

A machine learning model to predict death outcome in severe COVID-19 patients

Qiao Yang

The 941st Hospital of the PLA Joint Logistic Support Force

Jixi Li

Xinqiao Hospital

Zhijia Zhang

Xinqiao Hospital

Xiaocheng Wu

Xinqiao Hospital

Tongquan Liao

Xinqiao Hospital

Shiyong Yu

Xinqiao Hospital

Zaichun You

Army Medical University

Xianhua Hou

Southwest Hospital, Army Medical University

Jun Ye

Southwest Hospital, Army Medical University

Gang Liu

Xinqiao Hospital

Siyuan Ma

State Key Laboratory of Trauma, Burns and Combined Injury

Ganfeng Xie

Southwest Hospital, Army Medical University

Yi Zhou

Xinqiao Hospital

Mengxia Li

Cancer Center, Army Medical Center

Meihui Wu

Army Medical Center

Yimei Feng

Xinqiao Hospital

Weili Wang

Kidney Center of PLA, Army Medical University

Lufeng Li

Southwest Hospital, Army Medical University

Dongjing Xie

Xinqiao Hospital

Yunhui Hu

Southwest Hospital, Army Medical University

Xi Liu

Army Medical University

Bin Wang

Xinqiao Hospital

Songtao Zhao

Southwest Hospital, Army Medical University

Li Li

Army Medical Center

Chunmei Luo

Xinqiao Hospital

Tang Tang

Xinqiao Hospital

Hongmei Wu

Xinqiao Hospital

Tianyu Hu

Xinqiao Hospital

Guangrong Yang

Xinqiao Hospital

Bangyu Luo

Xinqiao Hospital

Lingchen Li

Xinqiao Hospital

Xiu Yang

Xinqiao Hospital

Qi Li

Xinqiao Hospital

Zhi Xu

Xinqiao Hospital

Hao Wu

Xinqiao Hospital


Jianguo Sun (✉ sunjg09@aliyun.com)

Xinqiao Hospital

Research Article

Keywords: COVID-19, machine learning model, decision tree, LDH, NLR, CRP

DOI: <https://doi.org/10.21203/rs.3.rs-96621/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

The novel coronavirus disease 2019 (COVID-19) spreads rapidly among people and causes a global pandemic. It is of great clinical significance to identify COVID-19 patients with high risk of death.

Results

Of the 2,169 COVID-19 patients, the median age was 61 years and male patients accounted for 48%. A total of 646 patients were diagnosed with severe illness, and 75 patients died. Obvious differences in demographics, clinical characteristics and laboratory examinations were found between survivors and non-survivors. A decision tree classifier, including three biomarkers, neutrophil-to-lymphocyte ratio, C-reactive protein and lactic dehydrogenase, was developed to predict death outcome in severe patients. This model performed well both in train dataset and test dataset. The accuracy of this model was 0.98 and 0.98, respectively.

Conclusion

The machine learning model was robust and effective in predicting the death outcome in severe COVID-19 patients.

Background

The novel coronavirus disease 2019 (COVID-19) has been a global pandemic. The most common symptoms of COVID-19 patients were fever, followed by dry cough, fatigue, dyspnea, etc[1, 2]. A small piece of patients had digestive symptoms, such as nausea, vomiting and diarrhoea[3, 4]. A study from the Chinese Center for Disease Control and Prevention reported that about 81% COVID-19 patients were considered as mild, and the proportion of severe and critical patients was 14% and 5%, respectively[5]. The mortality in overall population was 3.2%, but it increased to 49% in critical population[5]. Hence, how to use effective biomarkers to identify patients with poor clinical outcomes have caused extensive concern.

COVID-19 patients with comorbidities were considered to be prone to having poor clinical outcomes. A study reported that COVID-19 patients with chronic obstructive pulmonary disease, diabetes, hypertension and malignancy had a higher risk of admission to an intensive care unit (ICU), invasive ventilation or death[6]. Another study demonstrated that the risk factors included older age, high Sequential Organ Failure Assessment score, and higher D-dimer expression on admission[7]. To help physicians identify potential patients with critical illness, Liang and colleagues developed a risk score model based on ten characteristics of COVID-19 patients on admission to calculate the probability of developing critical illness[3].

It is of great clinical significance to identify COVID-19 patients with high risk of death. In this study, we aimed to develop a machine learning model to distinguish COVID-19 patients with high risk of death from those without.

Methods

Data sources

A total of 2,169 adult patients (aged ≥ 18 years) were enrolled from Wuhan, China between February 10th and April 15th, 2020. All patients were confirmed with COVID-19 infection by real-time reverse-transcription polymerase-chain-reaction (RT-PCR) assay. In addition, medical records, including demographics, clinical characteristics and laboratory examinations on admission of all patients were collected. This study was approved by the Ethics Committee of the Taikang Hospital (TKTJLL-007), and performed in accordance with the Declaration of Helsinki. The Ethics Committee of the Taikang Hospital waived the need for informed consent of each patients. This study was registered in the Clinical Trials Register (NCT04347369, <https://clinicaltrials.gov/>).

Study design

First of all, we performed a difference analysis of medical records between severe group and non-severe group. All the patients meeting the severity diagnosis criteria during hospitalization were assigned into the severe group. We defined disease severity according to the Seventh Revised Trial Version of the COVID-19 Diagnosis and Treatment Guidance (2020) of China[8]. Next, we performed difference analyses of medical records between survivors and non-survivors. The definition of survivors was patients who were discharged from hospital or still in hospital at the end of study. Last, we developed a decision tree classifier to identify risk factors for death outcome.

Statistical analysis

Continuous variables were described as median and interquartile range (IQR) and compared by the Mann-Whitney U test. Categorical variables were represented as frequencies and compared by Pearson's χ^2 test. All statistical analyses were performed and the decision tree classifier was developed using R software (version 3.5.2). The following R packages were used: CBCgrps, rpart, rpart.plot and pROC. A two-sided P value < 0.05 was considered statistically significant.

Results

Of the 2,169 COVID-19 patients confirmed by RT-PCR, the median age was 61 years (IQR, 50-70; range, 18-100 years). Male patients accounted for 48% (1036 cases) and female patients 52% (1133 cases). Approximately 8% of patients (184 cases) had smoking history. On admission, 117 (5%) patients had high body temperature, 270 (12%) had low oxygen saturation, 359 (17%) had abnormal heart rates and 596 had faster respiratory rates. In total, 1134 (52%) patients had at least one comorbidity, and the common comorbidities were hypertension, diabetes and coronary heart disease. In addition, 728 (34%) patients had one system symptom, 1130 (52%) patients had two systems symptoms and 218 (10%) patients had three or more systems symptoms. The most common systems symptoms were respiratory symptoms, systemic symptoms and digestive symptoms (Table 1).

A total of 646 (29.8%) patients were diagnosed with severe illness during hospitalization. Compared to non-severe group, severe group had a significantly higher median age (68 vs. 58 years, $p < 0.001$) and a higher proportion of male patients. On admission, a higher proportion of high body temperature, low oxygen saturation, abnormal heart rate and faster respiratory rate were found in severe group. Moreover, patients in severe group had a higher proportion of comorbidities and symptoms. No difference was found in the smoking history (Table 1). When comparing laboratory examinations between the two groups, we found that severe group had a significantly higher white blood cell (WBC) count, neutrophil count, neutrophil-to-lymphocyte ratio (NLR), and C-reactive protein (CRP), lactic dehydrogenase (LDH), interleukin-6 (IL-6), procalcitonin and D-dimer levels, but lower lymphocyte count, eosinophilia count, basophilia count, red blood cell (RBC) count, hemoglobin and platelet count. No difference was found in monocyte count (Table 1).

Until April 15th, 2020, 75 patients were dead. Differences in demographics and clinical characteristics between survivors and non-survivors were similar to the differences between severe and non-severe groups. By laboratory examinations comparison, a much higher WBC count, neutrophil count, NLR, and higher CRP, LDH, IL-6, procalcitonin and D-dimer levels were found in non-survivors (Table 2). RBC count and hemoglobin level showed no difference between the two groups. Other laboratory indexes were lower in non-survivors (Table 2).

To explore crucial predictive biomarkers of disease mortality in severe patients, we used a machine learning model, decision tree, to identify related biomarkers. Laboratory indexes with missing values over 20% were excluded, including IL-6, procalcitonin and D-dimer. Then all missing values were imputed with mean value of each laboratory index. All severe patients were randomly split into train dataset and test dataset in a ratio of 7:3. A total of 452 patients were included in train dataset, including 57 non-survivors. In this step, a decision tree classifier was developed to differentiate non-survivors from survivors. As shown in Figure 1, three biomarkers were included in the decision tree classifier, including LDH, NLR and CRP. The threshold of each biomarker helped to classify each patient into survivor group or non-survivor group. The area under the curve (AUC) of the receiver operating characteristic of this model was 0.96, which was higher than each single biomarker (Figure 2). The associated confusion matrix of train dataset was presented in Supplementary Table 1. The accuracy of this model was 0.98. The precision, recall and F1 score for survivor prediction was 0.97, 1.00 and 0.98, respectively. For non-survivors, the precision, recall and F1 score was 1.00, 0.81 and 0.90, respectively (Table 3).

To validate the performance of the decision tree classifier, we applied it to the test dataset. The associated confusion matrix of test dataset was presented in Supplementary Table 1. The accuracy in test dataset was also 0.98. The precision, recall and F1 score for survivor prediction in test dataset was 0.98, 0.99 and 0.98, respectively. For non-survivor prediction in test dataset, the precision, recall and F1 score was 0.94, 0.83 and 0.88, respectively (Table 3).

Discussion

In this study, we found that COVID-19 patients in severe group or non-survivor group had a higher median age. Also, these patients had a higher proportion of comorbidities and symptoms than their counterparts. Most importantly, we proposed a machine learning model to

quickly quantify the risk of death based on three biomarkers (LDH, NLR and CRP), which could be easily obtained on admission. This model had a high AUC of 0.96, and performed well in both train dataset and test dataset.

Zhang and colleagues[9] reported that the median age in a small cohort of COVID-19 non-survivors was 72.5, similar to our findings. In addition, both Zhang and we found that male COVID-19 patients accounted for the majority of the non-survivors. They also reported that approximately 76.8% of the non-survivors had comorbidities, including hypertension, heart disease, diabetes, cerebrovascular disease and cancer[9].

NLR is one of the research hotspots of inflammatory biomarkers in infectious diseases. It can comprehensively reflect the inflammatory response and immune status in patients with infectious diseases[10–12]. In COVID-19 patients, elevated NLR on admission was reported to be significantly associated with disease severity[13, 14]. Liu and colleagues proposed a simple model based on NLR and age to stratify COVID-19 patients into four groups[15]. COVID-19 patients with age < 50 years old and $NLR < 3.13$ or $NLR \geq 3.13$ had no risk of severity, and these patients should be treated in a community hospital, home isolation or general isolation ward. While COVID-19 patients with age ≥ 50 and $NLR < 3.13$ or $NLR \geq 3.13$ had a higher risk of severity, and these patients should be admitted to isolation ward or ICU with active treatment and care. In addition, Yang and coworkers found that approximately 46.1% of the mild COVID-19 patients could become severely ill if patients had an age ≥ 49.5 years old and $NLR \geq 3.3$ [14]. The dynamic change of NLR could also be used to distinguish severe patients from mild/moderate patients. A study demonstrated that NLR in severe group always kept a higher level on day 1, 4 and 14 compared with mild/moderate group[16]. In this study, we found that $NLR \geq 6.9$ was a high-risk factor for death in severe COVID-19 patients. In train dataset, 54 out of 113 patients with $NLR \geq 6.9$ died, while only 3 out of 339 patients with $NLR < 6.9$ died. In test dataset, 18 out of 43 patients with $NLR \geq 6.9$ died, while no patients with $NLR < 6.9$ died.

CRP reflects a persistent inflammatory activity state, and helps in assessing the severity of infectious patients[17]. A few studies have demonstrated that a higher CRP expression on admission was observed in severe COVID-19 patients compared with non-severe COVID-19 patients[17, 18]. LDH was a biomarker of severe illness and poor prognosis in COVID-19 patients[19]. Zeng et al. found that LDH decreased within 10 days after admission in non-critical COVID-19 patients, but did not decrease obviously in critical patients or non-survivors[20]. These findings indicate that CRP and LDH play important roles in COVID-19 patients. In this study, LDH and CRP were also risk factors for death in severe COVID-19 patients. For LDH expression, in train dataset, 53 out of 74 patients with $LDH \geq 330$ IU/L died, while only 4 out of 378 patients with $LDH < 330$ IU/L died. In test dataset, 16 out of 31 patients with $LDH \geq 330$ IU/L died, while only 2 out of 163 patients with $LDH < 330$ IU/L died. For CRP expression, in train dataset, 52 out of 126 patients with $CRP \geq 27$ mg/L died, while only 5 out of 326 patients with $CRP < 27$ mg/L died. In test dataset, 17 out of 53 patients with $CRP \geq 27$ mg/L died, while only 1 out of 141 patients with $CRP < 27$ mg/L died.

To date, a few prognostic models of COVID-19 have been developed. Dong et al. found that hypertension, higher NLR and increased NT-proBNP value were significantly poorer prognostic factors in hospitalized COVID-19 patients[21]. A nomogram for predicting 14-day and 21-day probability of in-hospital survival based on the three factors was developed, with high c-index and AUC in both training and validation cohorts. Moreover, Wang et al. constructed a nomogram based on age, chest tightness, AST and BUN on admission to predict the risk of death in critical COVID-19 patients[22]. The nomogram had a sensitivity of 96.0%, a specificity of 74.1% and an AUC of 0.893.

Conclusion

In summary, this study found that male COVID-19 patients were more prone to experience severe illness and death. Clinical characteristics and laboratory examinations were significantly different between severe and non-severe groups, as well as between survivors and non-survivors. Most importantly, we identified three biomarkers (LDH, NLR and CRP) on admission as risk factors for death, and developed a simple decision tree classifier to help clinicians rapidly identify patients at high risk of death and to give priority treatment and intensive care.

List Of Abbreviations

COVID-19 – novel coronavirus disease 2019; ICU – intensive care unit;

RT-PCR – real-time reverse-transcription polymerase-chain-reaction;

IQR – interquartile range; WBC – white blood cell;

NLR – neutrophil-to-lymphocyte ratio; CRP – C-reactive protein;

LDH – lactic dehydrogenase; IL-6 – interleukin-6;

RBC – red blood cell; AUC – area under the curve;

Declarations

Ethics approval and consent to participate:

This study was approved by the ethics committee of the Ethics Committee of the Taikang Hospital (TKTJLL-007).

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

None.

Authors' contributions

Contributions to the conception: Li Q, Xu Z, Wu H, Sun JG;

Design of the work: Liao TQ, Yu SY, You ZC;

The acquisition, analysis, and interpretation of data: Hou XH, Ye J, Liu G, Ma SY, Xie GF, Zhou Y, Li MX, Wu MH, Feng YM, Wang WL, Li LF, Xie DJ, Hu YH, Liu X, Wang B, Zhao ST, Li L, Luo CM, Tang T, Wu HM, Hu TY;

The preparation of figures and tables: Yang GR, Luo BY, Li LC, Yang X;

Draft the work and substantively revised it: Yang Q, Li JX, Zhang ZJ, Wu XC, Sun JG.

All authors have approved to submit this study and any substantially modified version that involves the author's contribution to the study. And all authors have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Acknowledgements

This study was supported by the Research Projects of the Joint Logistics Support Force of PLA (TKTJKY2020003, TKTJKY2020029 and TKTJKY2020136).

The authors thank all the patients for providing their clinical data.

References

[1] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* 2020;323:1061.

- [2] Tian J, Yuan X, Xiao J, Zhong Q, Yang C, Liu B, Cai Y, Lu Z, Wang J, Wang Y, Liu S, Cheng B, Wang J, Zhang M, Wang L, Niu S, Yao Z, Deng X, Zhou F, Wei W, Li Q, Chen X, Chen W, Yang Q, Wu S, Fan J, Shu B, Hu Z, Wang S, Yang X, Liu W, Miao X, Wang Z. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. *The Lancet Oncology* 2020;21:893.
- [3] 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. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA INTERN MED* 2020;180:1081.
- [4] 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. *The Lancet (British edition)* 2020;395:497.
- [5] 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 : the journal of the American Medical Association* 2020;323:1239.
- [6] Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, Liu X, Chen R, Tang C, Wang T, Ou C, Li L, Chen P, Sang L, Wang W, Li J, Li C, Ou L, Cheng B, Xiong S, Ni Z, Xiang J, Hu Y, Liu L, Shan H, Lei C, 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, Cheng L, Ye F, Li S, Zheng J, Zhang N, Zhong N, He J. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *EUR RESPIR J* 2020;55:2000547.
- [7] 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. *The Lancet* 2020;395:1054.
- [8] National Health Commission of the People's Republic of China.
Chinese management guideline for COVID-19 (version 7.0). March 3, 2020.
<http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf>
(accessed March 27, 2020; in Chinese).
- [9] Zhang B, Zhou X, Qiu Y, Song Y, Feng F, Feng J, Song Q, Jia Q, Wang J. Clinical characteristics of 82 cases of death from COVID-19. *PLOS ONE* 2020;15:e235458.
- [10] Chen XQ, Xue CR, Hou P, Lin BQ, Zhang JR. Lymphocyte-to-monocyte ratio effectively predicts survival outcome of patients with obstructive colorectal cancer. *World J Gastroenterol* 2019;25:4970.
- [11] Zhang HF, Ge YL, Wang HY, Zhang Q, Li WQ, Chen Y, Chen QC, Jin JJ, Xu J, Zhang S, Xu TT, Zhang X, Yu HL, Zhu XY, Liu CH, Bai J, Zhang JB, Xie BQ, Wang Y, Fu AS, Zhang PP. Neutrophil-to-Lymphocyte Ratio Improves the Accuracy and Sensitivity of Pneumonia Severity Index in Predicting 30-Day Mortality of CAP Patients. *CLIN LAB* 2019;65.
- [12] Ge YL, Zhang HF, Zhang Q, Zhu XY, Liu CH, Wang N, Zhang JB, Chen H, Chen Y, Li WQ, Li ZZ, Fu AS, Wang HY. Neutrophil-to-Lymphocyte Ratio in Adult Community-Acquired Pneumonia Patients Correlates with Unfavorable Clinical Outcomes. *CLIN LAB* 2019;65.
- [13] Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, Xie J, Guan W, Liang W, Ni Z, Hu Y, Liu L, Shan H, Lei C, 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, Zheng J, Zhang N, Li Y, He J, Li J, Li S, Zhong N. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J ALLERGY CLIN IMMUN* 2020;146:89.
- [14] Yang AP, Liu JP, Tao WQ, Li HM. The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *INT IMMUNOPHARMACOL* 2020;84:106504.
- [15] Liu J, Liu Y, Xiang P, Pu L, Xiong H, Li C, Zhang M, Tan J, Xu Y, Song R, Song M, Wang L, Zhang W, Han B, Yang L, Wang X, Zhou G, Zhang T, Li B, Wang Y, Chen Z, Wang X. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J TRANSL MED* 2020;18:206.

- [16] Fu J, Kong J, Wang W, Wu M, Yao L, Wang Z, Jin J, Wu D, Yu X. The clinical implication of dynamic neutrophil to lymphocyte ratio and D-dimer in COVID-19: A retrospective study in Suzhou China. *THROMB RES* 2020;192:3.
- [17] Danwang C, Endomba FT, Nkeck JR, Wouna D, Robert A, Noubiap JJ. A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). *Biomark Res* 2020;8:37.
- [18] Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, Akdis CA, Gao YD. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *ALLERGY* 2020;75:1730.
- [19] Ke C, Yu C, Yue D, Zeng X, Hu Z, Yang C. Clinical Characteristics of confirmed and clinically diagnosed patients with 2019 novel coronavirus pneumonia: a single-center, retrospective, case-control study. *Med Clin (Barc)* 2020.
- [20] Zeng Z, Yu H, Chen H, Qi W, Chen L, Chen G, Yan W, Chen T, Ning Q, Han M, Wu D. Longitudinal changes of inflammatory parameters and their correlation with disease severity and outcomes in patients with COVID-19 from Wuhan, China. *CRIT CARE* 2020;24:525.
- [21] Dong YM, Sun J, Li YX, Chen Q, Liu QQ, Sun Z, Pang R, Chen F, Xu BY, Manyande A, Clark TG, Li JP, Orhan IE, Tian YK, Wang T, Wu W, Ye DW. Development and Validation of a Nomogram for Assessing Survival in Patients with COVID-19 Pneumonia. *CLIN INFECT DIS* 2020.
- [22] Wang B, Zhong F, Zhang H, An W, Liao M, Cao Y. Risk factors analysis and nomogram construction of non-survivors in critical patients with COVID-19. *JPN J INFECT DIS* 2020.

Tables

Table 1: Demographics, clinical characteristics and laboratory findings of severe COVID-19 patients and non-severe COVID-19 patients

Variables	Total (N=2169)	Severe (N=646)	Non-severe (N=1523)	P value ^a
Age, years old	61 (50, 70)	68 (60, 76)	58 (47, 66)	< 0.001
Sex				< 0.001
Male	1036 (48)	360 (56)	676 (44)	
Female	1133 (52)	286 (44)	847 (56)	
Smoking history				0.582
Never	1985 (92)	587 (91)	1398 (92)	
Former	80 (4)	28 (4)	52 (3)	
Current	104 (4)	31 (5)	73 (5)	
Body temperature				< 0.001
< 37.3 °C	2052 (95)	589 (91)	1463 (96)	
37.3-38.0 °C	87 (4)	37 (6)	50 (3)	
>38.0 °C	30 (1)	20 (3)	10 (1)	
Oxygen saturation				< 0.001
≤93%	270 (12)	270 (42)	0 (0)	
>93%	1899 (88)	376 (58)	1523 (100)	
Heart rate				0.009
Normal	1810 (83)	518 (80)	1292 (85)	
Abnormal	359 (17)	128 (20)	231 (15)	
Respiratory rate				< 0.001
≤20 per minute	1573 (73)	343 (53)	1230 (81)	
21-29 per minute	528 (24)	235 (36)	293 (19)	
≥ 30 per minute	68 (3)	68 (11)	0 (0)	
No. of comorbidities				< 0.001
0	1035 (48)	196 (30)	839 (55)	
1	591 (27)	186 (29)	405 (27)	
≥2	543 (25)	264 (41)	279 (18)	
No. of systems symptoms				< 0.001
0	93 (4)	12 (2)	81 (5)	
1	728 (34)	164 (25)	564 (37)	
2	1130 (52)	386 (60)	744 (49)	
≥3	218 (10)	84 (13)	134 (9)	
White blood cell count, ×10⁹/L	5.8 (4.74, 7.06)	6.23 (4.98, 8)	5.68 (4.69, 6.83)	< 0.001
Neutrophil count, ×10⁹/L	3.48 (2.62, 4.59)	4.02 (3, 5.93)	3.28 (2.52, 4.25)	< 0.001
Lymphocyte count, ×10⁹/L	1.53 (1.13, 1.91)	1.2 (0.8, 1.69)	1.63 (1.27, 1.97)	< 0.001
Neutrophil-to-lymphocyte ratio	2.23 (1.62, 3.31)	3.21 (1.99, 6.62)	2.03 (1.5, 2.77)	< 0.001
Monocyte count, ×10⁹/L	0.47 (0.37, 0.59)	0.48 (0.36, 0.62); n=646	0.47 (0.38, 0.59); n=1516	0.906
Eosinophilia count, ×10⁹/L	0.11 (0.06, 0.18)	0.09 (0.03, 0.17); n=646	0.11 (0.07, 0.18); n=1516	< 0.001
Basophilia count, ×10⁹/L	0.03 (0.01, 0.04)	0.02 (0.01, 0.03); n=646	0.03 (0.02, 0.04); n=1517	< 0.001
Red blood cell count, ×10⁹/L	4.02 (3.63, 4.4)	3.85 (3.44, 4.24); n=644	4.08 (3.72, 4.46); n=1513	< 0.001
Hemoglobin, g/L	121 (110, 133)	117 (102, 128); n=644	123 (113, 134); n=1514	< 0.001
Platelet count, × 10⁹/L	221.5 (180, 272)	216 (164, 275); n=645	224 (185, 270); n=1519	0.008
C-reactive protein, mg/L	1.32 (0.5, 7.35)	6.38 (1, 32.14); n=638	0.77 (0.5, 3.36); n=1438	< 0.001
Lactic dehydrogenase, IU/L	176.92 (149.9, 216.95)	212 (172.2, 279.36); n=625	166.92 (144.97, 197.7); n=1437	< 0.001
Interleukin-6, pg/mL	2.48 (1.5, 6.36)	6.84 (2.73, 21.23); n=404	1.77 (1.5, 3.68); n=961	< 0.001
Procalcitonin, ng/mL	0.05 (0.03, 0.08)	0.07 (0.04, 0.14); n=514	0.04 (0.03, 0.06); n=947	< 0.001
D-dimer, µg/mL	0.39 (0.18, 0.8)	0.64 (0.31, 1.49); n=420	0.3 (0.14, 0.55); n=701	< 0.001

Data are n (%), n/N (%), or median (IQR), unless specified otherwise. Temperature, oxygen saturation, heart rate, respiratory rate were detected at rest when patients were admitted to hospital.

^a P values indicate differences between severe and non-severe patients.

Table 2: Demographics, clinical characteristics and laboratory findings of survivors and non-survivors

Variables	Non-survivor (N=75)	Survivor (N=2094)	P value ^a
Age, years old	72 (67, 82)	61 (50, 69)	< 0.001
Sex			0.003
Male	49 (65)	987 (47)	
Female	26 (35)	1107 (53)	
Smoking history			0.325
Never	66 (88)	1919 (92)	
Former	3 (4)	77 (4)	
Current	6 (8)	98 (5)	
Body temperature			< 0.001
< 37.3 °C	54 (72)	1998 (95)	
37.3-38.0 °C	12 (16)	75 (4)	
>38.0 °C	9 (12)	21 (1)	
Oxygen saturation			< 0.001
≤93%	59 (79)	211 (10)	
>93%	16 (21)	1883 (90)	
Heart rate			< 0.001
Normal	51 (68)	1759 (84)	
Abnormal	24 (32)	335 (16)	
Respiratory rate			< 0.001
≤20 per minute	26 (35)	1547 (74)	
21-29 per minute	35 (47)	493 (24)	
≥ 30 per minute	14 (19)	54 (3)	
No. of comorbidities			< 0.001
0	11 (15)	1024 (49)	
1	17 (23)	574 (27)	
≥2	47 (63)	496 (24)	
No. of systems symptoms			0.03
0	3 (4)	90 (4)	
1	14 (19)	714 (34)	
2	49 (65)	1081 (52)	
≥3	9 (12)	209 (10)	
White blood cell count, ×10⁹/L	10.2 (7.27, 14.09)	5.75 (4.71, 6.94)	< 0.001
Neutrophil count, ×10⁹/L	9.62 (6.8, 13.53)	3.42 (2.59, 4.46)	< 0.001
Lymphocyte count, ×10⁹/L	0.56 (0.4, 0.8)	1.56 (1.17, 1.92)	< 0.001
Neutrophil-to-lymphocyte ratio	16.98 (11.88, 26.16)	2.18 (1.6, 3.15)	< 0.001
Monocyte count, ×10⁹/L	0.38 (0.22, 0.64); n=75	0.48 (0.38, 0.59); n=2087	0.002
Eosinophilia count, ×10⁹/L	0.01 (0, 0.04); n=75	0.11 (0.06, 0.18); n=2087	< 0.001
Basophilia count, ×10⁹/L	0.01 (0, 0.02); n=75	0.03 (0.02, 0.04); n=2088	< 0.001
Red blood cell count, ×10⁹/L	4.02 (3.31, 4.34); n=75	4.02 (3.64, 4.4); n=2083	0.051
Hemoglobin, g/L	118 (101, 134); n=75	121 (110, 132); n=2083	0.213
Platelet count, × 10⁹/L	172 (89, 263); n=75	223 (182, 273); n=2089	< 0.001
C-reactive protein, mg/L	89.33 (38.26, 135.91); n=75	1.18 (0.5, 6.09); n=2001	< 0.001
Lactic dehydrogenase, IU/L	429.14 (366, 541.2); n=73	175.2 (148.9, 211.2); n=1989	< 0.001
Interleukin-6, pg/mL	65.28 (20.41, 154.1); n=27	2.42 (1.5, 5.91); n=1338	< 0.001
Procalcitonin, ng/mL	0.28 (0.14, 0.66); n=63	0.04 (0.03, 0.08); n=1398	< 0.001
D-dimer, µg/mL	2.43 (0.78, 5.61); n=47	0.37 (0.17, 0.74); n=1074	< 0.001

Data are n (%), n/N (%), or median (IQR), unless specified otherwise. Temperature, oxygen saturation, heart rate, respiratory rate were detected at rest when patients were admitted to hospital.

^a P values indicate differences between survivors and non-survivors.

Table 3: Performance of the decision tree classifier on the train and test datasets

Train dataset	Precision	Recall	F1 score	Support
Survivor	0.97	1.00	0.98	406
Non-survivor	1.00	0.81	0.90	46
Accuracy			0.98	452
Test dataset	Precision	Recall	F1 score	Support
Survivor	0.98	0.99	0.98	178
Non-survivor	0.94	0.83	0.88	16
Accuracy			0.98	452

Figures

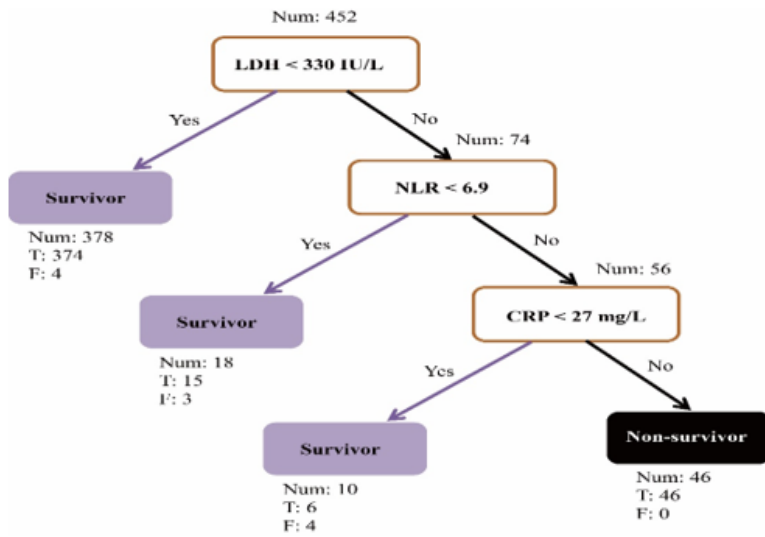


Figure 1

A decision tree classifier using three biomarkers and their thresholds in absolute value to predict death outcome in severe COVID-19 patients. Num, the number of patients in a class; T, the number of correctly classified patients; F, the number of misclassified patients; NLR, neutrophil-to-lymphocyte ratio; CRP, C-reactive protein; LDH, lactic dehydrogenase; COVID-19, novel coronavirus disease 2019.

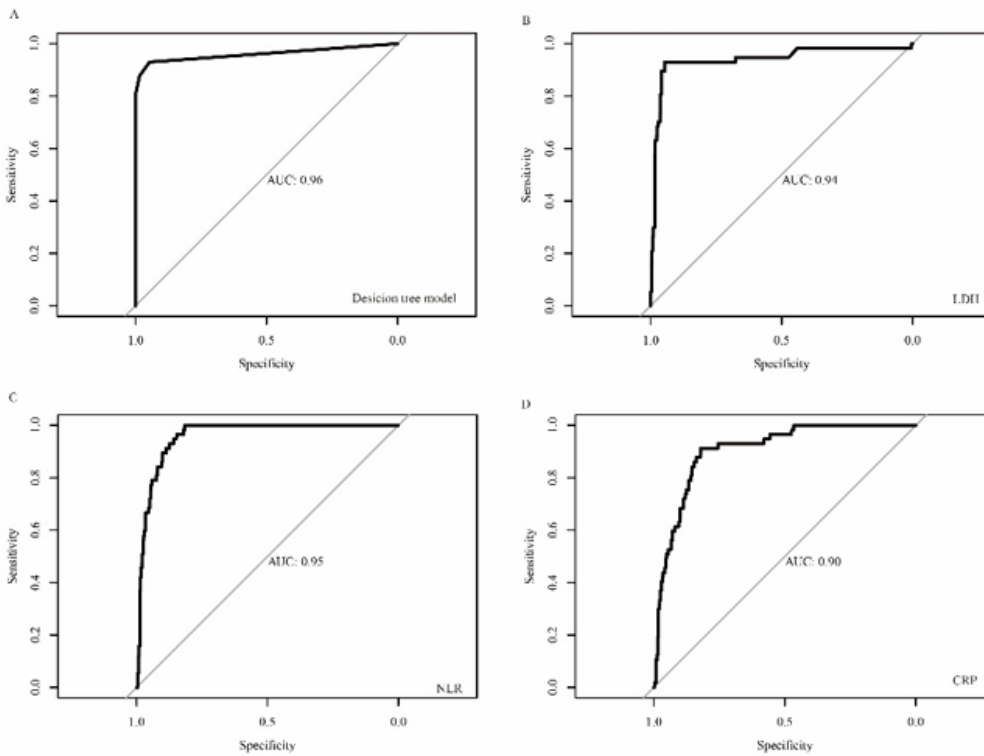


Figure 2

ROC curves for the decision tree classifier and each biomarker. (A) ROC curve for the decision tree classifier; (B) ROC curve for LDH; (C) ROC curve for NLR; (D) ROC curve for CRP. ROC, receiver operating characteristic; NLR, neutrophil-to-lymphocyte ratio; CRP, C-reactive protein; LDH, lactic dehydrogenase; AUC, area under the curve of ROC.

Supplementary Files

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

- [SupplementaryTable1.docx](#)