Laboratory Testing Implications of Risk-Stratification and Management of COVID-19 Patients

Caidong Liu
Nanjing First Hospital, Nanjing Medical University
Ziyu Wang
Nanjing Medical University
Wei Wu
Nanjing Medical University
Changgang Xiang
The First People's Hospital of Jiangxia District
Lingxiang Wu
Nanjing Medical University
Jie Li
Nanjing Medical University
Weiye Hou
Nanjing First Hospital, Nanjing Medical University
Huiling Sun
Nanjing First Hospital, Nanjing Medical University
Youli Wang
Department of laboratory medicine, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, Jiangsu
Zhenling Nie
Department of laboratory medicine, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, Jiangsu
Yingdong Gao
Department of laboratory medicine, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, Jiangsu
Ruisheng Zhang
Department of laboratory medicine, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, Jiangsu
Kening Li
Nanjing Medical University
Xinyi Xia
Nanjing University School of Medicine https://orcid.org/0000-0001-7552-3170
Qianghu Wang (wangqh@njmu.edu.cn)

Nanjing Medical University https://orcid.org/0000-0002-3488-1059

Shukui Wang

Nanjing First Hospital, Nanjing Medical University

Article

Keywords: COVID-19, admission stratification model, laboratory indicators, risk-stratification model

Posted Date: November 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-98408/v1

Abstract

The progression from mild to critical illness is the main reason leading to the death of COVID-19 patients. Rapid risk-stratification at admission is important for precise management of COVID-19. Here, we developed a practical admission stratification model to predict the severity during hospitalization of COVID-19 patients using laboratory data from 3563 patients, including 548 patients in the training dataset, and 3015 patients in the testing dataset. We first identified the significant laboratory indicators related to the severity of COVID-19 in the training dataset. Neutrophils percentage (NEUT%), lymphocytes percentage (LYMPH%), creatinine (CREA), and blood urea nitrogen (BUN) with AUC greater than 0.7 were included in the model. These indicators were further used to build a support vector machine model to classify patients into low-risk and high-risk at admission. Results showed that this model could stratify the patients in the testing dataset effectively (AUC=0.89). Moreover, laboratory indicators detected in the first week after admission were able to estimate the probability of death (AUC=0.95). Besides, we could diagnose COVID-19 and differentiated it from other kinds of viral pneumonia based on laboratory indicators (accuracy=0.97). Our risk-stratification model based on laboratory indicators could help to diagnose, monitor, and predict severity at an early stage of COVID-19.

Introduction

The coronavirus disease 2019 (COVID-19) has become a serious worldwide problem, which is caused by a novel coronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV-2). As of 18 August, 2020, there have been 3,356,205 confirmed cases of COVID-19, including 771635 deaths (https://who.sprinklr.com/). The global outbreak of COVID-19 highlights the importance of early and rapid diagnosis, monitoring, risk assessment, and medical resource management in the prevention and control of epidemics [1].

The death of COVID-19 patients is mainly caused by the progression from the mild to the critical illness [2]. Therefore, it is an urgent need for effective methods to predict prognosis early. At present, nucleic acid detection and antibody detection are the main technical approaches for clinical diagnosis of COVID-19 patients, but both of them are affected by many factors, such as sample location, type, quality and patient condition as well as sample storage, which caused a certain degree of false positives and false negatives [3]. Most importantly, they all failed to help to judge whether a patient will progress to severe illness [4, 5, 6]. Besides, CT imaging is also a common method, but it lacks specificity and requires a large number of professional technicians, thus easily exhausts resources when the epidemic is serious. The latest research shows that, based on artificial intelligence methods, CT can be used to diagnose or stratify COVID-19 quickly. However, the accuracy of using CT alone to predict patient severity is limited [7, 8, 9].

Previous studies have reported that in the early published 41 COVID-19 cases, five patients presented with varying degrees of myocardial injury, cardiovascular disease patients are more likely to develop into severe patients after COVID-19 infection, and the risk of death is higher [10]. The abnormal of different

laboratory indicators can represent damage to different organs. For example, NT-proBNP indicates cardiac dysfunction and Alkaline phosphatase (ALP) indicates liver dysfunction. In addition, other laboratory indicators are highly correlated with the risk of disease progression, such as the lymphocyte, IL-6, etc. [11,12]. These findings suggested that the laboratory indicators can be used to predict the severity of COVID-19 pneumonia patients. It is of significantly importance to perform risk-stratification and management of epidemic disease, especially to countries with shortage of medical resources. Using limited resources to a greater extent for more critically ill patients will help improve the utilization of medical resources. And performing more rigorous testing and clinical observation for patients who tend to be severer is necessary.

Therefore, our study aims to identify the laboratory indicators that could predict the severity as early as admission, and build a practical risk-stratification model for screening of severe COVID-19 patients, as well as predicting the risk of death. This prognostic model based on laboratory indicators could provide important informations for diagnosis, stratification, and monitoring for COVID-19 patients as early as possible.

Results

Study design

We collected the clinical data of 3563 COVID-19 patients and 18 non-COVID-19 viral pneumonia (designated as non-COVID-19) to build and validate the risk-stratification model. Specifically, data of 548 patients from First People's Hospital of Jiangxia District of Wuhan were used as training dataset (FPHJ-548); data of 3015 patients from Wuhan Huoshenshan Hospital were used as testing dataset (HSSH-3015); data of 18 non-COVID-19 viral pneumonia patients from First People's Hospital of Jiangxia District of Wuhan were used differentiating COVID-19 from non-COVID-19 (VPP-18).

The highest severity during hospitalization of each patient was recorded, and laboratory findings of blood test at admission were used to predict the progression of these patients. In the FPHJ-548 dataset, the average age of these patients was 52.4 (SD=14.2), and 49.8% were female. Notably, the median age of severe patients was significantly higher than that of moderate patients (Fisher's exact test, P<0.01, supplementary Table 1). The clinical information of 385 cases (including 329 moderate and 56 severe cases) that have detection data at admission were selected to do the following analysis. To predict the severity of COVID-19 patients at admission, we employed a risk-stratification model based on support vector machine (SVM) by laboratory indicators in the FPHJ-548 dataset. This model was further validated in an independent dataset (HSSH-3015) (**Figure 1**, details see Methods). 51 patients in the HSSH-3015 dataset that have detection data at admission were selected as testing dataset.

Then, in order to monitor the survival outcome of severe COVID-19 patients, we select 1448 survival patients and 55 deaths from HSSH-3015 dataset. 60 patients without laboratory findings within the first week since admission were excluded. We randomly split HSSH-3015 dataset into a leave-in training set

and a leave-out test set for data analysis at a ratio of approximately 1:1. We ensemble a logistic regression model (LR) based on laboratory findings in the training set and validated in the testing set (**Figure 1**, details see Methods). Besides, to distinguish COVID-19 from non-COVID-19 viral pneumonia, we compared the laboratory difference between COVID-19 datasets (FPHJ-548 or HSSH-3015) and non-COVID-19 dataset (VPP-18)(**Figure 1,** details see Methods).

A risk-stratification model of COVID-19 based on 4 laboratory findings at admission

According to the highest severity of each patient during hospitalization, we explored the difference in laboratory findings between moderate and severe COVID-19 cases in FPHJ-548 dataset. We found the high risk factors related to the progression of COVID-19 included procalcitonin (PCT), C-reactive protein (CRP), neutrophils percentage (NEUT%), lymphocytes percentage (LYMPH%), lactate dehydrogenase (LDH), (Wilcoxon rank-sum test, P<0.001, **Table 1**). We noted that most of the severe patients presented lymphopenia and elevated levels of inflammatory biomarkers. The levels of PCT in severe patients at the initial stage were higher than those in moderate patients (0.225 vs. 0.06, Wilcoxon rank-sum test, P<0.001), suggesting serial procalcitonin measurement may play a role in predicting evolution towards a more critical form of the disease [13]. The CRP showed a similar trend to PCT, which became significantly higher in severe patients (44.5 vs. 21.8, Wilcoxon rank-sum test, P<0.001). Lymphocyte percentage was significantly higher in the moderate COVID-19 patients than severe COVID-19 patients (22.4% vs 13.8%, Wilcoxon rank-sum test, P<0.001). The percentage of neutrophils was elevated along with the severity of COVID-19 (77.8 vs. 66.4, Wilcoxon rank-sum test, P<0.001). Besides, LDH (314 vs. 235, Wilcoxon rank-sum test, P<0.001) of severe patients were significantly higher than those of moderate patients. Considering most of these differential indicators are related to organ damage, we next explored the impact of the pre-existing diseases on the progression of COVID-19. Based on the FPHJ-548 dataset, we found only 9% of patients without pre-existing disease progressed to severe condition. In contrast, 16% of severe patients were diagnosed with at least one kind of pre-existing disease (Fisher's Exact Test, P=0.029, **Figure 2A**), suggesting that COVID-19 patients with pre-existing disease were prone to develop severe illness. Furthermore, we found the same trend in the HSSH-3015 dataset. Patients with multiple pre-existing diseases are more inclined to progress into severe (**Figure 2B**).

The difference in laboratory indicators between severe and moderate patients prompted us to develop a model based on laboratory indicators to predict the state of patients (**Figure 1,** details see Methods). To validate that whether laboratory findings could predict the progression of COVID-19, we performed t-distributed stochastic neighbor embedding (t-SNE) based on the laboratory indicators in the FPHJ-548 dataset. The result showed that there was an essential difference in laboratory indicators between moderate and severe patients. 95% of the samples were correctly classified (true positive rate:0.66, true negative rate:1, **Figure 3A**).

For each indicator in FPHJ-548, it's correspondent AUC was calculated using the detected value as predictors and the status of progression as an outcome. We selected features whose AUC is greater than 0.7 and only kept indicators that has detection data at admission in both FPHJ-548 dataset and HSSH-

3015 datasets.. Finally, our model incorporates 4 indicators, including LYMPH%, NEUT%, creatinine (CREA), and urea nitrogen (BUN) (**Figure 3B**). The NEUT% between moderate and severe patients showed a noticeable increase at about four days before the admission in the FPHJ-548 dataset (**Figure 3C**). On the contrary, the neutrophil of moderate patients was stable, and between the range of normal reference.

Next, we applied these four indicators to develop a support vector machine model, followed by five-fold cross-validations as internal validation. The average sensitivity and specificity of five cross-validations were 0.89 and 0.84, respectively. The average AUC of the five cross-validations was 0.86 (AUC 95% CI:0.84-0.88). The representative receiver operating characteristic (ROC) for the external validation (HSSH-3015 dataset) was shown in **Figure 3D**. It still achieved satisfying results in the testing dataset (sensitivity and specificity, 0.73 and 0.96, respectively, AUC:0.89). Lastly, to avoid the biases of age and sex, we divided patients into two groups by age or sex to test our model, the results showed that our model still has good performance when considering age and sex **(Figure 3E, F**).

Laboratory findings within the first week after admission could predict the risk of death of COVID-19

The progression of COVID-19 into severe illness increases the risk of death, so we predict the survival outcome of severe patients in HSSH-3015 dataset based on the laboratory findings within the first week after admission (Figure 1). Patients were randomly divided into training group and validation group at the ratio of 1:1. To avoid the deviation caused by the difference between the number of deaths and the number of survivors, we randomly selected the surviving patients so that the number of surviving patients equals the number of dead patients. We use the stepwise logistic regression to identify the important laboratory indicators. This process repeats 100 times (details see Methods). Thirteen indicators with statistically significant differences between survivors and deaths were identified. These were Albumin/Globulion, DD dimer, leukocyte, monocytes, Cystatin C, Creatinine, lymphocyte, Urea nitrogen, Thrombin time, Prothrombin time, Lactate dehydrogenase, Fibrinogen, Percentage of neutrophils. We performed multidimensional scaling in the training dataset based on these 13 markers. Results show that these indicators could distinguish deaths from survivors (accuracy=0.96, true positive rate: 0.82, true negative rate: 0.97, **Figure 4A**). Then, based on these 13 indicators, we develop a logistic model to predict the survival outcome in the training dataset. We found that the model predicts the survival outcome with high accuracy in the testing dataset (AUC = 0.95, **Figure 4B**). Besides, The average NEUT% of dead patients exceeded the maximum normal value during hospitalization. On the contrary, the neutrophil of survivors was stable, and between the range of normal reference (**Figure 4C**).

Distinguishing COVID-19 from non-COVID-19 viral pneumonia based on laboratory findings

Increasing studies showed that the infection of viral pneumonia might be associated with organ dysfunction [14, 15, 16, 17]. Hence, we explored the change of organ function-related indicators between FPHJ-548 and VVP-18. Interestingly, we found that some indicators related to organ dysfunction showing significant differences between the two groups (**Table 2**, Wilcoxon two-sided rank-sum test, P<0.05). Our studies showed that patients in non-COVID-19 group had higher levels of NT-proBNP than those of

COVID-19 group (1259.4pg / mL vs. 90.285pg / mL, *P*=0.045). Besides, the level of LDH in non-COVID-19 was higher than COVID-19 patients (594 vs. 242.85, *P*<0.001). The level of alanine aminotransferase (19 U / L vs. 40U / L, *P*<0.001) and aspartate aminotransferase (30.1s vs. 36s, *P*<0.001) were higher in non-COVID-19 group. The median activated partial thromboplastin time was longer than that in the COVID-19 group. The median level of albumin and hemoglobin decreased by more than 5g/L and 10 g/L in non-COVID-19 patients, respectively (albumin: 33.8 g / L vs. 38.95 g / L, *P*<0.001; hemoglobin: 121 g / L vs. 135.25 g / L, *P*=0.003). Hence, we use three laboratory findings with significant differences (alanine aminotransferase, aspartate aminotransferase, and lactate dehydrogenase) to perform multidimensional scaling on FPHJ-548 and VPP-18 (**Figure 5A**, details see Methods). We find that these indicators can distinguish COVID-19 and non-COVID-19 (accuracy=0.93, true positive rate: 0.94, true negative rate: 0.78). For verification, we perform the same method on HSSH-3015 and VPP-18 and found the similar results (accuracy=0.97, true positive rate: 0.97, true negative rate: 0.93, **Figure 5B**).

Considering that these indicators are related to liver disease and heart disease, we removed the patients with liver and heart disease in the HSSH-3015 dataset to exclude the impact of pre-existing disease. Results showed that these indicators still could differentiate COVID-19 from non-COVID-19 (accuracy=0.96, true positive rate: 0.97, true negative rate: 0.92, **supplementary Figure 1**). In summary, these findings demonstrated that laboratory findings can distinguish COVID-19 patients from non-COVID-19 patients.

Discussion

About 6.5% of COVID-19 patients were sudden progression to severe with a fatality rate of 49% in these patients [18]. It is an urgent need for effective methods to predict and monitor the progression of COVID-19 patients from moderate to severe conditions. First, based on the FPHJ-548 dataset, we systematically explored the difference of laboratory findings between severe and moderate patients. We found the high-risk factors related to the progression of COVID-19 included PCT, CRP, NT-proBNP, neutrophils percentage, LDH, and LYMPH%, etc. Most of these laboratory indicators were reported to be associated with the progression of COVID-19. Lymphocyte count was lower in non-survivors than survivors [11]. Severe cases presented lower lymphocyte counts and higher neutrophil levels [11, 19]. LDH was found as a risk factor associated with disease progression in patients infected with COVID-19[20]. Many types of research proved that elevated NT-proBNP was significantly correlated with critical disease [21]. Initial blood urea nitrogen and serum creatinine are related to increased mortality in COVID-19[22, 23].In addition, we found that the proportion of severe condition positively associated with the increase in the number of pre-existing diseases diagnosed in the patients. Many studies demonstrated that these pre-existing diseases might promote the expression of ACE2 [24, 25], leading to a high-risk of COVID-19 infection. Based on a set of laboratory indicators (NEUT%, LYMPH%, CREA, and BUN), we finally constructed a risk-stratification model by using a SVM model, achieve the AUC of 0.89 in an independent dataset. Then, we based on 13 laboratory findings ensemble a model to predict the survival outcome with high accuracy. At last, we proved that laboratory findings could distinguish COVID-19 patients from non-COVID-19 patients.In the

latest research, Zhang et.al. have developed an artificial intelligence (AI) tool, which could classify the severity and predict critical illness based on chest CT images and laboratory indicators[7]. Deep learning survival Cox model was also developed to predict the clinical outcome of COVID-19 patients with high accuracy. This model uses ten clinical variables, including common demographic and clinical characteristics, as well as laboratory results[26]. However, in the emergency of pandemic, the requirments of the professional devices and clinicians make these methods difficult to use in rapid way. Our model uses four laboratory indicators that are available at most hospitals and achieve comparable sensitivity and specificity. When the medical system is overloaded in a pandemic or in rural area, this risk-stratification model can help screen patients who may develop severe illness accurately by easy detect and low-cost testing, as early as admission.

Our model has some limitations. First, because of the emergency of the epidemic, some patients did not take the blood tests at admission, which limited the power of prediction and validation. Second, further studies on different populations with larger patient cohorts are required to verify our findings, especially regarding the distinguishment of COVID-19 from non-COVID-19 viral pneumonia based on laboratory findings. As the tendency of organ dysfunction of COVID-19 and other pneumonia is controversy currently, more extensive comparison analysis was needed to validate the difference[27, 28].

In conclusion, our practical prognostic model based on laboratory indicators is convenient and effective for risk-stratification for COVID-19 at admission, so as to ensure that severe patients receive treatment early, as well as medical resources could be allocated effectively. Our study would provide vital information for clinical practice in diagnosis and monitoring for COVID-19 patients.

Methods

Data collection

From December 1, 2019 to February 13, 2020, a total of 548 cases of confirmed COVID-19 patients were collected from the First People's Hospital of Jiangxia District of Wuhan, including 474 moderate COVID-19 patients and 74 severe COVID-19 patients (FPHJ-548 dataset). 385 COVID-19 patients who received blood tests at admission were included for the analysis. 18 non-COVID-19 viral pneumonia cases (designated as VPP-18) were also collected from December 1, 2019, to February 13, 2020, in the First People's Hospital of Jiangxia District of Wuhan. 1452 moderate and 1563 severe COVID-19 patients were collected from Wuhan Huoshenshan Hospital as a validated dataset (HSSH-3015) from February 4, 2020, to April 10, 2020. The diagnosis of COVID-19 in these datasets is based on the "New Coronavirus Pneumonia Diagnosis and Treatment Plan (provisional 6th Edition)" issued by the National Health and Health Commission. This study was approved by the hospital ethics committee (number of ethic committee, 2020029).

SVM approach for risk-stratification based on laboratory indicators at admission

Using the highest severity during hospitalization of each patient in training dataset (FPHJ-548) as labels, a SVM model was constructed to predict the severity of at admission based on the blood test results. COVID-19 patients The steps of the SVM risk-stratification method are described as follows: (1) the laboratory indicators with AUC>0.7 were selected. The indicators which have no detection data in testing dataset (HSSH-3015) were excluded. Finally, four laboratory findings (LYMPH%, NEUT%, CREA, and BUN) were used to develop a risk-stratification model. (2) Normalize the original value of each indicator according to the normal range. Normalized value greater than 1 means exceeding the maximum normal range. Normalized value less than 1 indicates that it is below the minimum normal range. Normalized value range 0 to 1 indicates that it is within the normal range. (Eq .1). (3) Predict the severity of each patients using SVM model. The basic principle of this method is to find a fractal hyperplane for the training set in the sample space, which will maximize the separation of categories. We define the distance from the sample to the hyperplane as the risk-stratification score (RSS) (Eq.2) W represents the coefficient of laboratory indicators trained by SVM. X represents the vector of laboratory indicators. We used five-fold cross-validation in training dataset (FPHJ-548) to prove the feasibility of the risk-stratification based on the four indicators. We validated the model the in HSSH-3015 dataset. Patients are grouped based on age and sex to validate the model. The prediction performances of AUC was calculated using the predicted values estimated by model with the combination of selected feature as predictors and the status of progression as an outcome.

$$normalizaed\ value\ = \frac{Detected\ value - min(normal\ range)}{max(normal\ range) - min(normal\ range)}\ (eq\ 1)$$

$$RSS = \sum_{i=1}^{N} W_i * X_i + B\ (eq\ .2)$$

LR approach for survival outcome based on laboratory indicators within one week since admission

To monitor the risk of deaths for severe COVID-19 patients, we randomly split HSSH-3015 into a leave-in training set and a leave-out test set for data analysis at a ratio of approximately 50%:50% (using a random number generator). To predict the survival outcome early, we only selected laboratory findings within the first week after admission. For the training set, 724 survival samples, and 28 deaths were selected. For the matched leave-out test set, 724 survivors, and 27 dead samples were selected. For training dataset, we random select 28 survivors. We incorporated group sizes of 28 dead individuals and 28 deaths to develop model by stepwise LR. This random process was repeated 100 times, leading to 100 different model-building. Indicators which were significant in over ten out of 100 models were considered as potential risk related factors. Thirteen indicators were involved for the next modeling, including Albumin/Globulion, DD dimer, leukocyte, monocytes, Cystatin C, Creatinine, lymphocyte, Urea nitrogen, Thrombin time, Prothrombin time, Lactate dehydrogenase, Fibrinogen, Percentage of neutrophils. Then, we trained the LR model by these indicators in the training dataset and validated in the leave-out test set. RSS was calculated as above (eq .2). W represents the coefficient of laboratory indicators trained by LR.

Distinguishing COVID-19 from non-COVID-19 based on laboratory indicators

Laboratory findings at admission were used to distinguish the COVID-19 from non-COVID-19 patients. 25 laboratory findings that shared between FPHJ-548 and VPP-18 datasets were included, in which three indicators that showed significant difference between COVID-19 and non-COVID-19 patients were selected for the further analysis ($p<0.05$ and Fold Change>1.5). As a result, alanine aminotransferase, aspartate aminotransferase, and lactate dehydrogenase were used to distinguish COVID-19 and non-COVID-19 patients. To eliminate the influence of missing values, we only consider patients with no missing values in the three indicators. We finally selected 212 patients from FPHJ-548, 14 patients from VPP-18, and 2828 patients from HSSH-3015. For FPHJ-548 and VPP-18, we cluster based on the maximum distance and perform multidimensional scaling and validated by HSSH-3015 and VPP-18.

Statistical Analysis

Continuous and categorical variables were presented as median (IQR) and n (%), respectively. We used the Wilcoxon rank-sum test (for continuous quantitative variables) or Fisher's exact test (for categorical variables) to compare differences between moderate and severe patients where appropriate. In the bilateral test, the index of $p <0.05$ is considered statistically significant. ROC curves and their correspondent AUC of RSS were calculated by R package pROC. The analysis was carried out using the statistical software R (version: 3.6.0). All figures were plotted by ggplot2 package.

DATA AVAILABILITY

All the data supporting the findings of this study are available within the article and its Supplementary information files.

Declarations

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant Nos. 81572893, 81972358, 91959113), Key Research& Development Program of Jiangsu Province (Grant Nos. BE2017733).

Author contributions

Shukui Wang, Qianghu Wang, Xinyi Xia, Kening Li had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of thedata analysis. Caidong Liu, Ziyu Wang, Wei Wu, and Changgang Xiang contributed equally. Concept and design: Wang Shukui, Qianghu Wang, Xinyi Xia, Kening Li. Data collection: Caidong Liu, Changgang Xiang, Weiye Hou, Huiling Sun, Youli Wang, Zhenling Nie, Yingdong Gao, Ruisheng Zhang, . Data analysis and interpretation: Ziyu Wang, Wei Wu, LingxiangWu, Jie Li.

Competing interests

The authors declare there is no conflict of interest.

References

1. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689-697 (2020).
2. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis*, (2020).
3. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *J Pharm Anal* **10**, 102-108 (2020).
4. Wang Y, Kang H, Liu X, Tong Z. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. *J Med Virol*, (2020).
5. Fang Y*, et al.* Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, 200432 (2020).
6. Xiao AT, Tong YX, Zhang S. False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J Med Virol*, (2020).
7. Zhang K*, et al.* Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell*, (2020).
8. Dadario AMV, Paiva JPQ, Chate RC, Machado BS, Szarf G. Regarding "Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT". *Radiology*, 201178 (2020).
9. Bai HX*, et al.* AI Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Etiology on Chest CT. *Radiology*, 201491 (2020).
10. Huang C*, et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506 (2020).
11. Zhou F*, et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054-1062 (2020).
12. Gao Y*, et al.* Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *J Med Virol* **92**, 791-796 (2020).
13. Lippi G, Plebani M. Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *Clin Chim Acta* **505**, 190-191 (2020).
14. Ronco C, Reis T. Kidney involvement in COVID-19 and rationale for extracorporeal therapies. *Nat Rev Nephrol*, (2020).
15. Liu PP, Blet A, Smyth D, Li H. The Science Underlying COVID-19: Implications for the Cardiovascular System. *Circulation*, (2020).

16. Feng G*, et al.* COVID-19 and Liver Dysfunction: Current Insights and Emergent Therapeutic Strategies. *J Clin Transl Hepatol* **8**, 18-24 (2020).
17. Lechien JR*, et al.* Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Oto-Rhino-L*, (2020).
18. Guan WJ*, et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708-1720 (2020).
19. Kong M, Zhang HM, Cao XC, Mao XL, Lu ZX. Higher level of neutrophil-to-lymphocyte is associated with severe COVID-19. *Epidemiol Infect* **148**, (2020).
20. Lo IL*, et al.* Evaluation of SARS-CoV-2 RNA shedding in clinical specimens and clinical characteristics of 10 patients with COVID-19 in Macau. *Int J Biol Sci* **16**, 1698-1707 (2020).
21. Chen C, Chen C, Yan JT, Zhou N, Zhao JP, Wang DW. [Analysis of myocardial injury in patients with COVID-19 and association between concomitant cardiovascular diseases and severity of COVID-19]. *Zhonghua Xin Xue Guan Bing Za Zhi* **48**, E008 (2020).
22. Cheng A*, et al.* Diagnostic performance of initial blood urea nitrogen combined with D-dimer levels for predicting in-hospital mortality in COVID-19 patients. *Int J Antimicrob Agents* **56**, 106110 (2020).
23. Cheng YC*, et al.* Kidney disease is associated with in -hospital death of patients with COVID-19. *Kidney Int* **97**, 829-838 (2020).
24. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Resp Med* **8**, E21-E21 (2020).
25. Zheng YY, Ma YT, Zhang JY, Xie X. COVID-19 and the cardiovascular system. *Nat Rev Cardiol* **17**, 259-260 (2020).
26. Liang W*, et al.* Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun* **11**, 3543 (2020).
27. Zhao D*, et al.* A Comparative Study on the Clinical Features of Coronavirus 2019 (COVID-19) Pneumonia With Other Pneumonias. *Clin Infect Dis* **71**, 756-761 (2020).
28. Inamura N*, et al.* Management of refractory Mycoplasma pneumoniae pneumonia: utility of measuring serum lactate dehydrogenase level. *J Infect Chemother* **20**, 270-273 (2014).

Tables

Table 1. Comparing laboratory findings between moderate and severe COVID-19 patients

	Total (N=385)	Moderate (N=329)	Severe (N=56)	P-value
Laboratory Findings				
Infection markers				
Procalcitonin, ng/ml	0.07 (0.04-0.14) (157)	0.06 (0.04-0.11) (130)	0.225 (0.14-0.53) (27)	<0.001
C-reactive protein	24 (12-52) (288)	21.8 (9.4-41) (190)	44.5 (25-120) (38)	<0.001
Lymphocytepercentage, %	21.07 (14-29) (372)	22.4 (16-30) (318)	13.8 (7-21) (54)	<0.001
Monocyte percentage, %	8.98 (6.7-12) (373)	9.6 (7-13) (319)	7.5 (5.4-9.1) (54)	<0.001
Neutrophil percentage, %	67.92 (58-77) (372)	66.4 (57-75) (318)	78.8 (69-86) (54)	<0.001
Liver injury markers				
Albumin, g/L	38.95 (36-42) (206)	39.2 (36-43) (167)	37.95 (33-42) (39)	0.039
Creatine Kinase MB	15 (12-19) (185)	14.5 (12-17) (148)	17.85 (18-21) (37)	0.018
Uric acid, umol/L	261 (200-360) (208)	250 (190-330) (169)	337 (270-430) (39)	<0.001
Cholinesterase	6903 (5900-8100) (198)	7088 (6000-8300) (160)	6296.55 (4600-7400) (38)	0.007
Heart injury markers				
Lactate dehydrogenase, U/L	242.85 (200-330) (215)	235 (200-300) (173)	314 (240-500) (42)	<0.001
N terminal pro B type natriuretic peptide	90.285 (49-410) (81)	80.34 (45-190) (55)	292.1 (78-11000) (26)	0.0021
High-sensitivity troponin T	0.009 (0.006-0.016) (179)	0.008(0.006-0.013) (145)	0.016 (0.0095-0.04) (34)	<0.001
Kidney injury markers				
Creatinine	66.45 (53-81) (208)	64.85 (51-78) (169)	81.5 (65-330) (39)	<0.001
Glomerular filtration rate	98.7 (80-110) (208)	100.7 (89-110) (169)	85.5 (14-100) (39)	<0.001
Homocysteine	13.75 (11-17) (141)	13 (11-17) (111)	17 (14-27) (30)	<0.001

Bloodexamination				
Hematocrit, %	39.85 (36-43) (373)	40.1 (37-43) (319)	37.75 (34-42) (54)	0.002
Hemoglobin, g/L	135.25 (120-150) (373)	136 (120-150) (319)	129.5(110-140) 54)	0.012
Platelet count, /L	177 (140-230) (373)	179.5 (150-230) (319)	153 (120-210) (54)	<0.001
Red blood cell count	4.37 (4-4.7) (373)	4.39 (4.1-4.7) (319)	4.23 (3.7-4.6) (54)	<0.001
White blood cell count, g/L	5.7 (4.6-7.4) (372)	5.7 (4.5-7.2) (318)	6.25 (4.9-9.5) (54)	0.01
Platelet volume distribution width	13.4 (12-16) (372)	13.3 (12-16) (318)	15.2 (12-16) (54)	0.023
CO2	21.7 (20-23) (208)	22.1 (20-24) (169)	19.85 (18-21) (39)	<0.001
γ-glutamyltranspeptidase,U/L	25 (17-51) (206)	24 (16-44) (166)	38.5 (24-64) (40)	0.0015
MG	0.9 (0.85-0.96) (204)	0.9 (0.84-0.94) (164)	0.93 (0.89-1) (40)	0.0018
Urea	4.6 (3.5-6) (208)	4.4 (3.4-5.6) (169)	6.45 (4.9-19) (39)	<0.001
Myoglobin	46.64 (24-99) (178)	40.25 (21-81) (144)	106.9 (54-200) (34)	<0.001

Table 2. Comparing laboratory findings between COVID-19 and non-COVID-19 patient

	Normal Range	Total (N=403)	non-COVID-19 (N=18)	COVID-19 (N=385)	P-value
Laboratory Findings					
Infection markers					
Eosinophil percentage, %	0.4~8	0.54 (0.2-1.3)	2 (0.1-3.5)	0.5 (0.2-1.2)	0.035
White blood cell count, g/L	3.5~9.5	5.7 (4.4-7.4)	4.42 (2.3-7.2)	5.7 (4.6-7.4)	0.005
Liver injury markers					
Albumin, g/L	35~55	38.5 (35-42)	33.8 (30-36)	38.95 (36-42)	<0.001
Alanine transaminase	9~50	20.1 (13-34)	40 (23-96)	19 (12-29)	<0.001
Heart injury markers					
Aspartate Transaminase	13~35	29.6 (21-45)	51 (30-82)	27.4 (20-41)	<0.001
Lactate Dehydrogenase, U/L	80~285	259.15 (210-380)	594 (470-830)	242.85 (200-330)	<0.001
N-terminal pro-brain natriuretic peptide	0~125	94.415 (51-440)	1259.4 (340-2300)	90.285 (49-410)	0.045
Blood examination					
Activated partial thromboplastin time	24~36	30.25 (28-32)	36 (35-42)	30.1 (28-32)	<0.001
Hematocrit, %	40~50	39.8 (36-43)	36.775 (26-40)	39.85 (36-43)	<0.001
Hemoglobin, g/L	130~175	135 (120-150)	121 (95-130)	135.25 (120-150)	0.003
Calcium, mmol/L	2.08~2.8	2.09 (2-2.2)	1.97 (1.9-2)	2.1 (2-2.2)	<0.001
Potassium, mmol/L	3.5~5.3	3.95 (3.6-4.3)	3.405 (3.2-3.8)	4.02 (3.7-4.4)	<0.001
Sodium, mmol/L	137~147	139 (136-141)	137.8 (135.2-139.5)	139 (136-141)	0.032

Figures

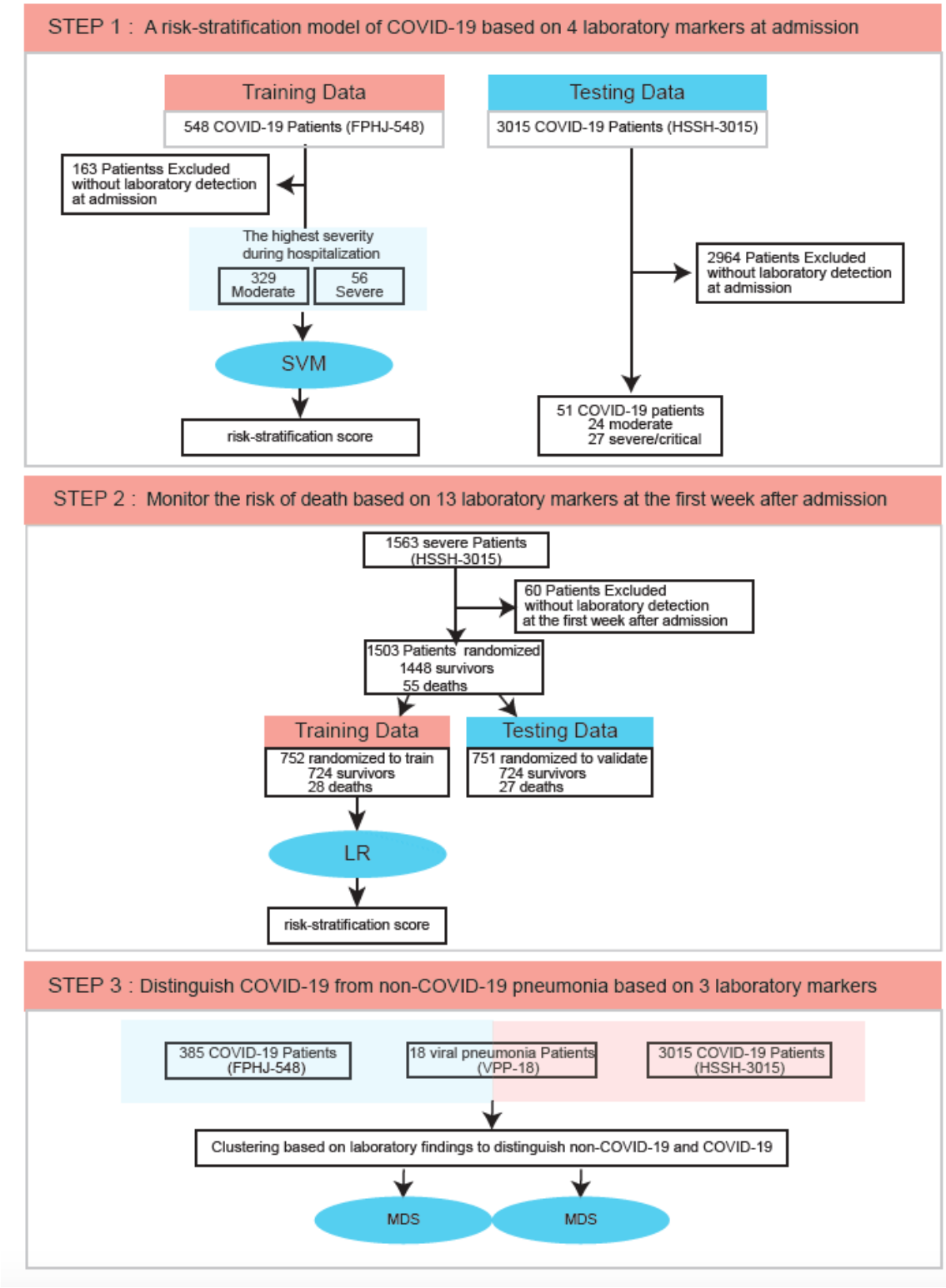

Figure 1

Flowchart of data processing. SVM: Support Vector Machine; LR: Logistic Regression; MDS: Multidimensional Scaling.

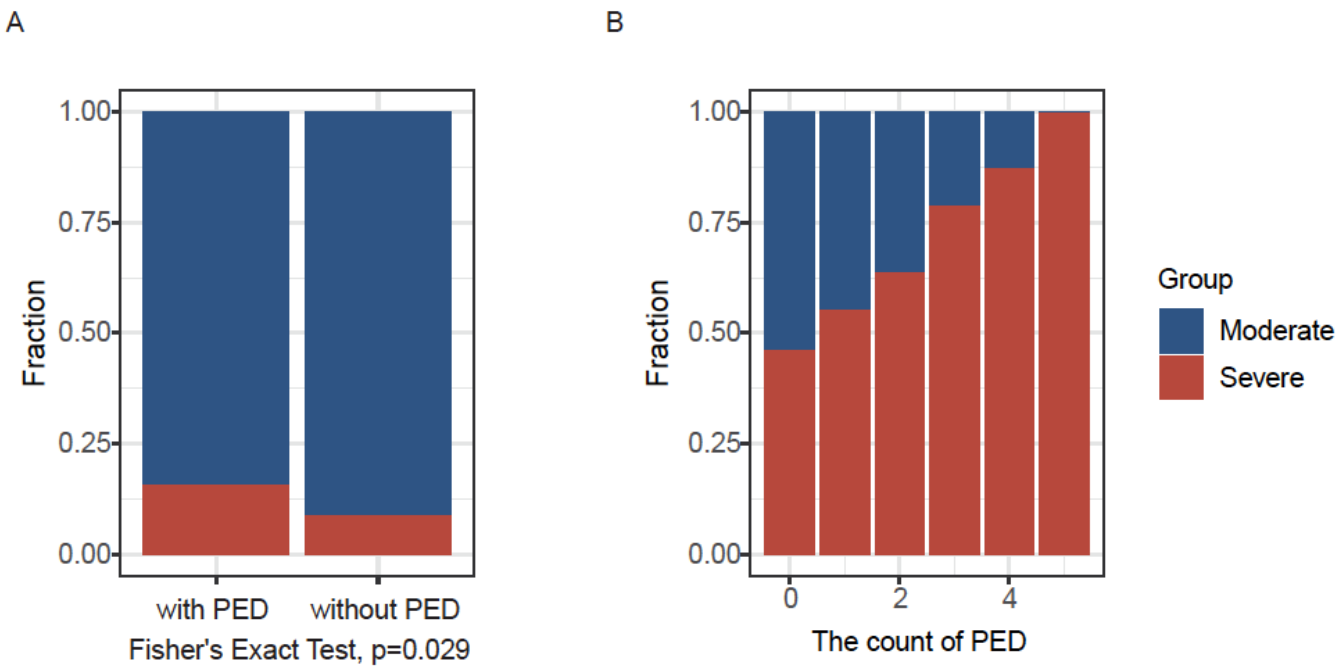

Figure 2

The impact of the pre-existing disease (PED) on the progression of COVID-19. A. The fraction of severe cases in patients with or without inPED in FPHJ-548 dataset. . The red represents the severe COVID-19 patients and blue represents moderate COVID-19 patients. B. The fraction of severe cases in patients with different numbers of PEDin HSSH-3015 dataset. The red represents the severe COVID-19 patients and blue represents moderate COVID-19 patients.

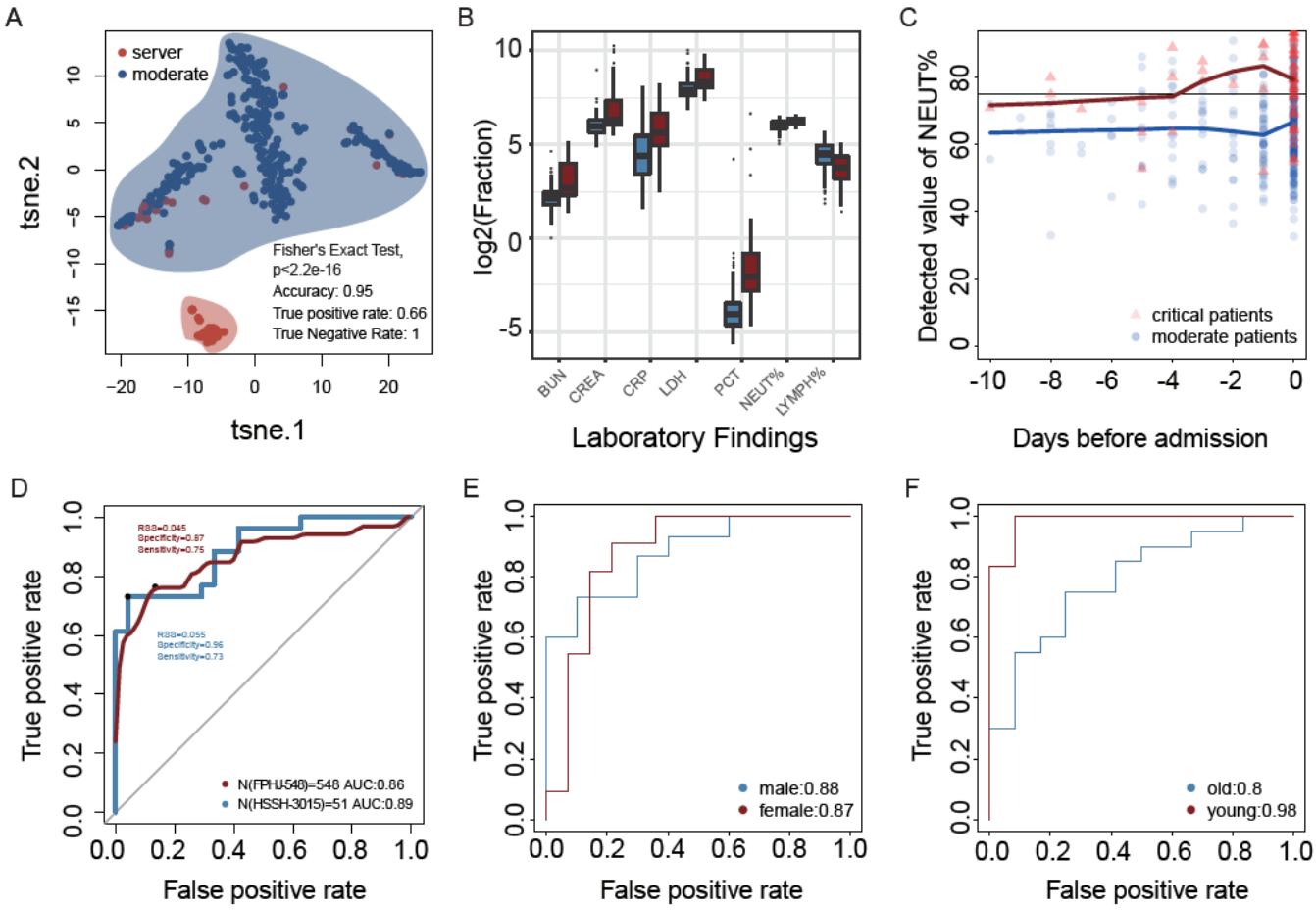

Figure 3

Laboratory findings for predicting progression of COVID-19 at admission. A. PCA was performed on 329 moderate COVID-19 patients and 56 severe COVID-19 patients based on 40 laboratory findings (FPHJ-548 dataset). The red represents the severe patients and blue represents moderate patients. B. The difference of representative clinical markers between moderate and severe patients in FPHJ-548 dataset. The red represents the severe patients and the blue represents the moderate patients. C. The change of neutrophils percentage during the period before admission. The black line is the maximum of reference value. The blue dot represents moderate patients and the red triangle represents severe patients. D. The ability of the model to distinguish severe from moderate patients based on 4 laboratory findings at admission. The x-axis is specificity and the y-axis represents sensitivity. The red solid line represents the mean of the five-fold cross-validation. The blue represents the AUC of HSSH-3015. E-F. The ability of the model in patients with different sex and age.

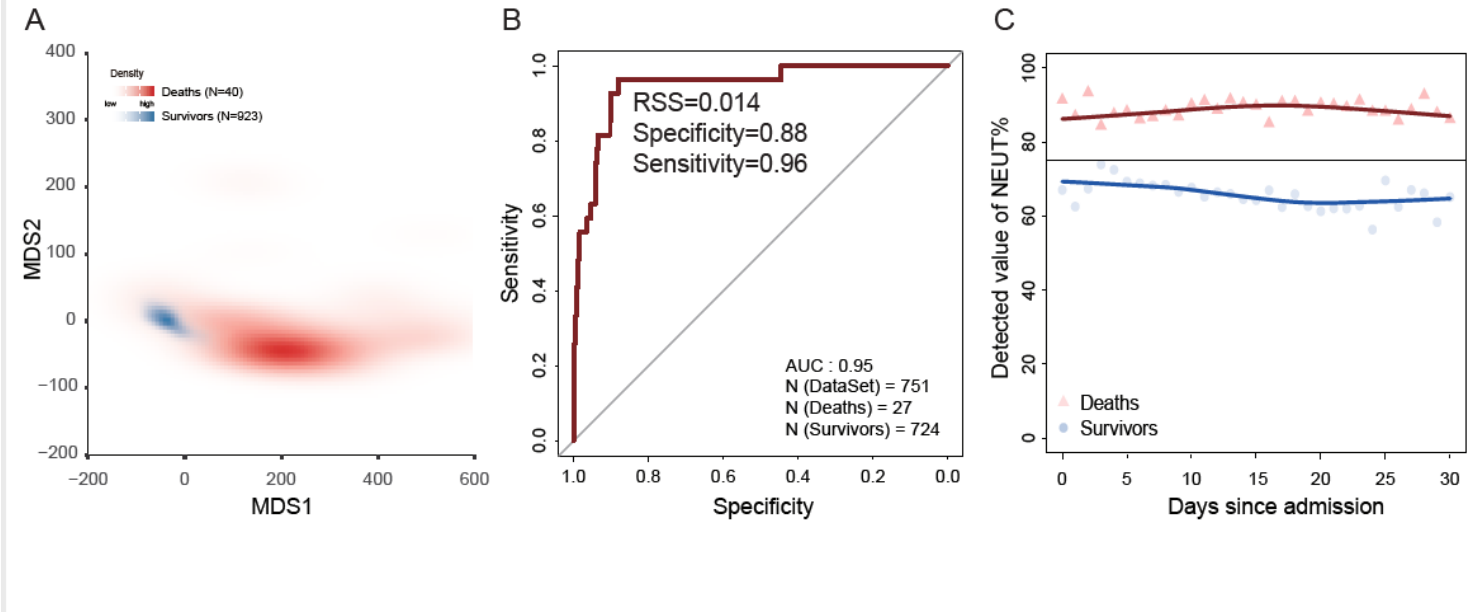

Figure 4

Laboratory findings to predict the clinical outcome of COVID-19 within the first week after admission. A. MDS plot for distinguishing deaths from survivors based on 13 laboratory findings in training dataset. Red represents deaths and blue represents survivors. B. The ability of the model for distinguishing deaths from survivors based on 13 laboratory findings in test dataset. The x-axis is specificity and the y-axis represents sensitivity. C. The change in the percentage of neutrophils since admission. The black line is the maximum reference value. The blue dot represents survivors and red represents deaths.

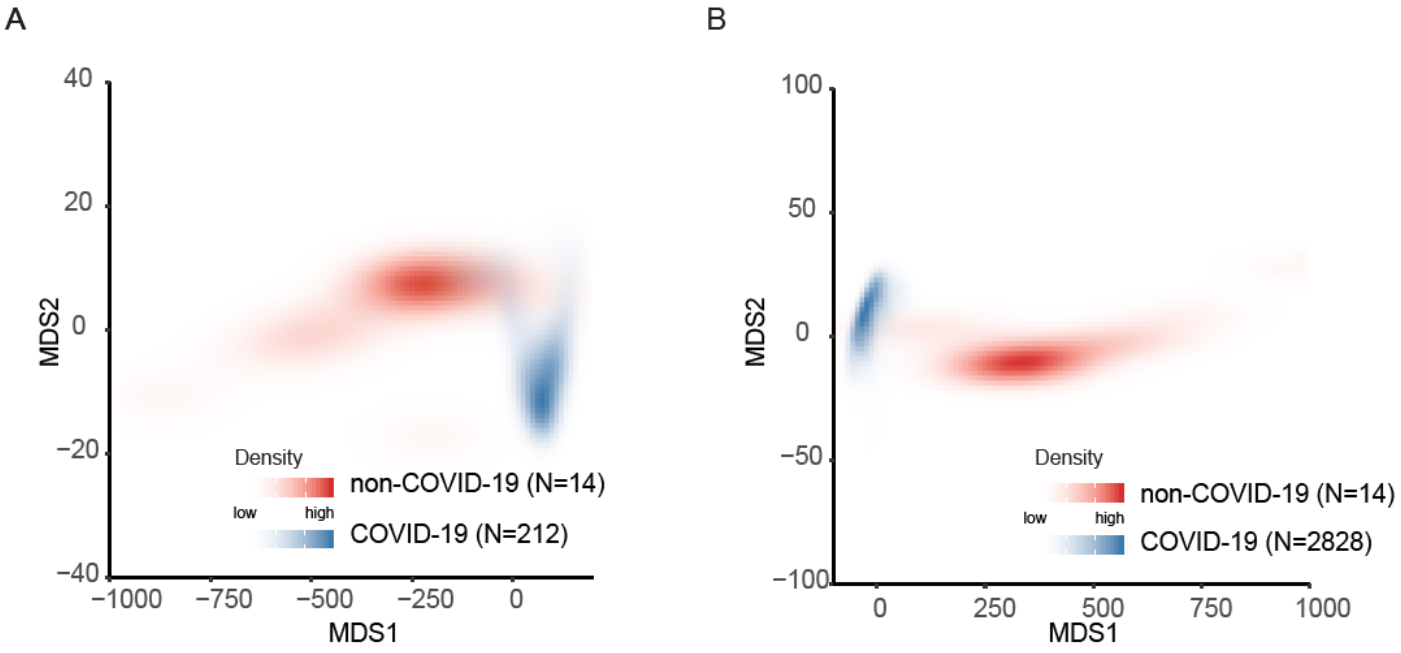

Figure 5

MDS plot for distinguishing non-COVID-19 from COVID-19 based on laboratory findings. Red represents non-COVID-19 and blue represents COVID-19. The depth of the color represents the density. Ashows the difference between FPHJ-548 and VPP-18 daatset. B shows the difference between VPP-18 and HSSH-3015 dataset.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx
- FigureS1.pdf