim, Ma Jie, Martin Glen P, McLernon David J, Andaur Constanza L, Reitsma Johannes B, Sergeant Jamie C, Shi Chunhu, Skoetz Nicole, Smits Luc J M, Snell Kym I E, Sperrin Matthew, Spijker René, Steyerberg Ewout W, Takada Toshihiko, Tzoulaki Ioanna, van Kuijk Sander M J, van Bussel Bas, van Royen Florien S, Verbakel Jan Y, Wallisch Christine, Wilkinson Jack, Wolff Robert, Hooft Lotty, Moons Karel G M, van Smeden Maarten. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ.* 2020 Apr 07;369:m1328. doi: 10.1136/bmj.m1328. http://www.bmj.com/lookup/pmidlookup?view=long&pmid=32265220. [PMCID: PMC7222643] [PubMed: 32265220] [CrossRef: 10.1136/bmj.m1328]

23. Shamsoddin E. Can medical practitioners rely on prediction models for COVID-19? A systematic review. *Evid Based Dent.* 2020 Sep 25;21(3):84–86. doi: 10.1038/s41432-020-0115-5. http://europepmc.org/abstract/MED/32978532. [PMCID: PMC7517064] [PubMed: 32978532] [CrossRef: 10.1038/s41432-020-0115-5]

24. Struyf Thomas, Deeks Jonathan J, Dinnes Jacqueline, Takwoingi Yemisi, Davenport Clare, Leeflang Mariska Mg, Spijker René, Hooft Lotty, Emperador Devy, Dittrich Sabine, Domen Julie, Horn Sebastiaan R A, Van den Bruel Ann, Cochrane COVID-19 Diagnostic Test Accuracy Group Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database Syst Rev.* 2020 Jul 07;7:CD013665. doi: 10.1002/14651858.CD013665. http://europepmc.org/abstract/MED/32633856. [PMCID: PMC7386785] [PubMed: 32633856] [CrossRef: 10.1002/14651858.CD013665]

25. Quinlan Jr. Induction of decision trees. *Mach Learn.* 1986 Mar;1(1):81–106. doi: 10.1007/BF00116251. [CrossRef: 10.1007/BF00116251]

26. Ho TK. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition; 3rd International Conference on Document Analysis and Recognition; 1995; Montréal, QC. 1995. pp. 278–282. [CrossRef: 10.1109/icdar.1995.598994]

27. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell.* 2020 Jan 17;2(1):56–67. doi: 10.1038/s42256-019-0138-9. http://europepmc.org/abstract/MED/32607472. [PMCID: PMC7326367] [PubMed: 32607472] [CrossRef: 10.1038/s42256-019-0138-9]

28. Tolosi Laura, Lengauer Thomas. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics.* 2011 Jul 15;27(14):1986–94. doi: 10.1093/bioinformatics/btr300. [PubMed: 21576180] [CrossRef: 10.1093/bioinformatics/btr300]

29. Strobl C, Boulesteix A, Augustin T. Unbiased split selection for classification trees based on the Gini Index. *Comput Stat Data Anal.* 2007 Sep;52(1):483–501. doi: 10.1016/j.csda.2006.12.030. [CrossRef: 10.1016/j.csda.2006.12.030]

30. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction. 2nd edition.* Berlin, Germany: Springer; 2009.

31. Bao F. Triaging moderate COVID-19 and other viral pneumonia from routine blood tests. *GitHub.* [2021-03-18]. https://github.com/forrestbao/covid-19.

32. Symptoms of coronavirus. *US Centers for Disease Control and Prevention.* [2021-03-18]. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

33. Garg Shikha, Kim Lindsay, Whitaker Michael, O'Halloran Alissa, Cummings Charisse, Holstein Rachel, Prill Mila, Chai Shua J, Kirley Pam D, Alden Nisha B, Kawasaki Breanna, Yousey-Hindes Kimberly, Niccolai Linda, Anderson Evan J, Openo Kyle P, Weigel Andrew, Monroe Maya L, Ryan Patricia, Henderson Justin, Kim Sue, Como-Sabetti Kathy, Lynfield Ruth, Sosin Daniel, Torres Salina, Muse Alison, Bennett Nancy M, Billing Laurie, Sutton Melissa, West Nicole, Schaffner William, Talbot H Keipp, Aquino Clarissa, George Andrea, Budd Alicia, Brammer Lynnette, Langley Gayle, Hall Aron J, Fry Alicia. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 - COVID-NET, 14 states, March 1-30, 2020. *MMWR Morb Mortal Wkly Rep.* 2020 Apr 17;69(15):458–464. doi: 10.15585/mmwr.mm6915e3. doi: 10.15585/mmwr.mm6915e3. [PMCID: PMC7755063] [PubMed: 32298251] [CrossRef: 10.15585/mmwr.mm6915e3] [CrossRef: 10.15585/mmwr.mm6915e3]

34. Mizuiri S, Ohashi Yasushi. ACE and ACE2 in kidney disease. *World J Nephrol.* 2015 Feb 06;4(1):74–82. doi: 10.5527/wjn.v4.i1.74. https://www.wjgnet.com/2220-6124/full/v4/i1/74.htm. [PMCID: PMC4317630] [PubMed: 25664248] [CrossRef: 10.5527/wjn.v4.i1.74]

35. Lieben L. Alport syndrome: ACE2 administration slows kidney damage. *Nat Rev Nephrol.* 2017 May 13;13(5):261–261. doi: 10.1038/nrneph.2017.36. [PubMed: 28287108] [CrossRef: 10.1038/nrneph.2017.36]

36. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y, Pan S, Zou X, Yuan S, Shang Y. 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 May;8(5):475–481. doi: 10.1016/s2213-2600(20)30079-5. [PMCID: PMC7102538] [PubMed: 32105632] [CrossRef: 10.1016/s2213-2600(20)30079-5]

37. Li Yanlei, Hu Yanbing, Yu Jinyan, Ma Tiangang. Retrospective analysis of laboratory testing in 54 patients with severe- or critical-type 2019 novel coronavirus pneumonia. *Lab Invest.* 2020 Jun;100(6):794–800. doi: 10.1038/s41374-020-0431-6. http://europepmc.org/abstract/MED/32341519. [PMCID: PMC7184820] [PubMed: 32341519] [CrossRef: 10.1038/s41374-020-0431-6]

38. Chen Liang, Li Xiangjie, Chen Mingquan, Feng Yi, Xiong Chenglong. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. *Cardiovasc Res.* 2020 May 01;116(6):1097–1100. doi: 10.1093/cvr/cvaa078. http://europepmc.org/abstract/MED/32227090. [PMCID: PMC7184507] [PubMed: 32227090] [CrossRef: 10.1093/cvr/cvaa078]

39. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, Gong W, Liu X, Liang J, Zhao Q, Huang H, Yang B, Huang C. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol.* 2020 Jul 01;5(7):802–810. doi: 10.1001/jamacardio.2020.0950. http://europepmc.org/abstract/MED/32211816. [PMCID: PMC7097841] [PubMed: 32211816] [CrossRef: 10.1001/jamacardio.2020.0950]

40. Li J, Han T, Woodward M, Anderson CS, Zhou H, Chen Y, Neal B. The impact of 2019 novel coronavirus on heart injury: a systematic review and meta-analysis. *Prog Cardiovasc Dis.* 2020 Apr;63(4):518–524. doi: 10.1016/j.pcad.2020.04.008. http://europepmc.org/abstract/MED/32305557. [PMCID: PMC7160645] [PubMed: 32305557] [CrossRef: 10.1016/j.pcad.2020.04.008]

41. Wu Chaomin, Chen Xiaoyan, Cai Yanping, Xia Jia'an, Zhou Xing, Xu Sha, Huang Hanping, Zhang Li, Zhou Xia, Du Chunling, Zhang Yuye, Song Juan, Wang Sijiao, Chao Yencheng, Yang Zeyong, Xu Jie, Zhou Xin, Chen Dechang, Xiong Weining, Xu Lei, Zhou Feng, Jiang Jinjun, Bai Chunxue, Zheng Junhua, Song Yuanlin. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med.* 2020 Jul 01;180(7):934–943. doi: 10.1001/jamainternmed.2020.0994. http://europepmc.org/abstract/MED/32167524. [PMCID: PMC7070509] [PubMed: 32167524] [CrossRef: 10.1001/jamainternmed.2020.0994]

42. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui DS, Du B, Li L, Zeng G, Yuen K, Chen R, Tang C, Wang T, Chen P, Xiang J, Li S, Wang J, Liang Z, Peng Y, Wei L, Liu Y, Hu Y, Peng P, Wang J, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S, Zhong N, China Medical Treatment Expert Group for Covid-19 Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* 2020 Apr 30;382(18):1708–1720. doi: 10.1056/NEJMoa2002032. http://europepmc.org/abstract/MED/32109013. [PMCID: PMC7092819] [PubMed: 32109013] [CrossRef: 10.1056/NEJMoa2002032]

43. Griffin DO, Jensen A, Khan M, Chin J, Chin K, Saad J, Parnell R, Awwad C, Patel D. Pulmonary embolism and increased levels of d-dimer in patients with coronavirus disease. *Emerg Infect Dis.* 2020 Aug;26(8):1941–1943. doi: 10.3201/eid2608.201477. doi: 10.3201/eid2608.201477. [PMCID: PMC7392455] [PubMed: 32348233] [CrossRef: 10.3201/eid2608.201477] [CrossRef: 10.3201/eid2608.201477]

44. Barnes Betsy J, Adrover Jose M, Baxter-Stoltzfus Amelia, Borczuk Alain, Cools-Lartigue Jonathan, Crawford James M, Daßler-Plenker Juliane, Guerci Philippe, Huynh Caroline, Knight Jason S, Loda Massimo, Looney Mark R, McAllister Florencia, Rayes Roni, Renaud Stephane, Rousseau Simon, Salvatore Steven, Schwartz Robert E, Spicer Jonathan D, Yost Christian C, Weber Andrew,

Zuo Yu, Egeblad Mikala. Targeting potential drivers of COVID-19: neutrophil extracellular traps. *J Exp Med.* 2020 Jun 01;217(6):e20200652. doi: 10.1084/jem.20200652. https://rupress.org/jem/article-lookup/doi/10.1084/jem.20200652. [PMCID: PMC7161085] [PubMed: 32302401] [CrossRef: 10.1084/jem.20200652]

45. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020 Mar;395(10229):1054–1062. doi: 10.1016/s0140-6736(20)30566-3. [PMCID: PMC7270627] [PubMed: 32171076] [CrossRef: 10.1016/s0140-6736(20)30566-3]

46. Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and multiorgan response. *Curr Probl Cardiol.* 2020 Aug;45(8):100618. doi: 10.1016/j.cpcardiol.2020.100618. http://europepmc.org/abstract/MED/32439197. [PMCID: PMC7187881] [PubMed: 32439197] [CrossRef: 10.1016/j.cpcardiol.2020.100618]

47. Wang T, Du Z, Zhu F, Cao Z, An Y, Gao Y, Jiang B. Comorbidities and multi-organ injuries in the treatment of COVID-19. *Lancet.* 2020 Mar;395(10228):e52. doi: 10.1016/s0140-6736(20)30558-4. [PMCID: PMC7270177] [PubMed: 32171074] [CrossRef: 10.1016/s0140-6736(20)30558-4]

Figures and Tables

Figure 1
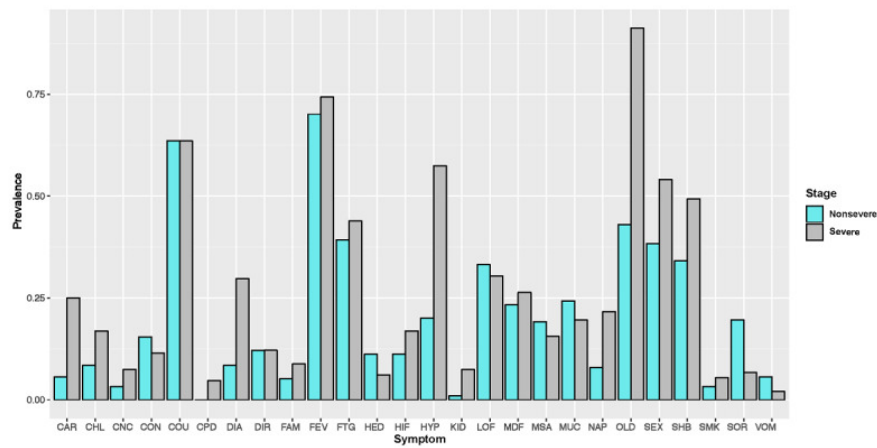
Comparison of clinical features of patients with nonsevere and severe COVID-19. Note that because these features were binary, the y-axis indicates the prevalence of a positive result. CAR: cardiovascular disease; CHL: chills and shaking; CNC: cancer; CON: contact with patients with COVID-19; COU: coughing; CPD: chronic obstructive pulmonary disease; DIA: diabetes; DIR: diarrhea; FAM: family members with COVID-19; FEV: fever; FTG: fatigue; HED: headache; HIF: high fever; HYP: hypertension; KID: renal disease; LOF: low fever; MOF: medium fever; MSA: muscle ache; MUC: phlegm; NAP: loss of appetite; OLD: older age; PREV: prevalence; SEX: male sex; SHB: chest congestion; SMK: history of smoking; SOR: sore throat; VOM: vomiting.

Figure 2
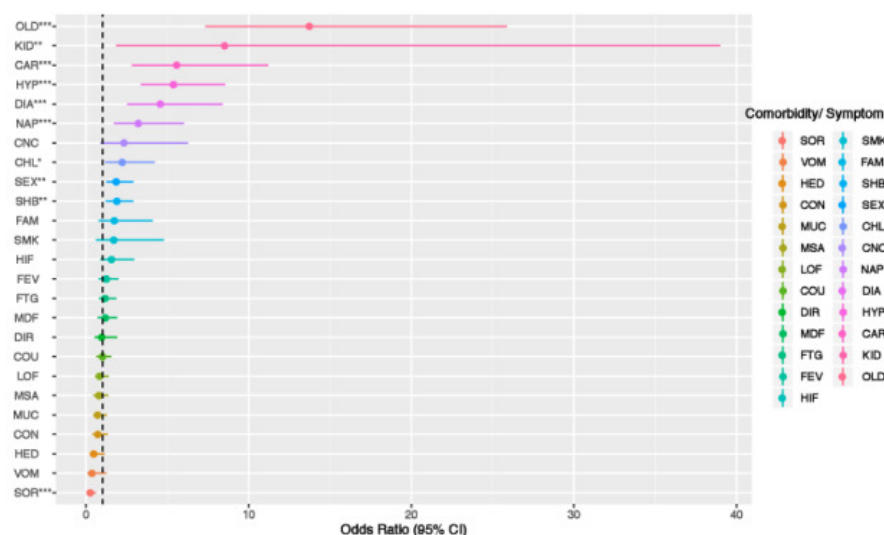
Forest plot of the importance of clinical features of patients with nonsevere and severe clinical types of COVID-19. Chronic obstructive pulmonary disease is not included because only patients with severe COVID-19 showed it as a comorbidity. The threshold for a feature to be positively or negatively associated with severe COVID-19 was 1 (dashed line), not 0. CAR: cardiovascular disease; CHL: chills and shaking; CNC: cancer; CON: contact with patients with COVID-19; COU: coughing; CPD: chronic obstructive pulmonary disease; DIA: diabetes; DIR: diarrhea; FAM: family members with COVID-19; FEV: fever; FTG: fatigue; HED: headache; HIF: high fever; HYP: hypertension; KID: renal disease; LOF: low fever; MOF: medium fever; MSA: muscle ache; MUC: phlegm; NAP: loss of appetite; OLD: older age; SEX: male sex; SHB: chest congestion; SMK: history of smoking; SOR: sore throat; VOM: vomiting. *P<.05, **P<.01; ***P<.001 from the 2×2 contingency table for each feature.

Figure 3
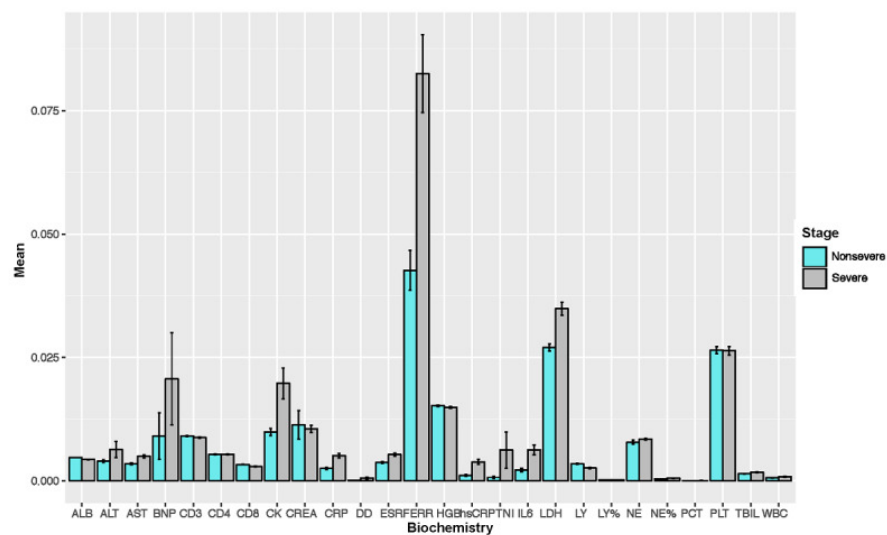
Comparison of laboratory testing features of patients with nonsevere and severe COVID-19. Values shown on the y-axis were obtained after feature scaling and are between 0 and 1. The error bars represent the standard error of each laboratory testing feature. ALB: albumin; ALT: alanine transaminase; AST: aspartate aminotransferase; BNP: Brain natriuretic peptide; CK: creatine kinase; CREA: creatinine; CRP: C-reactive protein; DD: dimerized plasmin fragment D; ESR: erythrocyte sedimentation rate; FERR: ferritin; HGB: hemoglobin; hsCRP: high-sensitivity C-reactive protein; TNI: troponin I; IL6: interleukin 6; LDH: lactate dehydrogenase; LY: lymphocyte; LY%: percent of lymphocytes; NE: neutrophil; NE% percent of neutrophils; PCT: procalcitonin; PLT: platelet; TBIL: total bilirubin; WBC: white blood cell.

Table 1

Performance of the random forest model with multimodal features. The results are based on 100 runs. In each run, 80% of the data was randomly selected as the training set and 20% as the testing set. The table shows the model performance on the testing set only.

Feature and performance metric	Median	Minim
Clinical (%)		
Accuracy	94.59	81.08
Sensitivity	>99	80.95
Specificity	93.75	75.00
F1 score	97.30	82.93
AUC[a]	90.20	82.90
Laboratory testing (%)		
Accuracy	97.22	93.06
Sensitivity	>99	94.59
Specificity	96.97	83.33
F1 score	97.89	94.74
AUC	97.10	92.90
Multimodal (%)		
Accuracy	97.22	91.67
Sensitivity	>99	90.00
Specificity	94.44	75.00
F1 score	97.78	97.22
AUC	97.40	92.20

[a]AUC: area under the curve

Figure 4
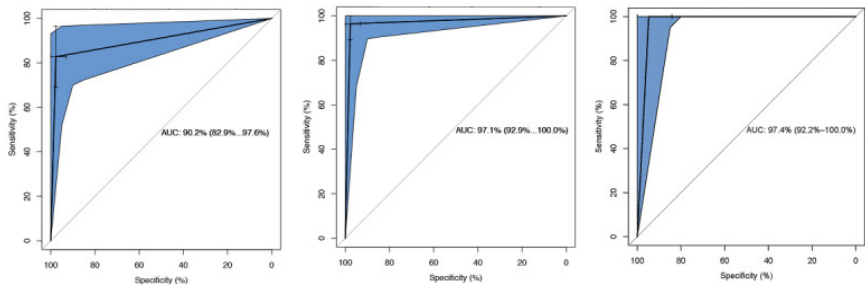
ROC curves from the random forest models based on clinical, laboratory testing, and multimodal features. Left: the symptom feature as the sole input; middle: the laboratory testing feature as the sole input; right: both features combined as the input. AUC: area under the curve; ROC: receiving operator characteristic.