

RESEARCH ARTICLE

The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity

Jiaqing Luo¹✉, Lingyun Zhou²✉*, Yunyu Feng³, Bo Li⁴, Shujin Guo⁵*

1 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, **2** Center of Infectious Diseases, West China Hospital of Sichuan University, Chengdu, China, **3** State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University and Collaborative Innovation Center, Chengdu, China, **4** Department of Otorhinolaryngology, Head & Neck Surgery, West China Hospital, Sichuan University, Chengdu, China, **5** The Geriatric Respiratory Department, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China

✉ These authors contributed equally to this work.

* 4423925@qq.com (LZ); shujinguo@126.com (SG)



OPEN ACCESS

Citation: Luo J, Zhou L, Feng Y, Li B, Guo S (2021) The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity. PLoS ONE 16(6): e0253329. <https://doi.org/10.1371/journal.pone.0253329>

Editor: Dragan Pamucar, University of Defence in Belgrade, SERBIA

Received: February 3, 2021

Accepted: June 3, 2021

Published: June 15, 2021

Copyright: © 2021 Luo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was funded by Natural Science Foundation of China 81802468, Sichuan Science and Technology Program 2019YFS0207 and China Postdoctoral Science Foundation 2020M670062ZX to Dr. Lingyun Zhou; Grant 81700044 from Natural Science Foundation of China to Dr. Shujin Guo; Sichuan Science and Technology Program 2020YFS0113 to Dr. Bo Li. The funders had no role

Abstract

The global pandemic of COVID-19 poses a huge threat to the health and lives of people all over the world, and brings unprecedented pressure to the medical system. We need to establish a practical method to improve the efficiency of treatment and optimize the allocation of medical resources. Due to the influx of a large number of patients into the hospital and the running of medical resources, blood routine test became the only possible check while COVID-19 patients first go to a fever clinic in a community hospital. This study aims to establish an efficient method to identify key indicators from initial blood routine test results for COVID-19 severity prediction. We determined that age is a key indicator for severity predicting of COVID-19, with an accuracy of 0.77 and an AUC of 0.92. In order to improve the accuracy of prediction, we proposed a Multi Criteria Decision Making (MCDM) algorithm, which combines the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Naïve Bayes (NB) classifier, to further select effective indicators from patients' initial blood test results. The MCDM algorithm selected 3 dominant feature subsets: {Age, WBC, LYMC, NEUT} with a selection rate of 44%, {Age, NEUT, LYMC} with a selection rate of 38%, and {Age, WBC, LYMC} with a selection rate of 9%. Using these feature subsets, the optimized prediction model could achieve an accuracy of 0.82 and an AUC of 0.93. These results indicated that Age, WBC, LYMC, NEUT were the key factors for COVID-19 severity prediction. Using age and the indicators selected by the MCDM algorithm from initial blood routine test results can effectively predict the severity of COVID-19. Our research could not only help medical workers identify patients with severe COVID-19 at an early stage, but also help doctors understand the pathogenesis of COVID-19 through key indicators.

in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: COVID-19, Coronavirus disease 2019; MCDM, Criteria Decision Making; TOPSIS, Technique for Order of Preference by Similarity to Ideal Solution; NB, Naïve Bayes; ARDS, Acute Respiratory Distress Syndrome; MODS, Multiple Organ Dysfunction Syndrome; CRP, C-reactive protein; WBC, White Blood Cell Count; LYMC, Lymphocyte Count; LYMPH, Lymphocyte Ratio; NEUT, Neutrophil Count; NEU, Neutrophil Ratio; NLR, Neutrophil to Lymphocyte Ratio; ACC, Accuracy; TPR, True Positive Rate; FPR, False Positive Rate; ROC, Receiver Operating Characteristic; AUC, Area Under the Curve.

Introduction

Currently, more than 40 million people worldwide are infected with the SARS-Cov-2 virus, and more than 10 million people are suffering from Coronavirus disease 2019 (COVID-19) and are receiving treatments [1]. This poses a huge threat to the health and lives of people all over the world, and brings unprecedented pressure to the medical system. Many infected patients cannot receive timely and effective treatment, and it will also reduce the treatment efficiency of other emergency patients [2].

Patients with suspicious symptoms and epidemiological history first visit the fever clinic of the community hospital [3]. They usually undergo three initial tests: SARS-Cov-2 RNA confirms SARS-Cov-2 infection, blood routine test, and chest CT scan to initially assess the severity of COVID-19 [4]. The timely and effective triage of COVID-19 patients based on the results of the three initial tests is of great significance for maintaining emergency capacity and optimizing treatment plans [2].

Although most COVID-19 patients are Mild-Moderate cases and can recover on their own, about 14% of patients are Severe cases, and 5% of patients are Critically Severe cases [5]. Severe-Critically Severe cases usually develop Acute Respiratory Distress Syndrome (ARDS) or Multiple Organ Dysfunction Syndrome (MODS) within two weeks of infection [6], which consumes most of the medical resources and leads to a high case fatality rate (up to 49%) [5, 6]. Early prediction of the severity of COVID-19 can help quickly triage patients (i.e., quarantine, hospital admission or ICU assignment, etc.) and optimize the use of medical resources and timely medical intervention [7, 8]. Blood routine test is the most basic examination. The blood routine test results include red blood cell count (RBC), hemoglobin (HGB), platelets (PLT), white blood cell count (WBC), lymphocyte count (LYMC), lymphocyte ratio (LYMPH), neutrophil count (NEUT), neutrophil ratio (NEU) neutrophil to lymphocyte ratio (NLR), etc. [9–11]. For infectious diseases, a substantial increase or decrease of WBC prompt the severity of the infection. The number and proportion of NEUT can be used to determine the presence or absence of bacterial infection. The rise or fall of LYMC is a characteristic of viral infection [12]. Decreasing of lymphocytes is one of the most critical features of SARS-Cov-2 infection [13]. Of all the initial tests for COVID-19 patients, blood routine test is the worldwide common test with good consistency, and the results are usually available within 2 hours. Due to the influx of a large number of patients into the hospital and the running of medical resources, blood routine test might be the only possible check while COVID-19 patients first go to a fever clinic in a community hospital [4].

When an emerging infectious disease breaks out, we need to quickly understand its pathogenic characteristics and independent risk factors that affect its progression [14]. At this time, the outbreak area is often limited, and the number of patients is small at the very beginning [3]. How to comprehensively analyze the high-risk factors leading to severe illness in a small sample is a serious clinical challenge [15]. Up to now, there have been many studies on predicting the severity of the COVID-19 (i.e., older age, pulmonary micro-thrombosis, increased inflammatory factors (C-reactive protein (CRP), IL-6), hyper-lactic acidemia, D-dimer progressive heightened, decreased lymphocyte count (especially CD8+ T cell count) and short-term progression of lung lesions, etc.) [7, 16–19]. However, the collection of these indicators requires multiple tests and takes a lot of time [19]. These studies certainly can help us improve the treatment, but can hardly help us quickly respond to emerging infectious disease outbreaks [20–22].

In this paper, we aimed to select features from initial blood test results to predict the severity of COVID-19 quickly and accurately. We first defined feature selection as a Multiple Criteria Decision Making (MCDM) problem that considers the correlation between input features and

the correlation between input and output features [23–26]. In MCDM, some methods provide the priority of indicators, while others provide the ranking of indicators. One of the MCDM ranking methods is the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), which has been used in the selection of significant risk factors for healthcare and prognosis [27–29]. Different from the existing TOPSIS methods [8, 27], we use maximum relevance and minimum redundancy [30–32] as the criteria for feature selection in order to select independent risk factors. The maximal relevance feature is to select the input features with the highest relevance to output features. The combinations of individually good features do not necessarily lead to good classification performance [30, 31]. The minimal redundancy is to reduce the redundancy among input features. We then used a series of intuitive measures of relevance and redundancy to select independent risk factors. Finally, we use Naïve Bayes (NB) classifier to achieve the highest prediction accuracy with the fewest input features. Using TOPSIS MCDM, we successfully screened out "independent risk factors" that predict the severity of COVID-19 [25].

Our research established an easy and accurate method for early predict the severity of COVID-19 based on the simple clinic characteristics, which could help medical workers identify patients with severe COVID-19 at an early stage, improve the efficiency of emergency triage of patients, and help doctors understand the pathogenesis of COVID-19 through key indicators.

Methods

Patient enrollment and study design

We performed this prospective cohort study from March 15 to March 20, 2020 in Wuhan Red Cross Hospital, a hospital designated to treat COVID-19 in Wuhan, China. We collected 196 COVID-19 patients diagnosed according to WHO guidance [33] from February 1, 2020 to March 15, 2020. The inclusion criteria were as follows: (1) diagnosis of COVID-19 pneumonia according to the WHO interim guidance published on 28 January 2020 (ref), and (2) availability of relevant medical record information, especially initial blood test results when patients first go to a fever clinic in a community hospital and patients' severity. Patients discharged within 24 h since admission were excluded.

Ethics

The study was approved by the ethics committee of Sichuan Provincial People's Hospital. Since it is not allowed to take any paper documents out of the quarantine area of Wuhan Red Cross Hospital, all participants have obtained oral informed consent, which is recorded by the doctor and kept in the medical record. Before building the predictive model, all data was completely anonymized and cleared.

Definitions

COVID-19 was confirmed by detecting SARS-CoV-2 RNA test. According to the 5th edition of the China Guidelines for the Diagnosis and Treatment Plan of COVID-19 Infection by the National Health Commission (Trial Version 5) [34], the cases were classified into Mild-Moderate and Severe- Critically Severe.

Data collection

The following information was extracted from each patient: Gender, Age and patients' initial blood routine test results including WBC, LYMC, LYMPH, NEUT, NEU and NLR. The

dataset contained 8 input features {Gender, Age, WBC, LYMC, LYMPH, NEUT, NEU, NLR}, and 1 output feature (Severity).

Statistical analysis

Quantitative variables were expressed as the mean \pm standard deviation or the median with interquartile ranges, while categorical variables were expressed as absolute and relative frequencies. The t test or Wilcoxon-test was performed to calculate differences between quantitative data; and χ^2 test was performed to calculate differences between qualitative data. According to the data characteristics, the correlation between clinic characteristics and COVID-19 severity was calculated according to Kendall correlation coefficient (Gender-severity) or Spearman correlation coefficient. Logistic regression analysis was performed for independent variables with collinearity. Wald test was used to determine the joint significance of variables. The standard deviation was used to measure dispersion degrees. Statistical procedures were performed with R statistical software. P values of ≤ 0.05 were considered significant.

The MCDM algorithm design and implementation

The proposed algorithm is basically designed for predicting COVID-19 severity, either Mild-Moderate or Severe-Critically Severe case. It reduces computation time, improves prediction performance, and a better understanding of the data in machine learning. It consists of 4 major stages: preprocessing, feature ranking, feature selection and performance evaluation. Preprocessing is the process to refine the collected raw data to de-noise it. Feature ranking is the process of ordering the features by the value of some scoring function, which usually measures feature-relevance. Feature selection aims to choosing a small subset of the relevant features from the original features by removing irrelevant, redundant, or noisy features. Performance evaluation is to measure the performance of the binary classification by statistical measures, i.e., Accuracy (ACC), True Positive Rate (TPR), False Positive Rate (FPR) and F1 score.

Preprocessing. We use stratified random sampling to divide the dataset into 2 subsets: training set (80%) and test set (20%). In these 4 stages, we only used the test set for performance evaluation. Suppose there are m input features and n output features. Let $X = \{x | 1 \leq x \leq m\}$ be the input feature set and $Y = \{y | m+1 \leq y \leq m+n\}$ be the output feature set. Elements x and y are indexes of features. The feature set is $F = X \cup Y = \{i | 1 \leq i \leq m+n\}$. We calculated and visualized a $(m+n) \times (m+n)$ correlation matrix R and a $(m+n) \times (m+n)$ p-value matrix P to show the correlations between all different feature pairs. To simplify the analysis, we then preprocess R in 2 steps. STEP1: We ignored the sign of $R[i,j]$. Let $R[i,j] = |R[i,j]|$ so that the range of $R[i,j]$ changes from $[-1,1]$ to $[0,1]$, where $i, j \in F$. STEP2: We filtered R through P . For $x \in X$ and $y \in Y$, if $P[x,y] = P[y,x] > 0.05$, $R[x,y]$ and $R[y,x]$ are not significant. We set $R[x,y] = R[y,x] = 0$ and $R[x,i] = R[i,x] = 1$ for $i \in X$.

Feature ranking. We defined a labeled feature set L and initialized with $L = \emptyset$. We iterated the procedure of ranking input features $x \in X$ and moved the first in each ranking from X to L . The ranking criteria includes 2 evaluations: EVAL1: The correlation between input feature $x \in X$ and output feature $y \in Y$, $R[x,y]$ or $R[y,x]$. EVAL2: The correlation between input feature $x \in X$ and labeled feature $v \in L$, $R[x,v]$ or $R[v,x]$. This explicitly evaluates multiple conflicting criteria in decision making. We proposed an algorithm to solve this Multiple Criteria Decision Making (MCDM) problem by using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), which is a compensatory aggregation method. The algorithm, called MCDM, creates an evaluation matrix E consisting of p criteria and q alternatives, to rank input features. According to Pareto's principle, the algorithm divided x into the following 2 types:

TYPE1: If $|X| > \min \{m-1, \lceil 0.8 \times m \rceil\}$, x to be labeled are core features (the top 20%), which should have the lowest $R[v,x]$ from EVAL2, and the highest $R[y,x]$ from EVAL1. The algorithm sorts the elements of sets $L \cup Y$ and X in ascending order to get sequences $(r_i)_{i=1}^{|L|+n}$ and $(c_j)_{j=1}^{|X|}$, respectively. Let $p = |L|+n$ and $q = |X|$, the algorithm extracts a $p \times q$ submatrix E from R such that $E[i, j] = R[r_i, c_j]$. The worst condition of $E[i, j]$ is $w_i = \begin{cases} 1, & i \leq |L| \\ 0, & i > |L| \end{cases}$, and the best condition of $E[i, j]$ is $b_i = \begin{cases} 0, & i \leq |L| \\ 1, & i > |L| \end{cases}$.

TYPE2: If $|X| \leq \min \{m-1, \lceil 0.8 \times m \rceil\}$, x to be labeled are auxiliary features (the rest 80%), which only need to have the lowest $R[v,x]$ from EVAL2. The algorithm sorts the elements of sets L and X in ascending order to get sequences $(r_i)_{i=1}^{|L|}$ and $(c_j)_{j=1}^{|X|}$, respectively. Let $p = |L|$ and $q = |X|$, E is a $p \times q$ matrix with $E[i, j] = R[r_i, c_j]$.

The algorithm calculates the L2-distance between the target alternative j and the worst condition:

$$d_{wj} = \sqrt{\sum_{i=1}^p (E[i, j] - w_i)^2} \tag{Eq 1}$$

It then calculates the distance between j 's condition and the best condition:

$$d_{bj} = \sqrt{\sum_{i=1}^p (E[i, j] - b_i)^2} \tag{Eq 2}$$

After that, it calculates the similarity to the worst condition:

$$s_j = \frac{d_{wj}}{d_{wj} + d_{bj}}, \quad 0 \leq s_j \leq 1 \tag{Eq 3}$$

$s_j = 1$ if and only if alternative j has the best condition, and $s_j = 0$ if and only if alternative j has the worst condition. Let $j^* = \arg \max_j \{s_j\}$, then $X = X \setminus \{c_{j^*}\}$ and $L = L \cup \{c_{j^*}\}$.

The pseudocode of the MCDM algorithm is as follows:

```

Algorithm MCDM is
Input: correlation matrix R, number of input features m, number of
input features n, input feature set X, output feature set Y
Output: labeled feature set L
initialize L = ∅
while X ≠ ∅ do
    if |X| > min {m-1, ⌈0.8×m⌉}
        (ri)i=1|L|+n ← sort LUY and X in ascending order
    else
        (ri)i=1|L| ← sort L in ascending order
        (cj)j=1|X| ← sort X in ascending order
    extract E from R such that E[i, j] ← R[ri, cj]
    for j = 1 to q do // q is the number of columns of E
        dwj ← Eq 2
        dbj ← Eq 3
        sj ← Eq 4
    j* ← arg maxj {sj}
    X ← X \ {cj*}
    L ← L ∪ {cj*}
print L
return L
    
```

Feature selection. The goal of feature subset selection is to find the optimal input feature subset. We gradually increased the number of labeled features, and trained the model with Naïve Bayes classifier in turn. To find the optimal subset, we sequentially tested the accuracy of trained models on the training set.

Performance evaluation. In order to test the stability of the algorithm and observe the influence of the dataset uncertainty on feature selection, we divided the data set 100 times (80% training set and 20% test set) and repeatedly run the algorithm. We used the test set to analyze the performance of feature selection from ACC, TPR, FPR and F1 score.

Evaluation of the predictive value of selected features

According to stratified random sampling, we divided the data set into 2 subsets: 80% of the “training set” and 20% of the “testing set”. We used Receiver Operating Characteristic (ROC) curve analysis to calculate the Area Under the Curve (AUC) and use “ROC” package in R to evaluate the prediction accuracy of our model.

Results

Baseline characteristics

We analyzed the data of 196 COVID-19 patients, of which 90 and 106 were male and female patients. After clearing the data set, there is no abnormal data (S1 Fig). Table 1 lists the detailed baseline characteristics. The mean age of patients was 57.74 ± 15.87 years old. The COVID-19 patients' initial blood routine test results showed that the WBC was $6.75 \pm 3.49 \times 10^9/L$; LYMC was $1.12 \pm 0.58 \times 10^9/L$; LYMPH was $19.91 \pm 11.52\%$; NEUT was $5.13 \pm 3.46 \times 10^9/L$; NEU was $71.34 \pm 15.24\%$; NLR was 7.45 ± 13.08 .

Difference in age and initial blood test results between Mild-Moderate and Severe-Critically Severe groups

According to the 5th edition of the China Guidelines for the Diagnosis and Treatment Plan of COVID-19 Infection by the National Health Commission, we divided patients into 2 groups: 67 cases in the Mild-Moderate group, and 129 cases in the Severe-Critically Severe group (Table 1). Comparing Mild-Moderate and Severe-Critically Severe groups, the basal features showed no differences in Gender ($p = 0.26$) (Fig 1A). The Severe-Critically Severe group was significantly older than the Mild-Moderate group ($p < 0.001$) (Fig 1B). The initial blood routine test seems to be important for predicting the severity of COVID-19: The Severe-Critically Severe group had a higher WBC level ($p = 0.02$) (Fig 1C). The Severe-Critically Severe group had extremely low LYMC ($p < 0.001$) and LYMPH ($p < 0.001$) (Fig 1D and 1E). In contrast,

Table 1. Clinical features of COVID-19 cases.

Clinic Characteristics		Total Data (n = 196)	Severity of COVID-19		
		Mean (SD)	Mild-Moderate	Severe-Critically Severe	
Gender	(Male/Female)	90/106	27/40	63/66	
Age	(Years old)	57.74 (15.87)	45.28 (10.33)	64.21 (14.33)	
Blood Routine Test Results	WBC	($1 \times 10^9/L$)	6.75 (3.49)	5.84(2.30)	7.23 (3.89)
	LYMC	($1 \times 10^9/L$)	1.12(0.58)	1.46(0.55)	0.94(0.51)
	LYMPH	(%)	19.91(11.52)	27.00 (11.07)	16.23 (9.96)
	NEUT	($1 \times 10^9/L$)	5.13(3.46)	3.85 (2.07)	5.79 (3.85)
	NEU	(%)	71.34(15.24)	62.27 (16.27)	76.05(12.32)
	NLR		7.45 (13.08)	3.07 (2.27)	9.72 (15.57)

<https://doi.org/10.1371/journal.pone.0253329.t001>

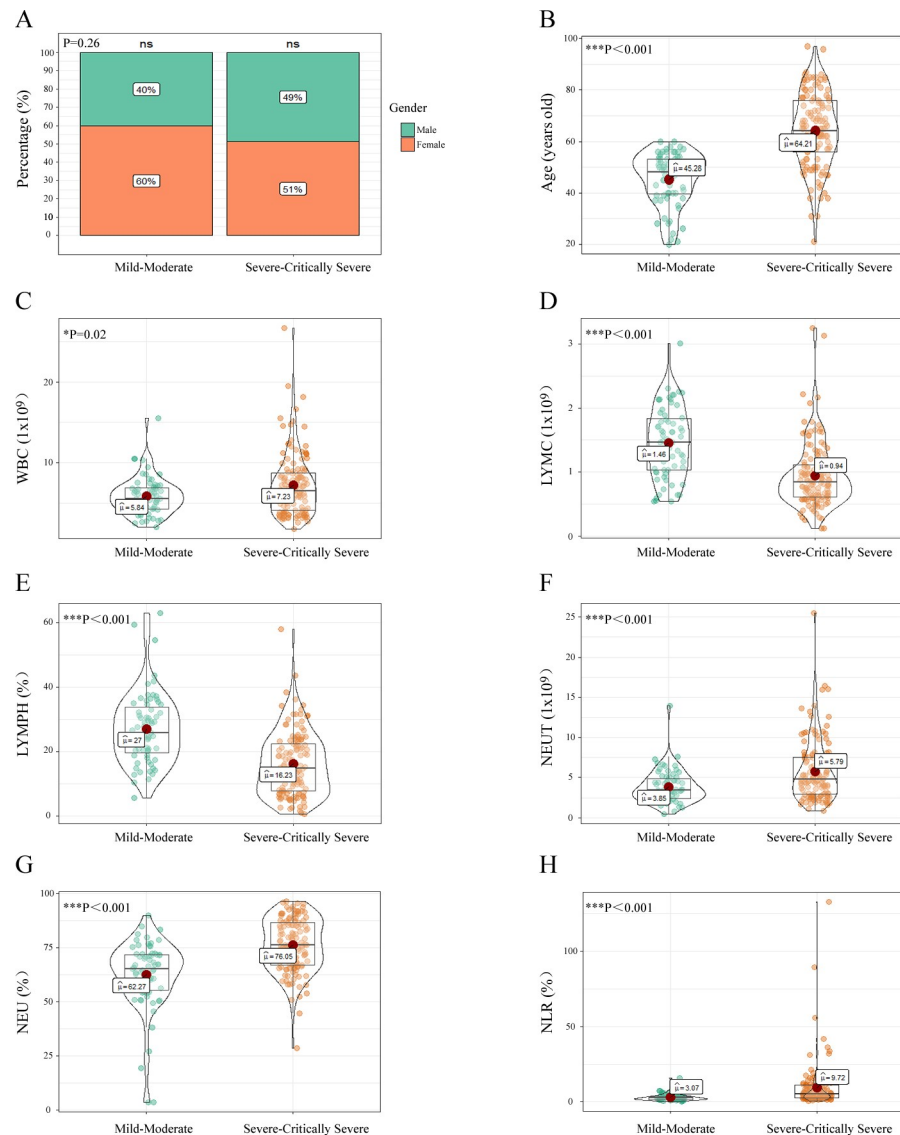


Fig 1. Comparison of clinic characteristics of COVID-19 patients in Mild-Moderate and Severe-Critically Severe groups (n = 196). COVID-19 were divided into Mild-Moderate and Severe-Critically Severe groups according to the 8th edition of the China Guidelines for the Diagnosis and Treatment Plan of COVID-19 Infection by the National Health Commission (Trial Version 8). (A). Gender differences between the two groups, P-value was calculated according to chi-square test. (B). Age differences between the two groups. Each plot graphically displays the central position and scatter/dispersion of the values of each group. P-value was calculated according to student t-test. (C). WBC differences between the two groups, P-value was calculated according to Wilcoxon-test. (D). LYMC differences between the two groups, P-value was calculated according to Wilcoxon-test. (E). LYMPH differences between the two groups, P-value was calculated according to Wilcoxon-test. (F). NEUT differences between the two groups, P-value was calculated according to Wilcoxon-test. (G). NEU differences between the two groups, P-value was calculated according to Wilcoxon-test. (H). NLR differences between the two groups, P-value was calculated according to Wilcoxon-test. *P < 0.05, **P < 0.01, ***P < 0.001.

<https://doi.org/10.1371/journal.pone.0253329.g001>

NEUT ($p < 0.001$) and NEU ($p < 0.001$) in the Severe-Critically Severe group were extremely high (Fig 1F and 1G). As a result, the Severe-Critically Severe group had a higher NLR ($p < 0.001$) (Fig 1H). These observations suggest that patients' age, and WBC, LYMC, LYMPH, NEUT, NEU, NLR from the initial blood routine test could be critical factors for predicting the severity of COVID-19.

Predictive value of age and initial blood test results for COVID-19 severity

By calculating the correlation between clinic characteristics and severity of COVID-19, we found that Age ($r = 0.73$, $p = 0.01$), WBC ($r = 0.24$, $p < 0.01$), NEUT ($r = 0.34$, $p < 0.01$), NLR ($r = 0.31$, $p < 0.01$) were significantly positively correlated with the severity of COVID-19, while LYMC ($r = -0.55$, $p = 0.01$) was significantly negatively correlated with the severity of COVID-19 (Fig 2A and 2B). These results indicated that Age and initial blood routine test results-WBC, LYMC, NEUT, NLR, might be important for predicting the severity of COVID-19. Wald test showed that only Age was the key indicator in predicting the severity of COVID-19 (Table 2). Using stratified random sampling, we generated the ROC curve to evaluate the predictive values: 80% for the “training set” and 20% for the “testing set”. Using [18] for prediction, we can obtain an accuracy of 0.77, and an AUC of 0.92 (Fig 2C). Through dispersion analysis, we found that WBC, LYMC and LYMPH may be able to optimize prediction performance (Tables 3 and 4). The ROC curve showed that {Age, WBC, LYMC} had an accuracy of 0.82 and an AUC of 0.93 (Fig 2D). These results suggested that it is a good predictor of COVID-19 severity, but the accuracy was only 0.77. Using WBC and LYMC from initial blood routine test could rise the accuracy to 0.82.

Details of the MCDM algorithm to predict the severity of COVID-19

The MCDM algorithm was conducted to further investigate whether there were other factors that could improve the accuracy of prediction. The MCDM algorithm and Logistic regression analysis have obtained consistent results: Age was a key indicator in predicting the severity of COVID-19. In addition, the MCDM algorithm verified that the {Age, WBC, LYMC} subset is one of the index sets with the highest prediction accuracy.

- Preprocessing (Fig 3A)—In the COVID-19 data set, $m = 8$ and $n = 1$. The 9×9 correlation matrix R , The 9×9 p-value matrix P and the range of $R[i,j]$ for $i, j \in F$ becomes $[0,1]$. Since $P[1,9] = P[9,1] = 0.1442 > 0.05$, $R[1,9]$ and $R[9,1]$ are not significant, $R[1,9] = R[9,1] = 0$, $R[1,1:8] = \text{ones}(1,8)$ and $R[1:8,1] = \text{ones}(8,1)$.
- Feature Ranking (Fig 3B)—When $|X| = 8 > \min\{8-1, \lceil 0.8 \times 8 \rceil\} = 7$, $L \cup Y = \emptyset \cup \{9\} = \{9\}$ and $X = \{1, \dots, 8\}$. Then, we have, $(r_i)_{i=1}^1 = (9)$ and $(c_j)_{j=1}^8 = (1, \dots, 8)$. Since $p = |L| + n = 1$ and $q = |X| = 8$, E is a 1×8 submatrix of R . When $|X| = 5 < 7$, $L = \{2,3,4\}$ and $X = \{1,5,6,7,8\}$. Then, we have $(r_i)_{i=1}^3 = (2, 3, 4)$ and $(c_j)_{j=1}^5 = (1, 5, 6, 7, 8)$. Since $p = |L| = 3$ and $q = |X| = 5$, E is a 3×5 submatrix of R . When $|X| = 8 > 7$, $w_i = 1$ and $b_j = 0$. By Eq 1 and Eq 2, we calculated $d_{w2} = 0.5913$ and $d_{b2} = 0.4087$. By Eq 3, we have $s_2 = 0.5913$. When $|X| = 5 < 7$, $w_i = 1$ and $b_j = 0$. By Eq 1 and Eq 2, we calculated $d_{w6} = 1.1871$ and $d_{b6} = 0.9912$. By Eq 3, we got $s_6 = 0.5450$.
- Feature Selection (Fig 3C)—When 4 features $\{2,5,8,4\}$ are selected, the accuracy of EVAL1 reached a peak of 0.803. Interestingly, with less features $\{2,3,4\}$, the accuracy of EVAL1 +EVAL2 can reach a higher 0.815.
- Performance Evaluation (Fig 3D)— $\{2,3,4\}$ has the lowest number of features, but the highest score among multiple performance metrics. We can see that the accuracy of $\{2,5,8,4,7,6,3\}$, $\{2,5,8,4\}$ and $\{2,3,4\}$ are 0.74, 0.82 and 0.87, respectively. We can also see that the F1 score of $\{2,5,8,4,7,6,3\}$, $\{2,5,8,4\}$ and $\{2,3,4\}$ are 0.67, 0.72 and 0.78, respectively.

Influence of dataset uncertainty on the feature selection of the MCDM algorithm

To test the stability of the algorithm and observe the influence of the dataset uncertainty on feature selection, we divided the data set 100 times (80% training set and 20% test set) and

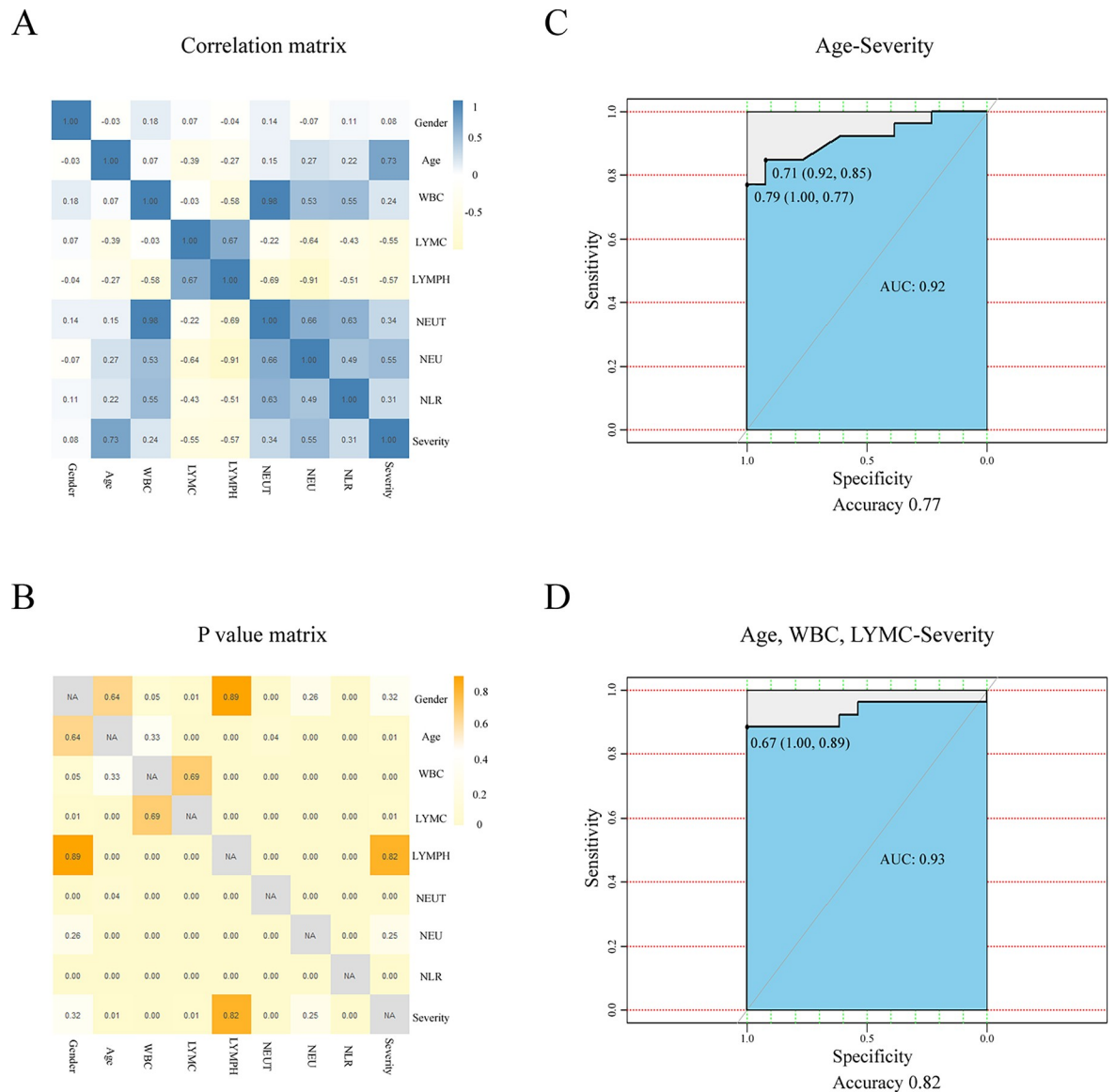


Fig 2. The correlation between clinic characteristics and severity of COVID-2019 and the predictive value of clinic characteristics for the severity of COVID-19. (A). Correlation analysis: Characteristics of the COVID-19 patient including Gender, Age WBC, LYMC, LYMPH, NEUT, NEU, NLR and Severity. WBC, LYMC, LYMPH, NEUT, NEU and NLR were extracted form patients' initial blood test results. Patients were divided into Mild-Moderate and Severe-Critically Severe groups according to the 8th edition of the China Guidelines for the Diagnosis and Treatment Plan of COVID-19 Infection by the National Health Commission (Trial Version 8). According to the characteristics of the data, the correlation was calculated based on the Kendall correlation coefficient (Gender-severity) or Spearman correlation coefficient. $P < 0.05$ was considered statistically significant. (B). P-values of correlation. (C). ROC curve used to evaluate the predictive value of Age for the severity of COVID-19 based on stratified random sampling: 80% as the training set and 20% as the testing set. (D). ROC curve used to evaluate the predictive value of Age, WBC, LYMC for the severity of COVID-19 based on stratified random sampling: 80% as the training set and 20% as the testing set.

<https://doi.org/10.1371/journal.pone.0253329.g002>

repeatedly run the algorithm. The average number of features selected by 3 different criteria, EVAL1, EVAL1 (subset) and EVAL1+EVAL2 (subset) are 6.58 (95% CI: 6.48–6.68), 3.26 (95% CI: 3.01–3.51) and 3.52 (95% CI: 3.40–3.64), respectively (Fig 4A). The criteria, EVAL1 +EVAL2 (subset), adopted by the MCDM algorithm improved most performance metrics. The metrics (ACC, TPR, FPR and F1 score) of EVAL1+EVAL2 (subset) are 0.81 (95% CI:

Table 2. The joint significance of clinical characteristics.

Clinic Characteristics	Estimate	Std. Error	Z Value	P value	Significant
(Intercept)	-9.55	10.90	-0.88	0.38	
Gender	0.76	0.54	1.42	0.16	
Age	0.13	0.03	4.97	0.00	***
WBC	0.61	2.10	0.29	0.77	
LYMC	-0.54	2.60	-0.21	0.83	
LYMPH	-0.06	0.14	-0.42	0.67	
NEUT	-0.74	2.24	-0.33	0.74	
NEU	0.06	0.11	0.57	0.57	
NLR	0.04	0.16	0.22	0.83	

Significant

**** p<0.001, *** p< 0.01, ** p< 0.05.

<https://doi.org/10.1371/journal.pone.0253329.t002>

0.80–0.82), 0.69 (95% CI: 0.67–0.71), 0.09 (95% CI: 0.08–0.11) and 0.75 (95% CI: 0.73–0.77) respectively, while those of EVAL1 are 0.75 (95% CI: 0.74–0.77), 0.60 (95% CI: 0.58–0.62), 0.07 (95% CI: 0.06–0.09) and 0.71 (95% CI: 0.70–0.73) respectively (Fig 4B). Although dataset uncertainties have an influence on feature selection, there were still 3 subsets: {Age, WBC, LYMC, NEUT} with a selection rate of 44%, {Age, NEUT, LYMC} with a selection rate of 38%, and {Age, WBC, LYMC} with a selection rate of 9%, which dominated EVAL1+EVAL2 (subset) feature selection. These 3 subsets can achieve high accuracy with a small number of features (Fig 4C).

Predictive value of the features selected by the MCDM algorithm

Using stratified random sampling, we generated ROC curves to evaluate the predictive values of the subsets selected by the MCDM algorithm: 80% for the “training set” and 20% for the “testing set”. Our analysis results showed that {Age, WBC, LYMC, NEUT} (Fig 5A), {Age, NEUT, LYMC} (Fig 5B) and {Age, WBC, LYMC} (Fig 5C) all achieved 0.82 accuracy and 0.93 AUC. The MCDM algorithm can steadily and accurately select Age and other features from initial blood routine test results to predict the severity of COVID-19.

Table 3. Dispersion analysis of clinical characteristics.

Clinic Characteristics	Df	Deviance resid	Df	resid. Dev	P value	Significant
NULL			156	202.10		
Gender	1	0.23	155	201.87	0.63	
Age	1	63.55	154	138.32	0.00	***
WBC	1	7.20	153	131.12	0.01	**
LYMC	1	15.65	152	115.47	0.00	***
LYMPH	1	4.95	151	110.52	0.03	*
NEUT	1	0.29	150	110.23	0.59	
NEU	1	0.38	149	109.86	0.54	
NLR	1	0.06	148	109.79	0.80	

Significant

**** p<0.001

*** p< 0.01

** p< 0.05.

<https://doi.org/10.1371/journal.pone.0253329.t003>

Table 4. The joint significance of different subsets.

Subset		Estimate	Std. Error	Z Value	P value	Significant
[18]	(Intercept)	-5.49	1.07	-5.10	0.00	***
	Age	0.11	0.02	5.62	0.00	***
{Age, WBC, LYMC, LYMPH}	(Intercept)	-4.43	1.80	-2.46	0.01	*
	Age	0.11	0.02	4.88	0.00	***
	WBC	0.19	0.18	1.09	0.28	
	LYMC	-1.42	0.87	-1.64	0.10	*
	LYMPH	-0.02	0.05	-0.41	0.68	
{Age, WBC, LYMC}	(Intercept)	-4.87	1.47	-3.32	0.00	***
	Age	0.11	0.02	4.89	0.00	***
	WBC	0.25	0.10	2.56	0.01	*
	LYMC	-1.73	0.45	-3.84	0.00	***

Significant

***' p<0.001

**' p< 0.01

'*' p< 0.05.

<https://doi.org/10.1371/journal.pone.0253329.t004>

Discussion

In this paper, we determined that age was the most critical indicator for predicting the severity of COVID-19. To improve the prediction accuracy, we proposed an MCDM algorithm, which combines the TOPSIS and NB classifier, to further select the indicators of patients' initial blood routine test. By ranking features, the MCDM algorithm selected three subsets, including {Age, WBC, LYMC, NEUT}, {AGE WBC, LMYC} and {Age, NEUT, LYMC}, all of which can achieve 0.82 accuracies and 0.93 AUC.

Previous studies have shown that elderly COVID-19 patients with multiple concomitant diseases tend to develop Multiple Organ Failure (MOFE), leading to high mortality in elderly patients infected by SARS-CoV-2 [7, 10]. According to the latest meta-analysis of the elderly in the European community, the prevalence of frailty is around 15% for the elderly 65 years and older [35], and the case fatality rate of patients over 85 years old is 1,000 times that of patients aged 5–17 years [36]. Our research indicated that age was the most important indicator for predicting the severity of COVID-19, with an accuracy 0.77 and an AUC of 0.92. However, some elderly patients had a good prognosis, so prognostic evaluation and medical decision-making based on age alone might not be accurate enough.

We found that WBC, LYMC and NEUT in initial blood routine test results other than age are also crucial for predicting the severity of COVID-19. Guo et al. [37] pointed out that the MuLBSTA score revealed that multi-lobar infiltrates, lymphocytes $\leq 0.8 \times 10^9/L$, bacterial infection, smoking status, hypertension, and age ≥ 60 years could help prognosticate outcomes in COVID-19 patients [38]. The elevated WBC/NEUT is an essential sign of bacterial infection. Bacterial co-infection in COVID-19 patients may develop a severe form of disease, complicating the clinical situation [39–41]. The control and elimination of viruses depend on humoral immunity. Viral infections usually lead to abnormal changes in lymphocyte subsets which further impaired immune system functionality. The decrease of LYMC is the most straightforward and most intuitive indicator to predict the humoral immune response, indicating that the patient's T cell function is defective [18, 42, 43]. The count of lymphocyte subsets (CD4+ and CD8+ T cell), especially CD8+ T cell, is directly proportional to the severity of COVID-19 [44, 45].

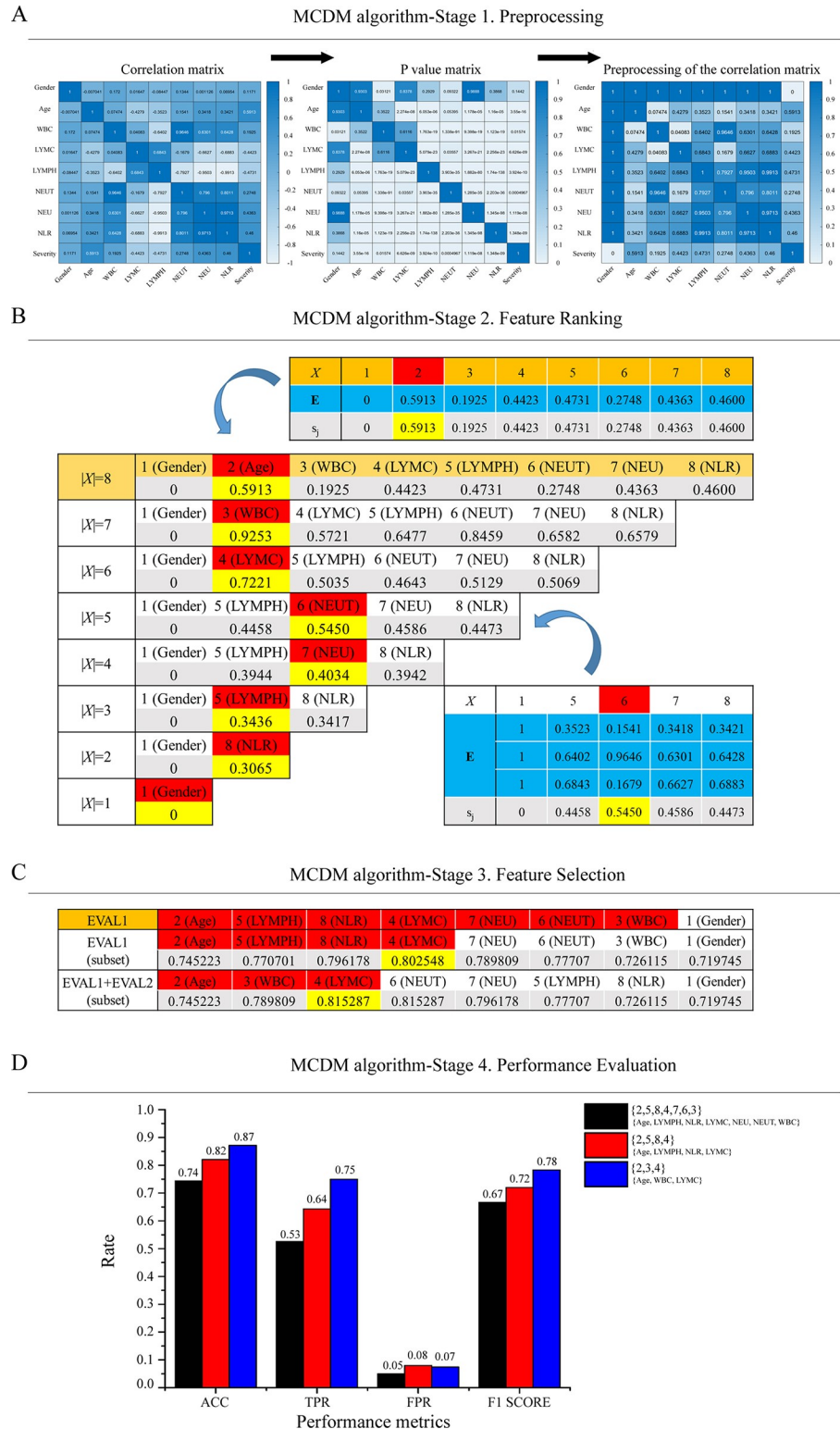


Fig 3. Design and implementation of the Multiple Criteria Decision Making (MCDM) algorithm for predicting the severity of COVID-19. (A). The MCDM algorithm-Stage 1. Preprocessing, this stage is the process of refining the collected raw data to eliminate noise, including correlation analysis and feature selection based on P values. Correlation was calculated according to Spearman correlation coefficient. $P < 0.05$ was considered statistically significant. (B). The MCDM algorithm-Stage 2. Feature Ranking, this stage is the process of using the TOPSIS method

to rank features. TOPSIS method: according to the severity-relevance, we defined the top 20% as the core features and the other 80% as the auxiliary feature. For key features: First, select the first feature that is most relevant to the severity; Second, select the remaining key features in turn by ranking. The ranking criteria are as relevant as possible to severity, and not relevant to the selected key features. For auxiliary features: score and rank auxiliary features according to the degree of irrelevance to key features. (C). The MCDM algorithm-Stage 3. Feature Selection, this stage is to select a subset of the features ranked by the TOPSIS method to remove irrelevant, redundant, or noisy features. EVAL1: The correlation between input features $x \in X$ and output features $y \in Y$, $R[x,y]$ or $R[y,x]$; EVAL2: The correlation between input features $x \in X$ and labeled features $v \in L$, $R[x,v]$ or $R[v,x]$; Subset: The optimal input feature subset. (D). The MCDM algorithm-Stage 4. Performance evaluation, this stage is to measure the performance of the binary classification by ACC, TPR, FPR and F1 score.

<https://doi.org/10.1371/journal.pone.0253329.g003>

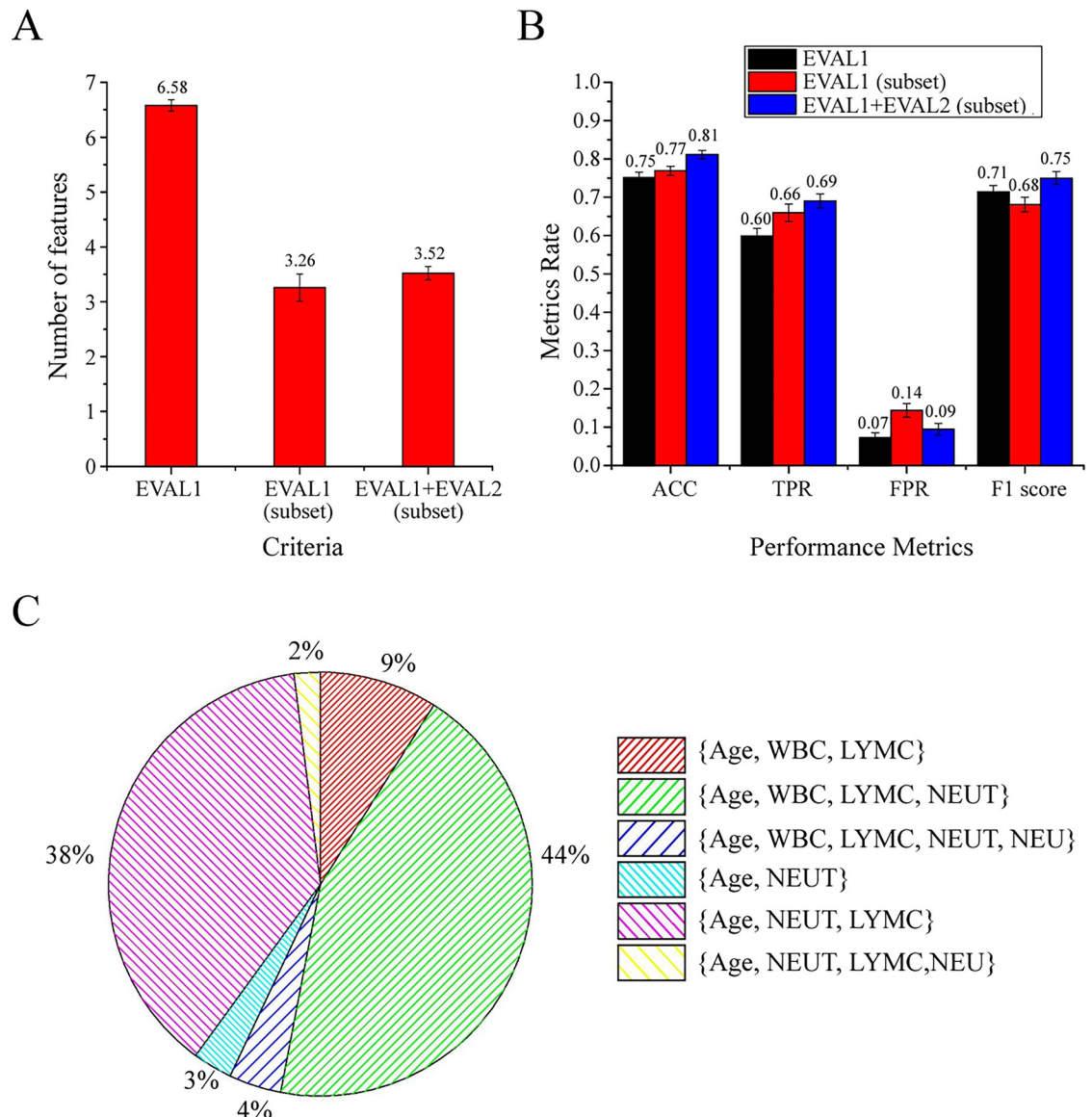


Fig 4. The subset of features selected by the MCDM algorithm to predict the severity of COVID-19. Data set was divided 100 times (80% training set and 20% test set) and repeatedly run the algorithm to test the stability of the algorithm and observe the influence of the dataset uncertainty on feature selection. (A). the average number of features selected by 3 different criteria. EVAL1: The correlation between input features $x \in X$ and output features $y \in Y$, $R[x,y]$ or $R[y,x]$; EVAL2: The correlation between input features $x \in X$ and labeled features $v \in L$, $R[x,v]$ or $R[v,x]$; Subset: The optimal input feature subset. Error bars represents 95% CI. (B). The metrics (ACC, TPR, FPR and F1 score) of 3 different criteria. Error bars represents 95% CI. (C). Different feature selection rates of EVAL1+ EVAL2 subsets.

<https://doi.org/10.1371/journal.pone.0253329.g004>

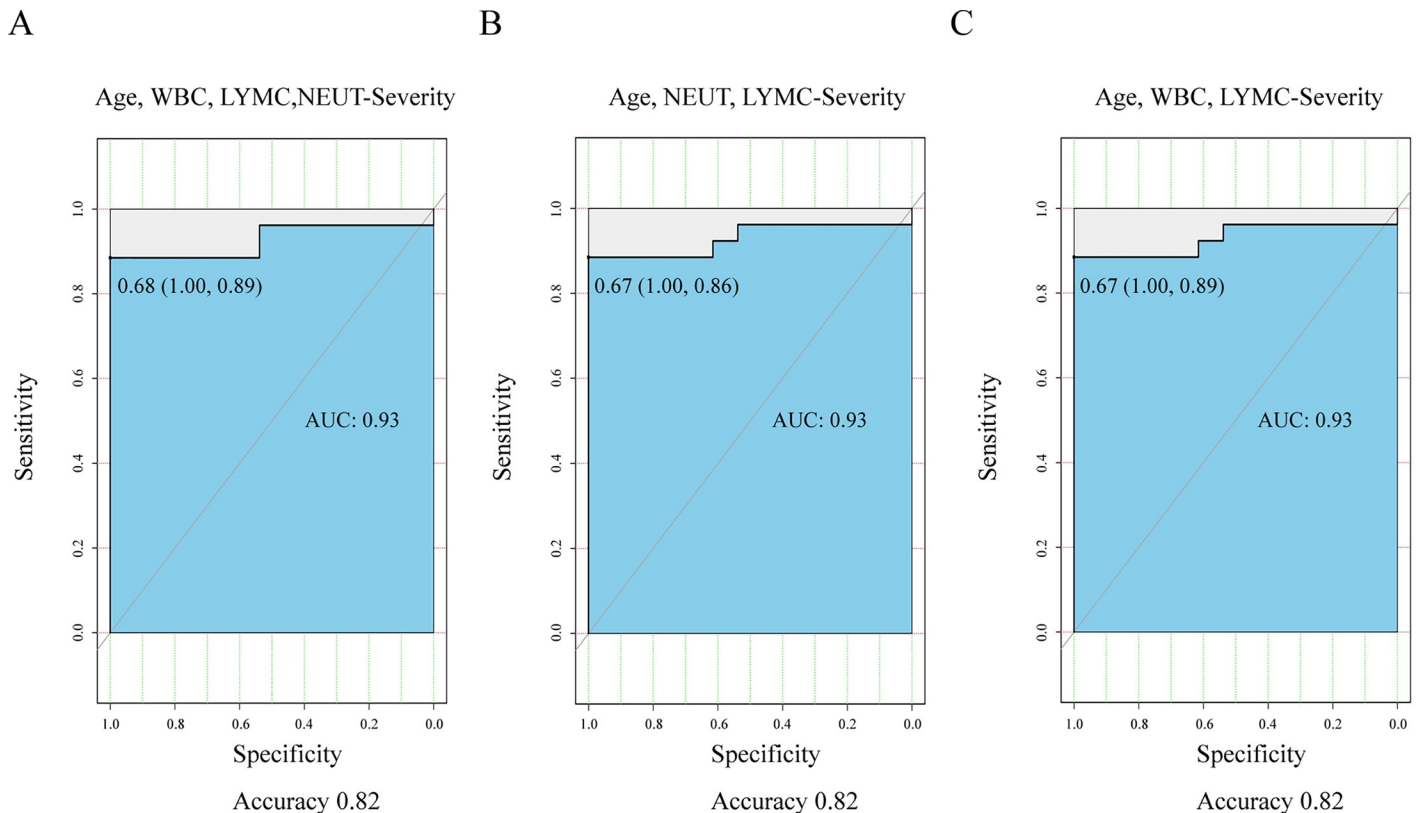


Fig 5. ROC curve used to evaluate the predictive value of the features selected by the MCDM algorithm for the severity of COVID-19. (A). ROC curve used to evaluate the predictive value of {Age, WBC, LYMC, NEUT} for the severity of COVID-19. (B). ROC curve used to evaluate the predictive value of {Age, NEUT, LYMC} for the severity of COVID-19. (C). ROC curve used to evaluate the predictive value of {Age, WBC, LYMC} for the severity of COVID-19. Stratified random sampling: 80% for the “training set” and 20% for the “testing set”.

<https://doi.org/10.1371/journal.pone.0253329.g005>

Although logistic regression can determine the key indicator Age, and discrete analysis can find a better subset {Age, WBC LYMC}, it is difficult to determine the best subset due to the small sample size or multicollinearity. Previous studies used the MCDM algorithm to evaluate diagnostic tests [46] and help doctors hasten COVID-19 treatment [47]. As far as we know, this is the first time the MCDM algorithm has been used to predict the severity of COVID-19. It first uses TOPSIS for feature ranking, and then combines the NB classifier for feature selection. Even if the sample size is small, the MCDM algorithm can select 3 effective subsets {Age, WBC, LYMC}, {Age, WBC, LYMC, NEUT} and {Age, NEUT, LYMC}. The selection process is visual and interpretable helping doctors find the features of the progress of emerging infectious diseases early, to make faster and better prevention and treatment plans. We used the ROC curve to evaluate the predictive value of the features selected by the MCDM algorithm. The results showed that the MCDM algorithm can not only find all effective subsets, but also predict stably and accurately.

Some recent studies point out that age [48–51], underlying diseases [17], systemic immune status [52], and blood test results can be used as key features to predict the severity of COVID-19. Although these features can improve the prediction accuracy (84%~93%), the tests are time-consuming, expensive, and labor-intensive. Our algorithm can select features from blood test results to achieve a prediction accuracy of 82%. During the COVID-19 pandemic, it is more in line with clinical needs and is easy to promote and use in areas with different medical levels.

Our research provides a possible and convenient strategy for the early prediction of COVID-19 severity. However, there are some limitations associated with it. First, there were only 196 cases, and all were from China. The sample size of the study was relatively small. We would like to collect more data and conduct multi-center evaluations. Second, the patient selection process may have been affected by referral bias because of the retrospective design. Third, the screening features are all derived from blood routine tests and are relatively simple. Other features, such as chest CT, absolute T cell count, etc., can be included during the therapy to further evaluate and predict patients' prognosis.

Conclusion

Our research revealed that using age and the indicators WBC/NEUT and LYMC selected by the MCDM algorithm from initial blood routine test results can effectively predict the severity of COVID-19. Advanced age, combined bacterial infections, and low immunity are the main reasons leading to the severity of COVID-19. We are considered feature selection as an MCDM problem so that the algorithm could provide a reference for clinical practice. Using the most common blood routine test, medical institutions could better determine the quarantine, hospital admission, ICU assignment of COVID-19 patients. The MCDM algorithm can be used for small sample data sets, and the prediction is accurate and stable. This study not only provided a reference for establishing a rapid response mechanism in the early stage of emerging infectious disease outbreaks but also help doctors understand the pathogenesis of new infectious disease through key indicators.

Supporting information

S1 Fig. Clean data detection. After cleaning the data, there is no missing data in the dataset. (TIF)

S1 Table. Anonymized data set. For gender, 0 represent female while 1 represent male; For Severity, 0 represent Mild-Moderate while 1 represent Severe-Critically Severe. (CSV)

Author Contributions

Conceptualization: Lingyun Zhou, Shujin Guo.

Data curation: Jiaqing Luo, Yunyu Feng, Bo Li, Shujin Guo.

Funding acquisition: Lingyun Zhou, Shujin Guo.

Investigation: Jiaqing Luo, Lingyun Zhou, Yunyu Feng, Bo Li.

Methodology: Jiaqing Luo, Yunyu Feng.

Project administration: Yunyu Feng, Bo Li.

Supervision: Lingyun Zhou, Shujin Guo.

Writing – original draft: Jiaqing Luo.

Writing – review & editing: Lingyun Zhou, Shujin Guo.

References

1. Organization WH. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update 2021. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.

2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med.* 2020; 8(5):475–81. Epub 2020/02/28. [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5) PMID: 32105632; PubMed Central PMCID: PMC7102538.
3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA.* 2020; 323(11):1061–9. Epub 2020/02/08. <https://doi.org/10.1001/jama.2020.1585> PMID: 32031570; PubMed Central PMCID: PMC7042881.
4. Xu Y, Chen Y, Tang X. Guidelines for the diagnosis and treatment of coronavirus disease 2019 (COVID-19) in China. *Glob Health Med.* 2020; 2(2):66–72. Epub 2020/12/18. <https://doi.org/10.35772/ghm.2020.01015> PMID: 33330780; PubMed Central PMCID: PMC7731342.
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. Epub 2020/02/25. <https://doi.org/10.1001/jama.2020.2648> PMID: 32091533.
6. Epidemiology Working Group for Ncip Epidemic Response CCFDC, Prevention. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi.* 2020; 41(2):145–51. Epub 2020/02/18. <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003> PMID: 32064853.
7. Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, et al. Predictors of COVID-19 severity: A literature review. *Rev Med Virol.* 2021; 31(1):1–10. Epub 2020/08/28. <https://doi.org/10.1002/rmv.2146> PMID: 32845042; PubMed Central PMCID: PMC7855377.
8. Mohammed MA, Abdulkareem KH, Al-Waisy AS, Mostafa SA, Al-Fahdawi S, Dinar AM, et al. Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods. *Ieee Access.* 2020; 8:99115–31. <https://doi.org/10.1109/Access.2020.2995597> WOS:000541127800015.
9. Kristensen M, Iversen AKS, Gerds TA, Ostervig R, Linnet JD, Barfod C, et al. Routine blood tests are associated with short term mortality and can improve emergency department triage: a cohort study of >12,000 patients. *Scand J Trauma Resusc Emerg Med.* 2017; 25(1):115. Epub 2017/11/29. <https://doi.org/10.1186/s13049-017-0458-x> PMID: 29179764; PubMed Central PMCID: PMC5704435.
10. Imam Z, Odish F, Gill I, O'Connor D, Armstrong J, Vanood A, et al. Older age and comorbidity are independent mortality predictors in a large cohort of 1305 COVID-19 patients in Michigan, United States. *J Intern Med.* 2020; 288(4):469–76. Epub 2020/06/05. <https://doi.org/10.1111/joim.13119> PMID: 32498135; PubMed Central PMCID: PMC7300881.
11. Piano S, Dalbeni A, Vettore E, Benfaremo D, Mattioli M, Gambino CG, et al. Abnormal liver function tests predict transfer to intensive care unit and death in COVID-19. *Liver Int.* 2020; 40(10):2394–406. Epub 2020/06/12. <https://doi.org/10.1111/liv.14565> PMID: 32526083; PubMed Central PMCID: PMC7307119.
12. Miller JM, Binnicker MJ, Campbell S, Carroll KC, Chapin KC, Gilligan PH, et al. A Guide to Utilization of the Microbiology Laboratory for Diagnosis of Infectious Diseases: 2018 Update by the Infectious Diseases Society of America and the American Society for Microbiology. *Clin Infect Dis.* 2018; 67(6):e1–e94. Epub 2018/06/30. <https://doi.org/10.1093/cid/ciy381> PMID: 29955859; PubMed Central PMCID: PMC7108105.
13. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020; 382(18):1708–20. Epub 2020/02/29. <https://doi.org/10.1056/NEJMoa2002032> PMID: 32109013; PubMed Central PMCID: PMC7092819.
14. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature.* 2020; 583(7816):437–40. Epub 2020/05/21. <https://doi.org/10.1038/s41586-020-2355-0> PMID: 32434211.
15. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020; 395(10223):497–506. Epub 2020/01/28. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) PMID: 31986264; PubMed Central PMCID: PMC7159299.
16. Mobley LR, Root E, Anselin L, Lozano-Gracia N, Koschinsky J. Spatial analysis of elderly access to primary care services. *Int J Health Geogr.* 2006; 5:19. Epub 2006/05/17. <https://doi.org/10.1186/1476-072X-5-19> PMID: 16700904; PubMed Central PMCID: PMC1482683.
17. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med.* 2020; 180(7):934–43. Epub 2020/03/14. <https://doi.org/10.1001/jamainternmed.2020.0994> PMID: 32167524; PubMed Central PMCID: PMC7070509.

18. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science*. 2020; 369(6508). Epub 2020/07/17. <https://doi.org/10.1126/science.abc8511> PMID: 32669297; PubMed Central PMCID: PMC7402624.
19. Ibanez-Samaniego L, Bighelli F, Uson C, Caravaca C, Fernandez Carrillo C, Romero M, et al. Elevation of Liver Fibrosis Index FIB-4 Is Associated With Poor Clinical Outcomes in Patients With COVID-19. *J Infect Dis*. 2020; 222(5):726–33. Epub 2020/06/21. <https://doi.org/10.1093/infdis/jiaa355> PMID: 32563190; PubMed Central PMCID: PMC7337860.
20. Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. *Int J Health Geogr*. 2004; 3(1):3. Epub 2004/02/28. <https://doi.org/10.1186/1476-072X-3-3> PMID: 14987337; PubMed Central PMCID: PMC394340.
21. Braithwaite J. Between-group behaviour in health care: gaps, edges, boundaries, disconnections, weak ties, spaces and holes. A systematic review. *BMC Health Serv Res*. 2010; 10:330. Epub 2010/12/08. <https://doi.org/10.1186/1472-6963-10-330> PMID: 21134295; PubMed Central PMCID: PMC3004899.
22. Guagliardo MF, Ronzio CR, Cheung I, Chacko E, Joseph JG. Physician accessibility: an urban case study of pediatric providers. *Health Place*. 2004; 10(3):273–83. Epub 2004/06/05. <https://doi.org/10.1016/j.healthplace.2003.01.001> PMID: 15177201.
23. Gharizadeh Beiragh R, Alizadeh R, Shafiei Kaleibari S, Cavallaro F, Zolfani S, Bausys R, et al. An integrated Multi-Criteria Decision Making Model for Sustainability Performance Assessment for Insurance Companies. *Sustainability*. 2020; 12(3). <https://doi.org/10.3390/su12030869> PMID: 33841922
24. Al-Waisy AS, Mohammed MA, Al-Fandawi S, Maashi MS, Garcia-Zapirain B, Abdulkareem KH, et al. COVID-DeepNet: Hybrid Multimodal Deep Learning System for Improving COVID-19 Pneumonia Detection in Chest X-ray Images. *Cmc-Comput Mater Con*. 2021; 67(2):2409–29. <https://doi.org/10.32604/cmc.2021.012955> WOS:000616713000029.
25. Mohammed MA, Abdulkareem KH, Garcia-Zapirain B, Mostafa SA, Maashi MS, Al-Waisy AS, et al. A Comprehensive Investigation of Machine Learning Feature Extraction and Classification Methods for Automated Diagnosis of COVID-19 Based on X-ray Images. *Cmc-Comput Mater Con*. 2021; 66(3):3289–310. <https://doi.org/10.32604/cmc.2021.012874> WOS:000604616100032.
26. Alizadeh R, Lund PD, Soltanisehat L. Outlook on biofuels in future studies: A systematic literature review. *Renewable and Sustainable Energy Reviews*. 2020;134. <https://doi.org/10.1016/j.rser.2020.110326>
27. Majumder P, Biswas P, Majumder S. Application of New TOPSIS Approach to Identify the Most Significant Risk Factor and Continuous Monitoring of Death of COVID-19. *Electron J Gen Med*. 2020; 17(6). ARTN em234 10.29333/ejgm/7904. WOS:000575918600006.
28. Majumder PMAKSM. A mathematical approach of exploration towards extreme risk factor in cancer of optimal condition. *International Journal of Pharmaceutical Sciences and Research*. 2018; 9(9). [https://doi.org/10.13040/ijpsr.0975-8232.9\(9\).3732-42](https://doi.org/10.13040/ijpsr.0975-8232.9(9).3732-42)
29. Alizadeh R, Soltanisehat L, Lund PD, Zamanisabzi H. Improving renewable energy policy planning and decision-making through a hybrid MCDM method. *Energy Policy*. 2020;137. <https://doi.org/10.1016/j.enpol.2019.111174>
30. Cover TM. The Best Two Independent Measurements Are Not the Two Best. *IEEE Transactions on Systems, Man, and Cybernetics*. 1974; SMC-4(1):116–7. <https://doi.org/10.1109/tsmc.1974.5408535>
31. Jain AK, Duin PW, Jianchang M. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(1):4–37. <https://doi.org/10.1109/34.824819>
32. Kohavi R, John GH. Wrappers for feature subset selection. *Artificial Intelligence*. 1997; 97(1–2):273–324. [https://doi.org/10.1016/s0004-3702\(97\)00043-x](https://doi.org/10.1016/s0004-3702(97)00043-x)
33. World Health O. Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance, 28 January 2020. Geneva: World Health Organization, 2020 2020. Report No.: Contract No.: WHO/nCoV/Clinical/2020.3.
34. Lin L, Li TS. [Interpretation of "Guidelines for the Diagnosis and Treatment of Novel Coronavirus (2019-nCoV) Infection by the National Health Commission (Trial Version 5)"]. *Zhonghua Yi Xue Za Zhi*. 2020; 100(0):E001. Epub 2020/02/09. <https://doi.org/10.3760/cma.j.issn.0376-2491.2020.0001> PMID: 32033513.
35. Lippi G, Plebani M. Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *Clin Chim Acta*. 2020; 505:190–1. Epub 2020/03/08. <https://doi.org/10.1016/j.cca.2020.03.004> PMID: 32145275; PubMed Central PMCID: PMC7094472.
36. Mo P, Xing Y, Xiao Y, Deng L, Zhao Q, Wang H, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin Infect Dis*. 2020. Epub 2020/03/17. <https://doi.org/10.1093/cid/ciaa270> PMID: 32173725; PubMed Central PMCID: PMC7184444.

37. Guo L, Wei D, Zhang X, Wu Y, Li Q, Zhou M, et al. Clinical Features Predicting Mortality Risk in Patients With Viral Pneumonia: The MuLBSTA Score. *Front Microbiol.* 2019; 10:2752. Epub 2019/12/19. <https://doi.org/10.3389/fmicb.2019.02752> PMID: 31849894; PubMed Central PMCID: PMC6901688.
38. Frontiers Production O. Erratum: Clinical Features Predicting Mortality Risk in Patients With Viral Pneumonia: The MuLBSTA Score. *Front Microbiol.* 2020; 11:1304. Epub 2020/06/26. <https://doi.org/10.3389/fmicb.2020.01304> PMID: 32582135; PubMed Central PMCID: PMC7296221.
39. Ma Y, Hou L, Yang X, Huang Z, Yang X, Zhao N, et al. The association between frailty and severe disease among COVID-19 patients aged over 60 years in China: a prospective cohort study. *BMC Med.* 2020; 18(1):274. Epub 2020/09/08. <https://doi.org/10.1186/s12916-020-01761-0> PMID: 32892742; PubMed Central PMCID: PMC7474968.
40. Stawicki SP, Jeanmonod R, Miller AC, Paladino L, Gaieski DF, Yaffee AQ, et al. The 2019–2020 Novel Coronavirus (Severe Acute Respiratory Syndrome Coronavirus 2) Pandemic: A Joint American College of Academic International Medicine-World Academic Council of Emergency Medicine Multidisciplinary COVID-19 Working Group Consensus Paper. *J Glob Infect Dis.* 2020; 12(2):47–93. Epub 2020/08/11. https://doi.org/10.4103/jgid.jgid_86_20 PMID: 32773996; PubMed Central PMCID: PMC7384689.
41. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents.* 2020; 55(3):105924. Epub 2020/02/23. <https://doi.org/10.1016/j.ijantimicag.2020.105924> PMID: 32081636; PubMed Central PMCID: PMC7127800.
42. Wang F, Nie J, Wang H, Zhao Q, Xiong Y, Deng L, et al. Characteristics of Peripheral Lymphocyte Subset Alteration in COVID-19 Pneumonia. *J Infect Dis.* 2020; 221(11):1762–9. Epub 2020/04/01. <https://doi.org/10.1093/infdis/jiaa150> PMID: 32227123; PubMed Central PMCID: PMC7184346.
43. Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, et al. Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). *Front Immunol.* 2020; 11:827. Epub 2020/05/20. <https://doi.org/10.3389/fimmu.2020.00827> PMID: 32425950; PubMed Central PMCID: PMC7205903.
44. Pallotto C, Suardi LR, Esperti S, Tarquini R, Grifoni E, Meini S, et al. Increased CD4/CD8 ratio as a risk factor for critical illness in coronavirus disease 2019 (COVID-19): a retrospective multicentre study. *Infect Dis (Lond).* 2020; 52(9):675–7. Epub 2020/06/17. <https://doi.org/10.1080/23744235.2020.1778178> PMID: 32543971.
45. Ganji A, Farahani I, Khansarinejad B, Ghazavi A, Mosayebi G. Increased expression of CD8 marker on T-cells in COVID-19 patients. *Blood Cells Mol Dis.* 2020; 83:102437. Epub 2020/04/24. <https://doi.org/10.1016/j.bcmd.2020.102437> PMID: 32325421; PubMed Central PMCID: PMC7194879.
46. Sayan M, Sarigul Yildirim F, Sanlidag T, Uzun B, Uzun Ozsahin D, Ozsahin I. Capacity Evaluation of Diagnostic Tests For COVID-19 Using Multicriteria Decision-Making Techniques. *Comput Math Methods Med.* 2020; 2020:1560250. Epub 2020/08/18. <https://doi.org/10.1155/2020/1560250> PMID: 32802146; PubMed Central PMCID: PMC7411452 publication of this paper.
47. Albahri OS, Al-Obaidi JR, Zaidan AA, Albahri AS, Zaidan BB, Salih MM, et al. Helping doctors hasten COVID-19 treatment: Towards a rescue framework for the transfusion of best convalescent plasma to the most critical patients based on biological requirements via ml and novel MCDM methods. *Comput Methods Programs Biomed.* 2020; 196:105617. Epub 2020/06/28. <https://doi.org/10.1016/j.cmpb.2020.105617> PMID: 32593060; PubMed Central PMCID: PMC7305916.
48. Guan WJ, Zhong NS. Clinical Characteristics of Covid-19 in China. Reply. *N Engl J Med.* 2020; 382(19):1861–2. Epub 2020/03/29. <https://doi.org/10.1056/NEJMc2005203> PMID: 32220206.
49. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet.* 2020; 395(10223):507–13. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7) PMID: 32007143
50. Chen R, Liang W, Jiang M, Guan W, Zhan C, Wang T, et al. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. *Chest.* 2020; 158(1):97–105. Epub 2020/04/19. <https://doi.org/10.1016/j.chest.2020.04.010> PMID: 32304772; PubMed Central PMCID: PMC7158802.
51. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet.* 2020; 395(10229):1054–62. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) PMID: 32171076
52. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet.* 2020; 395(10223):497–506. [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5) PMID: 31986264