Risk Factors Analysis of COVID-19 Patients with ARDS and Prediction Based on Machine Learning

Wan Xu[1#], Nan-Nan Sun[1#], Hai-Nv Gao[3], Zhi-Yuan Chen[1], Ya Yang[2], Bin Ju[1]*, Ling-Ling Tang [3]*

[1] Hangzhou Wowjoy Information Technology Co., Ltd, Hangzhou, China

[2] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Centre for Infectious Diseases, Collaborative Innovation Centre for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang Province, 310003, China

[3] Department of Infectious Diseases, ShuLan (Hangzhou) Hospital Affiliated to Zhejiang Shuren University Shulan International Medical College, Hangzhou, China

[#]Equal contributors.

*To whom correspondence should be addressed.

Bin Ju, Email：bin.ju@wowjoy.com; Phone Number:13516706325

Ling-Ling Tang, Email：lingling.tang@shulan.com; Phone Number:13588460108

Abstract

COVID-19 is a newly emerging infectious disease, which is generally susceptible to human beings and has caused huge losses to people's health. Acute respiratory distress syndrome (ARDS) is one of the common clinical manifestations of severe COVID-19 and it is also responsible for the current shortage of ventilators worldwide. This study aims to analyze the clinical characteristics of COVID-19 ARDS patients and establish a diagnostic system based on artificial intelligence (AI) method to predict the probability of ARDS in COVID-19 patients. We collected clinical data of 659 COVID-19 patients from 11 regions in China. The clinical characteristics of the two groups were elaborately compared and both traditional machine learning algorithms and deep learning-based methods were used to build the prediction models. Results indicated the median age of ARDS patients was 56.5 years old, which was significantly older than those with non-ARDS by 7.5 years. Male and patients with BMI>25 were more likely to develop ARDS. The clinical features of ARDS patients included cough (80.3%), polypnea (59.2%), lung consolidation (53.9%), secondary bacterial infection (30.3%),

and comorbidities such as hypertension (48.7%). Abnormal biochemical indicators such as lymphocyte count, leukocyte counting, CK, NLR, AST, LDH, and CRP were all strongly related to the aggravation of ARDS. Furthermore, through various AI methods for modeling and prediction effect evaluation based on the above risk factors, decision tree achieved the best AUC, sensitivity, and specificity in identifying the mild patients who were easy to develop ARDS, which undoubtedly helps to optimize the treatment strategy, reduce mortality, and relieve the medical pressure.

Keywords ARDS, COVID-19, Clinical features, Risk factors, Machine learning

1 Introduction

The coronavirus disease 2019 (COVID-19) is an acute infectious pneumonia caused by a severe acute respiratory sysdrom-coronavirus-2 (SARS-CoV-2) infection previously unknown to humans. Spreading mainly through the droplet route and close contact, the virus causes mild symptoms in the majority of cases, the most common being: fever, dry cough, and fatigue[1, 2].

The disease has the characteristics of fast transmission and strong infectivity[3]. Since the outbreak in early December 2019 in Wuhan, China, it has rapidly developed into a worldwide pandemic, with more than 3 million patients confirmed to have been diagnosed with the disease in more than 200 countries, and the number of infected people is probably much higher. As of April 30, 2020, 217769 people died of COVID-19 infection. Despite the public health responses aimed at containing the disease and delaying its spread; during the courses of treatment, due to the large increase in the demand for hospital beds and the shortage of medical equipment, coupled with the lack of specific medicine, patients with basic diseases or old age are more likely to progress to severe disease, with ARDS and inflammatory storms, leading to death. Recent reports show that 14.1-33.0% of COVID-19 infected patients are prone to develop into severe cases, and the mortality rate of critical cases is 61.5%, increasing sharply with age and underlying comorbidities[4-7]. Furthermore, medical staff may also be infected, which makes many countries face critical care crisis. COVID-19 poses an important and urgent threat to global health.

Acute Respiratory Distress Syndrome (ARDS) is a common and devastating critical illness[8]. It has been reported that 67% of COVID-19 patients with the severe illness

have developed ARDS, which is the main cause of death[9]. However, in the early stage of onset, quite a few patients have no obvious clinical symptoms, so it is difficult to judge until ARDS occurs. Predicting which patients are more likely to develop ARDS, and as such to face a greater risk of complications including death is particularly important in a novel and accelerating outbreak[10]. It would be useful in evaluation or prediction the public health burden or resources demand in a large scale e.g. in a city or a province.

Artificial intelligence (AI) has begun to tackle these difficult challenges in healthcare and it can provide clinical decision support if used carefully[11]. Currently, the prediction models of COVID-19 reported mainly focus on epidemics trend, early screening, CT diagnosis, and prognosis of COVID-19 patients[12-15]. Few models have been studied for early identification of patients who are most likely to develop ARDS and recommending interventions. Xiang Bai et al. established a Long Short-Term Memory (LSTM) model by combining 75 clinical features and a quantitative CT sequence data obtained at different times to predict the malignant progression of COVID-19, which achieved an AUC of 0.954[16]. Xiangao Jiang et al. used traditional machine learning methods such as decision tree(DT), random forest(RF), and support vector machine(SVM) to predict disease progression to ARDS in COVID-19 patients, with the overall accuracy of 70%-80%[10]. This study was a small sample prediction model of only 53 patients, so the prediction accuracy was slightly lower. The most-reported predictors of severe progression in patients with COVID-19 included age, sex, features derived from computed tomography scans, C reactive protein, lactic dehydrogenase, and lymphocyte count. C index estimates of these models ranged from 0.85 to 0.98[17]. However, most reports did not include a description of the study population or intended use of the models and were rated at high risk of bias at the same time. To reduce the mortality rate of COVID-19 and alleviate the shortage of medical resources, we developed the COVID-19 ARDS clinical decision support system using deep learning algorithms and transplanted it into electronic medical records(EMR) to assist doctors in identifying severe patients early and giving active treatment.

2 Results

2.1 Characteristics of COVID-19 patients

Tables 1 to 3 lists the distribution of various parameters including demographic, epidemiology and clinical characteristics of the COVID-19 ARDS and non-ARDS populations.

- Demographics and Epidemiology

In this study, we collected a total of 659 patients from Wuhan and non-Wuhan areas who were confirmed with COVID-19，of which 76 patients (11.5%) developed ARDS. 447 patients (70.9%) had contact with infected persons and 50.9% had a family infection. The median incubation period was 5 days (interquartile range, 3 to 9) and the average time from onset to ARDS and admission to ARDS were 10 days and 3 days, respectively. The median age of the patients was 50 years (interquartile range, 37 to 62) and 50.4% of the patients were male. Patients with ARDS were significantly older than those with non-ARDS by a median of 7.5 years (56.5 years vs. 49 years) and male patients (76.3%) were more likely to develop ARDS. More than 50% of ARDS patients had a BMI greater than 25. However, the exposure histories of the two groups were similar, (Table 1).

- Clinical Characteristics and Underlying Diseases

On severity evaluation at admission, 75.4% of COVID-19 patients were assessed as common type while among the patients with ARDS, 80.3% were evaluated as severe or critical. The most common clinical symptoms of COVID-19 patients at the time of onset were fever (66.6%), cough (68.7%), expectoration (39.6%), fatigue (34.2%) and dry cough (29.6%); encephalopathy (0.5%), hemoptysis (1.6%), vomiting (3.0%) and stuffy nose (3.8%) were uncommon. Compared with non-ARDS patients, ARDS patients had a higher frequency of coughing (80.3% vs. 67.2%) and dyspnea (59.2% vs. 11.6%). The median temperature was 37.4℃, and ARDS patients were 0.5 ℃ higher than that of non-ARDS patients (37.9℃vs.37.4℃), which was statistically significant ($P<0.001$).

Overall, the presence of any comorbidities was more common among ARDS patients than no-ARDS (56.6% vs. 39.8%). Patients with ARDS had a much higher incidence of hypertension (48.7% vs.23%) and diabetes (17.8% vs.9.5%). Two of the five patients infected with other viruses developed ARDS. ARDS also occurred in one patient treated with immunosuppressive agents, (Table 2).

- Radiologic, Laboratory Findings and Complications

Table 3 shows the radiologic, laboratory findings on admission and complications. 74.7% of the patients presented ground-glass shadows on chest CT images, and 28.3% of the patients presented consolidation. The above two imaging features accounted for a higher proportion of patients with ARDS than non-ARDS patients, which were 80.8% vs 73.9% and 53.9% vs 24.7%, respectively. The median number of consolidation in ARDS patients was two.

Within 48 hours of admission, lymphocytopenia was present in 23.4% of the patients and leukopenia in 24.8%. However, among ARDS patients, 19.8% had an increase in the white blood cell count, which indicated that ARDS patients had a secondary infection. The ratio of neutrophils to lymphocytes was greater than three in 45.3% of COVID-19 patients and 82.7% in ARDS patients with a Median of 6.11. 47.7% and 32.2% of patients had elevated levels of C-reactive protein and lactate dehydrogenase, respectively. In a small number of patients, levels of alanine aminotransferase (ALT), glutamate aminotransferase (AST), creatine kinase (CK) and D-dimer were elevated. Laboratory abnormalities were more severe in ARDS patients than in non-ARDS patients. Besides, the medians of myoglobin and fasting glucose in ARDS patients were 85.9μg/L and 8.1mmol/L respectively, which exceeded the normal reference range and was significantly different from the non-ARDS group.

During hospitalization, 91.3% of patients were diagnosed with pneumonia, and there was no statistical difference between the ARDS group and non-ARDS group. However, patients with ARDS had a higher incidence of shock and secondary bacterial infection (5.5% and 30.3%) than those with non- ARDS (0 and 4.3%), and 45.2% of them were admitted to ICU, (Table 3).

2.2 Prediction of risk factors for COVID-19 ARDS

After removal of variables with missing rate >20%, a total of 98 variables consisting of demographic, epidemiology, clinical symptoms, underlying diseases, complication, CT image features and laboratory results were extracted from the structured and unstructured data of electronic medical record (EMR) according to literature reviews and expert clinician opinions. Then, we selected 22 significant risk factors related to COVID-19 by means of SPSS single factor analysis. Among all risk factors, severity evaluation at admission (odds ratio [OR], 16.396; 95% CI, 6.882-39.065; P<0.001), gender (OR, 3.634; 95% CI, 2.090-6.219； P<0.001), age (≥70 year) (OR， 14.258; 95% CI, 3.032-67.049; P<0.001), BMI (<23 vs. >25) (OR, 3.145; 95% CI, 1.635-6.052；

P<0.001), temperature (>39℃) (OR, 6.496; 95% CI, 2.667-15.820; P<0.001), shortness of breath (OR, 11.302; 95% CI, 6.532-18.659; P<0.001), hypertension (OR, 3.527; 95% CI, 2.159-5.761; P<0.001), secondary bacterial infection (OR, 9.686; 95% CI, 5.146-18.323; P<0.001), lung consolidation (OR, 3.575; 95% CI, 2.187-5.845; P<0.001), lymphocyte count (OR, 0.149; 95% CI, 0.080-0.277; P<0.001), leukocyte counting (OR, 12.215; 95% CI, 5.303-28.135; P<0.001), neutrophils/lymphocytes ratio (NLR) (<3 vs. ≥3) (OR, 7.069; 95% CI, 3.798-13.158; P<0.001), AST (≤40 vs. >40 U/L) (OR, 5.898; 95% CI, 3.452-10.075; P<0.001), CK (≤185 vs.>185 U/L) (OR, 6.054; 95% CI, 3.222-11.374; P<0.001), lactate dehydrogenase (LDH) (≤250 vs. >250 U/L) (OR, 8.277; 95% CI, 4.643-14.755; P<0.001), C-reactive protein (CRP) (≤10 vs.>10 mg/L) (OR, 5.557; 95% CI, 3.033-10.182; P<0.001) were strongly correlated with ARDS, (Table 4).

2.3 Development and verification of predictive models

Based on the above results of univariate analysis and the previous clinically-related studies, we determined 14 risk factors including severity evaluation at admission, gender, age, temperature, shortness of breath, secondary bacterial infection, lung consolidation, lymphocyte count, leukocyte counting, CK, NLR, AST, LDH, and CRP as inputs to the model to evaluate whether COVID-19 patients would develop ARDS, (Table 5). We tried 5 models, including logistic regression (LR), random forest (RF), support vector machine (SVM), decision tree (DT) and deep neural networks (DNN), and the performance of five models was evaluated by assessing the receiver operating characteristic (ROC) curves (Figure 1), the classification accuracy, sensitivity and specificity on the external test set (Table 6). The results showed that decision tree (DT) performed best for ARDS prediction with AUC of 0.99, sensitivity of 0.92， and specificity of 1.00， respectively ; other models, less so.

3 Discussion

In this study, we comprehensively compared the clinical characteristics of all confirmed COVID-19 patients with and without ARDS， and determined 14 features for modeling. All included variables were strongly correlated with disease progression. Age (>70 years), gender, as well as severity evaluation are recognized risk factor for developing ARDS in COVID-19 patients [18]. Clinical manifestation such as fever, shortness of breath and lung consolidation reflect the progression of COVID-19[19]. Viral infections predispose patients to secondary bacterial infections, which often have a

more severe clinical course. Secondary bacterial infection has been considered as a critical risk factor for the severity and mortality rates of COVID-19 despite antimicrobial therapies [20,21]. Lymphopenia, leukocytosis, high concentrations of CRP and LDH may indicate severe acute lung inflammatory reaction and cell damage [22-24], which has been reported to be risk factors for severe patients with COVID-19 [25]. AST is a marker of acute liver injury. Studies have found that abnormal liver tests in patients with COVID-19 were associated with the progression to severe pneumonia and the detrimental effects on liver injury mainly related to lopinavir/ritonavir used during hospitalization, which should be monitored and evaluated frequently [26, 27]. NLR is an indicator of systemic inflammation [28], mainly seen in tumor-related diseases, autoimmune diseases, bacterial infectious pneumonia and tuberculosis [29-32]. It was reported that COVID-19 infection-triggered inflammation increased NLR and the increase in NLR was significantly associated with poor clinical outcomes of COVID-19 patients [33]. We found that CK is a high-risk factor for ARDS. On the one hand, it may be associated with heart injury in critically ill patients with COVID-19 [34]. On the other hand, this indicator was related to rhabdomyolysis [35-36].Several cases of rhabdomyolysis were reported in COVID-19 severe patients, with a marked increase of CK [37-39].

We tried five algorithms for modeling and finally the decision trees performed best. In clinical prediction research, decision tree is frequently designed to build binary classifiers, such as cancer prediction/prognosis [40]. As a method used in machine learning, it is nonparametric, makes fewer data assumptions and it can accommodate collinear independent variables [41]. It is also less sensitive to outliers and more robust to high-dimensional data, which possess many independent variables relative to outcomes [42]. And its main advantage is its simple structure, which allows for better extracting classification rules and interpretation. Our model currently consists of fourteen clinical variables, which are all relatively inexpensive and easy to be obtained directly from clinical symptoms and routine laboratory tests. And the system shows good sensitivity, specificity and AUC in the external test cohort. Compared with the results of Jiang et al.[10], the overall accuracy of our model is higher (70% vs 90%).

Our study has several strengths: First, we have successfully used a machine-learning algorithm to analyze clinical datasets and developed a diagnosis aid system, which has been deployed in electronic medical records for early identification of ARDS in

COVID-19 patients. By submitting clinical information online, medical staff can triage patients at hospital admission using the predicted risk factors and arrange patient treatment plans accordingly, ensuring patients receive treatment early and medical resources can be efficiently allocated. Secondly, to ensure the reliability of the conclusion, we use data from multi-centers with large samples for modeling and verification. Third, we find that CK (>185 U/L) and NLR is strongly correlated with ARDS, which might be the new potential early identification biomarkers in COVID-19 severe patients.

There are still some deficiencies in our study and we have a lot of works to do in the future. Firstly, although we collected data from 659 COVID-19 patients in multiple centers, samples available for ARDS were limited. Secondly, we did not collect CT images data, and the quantitative information of CT diagnostic data was not detailed enough. Thirdly, It has been reported that D-dimer is a risk factor for COVID-19 severity and the development of ARDS, but in our study, due to a large number of missing data, similar conclusions were not reached. Finally, it is of great clinical value to study the intervention measures and prognosis of COVID-19 patients before and after the development of ARDS and integrate them into the diagnostic system to achieve personalized recommendations of treatment measures.

We retrospectively analyzed the clinical characteristics of COVD-19 patients with and without ARDS from Zhejiang Province and Wuhan and identified fourteen risk factors. The traditional early warning scores and organ damage markers cannot help predict the development of ARDS in COVID-19 patients. Therefore, we have further developed an early warning model based on machine learning, which can predict the incidence of ARDS with an accuracy rate of 90%. We have deployed it to the infectious disease electronic medical record system to assist doctors in diagnosis and treatment.

4 Method

4.1 Patient Population and Clinical Data

Data on a total of 659 consecutive COVID-19 patients from January 22 to April 1, 2020, were retrospectively collected in hospitals from 11 regions：NingBo, ZhouShan, HuBei, Lishui, Jiaxin, HangZhou, TaiZhou, DongYang, ShaoXing, WenZhou. The age of the patient ranged from 14 and 90 years old. All patients were diagnosed by positive

tests of severe acute respiratory sysdrom-coronavirus-2(SARS-CoV-2) nucleic acids, according to WHO interim guidance. Clinical information, including demographic, comorbidities, epidemiological history of exposure to COVID-19, vital sign, clinical symptoms, biochemical indices, blood routine, infection-related biomarkers, CT findings, therapeutic measures, and all the time information from onset to admission were collected from routine clinical practice. The date of disease onset was defined as the day when symptoms (i.e. fever, dry cough, expectoration, polypnea, fatigue, myalgia, pharyngalgia, dyspnea, headache, vomiting) first appeared. ARDS was defined according to the Berlin definition. Severity evaluation criteria on admission was based on the *Guidelines for the Diagnosis and Treatment of Novel Coronavirus (2019-nCoV) Infection (Trial Version 7)*, which is a comprehensive evaluation index with important clinical diagnostic value. Patient with following conditions are judged as secondary bacterial infection: bacteria cultured in sterile sites, or fever unrelated to the initial disease, accompanied by elevated CRP. This study was approved by the Ethics Committee of Shulan Hangzhou Hospital.

4.2 Data Analysis

Continuous variables were expressed as medians and interquartile ranges or simple ranges, as defined by experts. Categorical variables were summarized as counts and percentages. We assessed differences between ARDS and non-ARDS using Two-Sample T test or Mann–Whitney U test depending on parametric or non-parametric data for continuous variables and the Chi-square for categorical variables. We considered a P-value less than 0.05 as statistically significant. All statistical analysis was performed using IBM SPSS Ver. 19.0.

4.3 Machine Learning Model Establishment and Evaluation

Datasets: All data was divided into three separate parts with no overlapping topics: training, validation, and external test sets.

For COVID-19 ARDS prediction

- Training and validation datasets: 167 subjects were assigned to the training and validation datasets following a 5:1 ratio, including 125 non-ARDS and 42 ARDS cases from 11 regions in Wuhan and Zhejiang, further cross-validated 5 times. These datasets were used to train model parameters.

- External test dataset: There were 73 non-ARDS and 17 ARDS cases from 11 regions in Wuhan and Zhejiang. This dataset was used to evaluate and analyze the performances of different models to select the best model for AI system.

Algorithms： Four traditional machine learning algorithms (decision trees, random forests, support vector machines and logistic regression) and one deep learning method (deep neural networks, DNN) were used to develop the ARDS prediction model in COVID-19 patients. The pipeline of the DNN model is shown in Figure S1. The input data is a 4-dimensional vector, containing the clinical data of patients. The DNN model employed in this study is a 3-layer network structure with the hidden neurons of 64, 32, 8 respectively. A softmax layer is added at the top of the network to output the probability of ARDS occurrence. The performance of these models was evaluated by assessing the receiver operating characteristic (ROC) curves, the classification accuracy, sensitivity and specificity on the external test set.

4.4 Application Development

The best algorithm for ARDS risk prediction was embedded into EMR and could be accessed and evaluated via a browser. The Anaconda Distribution (Anaconda Inc, Austin, Texas), Visual Studio Code version 1.45.1 (Microsoft, Redmond, Washington), and Python version 3.6 (Python Software Foundation, Wilmington, Delaware) were used for data analysis, model creation, and web application development.

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382(18):1708-1720.

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497-506.

3. Ding Q, Lu P, Fan Y, Xia Y, Liu M. The clinical characteristics of pneumonia patients coinfected with 2019 novel coronavirus and influenza virus in Wuhan, China. Journal of Medical Virology. 2020.

4. Yang X, Yu Y, Xu J, Shu H, Xia Ja, 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. The Lancet Respiratory Medicine. 2020;8(5):475-481.

5. Liu W, Tao Z-W, Wang L, Yuan M-L, Liu K, Zhou L, et al. Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. Chinese Medical Journal. 2020;133(9):1032-1038.

6. Zhao X-Y, Xu X-X, Yin H-S, Hu Q-M, Xiong T, Tang Y-Y, et al. Clinical characteristics of patients with 2019 coronavirus disease in a non-Wuhan area of Hubei Province, China: a retrospective study. BMC Infectious Diseases. 2020;20(1):311.

7. 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. Investigative Radiology. 2020;55(6):327-331.

8. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, Patterns of Care, and Mortality for Patients with Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. JAMA. 2016;315(8):788-800.

9. 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(5):475-481.

10. Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. Computers, Materials & Continua. 2020;62(3):537-551.

11. Shortliffe EH, Sepulveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA. 2018;320(21):2199-200.

12. 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.

13. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis. 2020. doi: 10.1183/13993003.00775-2020.

14. Meylan S, Akrour R, Regina J, Bart PA, Dami F, Calandra T. An Early Warning Score to predict ICU admission in COVID-19 positive patients. J Infect. 2020. doi: 10.1016/j.jinf.2020.05.047

15. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165-174.

16. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Chen Q, et al. Predicting COVID-19 malignant progression with AI techniques. medRxiv. 2020:2020.03.20.20037325.

17. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020;369:m1328.

18. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. J Infect. 2020, 81: e16-e25

19. 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. doi: 10.1001/jamainternmed.2020.0994

20. Mirzaei R, Goodarzi P, Asadi M, Soltani A, Aljanabi HAA, Jeda AS, et al. Bacterial co-infections with SARS-CoV-2. IUBMB Life. 2020. doi: 10.1002/iub.2356

21. Manna S, Baindara P, Mandal SM. Molecular pathogenesis of secondary bacterial infection associated to viral infections including SARS-CoV-2. J Infect Public Health. 2020. doi: 10.1016/j.jiph.2020.07.00

22. Agassandian M, Shurin GV, Ma Y, Shurin MR. C-reactive protein and lung diseases. Int J Biochem Cell Biol. 2014; 53:77-88.

23. Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci. 2020;63(3):364-74.

24. Li X, Wang L, Yan S, Yang F, Xiang L, Zhu J, et al. Clinical characteristics of 25 death cases with COVID-19: A retrospective review of medical records in a single medical center, Wuhan, China. Int J Infect Dis. 2020; 94:128-32.

25. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol. 2020. doi: 10.1016/j.jaci.2020.04.006

26. Cai Q, Huang D, Yu H, Zhu Z, Xia Z, Su Y, et al. COVID-19: Abnormal liver function tests. J Hepatol. 2020. doi: 10.1016/j.jhep.2020.04.006

27. Fan Z, Chen L, Li J, Cheng X, Yang J, Tian C, et al. Clinical Features of COVID-19-Related Liver Functional Abnormality. Clin Gastroenterol Hepatol. 2020;18(7):1561-6.

28. Guthrie GJ, Charles KA, Roxburgh CS, Horgan PG, McMillan DC, Clarke SJ. The systemic inflammation-based neutrophil-lymphocyte ratio: experience in patients with cancer. Crit Rev Oncol Hematol. 2013;88(1):218-30.

29. Shimoyama Y, Umegaki O, Inoue S, Agui T, Kadono N, Minami T. The Neutrophil to Lymphocyte Ratio Is Superior to Other Inflammation-Based Prognostic Scores in Predicting the Mortality of Patients with Pneumonia. Acta Med Okayama. 2018;72(6):591-3.

30. Jeon Y, Lee WI, Kang SY, Kim MH. Neutrophil-to-Monocyte-Plus-Lymphocyte Ratio as a Potential Marker for Discriminating Pulmonary Tuberculosis from Nontuberculosis Infectious Lung Diseases. Lab Med. 2019;50(3):286-91.

31. Ying HQ, Deng QW, He BS, Pan YQ, Wang F, Sun HL, et al. The prognostic value of preoperative NLR, d-NLR, PLR and LMR for predicting clinical outcome in surgical colorectal cancer patients. Med Oncol. 2014;31(12):305.

32. Uslu AU, Kucuk A, Sahin A, Ugan Y, Yilmaz R, Gungor T, et al. Two new inflammatory markers associated with Disease Activity Score-28 in patients with rheumatoid arthritis: neutrophil-lymphocyte ratio and platelet-lymphocyte ratio. Int J Rheum Dis. 2015;18(7):731-735.

33. Yang AP, Li HM, Tao WQ, Yang XJ, Wang M, Yang WJ, et al. Infection with SARS-CoV-2 causes abnormal laboratory results of multiple organs in patients. Aging (Albany NY). 2020;12. doi.org/10.18632/aging.103255

34. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, et al. Association of Cardiac Injury with Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. JAMA Cardiol. 2020. doi:10.1001/jamacardio.2020.0950

35. Chan KH, Farouji I, Abu Hanoud A, Slim J. Weakness and elevated creatinine kinase as the initial presentation of coronavirus disease 2019 (COVID-19). Am J Emerg Med. 2020. doi: 10.1016/j.ajem.2020.05.015

36. Cabral BMI, Edding SN, Portocarrero JP, Lerma EV. Rhabdomyolysis. Dis Mon. 2020:101015.

37. Suwanwongse K, Shabarek N. Rhabdomyolysis as a Presentation of 2019 Novel Coronavirus Disease. Cureus. 2020;12(4): e7561.

38. Jin M, Tong Q. Rhabdomyolysis as Potential Late Complication Associated with COVID-19. Emerg Infect Dis. 2020;26(7).

39. Gefen AM, Palumbo N, Nathan SK, Singer PS, Castellanos-Reyes LJ, Sethna CB. Pediatric COVID-19-associated rhabdomyolysis: a case report. Pediatr Nephrol. 2020. doi:10.1007/s00467-020-04617-0

40. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015; 13:8-17.

41. Visweswaran S, Ferreira A, Ribeiro GA, Oliveira AC, Cooper GF. Personalized Modeling for Prediction with Decision-Path Models. PLoS One. 2015;10(6): e0131022.

42. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. Adv Exp Med Biol. 2011; 696:191-9.

Acknowledgments

We would like to thank the medical staffs in the Department of Infectious Disease for their support in providing information about patients. We also thank the patients for their willingness to participate. Finally, we thank Prof. Peilin Yu and Dr. Lin Wei for their generous help. This work was supported by National Science and Technology Major Project (No.2017ZX10204401) and COVID-19 Emergency Research Project of Zhejiang Provincial Department of Science and Technology (2020C03123-1).

Author contributions

W.X. wrote the main manuscript text and prepared tables 1–4. N.N.S. and W.X. prepared tables5-6 and all figures. H.N.G. provided original dataset. Z.Y.C. and Y.Y.

modified the main manuscript. B.J. and L.L.T. were the corresponding author of the article and guided the research. All authors reviewed the manuscript.

Additional Information

- **Ethics approval and consent to participate**

This study has been approved by the ethics committee of ShuLan (Hangzhou) Hospital. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written informed consents were signed during hospitalization. The data used in this study were anonymized before its use.

- **Conflict of Interest Statement**

The authors declare that they have no conflict of interest.

- **Source of Funding**

The results reported herein correspond to specific aims of grant 2020C03123-1 to investigator Ling-Ling Tang from COVID-19 Emergency Research Project of Zhejiang Provincial Department of Science and Technology. This work was also supported by grant No.2017ZX10204401 from National Science and Technology Major Project.

Figure Legends

Figure 1 ROC of the five algorithms on the external testing dataset.

Figure S1 Deep Learning (DL) algorithm framework

Tables

Table 1 Demographic and Epidemiology of the Study Patients

Characteristics	Patients (N=659)	ARDS Patients (N=76)	Non-ARDS Patients (N=583)	P value
Demography				
Age				
Median(IQR)- years	50.0 (37.0-62.0)	56.5(47.5-67.8)	49.0(36.0-60.0)	0.000
Distribution - no./total no. (%)				0.000

14-30 years	70/659(10.6%)	2/76(2.6%)	68/583(11.7%)	
30-70 years	545/65(82.7%)	61/76(80.3%)	484/583(83%)	
≥70 years	44/659(6.7%)	13/76(17.1%)	31/583(5.3%)	
Gender				0.000
Male - no./total no. (%)	332/659(50.4%)	58/76(76.3%)	274/583(46.9%)	
Female - no./total no. (%)	327/659(49.6%)	18/76(23.7%)	309/583(53.1%)	
BMI				
Median(IQR)	23.9(21.5-25.9)	25.5(21.5-25.7)	23.7(23.2-27.7)	0.001
Distribution - no./total no. (%)				0.007
<23	229/591(38.7%)	14/65(21.5%)	215/526(40.9%)	
23-25	161/591(27.2%)	17/65(26.2%)	144/526(27.4%)	
>25	202/592(34.1%)	34/65(52.3%)	167/526(31.7%)	
Medical staff - no./total no. (%)	9/659(1.4%)	1/76(1.3%)	8/583(1.4%)	0.968
Pregnancy history - no./total no. (%)	11/659(16.7%)	3/76(3.9%)	8/583(1.4%)	0.123
Smoking history - no./total no. (%)	51/655(7.8%)	11/76(14.5%)	40/578(6.9%)	0.198
Epidemiology - no./total no. (%)				
History of exposure	447/630(70.9%)	38/70(54.3%)	409/560(73.0%)	0.002
Family members have disease	258/507(50.9%)	27/76(35.5%)	231/583(39.6%)	0.285
Number of cases in family members - Median(IQR)	1(0-13)	1(0-1)	1(0-1)	0.263
Have been to Wuhan	381/657(57.9%)	34/76(44.7%)	347/581(59.7%)	0.091
Stay in ICU	33/608(5.4%)	33/73(45.2%)	0/535(0%)	0.000
Interval between date of contact and date of onset - Median (IQR)	5(3-9)	6(3-8)	5(3-10)	0.901
The interval between the onset date and the visit date				
Median (IQR)	2(0-5)	3(1-6)	1(0-4)	0.772
Distribution - no./total no. (%)				0.035
≤2 days	385/651(59.1%)	34/75(45.3%)	351/576(60.9%)	
3-5 days	142/651(21.8%)	22/75(29.3%)	120/576(20.8%)	
>5 days	124/651(19.0%)	19/75(25.3%)	105/576(18.3%)	
The interval between the date of onset and the date of admission				

Median (IQR)	5(2-10)	6(3-8)	5(3-8)	0.759
Distribution - no./total no. (%)				0.250
≤2 days	182/656(27.7%)	15/76(19.7%)	167/580(28.8%)	
3-5 days	167/656(25.5%)	22/76(28.9%)	145/580(25.0%)	
>5 days	307/656(46.8%)	39/76(51.3%)	268/580(46.2%)	
The interval between the onset date and the antiviral date				
Median (IQR)	5(3-10)	4(4-9)	5(4-9)	0.278
Distribution - no./total no. (%)				0.250
≤2 days	182/656(27.7%)	15/76(19.7%)	167/580(28.8%)	
3-5 days	167/656(25.5%)	22/76(28.9%)	145/580(25.0%)	
>5 days	307/656(46.8%)	39/76(51.4%)	268/580(46.2%)	

Abbreviations: BMI, body mass index; ICU, intensive care unit; no., number
Data are presented as medians (interquartile ranges, IQR) and no./total no. (%)

Table 2 Clinical Characteristics and Underlying Diseases

Characteristics	Patients (N=659)	ARDS Patients (N=76)	Non-ARDS Patients (N=583)	P value
Clinical symptoms - no./total no. (%)				
Severity evaluation at admission				0.000
mild	59/643(9.2%)	0/76(0.0%)	59/567(10.4%)	
ordinary	485/643(75.4%)	15/76(19.7%)	470/567(82.9%)	
severe	67/643(10.4%)	29/76(38.2%)	38/567(6.7%)	
critical	32/643(5.0%)	32/76(42.1%)	0/567(0.0%)	
Fever on the first day of admission	439/659(66.6%)	65/76(85.5%)	374/583(64.2%)	0.000
Temperature				
Median (IQR) - ℃	37.5(36.8-38.2)	37.9(37.3-38.5)	37.4(36.8-38.1)	0.000
Distribution - no./total no. (%)				0.000
< 37.3℃	275/659(41.7%)	18/76(23.7%)	257/583(44.1%)	
37.3-38℃	198/659(30.0%)	24/76(31.6%)	174/583(29.8%)	
38.1-39℃	153/659(23.2%)	24/76(31.6%)	129/583(22.1%)	
>39℃	33/659(5.0%)	10/76(13.2%)	23/559(3.9%%)	

Cough	452/658(68.7%)	61/76(80.3%)	391/582(67.2%)	0.025
Expectoration	261/659(39.6%)	37/76(48.7%)	224/583(38.4%)	0.105
Dry cough	191/644(29.6%)	26/75(34.7%)	165/569(29.0%)	0.495
Yellow sputum	32/643(5.0%)	3/73(4.1%)	29/570(5.08%)	0.049
Hemoptysis	10/626(1.6%)	4/73(5.5%)	6/553(1.1%)	0.017
Sore throat	81/638(12.7%)	9/72(12.5%)	72/563(12.8%)	0.051
Stuffy nose	24/631(3.8%)	1/72(1.4%)	23/559(4.1%)	0.506
Muscle ache	108/633(17.1%)	7/72(9.7%)	101/561(18.0%)	0.095
Fatigue	221/646(34.2%)	27/73(36.9%)	194/573(33.8%)	0.366
Shortness of breath	110/635(17.3%)	45/76(59.2%)	65/559(11.6%)	0.000
Gastrointestinal symptoms	121/645(18.6%)	14/74(18.9%)	107/571(18.7%)	0.948
Diarrhea	91/607(15.0%)	15/73(20.5%)	76/534(14.2%)	0.179
Vomiting	18/603(3.0%)	5/73(6.8%)	13/530(2.4%)	0.054
Nausea	60/609(9.9%)	5/73(6.8%)	55/536(10.3%)	0.323
Encephalopathy	3/617(0.5%)	3/544(0.6%)	0/73(0.0%)	0.590
Headache	67/614(10.9%)	6/73(8.2%)	61/541(11.3%)	0.450
Underlying diseases - no./total no. (%)				
Basic disease	274/657(41.7%)	43/76(56.6%)	231/581(39.8%)	0.006
Hypertension	169/641(26.4%)	37/76(48.7%)	130/565(23.0%)	0.000
Heart disease	23/628(3.7%)	3/73(4.1%)	20/555(3.6%)	0.829
Diabetes	66/628(10.5%)	13/73(17.8%)	53/555(9.5%)	0.041
Fatty liver	65/610(10.7%)	12/73(16.4%)	53/537(9.9%)	0.088
COPD	7/614(1.1%)	1/73(1.4%)	6/541(1.1%)	0.590
Asthma	3/614(0.5%)	0/73(0.0%)	3/541(0.5%)	0.524
Malignancy	9/629(1.4%)	1/73(1.4%)	8/556(1.4%)	0.971
Immunosuppressor	1/614(0.2%)	1/73(1.4%)	0/541(0.0%)	0.119
Blood system diseases	2/613(0.3%)	1/73(1.4%)	1/540(0.2%)	0.096
Chronic hepatopath	29/631(4.6%)	5/73(6.8%)	24/558(4.3%)	0.328
Chronic nephrosis	6/616(1.0%)	1/73(1.4%)	5/543(0.9%)	0.756
Pneumonia on admission	591/647(91.3%)	75/76(98.7%)	516/571(90.4%)	0.056

Acute kidney injury on admission	9/652(1.4%)	3/76(3.9%)	6/576(1.0%)	0.076
With other respiratory virus infections	5/569(0.8%)	2/73(2.7%)	3/496(0.6%)	0.126

Abbreviations: COPD, chronic obstructive pulmonary disease; no., number
Data are presented as medians (interquartile ranges, IQR) and no./total no. (%)

Table 3 Radiologic, Laboratory Findings and Complications

Variables	Patients (N=659)	ARDS Patients (N=76)	Non-ARDS Patients (N=583)	P value
CT image features - no./total no. (%)				
Consolidation	175/619(28.3%)	41/76(53.9%)	134/543(24.7%)	0.000
Ground-glass opacity	467/625(74.7%)	59/73(80.8%)	408/552(73.9%)	0.251
Number of Consolidation quadrant - Median (IQR)	0(0-2)	2(1-4)	0(0-2)	0.000
Laboratory findings				
Leukocyte count within 48 hours of admission (10e9/l)				
Median (IQR)	5.0(4.0-6.6)	5.7(4.6-8.8)	4.9(3.9-6.4)	0.000
Distribution - no./total no. (%)				0.000
< 4	163/657(24.8%)	14/76(18.4%)	149/581(25.6%)	
4-10	468/657(71.2%)	47/76(61.8%)	421/581(72.5%)	
>10	26/657(4.0%)	15/76(19.8%)	11/581(1.9%)	
Neutrophil count (10e9/l) - Median(IQR)	3.2(2.3-4.4)	4.8(3.1-7.6)	3.1(2.3-4.2)	0.000
Lymphocyte count (10e9/l)				
Median (IQR)	1.1(0.8-1.6)	0.7(0.5-1.1)	1.2(0.8-1.7)	0.000
Distribution - no./total no. (%)				0.000
≤1.5	468/658(71.0%)	71/76(93.4%)	397/583(68.1%)	
>1.5	191/658(29.0%)	5/76(6.6%)	186/583(31.9%)	
Lymph%				
Median (IQR)	0.24(0.2-0.3)	0.13(0.1-0.2)	0.25(0.2-0.3)	0.403
Distribution - no./total no. (%)				
<0.2	237/657(36.0%)	56/76(73.6%)	181/581(31.2%)	0.000
>=0.2	420/657(63.9%)	20/76(29.9%)	400/581(68.8%)	
Neutrophils / lymphocytes				

Median (IQR)	2.73(1.8-4.5)	6.11(3.7-14.6)	1.2(1.7-4.0)	0.000
Distribution - no./total no. (%)				0.000
<3	348/636(54.7%)	13/75(17.3%)	335/561(59.7%)	
≥3	288/636(45.3%)	62/75(82.7%)	226/561(40.3%)	
Hemoglobin (g/l)- Median (IQR)	134.0(123.0-146.0)	135.5(123.3-150.0)	134(123.0-144.0)	0.462
Hematocrit (%)- Median (IQR)	0.4(0.0-0.4)	0.4(0.3-0.4)	0.39(0.4-0.5)	0.968
Platelet (10e9/l)				
Median (IQR)	195(152-243)	174(138-222)	197(154-243)	0.012
Distribution - no./total no. (%)				0.509
≤100	24/658(3.6%)	4/76(5.3%)	20/582(3.4%)	
>100	634/658(96.4%)	72/76(94.7%)	562/582(96.6%)	
Alanine aminotransferase (U/L)				
Median (IQR)	21.0(15-34.0)	30.3(21.3-46.0)	20.0(14.0-32.0)	0.000
Distribution - no./total no. (%)				0.000
≤40	533/654(81.5%)	48/74(64.9%)	485/580(83.6%)	
>40	121/654(18.5%)	26/74(35.1%)	95/580(16.4%)	
Aspartate aminotransferase (U/L)				
Median (IQR)	24.0(18.0-31.0)	31.5(24.0-46.0)	23(18.0-30.0)	0.000
Distribution - no./total no. (%)				0.000
≤40	563/653(86.2%)	44/74(59.5%)	519/579(89.6%)	
>40	90/653(13.8%)	30/74(40.5%)	60/579(10.4%)	
Potassium ion (mmol/L) - Median(IQR)	3.8(3.6-4.2)	3.8(3.5-4.1)	3.9(3.6-4.2)	0.188
Sodium ion (mmol/L) - Median(IQR)	138.8(136.7-146.7)	138.0(135.0-140.6)	139.0(137.0-140.7)	0.052
Creatinine (umol/l)				
Median (IQR)	65.3(54.0-78.0)	70.5(59.5-92.0)	65.0(53.4-77.4)	0.002
Distribution - no./total no. (%)				0.072
≤133	648/657(98.6%)	72/75(96.0%)	576/582(99.0%)	
>133	9/657(1.4%)	3/75(4.0%)	6/582(1.0%)	
Creatine Kinase (U/L)				
Median (IQR)	74.0(50.8-107.2)	112.5(63.0-245.0)	70.6(49.0-96.3)	0.000

Distribution - no./total no. (%)				0.000
<=185	461/513(89.9%)	47/68(69.1%)	414/445(93.0%)	
>185	52/657(10.1%)	21/68(30.9%)	31/445(7.0%)	
Lactate dehydrogenase (U/L)				
Median (IQR)	215.0(177.0-275.3)	316.0(253.8-394.0)	211.0(173.5-257.5)	0.000
Distribution - no./total no. (%)				0.000
<=250	431/636(67.8%)	17/69(24.6%)	414/567(73.0%)	
>250	205/636(32.2%)	52/69(75.4%)	153/567(27.0%)	
Troponin I (ng/ml - Median (IQR)	0.0(0.01-0.06)	0.0(0.00-0.03)	0.0(0.01-0.09)	0.013
Brain Natriuretic Peptide (<300pg/ml) - Median (IQR)	82.0(27-144.5)	77.0(36.0-164.0)	82.3(26.5-139.7)	0.566
Myoglobin - Median (IQR)	25.1(18.5-40.6)	85.9(43.2-205.9)	23.6(18.2-33.1)	0.000
Glucose (mmol/l) - Median (IQR)	5.9(5.1-7.7)	8.1(6.6-9.9)	5.7(5.1-7.4)	0.000
C-reactive protein (mg/L)				
Median (IQR)	10(3.0-27.3)	34.1(15.3-76.0)	8.8(2.8-24.1)	0.000
Distribution - no./total no. (%)				0.000
<=10	341/648(32.4%)	14/73(19.2%)	327575(56.9%)	
>10	307/648(47.7%)	59/73(80.8%)	248/575(43.1%)	
Procalcitonin (ng/L)				
Median (IQR)	0.1(0.0-0.1)	0.1(0.1-0.2)	0.1(0.0-0.1)	0.008
Distribution - no./total no. (%)				0.195
<0.1	326/510(63.9%)	37/68(54.4%)	289/442(65.4%)	
0.1-0.5	164/510(32.2%)	27/68(39.7%)	137/442(31.0%)	
≥0.5	20/510(3.9%)	4/68(5.9%)	16/442(3.6%)	
Highest CRP within one week of admission (mg/L) - Median (IQR)	15.5(4.0-39.6)	45.8(17.7-83.1)	12.2(3.8-34.2)	0.000
Highest PCT within one week of admission (ng/L) - Median (IQR)	0.1(0.0-0.1)	0.1(0.1-0.2)	0.1(0.0-0.1)	0.000
FiO2 on the first day of admission - Median (IQR)	0.3(0.2-0.3)	0.0(0.3-0.5)	0.0(0.2-0.3)	0.000
PaO2 on the first day of admission (mmHg) - Median (IQR)	90.7(78.0-111.0)	79.0(67.9-95.2)	94.0(80.2-117.0)	0.019

PaCO2 on the first day of admission (mmHg) - Median (IQR)	37.5(34.0-41.5)	33.0(32.4-37.5)	38.0(34.8-41.9)	0.699
Oxygenation index on the first day of admission (mmHg)				
Median (IQR)	292.2(144.4-424.3)	240(160.8-261.2)	352.0(64.2-466.2)	0.000
Distribution - no./total no. (%)				0.000
≤200	106/325(32.6%)	35/69(50.7%)	71/256(27.7%)	
> 200	219/325(67.4%)	34/69(49.3%)	185/256(72.3%)	
D-dimer (μg/L) - Median (IQR)	250.0(140.0-525.3)	819.0(276.0-1212.0)	238.0(130.0-458.0)	0.000
Complications - no./total no. (%)				
Shock	4/647(0.6%)	4/73(5.5%)	0/574(0.0%)	0.000
Secondary bacterial infection	48/659(7.3%)	23/76(30.3%)	25/583(4.3%)	0.000

Abbreviations: CT, computerized tomography; CRP, c-reactive protein; PCT, procalcitonin; no., number
Data are presented as medians (interquartile ranges, IQR) and no./total no. (%)

Table 4 Risk Factor Analysis for COVID -19

Characteristics	OR	95% CI	P value
Gender (Male vs. Female)	3.634	2.090-6.219	0.000
Age			0.000
14-30 vs. 30-70 years	4.285	1.024-17.926	0.046
14-30 vs. ≥70 years	14.258	3.032-67.049	0.001
BMI (<23 vs.>25)	3.145	1.635-6.052	0.001
The interval between the onset date and the visit date			
≤2 vs. 3-5 days	1.893	1.065-3.363	0.030
≤2 vs. >5 days	1.868	1.023-3.411	0.042
Severity evaluation at admission	30.232	16.166-56.534	0.000
Temperature			
<37.3 vs.37.3-38 ℃	2.061	1.075-3.950	0.029
<37.3 vs.38.1-39 ℃	2.780	1.442-5.359	0.002
<37.3 vs.>39 ℃	6.496	2.667-15.820	0.000
Cough	1.978	1.100-3.586	0.025
Polypnea	11.302	6.532-18.659	0.000
Basic disease	1.974	1.218-3.200	0.006
Hypertension	3.527	2.159-5.761	0.000
Diabetes	2.052	1.057-3.983	0.041

Secondary bacterial infection	9.686	5.146-18.323	0.000
Lung consolidation	3.575	2.187-5.845	0.000
Neutrophil count (10e9/l)	1.370	1.250-1.502	0.000
Lymphocyte count (10e9/l)	0.149	0.080-0.277	0.000
Leukocyte count (4-10 vs. >10 10e9/l)	12.215	5.303-28.135	0.000
Neutrophils / lymphocytes (<3 vs. ≥3)	7.069	3.798-13.158	0.000
Alanine aminotransferase (≤40 vs.>40 U/L)	2.842	1.690-4.779	0.000
Aspartate aminotransferase (≤40 vs.>40 U/L)	5.898	3.452-10.075	0.000
Creatine Kinase (≤185 vs.>185 U/L)	6.054	3.222-11.374	0.000
Lactate dehydrogenase (≤250 vs.>250 U/L)	8.277	4.643-14.755	0.000
C-reactive protein (≤10 vs.>10 mg/L)	5.557	3.033-10.182	0.000
Oxygenation index on the first day of admission (≤200 vs. > 200 mmHg)	0.373	0.216-0.643	0.000

Table 5 The final multivariable model in development dataset

Variables	Odds ratio (95% CI)
Severity evaluation at admission	30.232 (16.166-56.534)
Gender	3.634 (2.090-6.219)
Age (≥70 years)	14.258 (3.032-67.049)
Temperature (>39 ℃)	6.496 (2.667-15.820)
Shortness of breath	11.302(6.532-18.659)
Secondary bacterial infection	9.686(5.146-18.323)
Lung consolidation	3.575(2.187-5.845)
Lymphocyte count	0.149(0.080-0.277)
Leukocyte counting	12.215(5.303-28.135)
CK	6.054(3.222-11.374)
NLR	7.069(3.798-13.158)
AST	5.898(3.452-10.075)
LDH	8.277(4.643-14.755)
CRP	5.557(3.033-10.182)

Abbreviations: CK, Creatine Kinase; NLR, Neutrophils / lymphocytes; AST, Aspartate aminotransferase; LDH, Lactate dehydrogenase; CRP, c-reactive protein

Table 6 Performance of the five algorithms on external testing dataset in predicting the occurrence of COVID-19 ARDS

Models	AUC	Accuracy	Sensitivity	Specificity

Decision Tree	0.99	0.91	0.92	1.00
Logistic Regression	0.98	0.88	0.83	0.97
Random Forest	0.99	0.90	0.83	1.00
Support Vector Machine	0.72	0.92	0.33	0.92
Deep Neural Networks	0.50	0.75	0.00	1.00

Figures

Figure 1

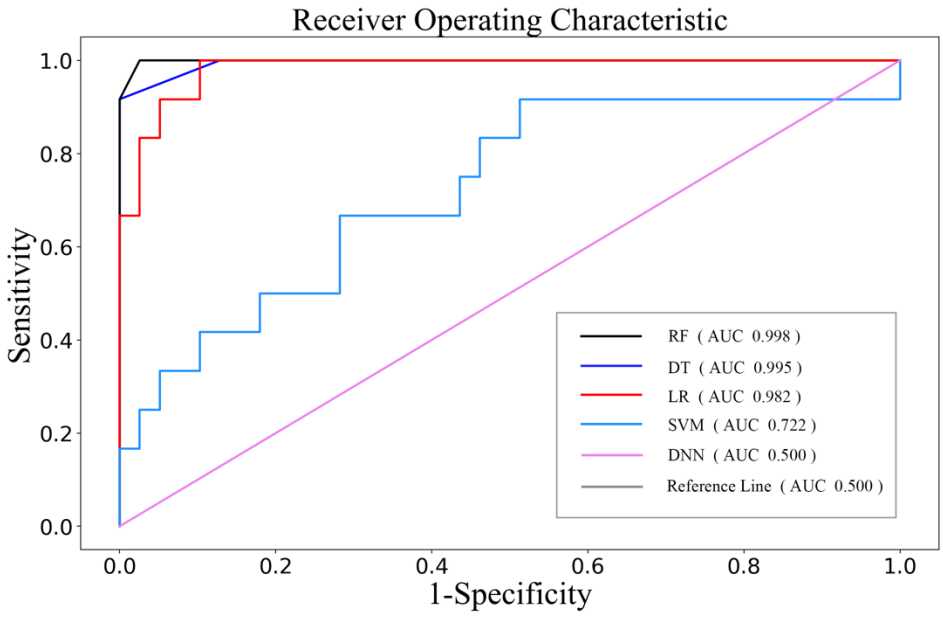

Figure S1

Correlation feature list:
(1): Severity evaluation at admission (X_1)
(2): Polypnea (X_2)
(3): CK (>185 U/L) (X_3)
(4): Oxygenation index on the first day of admission (> 200 mmHg) (X_4)