


Risk factors for mortality in critically ill patients with COVID-19 in Huanggang, China: A single-center multivariate pattern analysis

Yinyin Chen MD, PhD^{1,2,3} | Zeqiang Linli BS^{4,5}  | Yuting Lei M Med¹ |
Yiya Yang M Med^{1,2,3} | Zhipeng Liu B Med⁶ | Youchun Xia B Med⁶ |
Yumei Liang B Med^{1,2,3} | Huabo Zhu B Med¹ | Shuixia Guo PhD^{4,5}

¹Department of Nephrology and Laboratory of Kidney Disease, Hunan Provincial People's Hospital, Hunan Normal University, Changsha, China

²Changsha Clinical Research Center for Kidney Disease, Changsha, China

³Hunan Clinical Research Center for Chronic Kidney Disease, Changsha, China

⁴MOE-LCSM, School of Mathematics and Statistics, Hunan Normal University, Changsha, China

⁵Key Laboratory of Applied Statistics and Data Science, College of Hunan Province, Hunan Normal University, Changsha, China

⁶Dabie Mountain Regional Medical Center, Huanggang, China

Correspondence

Huabo Zhu, B Med, Hunan Provincial People's Hospital, Hunan Normal University, 1 Jiefang Xi Lu, Furong District, Changsha City, 410000, Hunan Province, PR China.
Email: hechenghecheng1212@163.com

Shuixia Guo, PhD, School of Mathematics and Statistics, Hunan Normal University, 36 Lushan Road, Yuelu District, Changsha City, 410000, Hunan Province, PR China.
Email: guoshuixia75@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Number: 11671129; Health and Family Planning Commission of Hunan Province, Grant/Award Number: B20180428; Hunan Provincial Natural Science Foundation for Outstanding Youth, Grant/Award Number: 2020JJ2020

Abstract

To date, the coronavirus disease 2019 (COVID-19) has a worldwide distribution. Risk factors for mortality in critically ill patients, especially detailed self-evaluation indicators and laboratory-examination indicators, have not been well described. In this paper, a total of 192 critically ill patients (142 were discharged and 50 died in the hospital) with COVID-19 were included. Self-evaluation indicators including demographics, baseline characteristics, and symptoms and detailed lab-examination indicators were extracted. Data were first compared between survivors and non-survivors. Multivariate pattern analysis (MVPA) was performed to identify possible risk factors for mortality of COVID-19 patients. MVPA achieved a relatively high classification accuracy of 93% when using both self-evaluation indicators and laboratory-examination indicators. Several self-evaluation factors related to COVID-19 were highly associated with mortality, including age, duration (time from illness onset to admission), and the Barthel index (BI) score. When the duration, age increased by 1 day, 1 year, BI decreased by 1 point, the mortality increased by 3.6%, 2.4%, and 0.9% respectively. Laboratory-examination indicators including C-reactive protein, white blood cell count, platelet count, fibrin degradation products, oxygenation index, lymphocyte count, and D-dimer were also risk factors. Among them, duration was the strongest predictor of all-cause mortality. Several self-evaluation indicators that can simply be obtained by questionnaires and without clinical examination were the risk factors of all-cause mortality in critically ill COVID-19 patients. The prediction model can be used by individuals to improve health awareness, and by clinicians to identify high-risk individuals.

KEYWORDS

clinical indicators, COVID-19, machine learning, risk factor, self-evaluation

1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19) has been reported in the Wuhan area of China's Hubei Province since the end of December 2019. Huanggang was the second city in China to be locked down only next to Wuhan. As of April 10, more than 1.5 million confirmed cases of COVID-19, including more than 92,000 deaths, had been reported to the World Health Organization (WHO).¹ COVID-19 not only causes pneumonia, but can also damage other organs such as the heart, liver, kidney, blood system, and immune system. Patients eventually die of multiple organ failure (MOF), shock, acute respiratory distress syndrome (ARDS), heart failure, arrhythmia, or renal failure.²⁻⁴ With the increasing number of deaths, clinicians are most concerned about the following questions: (1) What are the differences in the clinical characteristics between surviving and nonsurviving critically ill patients? (2) What are the risk factors for mortality in critically ill COVID-19 patients? (3) Is there a quick and feasible way for COVID-19 patients to perform risk self-assessments?

To date, there have been many studies on the clinical courses, outcomes, mortality of and risk factors for critically ill patients, but those studies were retrospective observational studies or cohort studies using descriptive analyses. As reported by Huang et al.,⁵ older patients and those with chronic underlying conditions may have worse outcomes. Patients with severe illness may develop dyspnea and hypoxemia within 1 week after disease onset, which may quickly progress to ARDS or MOF. Yang reported that compared with survivors, nonsurvivors were older, more likely to develop ARDS, and more likely to receive mechanical ventilation.⁶ Zhou proposed three clinical features of COVID-19 nonsurvivors: older age, higher sequential organ failure assessment (SOFA) score, and high D-dimer levels.⁷ Although these studies indicated the clinical courses of and risk factors for severe patients, they did not provide a quantitative ranking of the importance of risk factors, so for a given severe patient, it was impossible to predict the probability of mortality.

Multivariate pattern analysis (MVPA), based on machine learning, decodes different patterns by learning a discrimination rule from a dataset and can subsequently categorize new samples. Compared with the univariate approach, MVPA considers multiple variables together to take better advantage of the inherent multivariate nature of high-dimensional data, and it can identify the features that contribute the most to the classifier.^{8,9} Among many MVPA methods, support vector machine (SVM) method is widely used for its excellent classification performance and its ability to handle very high-dimensional data.

In this paper, the SVM method is performed to establish a mortality prediction model for critically ill patients with COVID-19. The main aims were to: (1) assess the differences in the clinical characteristics between survivors and nonsurvivors based on systematic comparison of predictors; (2) detect the possible risk factors associated with mortality in critically ill COVID-19 patients and (3) identify risk factors that were consistently good discriminators for mortality. The identified characteristics or risk factors can help clinicians to identify high-risk patients quickly and patients to do self-monitoring through risk assessment.

2 | METHODS

2.1 | Study design and participants

This single-center study was performed at Dabie Mountain Regional Medical Center (Huanggang, Hubei Province), which is a designated hospital for the treatment of patients with COVID-19. We screened all adult patients who had been diagnosed with COVID-19, according to WHO interim guidance, and those who were critically ill, who died or who were discharged between January 28, 2020 (when the first patients were admitted) and March 13, 2020, were included in our study. Identification of critically ill patients was achieved by reviewing and analyzing admission logs and histories from all available electronic medical records and patient care resources.

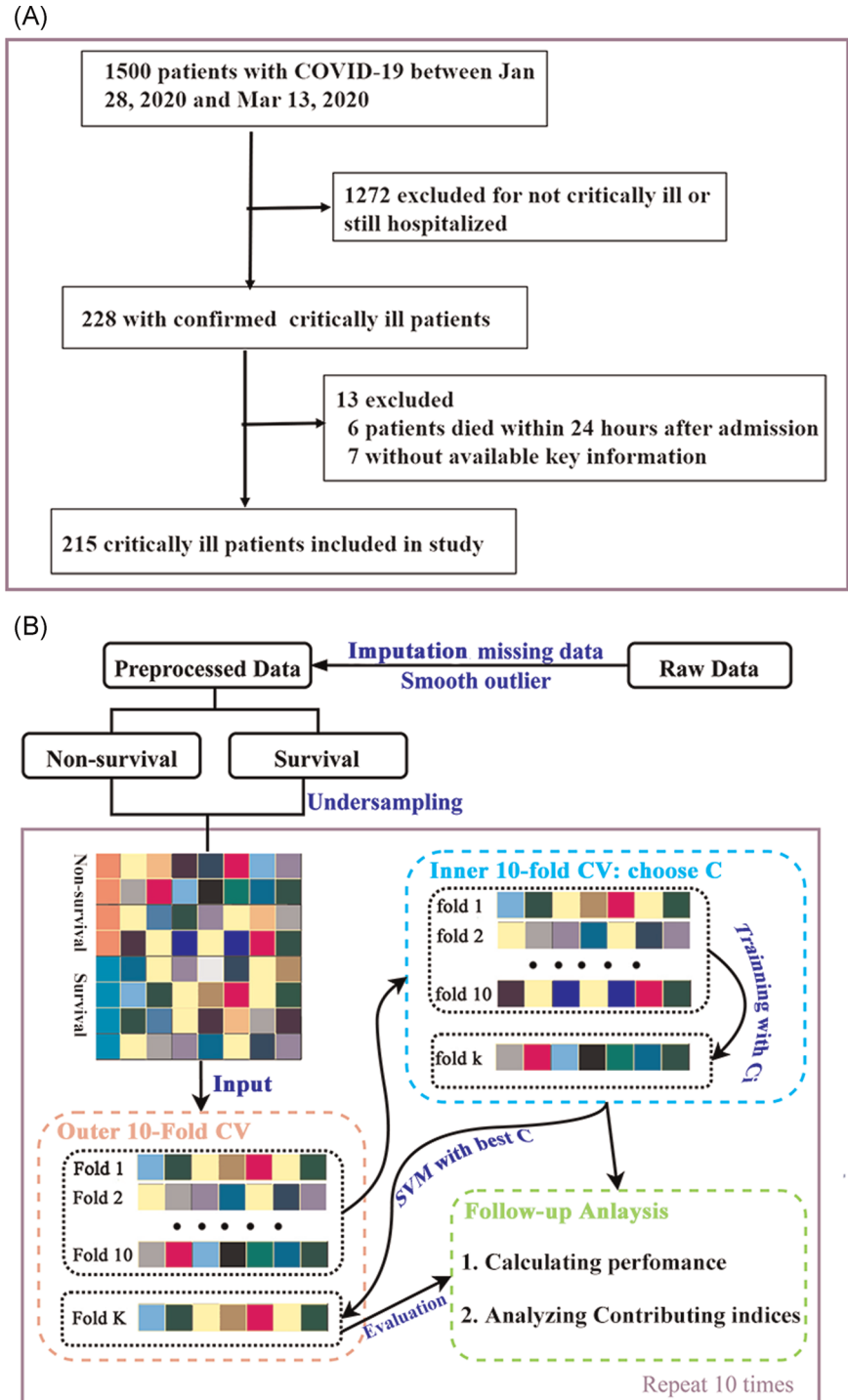
A total of 1500 adult COVID-19 patients were discharged or died in the Dabie Mountain Regional Medical Center before March 13, 2020. After excluding 1272 patients who were not critically ill or who were still hospitalized, six patients who died within 24 h after admission, and seven inpatients without available key information in their medical records, we included the maximum number of patients who met the inclusion criteria. In total, 215 inpatients were included in the following analysis. Of them, 51 patients died during hospitalization and 164 were discharged (Figure 1A). Two kinds of indicators were collected on admission: self-assessment indicators and laboratory-examination indicators. Self-assessment indicators included information on demographics, baseline characteristics, and symptoms. Laboratory-examination indicators included results of routine tests such as routine blood indicators, and so forth.

2.2 | Data preprocessing

Because not all patients had undergone all the laboratory-examinations, we needed to preprocess the raw data. First, subjects for whom more than 15% of information was missing and those for whom more than 15% of variable values were missing were excluded. Accordingly, 192 patients (50 nonsurvivors vs. 142 survivors) and 47 indicators (17 self-assessment indicators and 30 laboratory-examination indicators) were finally included in the following analysis, see Table S1 for abbreviations of the laboratory-examination indicators.

Imputation of the rest missing values was then conducted utilizing multivariate imputation by chained equations (MICE) implemented in the R package "mice." The method is based on fully conditional specification, where each incomplete variable is imputed by a separate model.¹⁰ The MICE algorithm can impute mixes of continuous, binary categorical data. MICE is a robust method for imputation compared with other methods such as mean imputation or complete case analysis which can bias results. We imputed missing data for C-reactive protein (CRP; 2.11%), procalcitonin (PCT; 2.11%), erythrocyte sedimentation rate (ESR; 4.74%), and so forth (See Table S2). Furthermore, the outliers were reset as the boundaries of the data. Specifically, the quantile

FIGURE 1 The study flow diagram of the participants (A) and the SVM analysis (B). SVM, support vector machine



of 1% and 99% of each continuous variable was calculated, replacing the data less than 1% quantile and more than 99% quantile, respectively.

2.3 | SVM modeling

SVM is a popular classifier among many machine learning technologies because of its ability to handle small samples with high-

dimensional features. To alleviate the risk of overfitting and facilitate follow-up characterization of contributing features, we chose linear-SVM as the classifier.¹¹

Our dataset contained unbalanced sample sizes (50 non-survivors vs. 142 survivors), which could cause serious bias in our classification model if we trained the model with the unbalanced data. To overcome this challenge, we used the random undersampling algorithm for the majority class group (i.e., the survival group). Specifically, for each repetition, we randomly selected the

under-sampling dataset from the survival group with the same sample size (50 subjects) as the nonsurvival group. We repeated the under-sampling procedure 10 times to obtain the final result. We applied a nested 10-fold cross-validation (CV), with the outer 10-fold CV loop estimating the generalizability of the model and the inner 10-fold CV loop determining the optimal parameter set C for the linear-SVM model. The outer 10-fold CV served as the primary mechanism to prevent overfitting, with the inner 10-fold CV for model selection. We applied the R package e1071 to implement the SVM in the present study.

In the outer 10-fold CV, all subjects were randomly divided into 10 subsets according to their group labels. Each feature was linearly scaled to zero mean and unit variance across the training dataset; and the scaling parameters were also used for later predictions. The training and testing procedures were repeated 10 times, with each subset used once as the testing set. For all the testing subjects for each fold, the accuracy sensitivity, specificity, precision, recall, and F1 were computed to quantify the performance of the classifier.

Within each loop of the outer 10-fold CV, we applied an inner 10-fold CV to determine the optimal cost C. Specifically, the training set for each loop of the outer 10-fold CV was further partitioned into 10 subsets according to group labels. Nine subsets were selected to train the model under a given parameter set of C values ($C = [0.01, 0.1, 1, 10, 100]$) and the remaining subset was used to test the model. For each C value, accuracies were measured for each inner 10-fold CV loop, and a mean value for all the 10-fold inner loops was then obtained to indicate the inner prediction accuracy. The C with the highest inner prediction accuracy among the 5 inner 10-fold CVs was chosen as the optimal parameter. Finally, all data of the inner 10-fold CV were trained with the best parameter C, and the testing subjects of the outer 10-fold CV were predicted.

Accordingly, for each under-sampling repetition, each loop of the outer 10-fold CV resulted in a specific optimal parameter C, and a corresponding SVM model. After establishing a 10-fold outer CV, 10 SVM models were generated, and the performance of each SVM model was evaluated. The final performance measures were averaged across the 10-fold performances to produce the under-sampling performance metrics.

2.4 | Contributing indices

A crucial aspect of linear SVM is displaying and analyzing the features that drive the multivariate classifier. Here, we calculated the feature weights ω (i.e., the coefficient for each indicator) for each SVM model, and the sum of the feature weights of all 10 SVM models was defined as the final feature weight for each data set. Then, the feature weights were linearly scaled to (0, 1). The indices with the highest weights were deemed the contributing indices for the nonsurvival and survival predictions. The flow diagram of the SVM analysis is shown in Figure 1B.

3 | RESULTS

3.1 | Statistical analysis

We expressed descriptive data as means (SDs) for continuous variables and numbers (%) for categorical variables. Of the 192 critically ill patients with COVID-19, the mean age was 59.57 ± 16.82 (mean \pm SD), and 124 were male (64.6%). The most common clinical symptoms were fever (155 patients, 80.7%), cough (134 patients, 69.8%), and chest pain (77 patients, 40.1%). A total of 134 (69.8%) patients had at least one basic disease. Among the nonsurvivors, 14 (28%) patients died of respiratory failure, 3 (6%) patients died of circulatory failure, 35 (70%) patients died of MOF, and 5 (10%) patients died of other causes, such as carcinoma, and severe immunological disorders secondary to rheumatic diseases.

We assessed the differences between survivors and nonsurvivors using a two-sample *t* test for continuous variables and a test of independence in a contingency table for categorical variables. The obtained test statistics and corresponding two-sided *p* values are listed in Tables 1 and 2. Table 1 demonstrates the statistical analysis results of all self-evaluation indicators. Compared with survivors, nonsurvivors were older ($t = -4.65, p = 5.73 \times 10^{-6}$), had a long time from illness onset to admission ($t = -13.2, p = 1.2 \times 10^{-28}$) and had a lower Barthel index (BI) score ($t = 7.75, p = 2.1 \times 10^{-13}$). There were significant differences between survivors and non-survivors in the distribution of age ($\chi^2 = 33.8, p = 2.5 \times 10^{-6}$; Figure S1). Patients with chronic underlying conditions had worse outcomes, such as digestive disease ($\chi^2 = 6.93, p = .0085$), cardiovascular disease ($\chi^2 = 6.66, p = .01$), cerebrovascular disease ($\chi^2 = 15.3, p = 8.9 \times 10^{-5}$) and chronic obstructive pulmonary disease ($\chi^2 = 5.6, p = .018$) than those without underlying conditions. The nonsurvivors were more likely to have dyspnea ($\chi^2 = 7.82, p = 5.1 \times 10^{-4}$) than the survivors among all critically ill patients with COVID-19.

Table 2 demonstrates the statistical analysis results of all laboratory-examination indicators. As summarized in Table 2, the levels of inflammation or infection-related indices (CRP, WBC, PCT) and hypercoagulability-related indicators (D-dimer, fibrin degradation products [FDP]) were significantly higher in the nonsurvivors than in the survivors. More prominent thrombocytopenia was found in nonsurvivors. Blood gas analysis and respiratory parameters revealed higher rates of hypoxia (lower PO_2, SO_2) and acidosis (lower pH values); higher CO_2 levels; and greater lactate accumulation (higher $PCO_2, lat.$) and respiratory impairment (lower oxygenation index [OI] values) in the nonsurvivors than in the survivors. The levels of albumin and calcium ions and lymphocyte counts (LYMPH) were significantly lower, and aspartate aminotransferase, the international normalized ratio, blood sugar, and sodium ions were higher in the survivors than in the nonsurvivors. There were no significant differences in alanine aminotransferase (ALT), total bilirubin (TBIL), creatinine, blood urea nitrogen (BUN), or markers of hepatic or renal function between the survivors and nonsurvivors. There were no differences in the hemoglobin level, ESR, prothrombin time, activated partial thromboplastin time, fibrinogen level, or potassium ions between the two groups.

TABLE 1 Demographics, baseline characteristics, and symptoms of patients with critically ill COVID-19 patients

	Survivors (n = 142)	Nonsurvivors (n = 50)	Statistics/p value
Demographics			
Age, years (SD)	56.4 (17.2)	68.45 (11.7)	$t = -4.65$, $p = 5.73 \times 10^{-6}$
Sex (female/male)	52/90	16/34	$\chi^2 = 0.34$, $p = .56$
Duration (SD)	7.9 (5.3)	21.1 (7.8)	$t = -13.2$, $p = 1.2 \times 10^{-28}$
Barthel index (SD)	80.3 (27.6)	44 (30.8)	$t = 7.75$, $p = 2.1 \times 10^{-13}$
Baseline characteristics: n (%)			
Digestive disease	5 (3.5%)	7 (14%)	$\chi^2 = 6.93$, $p = .0085$
Cardiovascular disease	58 (40.8%)	31 (62%)	$\chi^2 = 6.66$, $p = .01$
Cerebrovascular disease	16 (11.3%)	18 (36%)	$\chi^2 = 15.3$, $p = 8.9 \times 10^{-5}$
COPD	21 (14.8%)	15 (30%)	$\chi^2 = 5.6$, $p = .018$
Chronic kidney disease	29 (20.4%)	7 (14%)	$\chi^2 = 0.98$, $p = .31$
Diabetes	29 (20.4%)	6 (12%)	$\chi^2 = 1.76$, $p = .18$
Other	26 (18.3%)	8 (16%)	$\chi^2 = 0.13$, $p = .71$
Symptoms: n (%)			
Fever	113 (79.6%)	42 (84%)	$\chi^2 = 0.46$, $p = .49$
Cough	96 (67.6%)	38 (76%)	$\chi^2 = 1.24$, $p = .27$
Myalgia	38 (26.7%)	9 (18%)	$\chi^2 = 1.54$, $p = .22$
Diarrhea	6 (4.2%)	5 (10%)	$\chi^2 = 2.28$, $p = .13$
Chest pain	51 (35.9%)	26 (52%)	$\chi^2 = 3.98$, $p = .046$
Dyspnea	22 (15.5%)	17 (34%)	$\chi^2 = 7.82$, $p = 5.1 \times 10^{-4}$

Note: Results for continuous variables are reported as mean (SD). Results for categorical variables are reported as n (%). Duration: time from illness onset to admission.

Abbreviations: COPD: Chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019.

3.2 | Classification performance and risk factors for mortality

To predict the prognosis of critically ill patients with COVID-19 according to indicators at admission, we constructed an SVM model with all the self-evaluation indicators and laboratory-examination indicators as input features. Over the 10 undersampling iterations, the classifiers achieved a mean test accuracy of $93 \pm 2\%$ between survivors and non-survivors from 10-fold CV, as shown in Figure 2A (inner plot). Permutation testing showed that the accuracies reported by real-label samples were significantly different from those reported by random-label samples ($p = .019$, Table S3).

Figure 2A (outer plot) is the plot of all indicators sorted by weight. The greater the weight, the more likely the indicator was to be a risk factor for mortality. Duration (time from illness onset to admission) had the highest weight of 1, which meant it was the strongest predictor of all-cause mortality. The weights (ω) of 11 indicators were greater than 0.3, including three self-evaluation indicators: duration ($\omega = 1 \pm 0$), the BI score ($\omega = 0.45 \pm 0.11$), and age ($\omega = 0.33 \pm 0.12$). For

the nonsurvivors, the median time from illness onset to admission was 22.5 (interquartile range [IQR]: 15–28) days, and the median age was 71 (IQR: 61–78) years, the median BI score was 45 (IQR: 10–70). When the duration, age, and the BI score increased by 1 day, 1 year, and 1 point, the mortality increased by 3.6%, 2.4%, and 0.9% respectively, as shown in black dashed lines in Figure 2B.

Eight laboratory-examination indicators with weights greater than 0.3, including CRP ($\omega = 0.54 \pm 0.08$), WBC ($\omega = 0.5 \pm 0.09$), Ca^+ ($\omega = 0.40 \pm 0.07$), PLT ($\omega = 0.39 \pm 0.08$), FDP ($\omega = 0.33 \pm 0.05$), OI ($\omega = 0.33 \pm 0.12$), LYMPH ($\omega = 0.31 \pm 0.08$), and D-dimer ($\omega = 0.30 \pm 0.06$), were considered to be important risk factors. The detailed weights are listed in Table S4.

3.3 | Predictive ability of self-evaluation indicators and laboratory-examination indicators

To estimate the predictive ability of the self-evaluation indicators and laboratory-examination indicators in predicting a poor prognosis, we

TABLE 2 Differences in clinical medical records between survivors and non-survivors of critically ill COVID-19 patients

	Survivors (n = 142)	Nonsurvivors (n = 50)	t Statistics/p value
WBC (4–10 × 10 ⁹ /L)	6.3 (2.8)	12.2 (8.09)	t = -8, p = 6.1 × 10 ⁻¹⁴
Hb (120–160 g/L)	117.1 (22.1)	122.6 (23.6)	t = -1.54, p = .12
PLT (100–400 × 10 ⁹ /L)	202.7 (102.4)	161.6 (92)	t = 2.56, p = .011
LYMPH (0.8–4 × 10 ⁹ /L)	1.23 (0.67)	0.73 (0.79)	t = 4.43, p = 1.5 × 10 ⁻⁵
Alb (35–55 g/L)	35.8 (6.5)	31.7 (5.4)	t = 4.07, p = 1.5 × 10 ⁻⁵
ALT (0–50 U/L)	29.1 (34.1)	32.3 (41.5)	t = -0.55, p = .58
AST (0–50 U/L)	25.4 (25.8)	37.2 (19.9)	t = -2.99, p = .0031
TBIL (0–20 μmol/L)	15.9 (13.2)	19.6 (12.4)	t = -1.74, p = .083
Bun (1.7–8.2 mmol/L)	8.98 (10.4)	11.2 (9.2)	t = -1.37, p = .17
Crea (38–120 μmol/L)	223.6 (417.2)	124.9 (123.4)	t = 1.66, p = .098
UA (204–428 μmol/L)	228 (131.9)	310.3 (163.9)	t = -0.99, p = .32
CRP (0–10 mg/L)	29.9 (29.3)	73.4 (42.3)	t = -7.7, p = 8.1 × 10 ⁻¹³
PCT (0–0.1 ng/ml)	0.47 (1.4)	3.4 (6.7)	t = -4.5, p = 1.4 × 10 ⁻⁵
ESR (0–25 mm/h)	39.6 (28.2)	48.4 (26.7)	t = -1.87, p = .06
PT (8.6–12 s)	12.6 (10.1)	13.2 (1.66)	t = -0.4, p = .68
INR (0.8–1.1)	1.1 (0.16)	1.2 (0.16)	t = -4.3, p = 3.2 × 10 ⁻⁵
APTT (26–42 s)	31.7 (8.3)	31.5 (8.2)	t = 0.1, p = .91
D-Dimer (0–243 ng/ml)	605.4 (2162)	4341.8 (7338)	t = -5.5, p = 1.7 × 10 ⁻⁷
FDP (0–5 μg/ml)	4.5 (12.8)	39.4 (70.9)	t = -5.8, p = 2.6 × 10 ⁻⁸
Fib (1.9–4.6 g/L)	3.98 (1.1)	4.3 (1.7)	t = -1.6, p = .11
PH	7.4 (0.08)	7.3 (0.17)	t = 3.06, p = .0025
PCO ₂ (mmHg)	38.6 (5.4)	47.1 (23.9)	t = -3.96, p = .0001
PO ₂ (mmHg)	76.4 (27.6)	59.9 (27.3)	t = 3.55, p = .0005
SO ₂ (%)	93.4 (8.3)	85.1 (15.2)	t = 4.7, p = 4.9 × 10 ⁻⁶
Lat (0.18–3 mmol/L)	2.7 (1.3)	3.7 (3.2)	t = -4.5, p = 1.4 × 10 ⁻⁵
K ⁺ (3.8–5.4 mmol/L)	4.03 (0.7)	4.2 (1.4)	t = -1.37, p = .17
Na ⁺ (135–148 mmol/L)	138.6 (5.4)	142.2 (8.7)	t = -3.38, p = .0009
Ca ⁺ (2.25–3 mmol/L)	2.01 (0.29)	1.8 (0.44)	t = 3.7, p = .0003
BG (3.9–11.1 mmol/L)	9.3 (3.4)	10.5 (4.15)	t = -1.99, p = .05
OI (400–500 mmHg)	327.2 (134.8)	255.8 (66.8)	t = 3.63, p = .0004

Note: Brackets represent the range of reference values.

Abbreviations: ALT, alanine aminotransferase; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FDP, fibrin degradation products; Hb, hemoglobin; LYMPH, lymphocyte counts; OI, oxygenation index; PCT, procalcitonin; PT, prothrombin time; TBIL, total bilirubin.

constructed an SVM model with the self-evaluation indicators and laboratory-examination indicators as individual input features. Over the 10 under-sampling iterations, the classifiers achieved a mean test accuracy of 87.8 ± 1.8% between survivors and nonsurvivors from 10-fold CV when using the self-evaluation indicators and 85.6 ± 2.5% when using the laboratory-examination indicators. Duration ($\omega = 1 \pm 0$), the BI score

($\omega = 0.57 \pm 0.14$), and age ($\omega = 0.37 \pm 0.11$) were still the most important risk factors when considering the self-evaluation indicators, as shown in Figure 3A. WBC ($\omega = 0.96 \pm 0.08$), CRP ($\omega = 0.91 \pm 0.1$), LYMPH ($\omega = 0.65 \pm 0.1$), Ca⁺ ($\omega = 0.59 \pm 0.15$), FDP ($\omega = 0.56 \pm 0.08$), D-dimer ($\omega = 0.56 \pm 0.12$), and PLT ($\omega = 0.55 \pm 0.18$) were risk factors when considering the laboratory-examination indicators, as shown in Figure 3B.

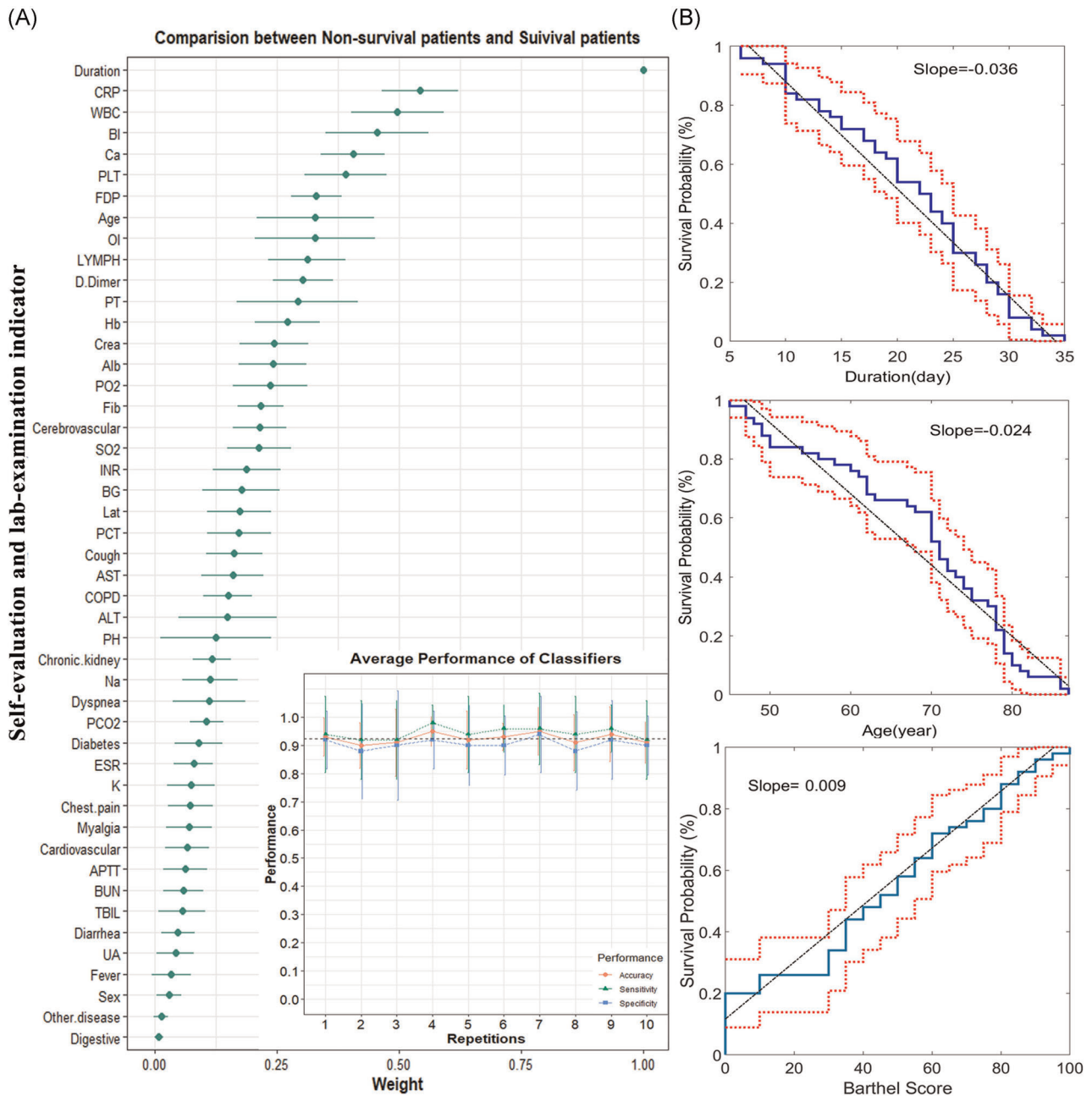


FIGURE 2 (A) Performance of support vector machine results. Mean accuracy, sensitivity, and specificity were shown in the inner plot for all 10 repetitions. The risk factors were shown in the outer plot in order of weight. (B) Kaplan–Meier plot was drawn for the duration, age, and Barthel index. Red dashed lines representing 95% confidence intervals. When the duration (time from illness onset to hospital), age, and the Barthel index score increased by 1 day, 1 year, and 1 point, the mortality increased by 3.6%, 2.4%, and 0.9%, respectively, as shown in black dashed lines in (B)

4 | DISCUSSION

In this study, MVPA was used to identify risk factors for mortality in critically ill COVID-19 patients. Relatively high classification accuracy of 93% was achieved when using both self-evaluation indicators and laboratory-examination indicators. Self-evaluation indicators, such as time from illness onset to hospital, age, and the BI score were risk indicators that can simply be obtained by questionnaires and they helped preliminarily and conveniently

assess disease severity, even by the patients themselves. Laboratory-examination indicators, such as CRP, WBC, Ca⁺, PLT, FDP, OI, LYMPH, and D-dimer, were risk factors as well; these factors may help us stratify patients based on possible requirements for the level of care to improve outcomes and find potential targets for therapeutic interventions.

Many studies have noted that age is a risk factor for COVID-19 severity, but for the first time, our results showed that the time from symptom onset to admission and the BI score were risk factors.^{3,12,13}

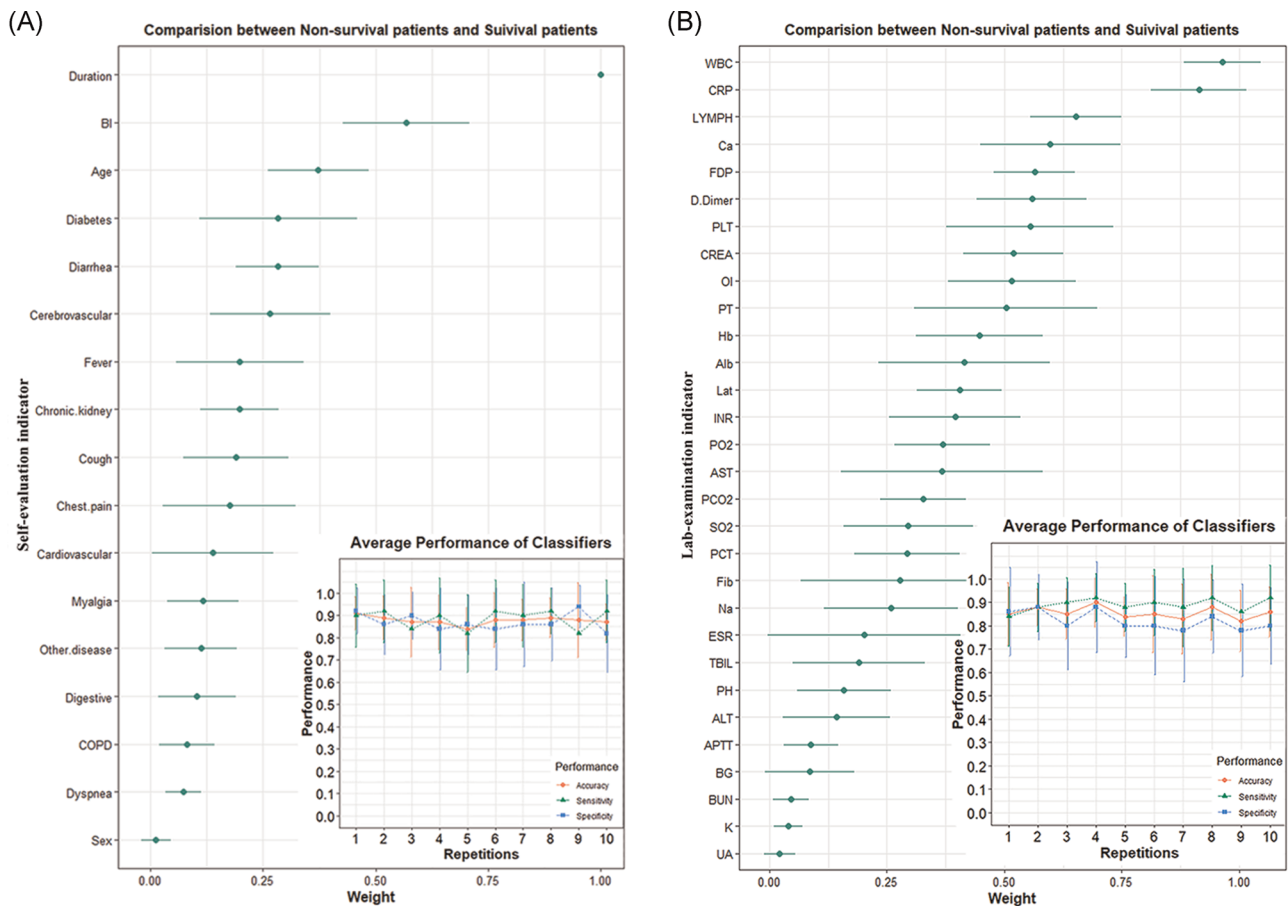


FIGURE 3 Performance of support vector machine results when using self-evaluation indicators (A) and laboratory-examination indicators (B) as input features individually

The longer time from symptom onset to admission in the non-survivors than in the survivors suggests that these patients tend to wait longer before admission, voluntary or involuntary, and might miss the best treatment opportunity. In Italy, nearly 10% of infected patients required intensive care management, which placed considerable strain on the health care system.^{14–16} Some critically ill patients could not get prompt and effective treatments due to limited resources.¹⁷ This may be one of the reasons for the high mortality (12.6%) in patients in Italy. The situation is worse in Iran because of the large epidemic scale of COVID-19, and greatly reduced access to medical care, which is partly influenced by US punitive policies.¹⁸ All these data suggest that early and effective access to medical care is vital for improving the chance of survival, especially for high-risk or critically ill patients. However, there are, of course, other factors that might contribute to the mortality differences, such as medical conditions and treatment methods which deserve further exploration. The BI was a 10-item ordinal scale (range: 0–100) which was widely used for evaluating patients' independence in activities of daily living.¹⁹ We found that BI was a strong predictor of mortality in critically ill patients with COVID-19 and might be applied to risk stratification, therapy optimization, and patients' self-assessment.

High PCT levels ($p = 1.4 \times 10^{-5}$) combined with increased WBC counts ($p = 6.1 \times 10^{-14}$), which were strongly suggestive of bacterial coinfection, were observed in the nonsurvivors and correlated with a high mortality rate. This was consistent with our clinical observations and some reports on other viral diseases. Among critically ill patients with Middle East respiratory syndrome, 18% had bacterial coinfections.²⁰ As mentioned in previous studies, nearly 100% of COVID-19 patients died in the intensive care unit (ICU) had sepsis.⁷ The nonsurvivors tended to suffer from severe coinfection, partly because of the need for invasive treatments such as tracheal intubation, tracheotomy, mechanical ventilation, central venous catheterization. Though the treatment of COVID-19 with glucocorticoids is controversial,^{2,13,21} it is still used to suppress cytokine storms in critically ill patients based on existing therapeutic strategies for community-acquired pneumonia, ARDS, and other viral infections.^{22–25} However, these patients are at a high risk of glucocorticoid-related secondary infection. Empiric antimicrobial treatment is recommended on the basis that superinfection is reasonably common in this population, as in pandemic influenza.^{26–28} However, the high rates (100%), long courses (19.2 days), and combined use (94.3%) of broad-spectrum antibiotics might lead to the emergence of drug-resistant bacteria that render combating bacterial infections increasingly difficult. The positive

rate on etiological examinations was extremely low (5.6%) in our center. This might be influenced by the prior use of antibiotics before sample collection and relatively limited medical human resources at the time of opening of the medical center. How to optimize antibiotic usage in this population is still a challenge. Timely etiological testing may be essential.

The D-dimer ($p = 1.7 \times 10^{-7}$) and FDP ($p = 2.6 \times 10^{-8}$) levels in nonsurvivors were dramatically elevated, reflecting the hypercoagulable and hyperfibrinolytic states of blood in vivo, which have been previously reported by Zhou.⁷ A recent study reported that anticoagulant therapy appears to be associated with an improved prognosis in severe COVID-19 patients with markedly elevated D-dimer levels.²⁹ Furthermore, we found an increased rate of obvious thrombocytopenia in nonsurvivors ($p = .011$), which might result from consumption by the thrombus, decreased production, and accelerated destruction. Thrombocytopenia occurred in some patients with viral infections and was identified as a risk factor for mortality,^{30,31} but this relationship was not previously clear in COVID-19 patients. Our study is the first to report that thrombocytopenia can be used as a risk factor for mortality in critically ill patients with COVID-19.

MOF was observed in both groups (survivor 20.1% vs. nonsurvivor 23.5%). Respiratory failure was often the earliest and most severe form of organ dysfunction,^{2,6} followed by renal, cardiac, and liver function impairment. The reported prevalence of MOF is highly variable (from 3% to 35%).^{4,22} We infer that acute organ failure or exacerbation of chronic dysfunction could partly result from prolonged hypoxia secondary to respiratory insufficiency. However, we did not find any statistically significant differences in indicators of renal, cardiac, and liver functions (BUN, creatinine, B-type natriuretic peptide, ALT, TBIL) between the two groups or myocardial injury indicators (creatinine kinase); these results were unexpected. OI, a marker of oxygenation, was significantly higher in survivors than in nonsurvivors. Our data showed that respiratory function, not comorbidities, was the most important determinant of prognosis in critically ill COVID-19 patients. A recent study in Italy reported that among 1591 patients, the majority were admitted to the ICU because of respiratory failure had a mortality rate of 26% at 5 weeks after the first admission; and could eventually be higher.¹⁶ This study has several limitations. First, only patients classified as having severe COVID-19 were included, and all the others were excluded from the analyses. Second, this study was conducted at a single-center hospital with a limited sample size. As such, this study may have included disproportionately more patients with poor outcomes. There may also be selection bias when identifying factors that influence clinical outcomes. A large-cohort study of patients with COVID-19 pneumonia from Wuhan, China, other cities in China, and other countries would help to further define the clinical characteristics of and risk factors for the disease. We collected and analyzed indexes most relevant to the pathogenesis and therapy according to clinical experience, and missed some other ones, such as LDH which is identified as a key prognostic marker of lung injury.³²

In conclusion, while there have been some studies on the risk factors of COVID-19, there are few studies that provide a

quantitative ranking of the importance of risk factors. We proposed a predictive model and found that self-evaluation indicators including time from illness onset to hospital, age, and the BI score were risk indicators and that laboratory-examination indicators including CRP, WBC, Ca⁺, PLT, FDP, OI, LYMPH, and D-dimer were risk factors as well. We suggested that for those with abnormal values at these indexes, close monitoring, and early intervention might be very important and could help to reduce mortality.

ACKNOWLEDGMENTS

Dr Guo was supported by the National Natural Science Foundation of China (Grant No.: 11671129). Dr Chen was supported by the Health and Family Planning Commission of Hunan Province, China (Grant No.: B20180428), Hunan Provincial Natural Science Foundation for Outstanding Youth (Grant No.: 2020JJ2020).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Drs Chen and Linli equally contributed to this work. The study was designed by Chen, Guo and Linli. Chen, Lei, Yang, Liu, Xia, Liang, Zhu collected the data, had full access to all of the data in the study and takes responsibility for the integrity of the data. Linli, Guo developed the analytical plan. Guo and Linli undertook the statistical analyses and took responsibility for its accuracy. Guo, Chen, and Linli wrote the first draft of the paper. All authors contributed to writing the manuscript.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The Ethics Commission of Hunan Provincial People's Hospital approved this study. Registration number: ChiCTR2000030941. The Ethics Commission exempted from the need for patient consent.

ORCID

Zeqiang Linli  <http://orcid.org/0000-0003-4181-3959>

REFERENCES

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19; 10 April/2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-10-april-2020>
2. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061-1069. <https://doi.org/10.1001/jama.2020.1585>
3. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA internal medicine*. 2020;180:934.

4. Guan W-J, Ni Z-Y, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *New England J Med*. 2020;382:1708-1720.
5. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497-506.
6. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8(5):475-481. [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5)
7. Zhou F, Yu T, Du R, 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):1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
8. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI. *data. Trends Cogn Sci (Regul Ed)*. 2006;10(9):424-430. <https://doi.org/10.1016/j.tics.2006.07.005>
9. Meier TB, Desphande AS, Vergun S, et al. Support vector machine classification and characterization of age-related reorganization of functional brain networks. *Neuroimage*. 2012;60(1):601-613. <https://doi.org/10.1016/j.neuroimage.2011.12.052>
10. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2010;45(3):1-68.
11. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*. 2009;45(1 Suppl):S199-S209. <https://doi.org/10.1016/j.neuroimage.2008.11.007>
12. Wang Y, Jiang W, He Q, et al. Early, low-dose and short-term application of corticosteroid treatment in patients with severe COVID-19 pneumonia: single-center experience from Wuhan, China. *medRxiv*. 2020;1-16. <https://doi.org/10.1101/2020.03.06.20032342>
13. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
14. Italian Government measures. Legislative decree for COVID-19 emergency. <http://www.governo.it/it/articolo/decreto-legge-17-marzo-2020/14333>. Accessed March 22, 2020.
15. Italian National Institute of Health. Report of COVID-19 patients. <http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus>
16. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323(16):1574-1581. <https://doi.org/10.1001/jama.2020.5394>
17. Remuzzi A, Remuzzi G. COVID-19. and Italy: what next? *Lancet*. 2020;395(10231):1225-1228. [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9)
18. Takian A, Raoufi A, Kazempour-Ardebili S. COVID-19 battle during the toughest sanctions against Iran. *Lancet*. 2020;395(10229):1035-1036. [https://doi.org/10.1016/S0140-6736\(20\)30668-1](https://doi.org/10.1016/S0140-6736(20)30668-1)
19. Loyd C, Markland AD, Zhang Y, et al. Prevalence of hospital-associated disability in older adults: a meta-analysis. *J Am Med Dir Assoc*. 2020;21(4):455-461. <https://doi.org/10.1016/j.jamda.2019.09.015>
20. Arabi YM, Mandourah Y, Al-Hameed F, et al. Corticosteroid therapy for critically ill patients with Middle East respiratory syndrome. *Am J Respir Crit Care Med*. 2018;197(6):757-767. <https://doi.org/10.1164/rccm.201706-1172OC>
21. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet*. 2020;395(10229):1033-1034. [https://doi.org/10.1016/S0140-6736\(20\)30628-0](https://doi.org/10.1016/S0140-6736(20)30628-0)
22. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020;323(13):1239-1242. <https://doi.org/10.1001/jama.2020.2648>
23. Villar J, Ferrando C, Martínez D, et al. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir Med*. 2020;8(3):267-276. [https://doi.org/10.1016/S2213-2600\(19\)30417-5](https://doi.org/10.1016/S2213-2600(19)30417-5)
24. Lewis SR, Pritchard MW, Thomas CM, Smith AF. Pharmacological agents for adults with acute respiratory distress syndrome. *Cochrane Database Syst Rev*. 2019;7:CD004477. <https://doi.org/10.1002/14651858.CD004477.pub3>
25. Lansbury L, Rodrigo C, Leonardi-Bee J, Nguyen-Van-Tam J, Lim WS. Corticosteroids as adjunctive therapy in the treatment of influenza. *Cochrane Database Syst Rev*. 2019;2(2):CD010406. <https://doi.org/10.1002/14651858.CD010406.pub3>
26. McCullers JA. Do specific virus-bacteria pairings drive clinical outcomes of pneumonia? *Clin Microbiol Infect*. 2013;19(2):113-118. <https://doi.org/10.1111/1469-0691.12093>
27. Shieh W-J, Blau DM, Denison AM, et al. 2009 pandemic influenza A (H1N1): pathology and pathogenesis of 100 fatal cases in the United States. *Am J Pathol*. 2010;177(1):166-175.
28. Rice TW, Rubinson L, Uyeki TM, et al. Critical illness from 2009 pandemic influenza A virus and bacterial coinfection in the United States. *Crit Care Med*. 2012;40(5):1487-1498. <https://doi.org/10.1097/CCM.0b013e3182416f23>
29. Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J Thromb Haemost*. 2020;18(5):1094-1099. <https://doi.org/10.1111/jth.14817>
30. Yang M, Ng MHL, Li CK. Thrombocytopenia in patients with severe acute respiratory syndrome (review). *Hematology*. 2005;10(2):101-105. <https://doi.org/10.1080/10245330400026170>
31. Yin S, Huang M, Li D, Tang N. Difference of coagulation features between severe pneumonia induced by SARS-CoV2 and non-SARS-CoV2. *J Thromb Thrombolysis*. 2020;1-4. <https://doi.org/10.1007/s11239-020-02105-8>
32. Yan L, Zhang H-T, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell*. 2020;2(5):283-288. <https://doi.org/10.1038/s42256-020-0180-7>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Chen Y, Linli Z, Lei Y, et al. Risk factors for mortality in critically ill patients with COVID-19 in Huanggang, China: A single-center multivariate pattern analysis. *J Med Virol*. 2021;93:2046–2055. <https://doi.org/10.1002/jmv.26572>