

RESEARCH

Open Access



Clinical characteristics and a decision tree model to predict death outcome in severe COVID-19 patients

Qiao Yang^{1†}, Jixi Li^{2†}, Zhijia Zhang^{3†}, Xiaocheng Wu^{4†}, Tongquan Liao⁵, Shiyong Yu⁶, Zaichun You⁷, Xianhua Hou⁸, Jun Ye⁹, Gang Liu¹⁰, Siyuan Ma¹¹, Ganfeng Xie¹², Yi Zhou², Mengxia Li¹³, Meihui Wu¹⁴, Yimei Feng¹⁵, Weili Wang¹⁶, Lufeng Li¹⁷, Dongjing Xie¹⁸, Yunhui Hu¹⁹, Xi Liu²⁰, Bin Wang¹⁰, Songtao Zhao¹⁷, Li Li²¹, Chunmei Luo²², Tang Tang²³, Hongmei Wu¹⁰, Tianyu Hu²⁴, Guangrong Yang², Bangyu Luo², Lingchen Li², Xiu Yang², Qi Li^{10*}, Zhi Xu^{10*}, Hao Wu^{5*} and Jianguo Sun^{2*}

Abstract

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

Methods: A total of 2169 adult COVID-19 patients were enrolled from Wuhan, China, from February 10th to April 15th, 2020. Difference analyses of medical records were performed between severe and non-severe groups, as well as between survivors and non-survivors. In addition, we developed a decision tree model to predict death outcome in severe patients.

Results: Of the 2169 COVID-19 patients, the median age was 61 years and male patients accounted for 48%. A total of 646 patients were diagnosed as severe illness, and 75 patients died. An older median age and a higher proportion of male patients were found in severe group or non-survivors compared to their counterparts. Significant differences in clinical characteristics and laboratory examinations were found between severe and non-severe groups, as well as between survivors and non-survivors. A decision tree, 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 training and test datasets. The accuracy of this model were 0.98 in both datasets.

Conclusion: We performed a comprehensive analysis of COVID-19 patients from the outbreak in Wuhan, China, and proposed a simple and clinically operable decision tree to help clinicians rapidly identify COVID-19 patients at high risk of death, to whom priority treatment and intensive care should be given.

*Correspondence: liqioliver@sina.com; xu_zhi999@163.com; xuzhihxk@tmmu.edu.cn; ewuhao@163.com; sunjg09@aliyun.com

[†]Qiao Yang, Jixi Li, Zhijia Zhang, and Xiaocheng Wu have contributed equally

² Cancer Institute, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China

⁵ Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China

¹⁰ Pulmonary and Critical Care Medicine Center, Chinese PLA Respiratory Disease Institute, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China

Full list of author information is available at the end of the article



Keywords: COVID-19, Decision tree, Neutrophil-to-lymphocyte ratio, C-reactive protein, Lactic dehydrogenase

Background

The novel coronavirus disease 2019 (COVID-19) has become a pandemic. The most common symptoms of COVID-19 patients were fever, dry cough, fatigue, dyspnea, etc. [1, 2]. A small part of patients had digestive symptoms, such as nausea, vomiting and diarrhea [3, 4]. A study [5] by the Chinese Center for Disease Control and Prevention showed that about 81% COVID-19 patients were considered as mild. The proportion was 14% and 5% respectively, for severe and critical patients, who should be hospitalized or transferred to intensive care unit (ICU) for urgent treatment. The mortality in overall population was 3.2%, but it increased to 49% in critical population. Hence, how to use effective biomarkers to identify patients who are at high risk of poor clinical outcomes have caused extensive concern.

COVID-19 patients with comorbidities were considered to be prone to having poor clinical outcomes. A study revealed that COVID-19 patients with chronic obstructive pulmonary disease, diabetes, hypertension and malignancy had a higher risk of admission to an 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].

During the early outbreak of COVID-19 in Wuhan, centre of early stage of the pandemic, medical resources were extremely scarce. It is of great clinical significance to use effective biomarkers to quickly identify patients with high risk of death, to whom should be given priority in accessing medical resources. In this study, we retrospectively enrolled patients from Taikang hospital and other temporary hospitals during the outbreak of COVID-19 in Wuhan, China. We analyzed the differences in clinical characteristics between severe and non-severe patients, as well as survivors and non-survivors. Furthermore, we developed a clinically operable and easy-to-interpret decision tree model to distinguish COVID-19 patients with high risks of death from those without.

Methods

Data sources

A total of 2169 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 test results on admission of all

patients were also collected. All our data were independent from other hospitals or different in periods from other studies, rather than a repetitive analysis. This study was approved by the Ethics Committee of the Taikang Hospital (TKTJLL-005, 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 patient. 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. Disease severity was defined according to the Seventh Revised Trial Version of the COVID-19 Diagnosis and Treatment Guidance (2020) of China [8]. In detail, COVID-19 patients with respiratory rate more than 30 breaths per minute, or oxygen saturation lower than 93% in rest state, or oxygenation index less than 300 mmHg, or rapid progression in lung images within 24–48 h were regarded as severe patients. Next, we performed difference analyses of medical records between survivors and non-survivors. Survivors were defined as patients who were discharged from hospital or transferred to other local hospitals due to advanced age or other basic diseases, instead of COVID-19, at the end of study. Last, we developed a decision tree to predict death outcome.

Development of a clinically operable decision tree

Many machine learning methods are available to develop a helpful predictive model. However, most of them are difficult to interpret because of their internal model mechanisms of black-box modelling strategies. In this study, we chose the decision tree as the predictive model because it's visible, clinically operable and easy to interpret due to its recursive tree-based decision system.

Before developing a decision tree, an appropriate data processing is needed. First, laboratory indexes with missing values over 20% were excluded, including interleukin-6 (IL-6), procalcitonin and D-dimer. We also excluded neutrophil count and lymphocyte count but retained neutrophil-to-lymphocyte ratio (NLR) because of a strong correlation. Then all missing values were input with mean value of each remaining laboratory index. Finally, factors including age, sex, smoking status, body temperature, oxygen saturation, heart rate, respiratory rate, number of comorbidities, number of system

symptoms, white blood cell (WBC), NLR, monocyte count, eosinophilia count, basophilia count, red blood cell (RBC), hemoglobin, platelet count, lactic dehydrogenase (LDH) and C-reactive protein (CRP) were used in the development of decision tree.

All severe patients were randomly split into training dataset and test dataset with a ratio of 7:3. The training dataset, including 452 severe COVID-19 patients, was used to build the decision tree. And the test dataset, including 194 severe COVID-19 patients, was used to validate the decision tree.

The decision tree is built by a two-stage process and the resulting models can be represented as binary trees. First of all, we explore to find each variable which could best split the data into two groups. The data is separated by related variables recursively until the subgroups either reach a minimum size or until no improvement can be made. The impurity function we used was "Information". In this step, a certain but complex tree model was built. But not all the target variables in the complex model are essential. Hence, secondly, we used cross-validation with the 1-SE rule to trim back the full tree. In the next step, we set the max nodes of split no more than 4 and chose the smallest complexity parameter in order to obtain a simple and meaningful decision tree.

The performance of the model was evaluated by the area under the curve (AUC), accuracy and a confusion matrix which could describe how many results were correctly and incorrectly classified. These indexes were calculated both in the training dataset and the test dataset.

Statistical analysis

Continuous variables were described as median with interquartile range (IQR), the comparison was analyzed 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 model was developed using R software (version 3.5.2). The following R packages were used: CBCgrps, rpart, rpart.plot, MICE and pROC. A two-sided p value < 0.05 was considered statistically significant.

Results

Of the 2169 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 (≥ 37.3 °C), 270 (12%) had low oxygen saturation ($\leq 93\%$), 359 (17%) had abnormal heart rates and 596 (27%) had faster respiratory rates (> 20 per minute). 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 system symptoms and 218 (10%) patients had three or more system symptoms. The most common system symptoms were respiratory symptoms, systemic symptoms and digestive symptoms (Table 1).

A total of 646 (29.8%) patients were diagnosed as 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 (56% vs. 44%, $p < 0.001$). On admission, higher proportions of high body temperature (9%), low oxygen saturation (42%), abnormal heart rate (20%) and faster respiratory rate (47%) were found in severe group. Moreover, patients in severe group had higher proportions of comorbidities (70%) and system symptoms (98%). No difference was found in smoking history (Table 1). When comparing laboratory test results between the two groups, we found that the severe group had significantly higher WBC count, neutrophil count, NLR, CRP, LDH, IL-6, procalcitonin and D-dimer levels, but lower lymphocyte count, eosinophilia count, basophilia count, RBC count, hemoglobin and platelet count. No difference was found in monocyte count (Table 1).

From February 10th to April 15th, 2020, 75 patients died of COVID-19. Differences in demographics and clinical characteristics between survivors and non-survivors were similar to the differences between severe and non-severe groups. For laboratory test comparison, much higher WBC count, neutrophil count, NLR, 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. A total of 452 patients were included in the training dataset, including 57 non-survivors. In this step, a decision tree model was developed to differentiate non-survivors from survivors. As shown in Fig. 1, three biomarkers were included in the decision tree model, including LDH, NLR and CRP. The threshold of each biomarker helped to classify each patient into survivor group or non-survivor group. The AUC of the receiver operating characteristic of this model was 0.96, which was higher than each single biomarker (Fig. 2). The associated confusion matrix of training dataset was presented in Additional file 1: Table S1. 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

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

Variables	Total (N = 2169)	Severe (N = 646)	Non-severe (N = 1523)	p value ^a
Age, years	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 system 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, × 10 ⁹ /L	5.8 (4.74, 7.06)	6.23 (4.98, 8)	5.68 (4.69, 6.83)	< 0.001
Neutrophil count, × 10 ⁹ /L	3.48 (2.62, 4.59)	4.02 (3, 5.93)	3.28 (2.52, 4.25)	< 0.001
Lymphocyte count, × 10 ⁹ /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, × 10 ⁹ /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, × 10 ⁹ /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, × 10 ⁹ /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, × 10 ⁹ /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, × 10 ⁹ /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 and respiratory rate were detected at rest when patients were admitted to hospital

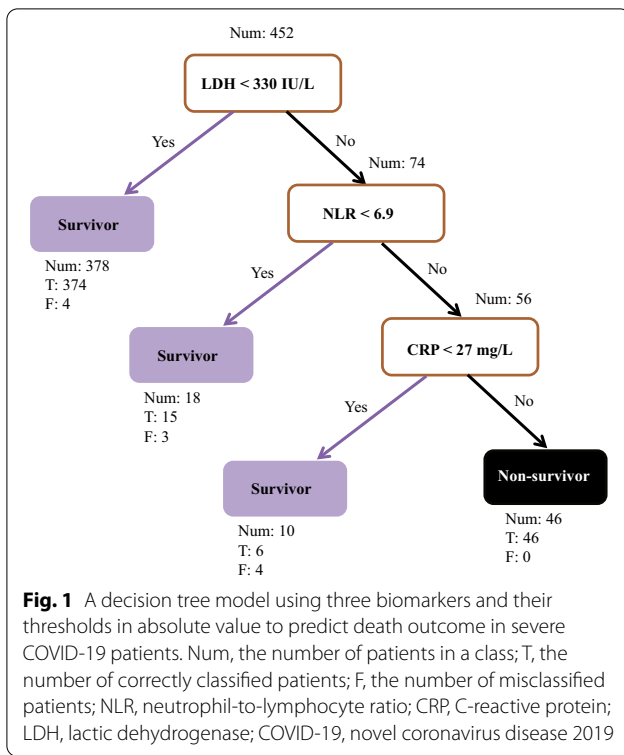
^a p values indicate differences between severe and non-severe patients

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	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 system 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, × 10 ⁹ /L	10.2 (7.27, 14.09)	5.75 (4.71, 6.94)	< 0.001
Neutrophil count, × 10 ⁹ /L	9.62 (6.8, 13.53)	3.42 (2.59, 4.46)	< 0.001
Lymphocyte count, × 10 ⁹ /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, × 10 ⁹ /L	0.38 (0.22, 0.64); n = 75	0.48 (0.38, 0.59); n = 2087	0.002
Eosinophilia count, × 10 ⁹ /L	0.01 (0, 0.04); n = 75	0.11 (0.06, 0.18); n = 2087	< 0.001
Basophilia count, × 10 ⁹ /L	0.01 (0, 0.02); n = 75	0.03 (0.02, 0.04); n = 2088	< 0.001
Red blood cell count, × 10 ⁹ /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, × 10 ⁹ /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 and respiratory rate were detected at rest when patients were admitted to hospital

^a p values indicate differences between survivors and non-survivors



precision, recall and F1 score was 1.00, 0.81 and 0.90, respectively (Table 3).

To validate the performance of the decision tree, we applied it to the test dataset, which included 194 severe patients. The associated confusion matrix of test dataset was presented in Additional file 1: Table S1. The accuracy in test dataset was 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 higher proportions of comorbidities and symptoms than their counterparts. Zhang et al. [9] reported that the median age in a small cohort of COVID-19 non-survivors was 72.5 years, similar to our findings. In the early outbreak in China, the case fatality ratio (CFR) of COVID-19 was 0.4%, 1.3%, 3.6%, 8% and 14.8% among patients aged 40 s or younger, 50 s, 60 s, 70 s and 80 s or older, respectively [10]. Some studies outside China also showed that the CFR of older

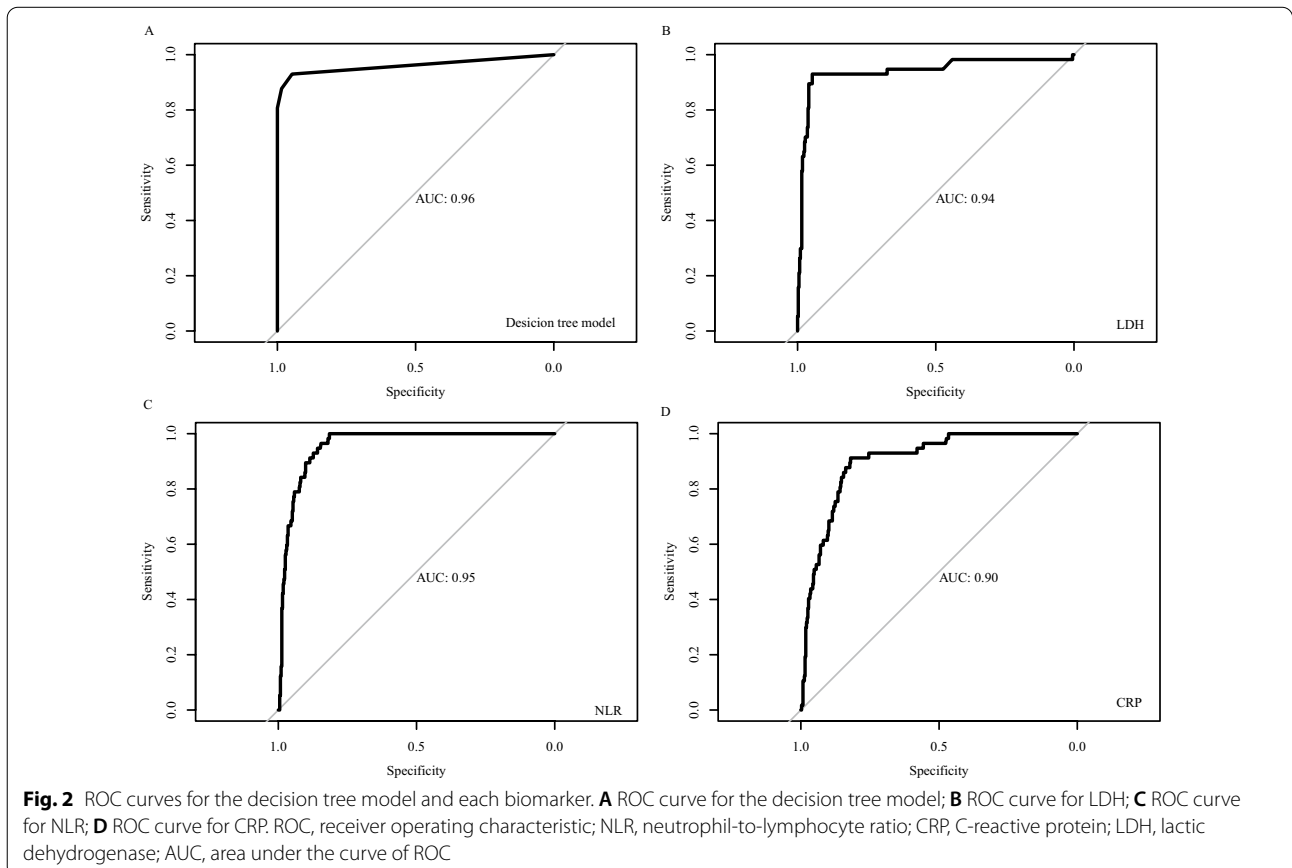


Table 3 Performance of the decision tree on the training and test datasets

	Precision	Recall	F1 score	Support
Training dataset				
Survivor	0.97	1.00	0.98	406
Non-survivor	1.00	0.81	0.90	46
Accuracy			0.98	452
Test dataset				
Survivor	0.98	0.99	0.98	178
Non-survivor	0.94	0.83	0.88	16
Accuracy			0.98	194

patients was much higher than that of younger patients [11–13]. Impairment of immune defense against COVID-19 infection, immunosenescence, and increased risk for immunopathology were thought to be related to higher severity and mortality in older patients [14]. Other proposed hypothesis regarding the vulnerability to COVID-19 among aged patients including age-related chronic inflammation [15] or immunosenescence secondary to cytomegalovirus infection [16, 17]. Fortunately, COVID-19 vaccines might have high efficacy and safety to protect older people from COVID-19 infection [18].

We found that male COVID-19 patients accounted for the majority of severe patients and non-survivors. Previous study also demonstrated that approximately 60% of patients died of COVID-19 were male all over the world [19]. Male had a hazard ratio of 1.59 for COVID-19 related death compared to female [20]. The probable reason might be higher levels of several important pro-inflammatory innate immune chemokines and cytokines, such as IL-8, IL-18, and CCL5, but weaker T cell response in male patients in comparison with female patients [21]. Besides, behavioral/lifestyle risk factors, prevalence of co-morbidities, aging, and underlying biological sex differences might also contribute to the differences of CFR and severity between male and female patients [22].

Above all, this study proposed a simple and clinically operable decision tree model to quickly quantify the risk of COVID-19 related death based on three biomarkers (LDH, NLR and CRP), which could be easily obtained on admission. Take the training dataset as example (Fig. 1), the first biomarker LDH could divide all 452 patients with severe COVID-19 into two subgroups. Only 4 out of 378 (1.1%) patients with $LDH < 330$ IU/L died, while 53 out of 74 (71.6%) patients with $LDH \geq 330$ IU/L died. Then next biomarker NLR could further stratify the subgroup of $LDH \geq 330$ IU/L. Among this subgroup, those with $NLR < 6.9$ had relatively low risk of death compared to those with $NLR \geq 6.9$ (16.7% vs. 89.3%). Moreover,

among patients with $LDH \geq 330$ IU/L and $NLR \geq 6.9$, all those with $CRP \geq 27$ mg/L died, 4 out of 10 of those with $CRP < 27$ mg/L died. In short, we recommend COVID-19 patients with $LDH \geq 330$ IU/L and $NLR \geq 6.9$ should be closely monitored or transfer to ICU. Those with $LDH \geq 330$ IU/L but $NLR < 6.9$ also need to be carefully observed. This simple decision tree model helps physician quickly identify patients with high risk of death and priority of healthcare should be allocated accordingly, which is especially important in crowded hospital or during COVID-19 outbreak with shortage of medical resources.

Separately, these three biomarkers also have important clinical significance. The increase of LDH is a marker of tissue/cell damage. In patients with idiopathic pulmonary fibrosis, the LDH level could reflect the extent of lung injury [23]. For patients with severe COVID-19, the rise in LDH might indicate the activity of lung injury. Evidence proved that LDH was a biomarker of severe illness and poor prognosis in COVID-19 patients [24]. 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 [25]. 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 [26–28]. In COVID-19 patients, elevated NLR on admission was reported to be significantly associated with disease severity [29, 30]. Liu and colleagues proposed a simple model based on NLR and age to stratify COVID-19 patients into four groups [31]. COVID-19 patients with age < 50 years 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 in patients with age ≥ 49.5 and $NLR \geq 3.3$ [30]. 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 [32]. CRP reflects a persistent inflammatory activity state, and helps in assessing the severity of infectious patients [33]. 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 [33, 34].

Some certain limitations should be acknowledged in this study. First, because of the limited data source, an

external validation needs to be performed in further studies. Second, the dynamic changes of some important biomarkers should be followed up to better and timely identify patients at higher risks of death. Third, because some markers, such as IL-6, procalcitonin, D-dimer, etc. were not enough in the study, further study should consider more markers in the development of decision tree.

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 proposed a simple, clinically operable and easy-to-interpret decision tree based on three biomarkers (LDH, NLR and CRP) on admission which could easily be obtained in clinical, to help clinicians rapidly identify COVID-19 patients at high risks of death, to whom priority treatment and intensive care should be given.

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.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-021-06478-w>.

Additional file 1: Table S1. Confusion matrixes of train and test datasets in the decision tree model

Acknowledgements

This study was supported by the Research Projects of the Joint Logistics Support Force of PLA (Grant Nos. TKTJKY2020003, TKTJKY2020029 and TKTJKY2020136) and Chongqing Key Clinical Specialty Construction Project (Respiratory and Critical Care Medicine Department, Xinqiao Hospital).

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.

Funding

None.

Availability of data and materials

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

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Taikang Hospital (TKTJLL-005, 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.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Ultrasound, The 941st Hospital of the PLA Joint Logistic Support Force, Xining, People's Republic of China. ²Cancer Institute, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ³Department of Clinical Laboratory, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ⁴Department of Emergency, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ⁵Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ⁶Department of Cardiology, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ⁷Department of General Medicine, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ⁸Department of Neurology, Southwest Hospital, Army Medical University, Chongqing, People's Republic of China. ⁹Department of Gastroenterology, Southwest Hospital, Army Medical University, Chongqing, People's Republic of China. ¹⁰Pulmonary and Critical Care Medicine Center, Chinese PLA Respiratory Disease Institute, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ¹¹Institute of Burn Research, State Key Laboratory of Trauma, Burns and Combined Injury, Army Medical University, Chongqing, People's Republic of China. ¹²Department of Oncology, Southwest Hospital, Army Medical University, Chongqing, People's Republic of China. ¹³Cancer Center, Army Medical Center, Chongqing, People's Republic of China. ¹⁴Nursing Department, Army Medical Center, Chongqing, People's Republic of China. ¹⁵Department of Hematology, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ¹⁶Department of Nephrology, The Key Laboratory for the Prevention and Treatment of Chronic Kidney Disease of Chongqing, Kidney Center of PLA, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ¹⁷Department of Infectious Diseases, Southwest Hospital, Army Medical University, Chongqing, People's Republic of China. ¹⁸Department of Neurology, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ¹⁹Department of Cardiology, The 958th Hospital, Southwest Hospital, Army Medical University, Chongqing, People's Republic of China. ²⁰Department of Gastroenterology, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ²¹Department of Respiratory Medicine, Army Medical Center, Chongqing, People's Republic of China. ²²Department of Orthopedics, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ²³Department of Obstetrics and Gynecology, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China. ²⁴Department of Nosocomial Infection Control, Xinqiao Hospital, Army Medical University, Chongqing, People's Republic of China.

Received: 28 February 2021 Accepted: 18 July 2021

Published online: 09 August 2021

References

1. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, et al. Clinical characteristics of 138 hospitalized patients with

- 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061–9.
2. Tian J, Yuan X, Xiao J, Zhong Q, Yang C, Liu B, Cai Y, Lu Z, Wang J, Wang Y, et al. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. *Lancet Oncol*. 2020;21(7):893–903.
 3. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, Li Y, Guan W, Sang L, Lu J, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. 2020;180(8):1081–9.
 4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506.
 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 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020;323(13):1239–42.
 6. Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, Liu XQ, Chen RC, Tang CL, Wang T, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J*. 2020;55:5.
 7. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054–62.
 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(7):e0235458.
 10. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727–33.
 11. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA*. 2020;323(18):1775–6.
 12. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, Andronico A, Hoze N, Richez J, Dubost CL, et al. Estimating the burden of SARS-CoV-2 in France. *Science*. 2020;369(6500):208–11.
 13. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA*. 2020;323(20):2052–9.
 14. Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, Li T, Margolick JB, Pawelec G, Leng SX. Aging in COVID-19: Vulnerability, immunity and intervention. *Ageing Res Rev*. 2021;65:101205.
 15. Akbar AN, Gilroy DW. Aging immunity may exacerbate COVID-19. *Science*. 2020;369(6501):256–7.
 16. Kadambari S, Klenerman P, Pollard AJ. Why the elderly appear to be more severely affected by COVID-19: The potential role of immunosenescence and CMV. *Rev Med Virol*. 2020;30(5):e2144.
 17. Moss P. "The ancient and the new": is there an interaction between cytomegalovirus and SARS-CoV-2 infection? *Immun Ageing*. 2020;17:14.
 18. Soiza RL, Scicluna C, Thomson EC. Efficacy and safety of COVID-19 vaccines in older people. *Age Ageing*. 2021;50(2):279–83.
 19. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol Sex Differ*. 2020;11(1):29.
 20. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584(7821):430–6.
 21. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, Silva J, Mao T, Oh JE, Tokuyama M, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature*. 2020;588(7837):315–20.
 22. Haitao T, Vermunt JV, Abeykoon J, Ghamrawi R, Gunaratne M, Jayachandran M, Narang K, Parashuram S, Suvakov S, Garovic VD. COVID-19 and sex differences: mechanisms and biomarkers. *Mayo Clin Proc*. 2020;95(10):2189–203.
 23. Kishaba T, Tamaki H, Shimaoka Y, Fukuyama H, Yamashiro S. Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. *Lung*. 2014;192(1):141–9.
 24. 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;155(8):327–34.
 25. Zeng Z, Yu H, Chen H, Qi W, Chen L, Chen G, Yan W, Chen T, Ning Q, Han M, et al. 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(1):525.
 26. 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(33):4970–84.
 27. Zhang HF, Ge YL, Wang HY, Zhang Q, Li WQ, Chen Y, Chen QC, Jin JJ, Xu J, Zhang S, et al. Neutrophil-to-lymphocyte ratio improves the accuracy and sensitivity of pneumonia severity index in predicting 30-day mortality of CAP patients. *Clin Lab*. 2019;65:10.
 28. Ge YL, Zhang HF, Zhang Q, Zhu XY, Liu CH, Wang N, Zhang JB, Chen H, Chen Y, Li WQ, et al. Neutrophil-to-lymphocyte ratio in adult community-acquired pneumonia patients correlates with unfavorable clinical outcomes. *Clin Lab*. 2019;65:5.
 29. Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, Xie J, Guan W, Liang W, Ni Z, et al. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J Allergy Clin Immunol*. 2020;146(1):89–100.
 30. 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.
 31. Liu J, Liu Y, Xiang P, Pu L, Xiong H, Li C, Zhang M, Tan J, Xu Y, Song R, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J Transl Med*. 2020;18(1):206.
 32. 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;192:3–8.
 33. Danwang C, Endomba FT, Nkeck JR, Wouna DLA, Robert A, Noubiap JJ. A meta-analysis of potential biomarkers associated with severity of coronavirus disease 2019 (COVID-19). *Biomark Res*. 2020;8:37.
 34. 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(7):1730–41.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.