

Identification of Risk Factors for COVID-19-related Death using Machine Learning Methods

Saeid Bashirian¹, Maryam Mohammadi-Khoshnoud², Salman Khazaei³, Elham Talebighane⁴, Fariba Keramat⁵, Fatemeh Bahreini⁶, Sepideh Zareeian⁴, Ali Reza Soltanian⁷

¹ Social Determinants of Health Research Center, Hamadan University of Medical Sciences, Hamadan, Iran, ² Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran, ³ Health Science Research Center, Hamadan University of Medical Sciences, Hamadan, Iran, ⁴ Hamadan University of Medical Sciences, Hamadan, Iran, ⁵ Department of Infectious Disease, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran, ⁶ Department of Molecular Medicine and Genetics, Faculty of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran, ⁷ Modeling of Non-Communicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.

Received: 2 February 2021

Accepted: 7 May 2021

Correspondence to: Soltanian AR

Address: Modeling of Non-Communicable

Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

Email address: soltanian@umsha.ac.ir

Background: Unknown cases of pneumonia appeared in late 2019 in Wuhan, China. Following the worldwide spread of the disease, the World Health Organization declared it a pandemic on March 11, 2020. The total number of infected people worldwide as of December 16, 2020, was more than 74 million, more than one million and six hundred thousand of whom died from Coronavirus Disease 2019 (COVID-19). This study aimed to identify the risk factors for the mortality of COVID-19 in Hamadan, west of Iran.

Materials and Methods: This cross-sectional study used the information of all patients with COVID-19 admitted to Shahid Beheshti and Sina hospitals in Hamadan during January 2020-November 2020. Logistic regression model, decision tree, and random forest were used to assess risk factors for death due to COVID-19.

Results: This study was conducted on 1853 people with COVID-19. Blood urea nitrogen change, SPO₂ at admission, the duration of hospitalization, age, neutrophil count, lymphocyte count, number of breaths, complete blood count, systolic blood pressure, hemoglobin, and sodium were effective predictors in both methods of decision tree and random forest.

Conclusion: The risk factors identified in the present study may serve as surrogate indicators to identify the risk of death due to COVID-19. The proper model to predict COVID-19-related mortality is random forest based on sensitivity.

Key words: Data mining; Emerging disease; Mortality; Risk factors; SARS-CoV-2

INTRODUCTION

Unknown cases of pneumonia appeared in late 2019 in Wuhan, China (1, 2). In January 2020, a new beta-coronavirus subtype was identified in China using a pharyngeal swab sample from patients with the disease, later renamed Coronavirus Disease 2019 (COVID-19) (3). On January 30, 2020, the World Health Organization

(WHO) declared the COVID-19 epidemic a public health emergency of international concern (4). The WHO announced the disease as a pandemic on March 11, 2020 (5).

According to Worldometer, the total number of infected people worldwide as of December 16, 2020, was more than 74 million, more than one million and six

hundred thousand of whom died from COVID-19 (6). The United States tops the world with more than 17 million cases and more than 312,000 deaths. India and Brazil are next in the world in terms of infection rates. Iran ranks 15th in the world with over one million infected cases and more than 52,000 deaths (6). In the United States and Iran, there are 941 and 626 deaths per million population, respectively. Moreover, the average number of deaths per million population in the world is 211.6.

The results of epidemiological studies indicate that this disease is more prevalent in men, the elderly, and patients with a history of chronic diseases, such as diabetes and hypertension (4, 6, 7). In addition, obesity and cardiovascular diseases have been mentioned as factors that aggravate the severity of symptoms (8, 9). Understanding disease epidemiology is essential for better management of the epidemic. Therefore, we need to know the risk of death for people with the disease and identify risk factors associated with the increased risk of death (10).

Although some studies have examined the effect of various pharmacological factors, demographic variables, chronic diseases, and serum biochemical parameters on mortality, few studies have used new and flexible methods to identify death predictors in COVID-19 patients. In this study, a statistical method known as classification and regression trees analysis (CART) was used to detect death predictors in COVID-19 patients. The CART analysis creates decision trees that perform the best in selecting the predictors that play a role in identifying high-risk groups and clinical decision-making.

MATERIALS AND METHODS

This cross-sectional study used the data of all patients with COVID-19 admitted to Shahid Beheshti and Sina hospitals in Hamadan, west of Iran, during January 2020–November 2020. The diagnosis of COVID-19 was based on a reverse transcription-polymerase chain reaction. The Ethics Committee of Hamadan University of Medical Sciences approved the study (IR.UMSHA.REC.1399.682). Predictor variables in this study included gender, age, marital status, habitat, smoking, hookah use, injecting drug

users (IDU), immunodeficiency, cardiovascular diseases, history of kidney diseases, diabetes, hypertension, cancer, history of pulmonary diseases, history of liver disease, history of blood disease, fever, chills, dry cough, sore throat, shortness of breath (dyspnea), sputum cough, headache, nausea, fatigue, vomiting, diarrhea, runny nose, constipation, weakness, decreased level of consciousness, anorexia, loss of taste and smell, vertigo, urinary disorders, hemoptysis, abdominal pain, chest pain, perspire, stomach pain, number of breaths, temperature at admission, systolic and diastolic blood pressure at admission, SPO₂ at admission, heart rate, auscultation of lungs, blood creatinine, sodium (Na), blood platelets, hematocrit (HCT), hemoglobin (Hb), complete blood count (CBC), lymphocytes (LYM), blood urea nitrogen (BUN), neutrophils (NEUT), and the duration of hospitalization. In the present study, the response variable was considered recovery=1 or death=0 from COVID-19 disease. The mentioned information was extracted from the files of patients using a researcher-made checklist by two trained nursing experts under the supervision of an infectious disease specialist.

Statistical Analysis

Logistic regression model, decision tree, and random forest were used for the analyses in this study. First, logistic regression models were fitted to the data separately. These models included each of the predictor variables and BUN. BUN was included in all fitted logistic regression models predictor variables due to its importance, along with the desired variables. After fitting logistic regression models, variables whose significance level was less than 0.2 were selected to enter the decision tree and random forest. Data in random forest and decision tree were divided into two parts of training and test, with 80% of the data being used for training and 20% for test.

In addition, the CART algorithm can be considered one of the well-known classification models for diagnosis and prediction in medical sciences (11). This method produces decision trees that divide a sample into several non-overlapping subsamples that differ in response variability (11). In the present study, the Gini index is used to assess

node impurity (12), which can be calculated by the following formula:

$$\emptyset(p) = p(1 - p)$$

Furthermore, in the CART model, pruning of a classification tree is completed according to cost-complexity and can be calculated based on the following formula (12):

$$R_{\alpha}(\tau) = R(\tau) + \alpha|\bar{T}|$$

Random forest is an ensemble effect of unpruned regression or classification trees (13). A random forest, unlike the tree, is extremely large for interpretation. One way to summarize or quantify information is to identify important predictors in the forest. In the current study, the importance of Gini was used to examine the importance of the variable. The training and test samples were evaluated using sensitivity, specificity, the area under the curve (AUC), and the F1 index. The F1 score is the harmonic mean of precision and recall and provides a better measure of accuracy for items classified incorrectly. This index uses the harmonic mean because it penalizes the extreme values. Accuracy is used when true positive and true negative results are more important, while F1 is used when false negative and false positive are critical. F1 is a better metric when there are unbalanced items. There is an unbalance class distribution in real classification problems. Therefore, F1 is a better index for evaluating the model.

Data were analyzed using tree, random forest, and the pROC packages of R software version 4.0.2 (14-16). The control argument in the tree function was used to determine the minimum number of observations for each node, the smallest size of each node, and deviance within the node. The minimum number of observations for each node, smallest size of each node, and deviance within the node were set at 10, 20, and 0.005, respectively.

RESULTS

The present study was conducted on 1853 people with COVID-19 hospitalized in Shahid Beheshti and Sina hospitals in Hamadan, Iran, during January 2020-November 2020. Of these, 264 (14.2%) died of Coronavirus heart disease, and 969 (52.3%) were male. Hypertension

(33.8%) and diabetes (18.2%) were the most common underlying diseases. On the other hand, the rarest underlying diseases were immunodeficiency disease (0.2%), liver disease (0.8%), and blood disease (0.8%). The most common clinical symptoms among these patients were shortness of breath (60.8%), fever (56.6%), chills (48.3%), myalgia (45.1%), dry cough (45.1%), and weakness (42.8%). The rarest symptoms were decreased consciousness (2.4%), stomach pain (1.3%), constipation (1.3%), and runny nose (0.5%) (Table 1).

The mean \pm SD of the age of patients was 59.45 \pm 16.92 years (range: 16-98 years). The mean \pm SD of diastolic and systolic blood pressure at admission were 76.68 \pm 11.12 and 68.121 \pm 17.93 mm Hg, respectively (Table 2). Two decision tree and random forest methods were used to perform the analyses. The results of the used methods are described separately.

Decision Tree Results

The decision tree for the death of COVID-19 patients had 34 final nodes. The modeling results showed that out of 33 variables entered into the decision tree, only BUN, length of stay, SPO₂ at admission, age, NEUT, diastolic blood pressure, dry cough, LYM, number of breaths, shortness of breath, CBC, systolic blood pressure, auscultation of lungs, Hb, Na, and cardiovascular disease were identified as variables affecting the mortality of patients with COVID-19 (Figure 1). The fitted decision tree was valid and had the necessary credibility so that the classification error in this decision tree was calculated as 0.1.

Random Forest Results

Overall, 500 trees were used to complete the process of modeling and constructing random forest. In addition, the Gini index was used to evaluate the importance of variables in this method. According to the Gini index, the most important predictors in the random forest method were BUN (43.74), SPO₂ at the time of admission (33.21), the length of stay (27.6), age (11.26), HCT (19.55), LYM (19.45), CBC (19.26), NEUT (18.77), Hb (18.25), the number of breaths (17.21), heart rate (15.7), Na (15.63), potassium (K) (15.38), systolic blood pressure (13.51) (Figure 2).

Table1. Demographic and clinical characteristics of patients with COVID-19 admitted to Sina and Shahid Beheshti teaching hospitals in Hamadan, western Iran

Variables		Frequency	Percent	Variables		Frequency	Percent
Treatment output	Recovery	1589	85.8	History of Blood diseases	No	1839	99.2
	Died	264	14.2		Yes	14	0.8
Gender	Female	844	47.7	Hemoptysis	No	1841	99.4
	Male	969	52.3		Yes	12	0.6
Auscultation of lungs	Normal	1326	71.6	Decreased level of consciousness	No	1808	97.6
	Crackles	479	25.8		Yes	45	2.4
History of liver diseases	Wheeze	48	2.6	Chest pain	No	1694	91.4
	No	1839	99.2		Yes	159	8.6
Diarrhea	Yes	14	0.8	Perspire	No	18.3	97.3
	No	1647	88.9		Yes	50	2.7
Vomiting	Yes	206	11.1	Stomach ache	No	1828	98.7
	No	1460	78.8		Yes	25	1.3
Sputum cough	Yes	393	21.2	Weakness	No	1059	57.2
	No	1509	81.4		Yes	794	42.8
Shortness of breath (dyspnea)	Yes	344	18.6	History of pulmonary diseases	No	1610	86.9
	No	726	39.2		Yes	243	13.1
Headache	Yes	1127	60.8	Married status	Married	1601	86.4
	No	1510	81.5		Single	74	4.0
Chills	Yes	343	18.5	Widow	167	9.0	
	No	958	51.7	Divorced	11	0.6	
Nausea	Yes	895	48.3	Habitat	Village	302	16.3
	No	1408	76.0		City	1551	83.7
Fatigue	Yes	445	24.0	Smoking	No	1712	92.4
	No	1800	97.1		Yes	141	7.6
Fever	Yes	53	2.9	Hookah use	No	1827	98.6
	No	804	43.4		Yes	26	1.4
Sore throat	Yes	1049	56.6	IDU	No	1719	92.8
	No	1776	95.8		Yes	134	7.2
Constipation	Yes	77	4.2	Diabetes	No	1515	81.8
	No	1828	98.7		Yes	338	18.2
Abdominal pain	Yes	25	1.3	History of hypertension diseases	No	1227	66.2
	No	1771	95.6		Yes	626	33.8
Loss of sense of taste and smell	Yes	82	4.4	History of Kidney diseases	No	1777	95.9
	No	1649	89.0		Yes	76	4.1
Urinary disorders	Yes	204	11.0	History of Cardiovascular disease	No	1528	82.5
	No	1801	97.2		Yes	325	17.5
Immunodeficiency	Yes	52	2.8	Myalgia	No	1017	54.9
	No	1849	99.8		Yes	836	45.1
Anorexia	Yes	4	0.2	Runny nose	No	1844	99.5
	No	1219	65.8		Yes	9	0.5
Cancer	Yes	634	34.2	Dry cough	No	1017	54.9
	No	1820	98.2		Yes	836	45.1
	Yes	33	1.8	Vertigo	No	1785	96.3
					Yes	68	3.7

Table 2. Descriptive statistics for clinical characteristics of patients with COVID-19

Variables	Mean	SD
Age	59.45	16.92
Diastolic blood pressure at admission (mm Hg)	76.68	11.12
Systolic blood pressure at admission (mm Hg)	68.121	17.93
Patients' body temperature at admission(° C)	37.30	0.83
Number of breaths	19.87	4.05
Spo2 at time of admission	84.69	9.88
Blood urea nitrogen(mg/dl)	18.81	13.01
Heart rate	92.32	14.88
Creatinine level (mg/dl)	1.24	2.11
Potassium (K) (mEq/l)	4.17	0.92
Sodium (NA) (mEq/l)	137.61	3.80
Platelets (mm)	198.78	84/26
Hematocrit (HCT) (percent)	42.51	5.47
Hemoglobin (g/dl)	13.91	1.97
Blood cell count (CBC)	7.07	4.54
Lymphocyte (percent)	23.77	11.77
Neutrophil (percent)	71.97	12.39
Duration of hospitalization (day)	7.36	5.99

There are similarities and differences between the critical predictions known in the two methods. The BUN, SPO₂ at admission, the length of stay, age, NEUT, LYM, the number of breaths, CBC, systolic blood pressure, Hb, and Na in both methods are known to be effective predictors. However, the variables diastolic blood pressure, dry cough, the shortness of breath, the auscultation of lungs, and cardiovascular diseases in the decision tree and the variables HCT, heart rate, and Kin random forest were identified as influential variables.

Good Fitting Results

Comparing the results shown in Table 3, we found that AUC for the decision tree in the training sample was higher than the test sample, while for the random forest, AUC in the test sample was better than the training sample. The sensitivity of both decision tree and random forest methods was higher than specificity in the training

and test samples. The F1 index for both training and test samples methods was higher than 0.9. Specificity in both training and test samples of the two methods was almost low. The specificity of the decision tree in the training sample was higher than 0.5, and the other specificity was lower than 0.5.

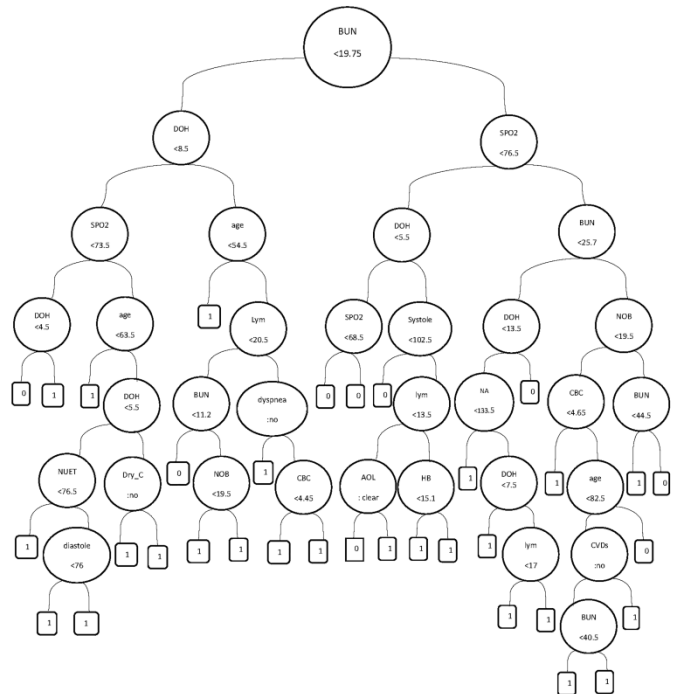


Figure1. Decision tree with allocating the response to each node (DOH: duration of hospitalization, NOB: number of breaths, AOL: Auscultation of lungs, CVDs: Cardiovascular disease, Dry_C: Dry cough 0: died, 1: recovery)

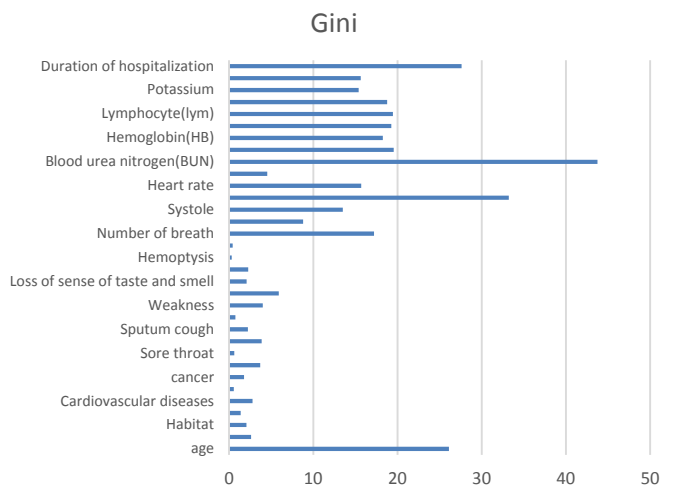


Figure 2. Importance of variables in random forest with Gini index

Table 3. Comparison of sensitivity, specificity, F1-index and area under the for training and test samples

Model		Sensitivity	specificity	F1 score	AUC(CI)
Random Forest	Train	0.98	0.27	0.94	0.86(0.84,0.88)
	Test	0.99	0.23	0.93	0.89(0.84,0.94)
Decision Tree	Train	0.97	0.52	0.95	0.94(0.92,0.95)
	Test	0.95	0.36	0.92	0.73(0.64,0.82)

DISCUSSION

This study aimed to identify the risk factors of death due to COVID-19 and determine its predictors. According to the decision tree and random forest results, BUN is one of the variables affecting the mortality of patients hospitalized with COVID-19. The relationship between the indicators of renal involvement and increased risk of death in COVID-19 patients has been shown in previous studies. Consistent with the results of the present study, Cheng et al. showed that BUN and D-dimer levels were significantly higher in patients who died due to coronavirus than in patients who recovered (17).

Mechanisms of renal involvement in COVID-19 include rhabdomyolysis, renal hypoperfusion, direct kidney damage by cytokines in cytokine release syndrome, hemophagocytic lymphohistiocytosis, or elevated cytokine production due to extracorporeal membrane oxygenation and mechanical ventilator. Furthermore, a cardiovascular syndrome caused by viral myocarditis, renal medullary hypoxia due to alveolar damage, renal compartment syndrome because of high peak airway pressure or intra-abdominal hypertension, and septic acute kidney injury by endotoxins may occur (18).

The SPO₂ at admission was another influential variable on the mortality of COVID-19 patients in this study. Bahl also noted the importance of SPO₂ in the mortality of COVID-19 patients in his study. In the latter research, older age, low oxygen saturation at admission, and early laboratory abnormalities, such as renal and hepatic impairments, were the risk factors for COVID-19 death in hospitals (19).

Our findings demonstrated the age difference between patients who died and recovered from COVID-19. This result has been confirmed in other countries as well. The Yanez study shows that in 16 surveyed countries, people aged 65 and older had a higher risk of death than younger people (20). It seems that defective T-cell and B-cell function and the overproduction of type 2 cytokines due to aging can lead to poor control of virus replication and long-term proinflammatory responses, associated with poor treatment outcomes in patients (21).

The effect of NEUT on patient mortality was identified by decision tree fitting and random forest. Liu stated that the risk of mortality augments with increasing NEUT counts (22). The permeability of capillary endothelial cells into the lungs leads to fluid penetration into the pulmonary parenchyma. These factors describe the pathogenicity of acute respiratory distress syndrome (ARDS) in the presence of an inflammatory response associated with NEUT and cytokines (23). The NEUT plays an important role in developing pulmonary edema associated with acute lung injury or ARDS (24). Endothelial damage occurs in minutes to hours after ARDS, and endothelial cell gaps allow fluid, NEUT, and cytokines to penetrate the parenchymal space of the lungs (24).

Based on the results of the present study, LYM is the next influential factor in COVID-19 mortality. The meta-analysis by Huang and Pranata showed that LYM is associated with severe COVID-19. This meta-analysis revealed that LYM is correlated with a poor response in COVID-19 patients (25). The LYM plays a key role in maintaining immune homeostasis and inflammatory response in the body. Understanding the mechanism of blood LYM depletion can provide an essential strategy for treating COVID-19 (26). Four potential mechanisms have been identified to result in LYM deficiency (26). The virus can directly infect LYM, resulting in LYM death. The LYM expresses the ACE2 receptor for coronavirus and may be the main target of viruses (27). The virus may directly destroy lymphatic organs, and acute LYM depletion may

be associated with LYM dysfunction. Inflammatory cytokines remain abnormal, leading to lymphocyte apoptosis (28). The last mechanism of inhibiting the LYM produced by metabolic molecules is the metabolic disorders, such as hyperlactic acidemia (26). Severe COVID-19 patients have high blood lactic acid levels that may suppress LYM proliferation (29).

We observed the influence of systolic blood pressure at admission on the mortality of COVID-19 patients. This result is inconsistent with the findings of a study by Trabulus et al. These authors indicated no significant difference between the systolic or diastolic blood pressure at the time of admission of patients who died or survived (30). However, in line with the results of the present study, in a meta-analysis by Lippi et al., COVID-19 patients with hypertension were 2.5 times more likely to die (31).

The decision tree and random forest have confirmed the relationship between Hb and COVID-19 mortality. The results of the current study are consistent with the investigation performed by Algassim et al. (32). They showed that low Hb levels are associated with more severe disease progression and higher mortality rates (32). This decrease in Hb can be attributed to inflammation associated with COVID-19 (32). Due to many complex mechanisms in acute inflammation, a decline in Hb levels is expected. The best-known mechanism is cytokine-induced iron metabolism and the inhibition of erythropoietin formation (33). In addition, COVID-19 patients are more prone to bleeding due to iatrogenic coagulopathy or disseminated intravascular coagulation. These factors can be a double reason for the diminished Hb (32).

COVID-19 causes more death in people with weakened or defective immune systems than immunocompetent individuals. Consequently, immunodeficiency diseases were also expected to be among the effective predictors. However, the impact of this variable was not significant in the logistic regression model along with the BUN variable. Perhaps the insignificance of this variable results from the

small number of patients because only 4 (0.2%) of the participants had immunodeficiency. The decision tree performs better in the training sample than in the test sample. This performance difference was more remarkable in specificity and AUC. However, it was not evident in the two criteria of sensitivity and F1.

In general, specificity and F1 criteria in the training sample were better than the test sample. The difference between the AUC and specificity of the decision tree was evident between the two training and test samples. In random forest, sensitivity and AUC in the test sample were higher than the training sample. The two criteria of specificity and F1 in the training sample were better than in the test sample. The sensitivity and F1 of both training and test samples methods were higher than 0.9%. The lowest AUC was 0.73 (decision tree test sample), and the highest AUC was 0.94 (decision tree training sample). The highest specificity was related to the training sample of the decision tree, and the rest of the specificity values were below 0.5.

The present study had several limitations that should be considered in interpreting the results. First, due to the retrospective nature of our study, we were unable to review the effect of some clinically important patient data, such as D-dimer, on disease outcome. Second, the information obtained from this study, which is related to two specialized hospitals in the province, may not be generalizable to patients admitted to other hospitals in the province. Third, in the current investigation, only severe cases that led to the hospitalization of the patient were evaluated. Therefore, it cannot be generalized to patients with moderate or mild symptoms. Fourth, the clinical characteristics and the outcome of some patients had not yet been determined and were excluded from the analysis. Finally, changes in the treatment regimen during the study period were a highly effective variable in disease outcome that has not been studied in the present study. It is recommended that the treatment regimen be considered a significant predictor in future studies.

CONCLUSION

Among the demographic variables of patients, age was an essential factor in COVID-19 mortality. In terms of clinical characteristics, the length of stay, SPO2 at admission, number of breaths, systolic blood pressure, diastolic blood pressure, the auscultation of lungs, and a history of cardiovascular disease were factors influencing COVID-19 mortality. We found that BUN, NEUT, LYM, CBC, Hb, Na, HCT, and K were associated with the elevated risk of death among blood factors. Moreover, patients with shortness of breath and dry cough were at a higher risk of death from COVID-19.

Acknowledgments

This study was supported by Vice chancellor for Research and Technology of Hamadan University of Medical Sciences (Grant number 9908276021).

REFERENCES

- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020;382(13):1199-207.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727-33.
- Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol* 2020;92(4):401-2.
- Alimohamadi Y, Holakouie-Naieni K, Sepandi M, Taghdir M. Effect of Social Distancing on COVID-19 Incidence and Mortality in Iran Since February 20 to May 13, 2020: An Interrupted Time Series Analysis. *Risk Manag Healthc Policy* 2020;13:1695-700.
- Azizmohammadi S, Safari A, Azizmohammadi S, Kaghazian M, Sadrkhanlo M, Yahaghi E, et al. Molecular identification of miR-145 and miR-9 expression level as prognostic biomarkers for early-stage cervical cancer detection. *QJM* 2017;110(1):11-5.
- Wang CJ, Zhou ZG, Wang L, Yang L, Zhou B, Gu J, et al. Clinicopathological significance of microRNA-31, -143 and -145 expression in colorectal cancer. *Dis Markers* 2009;26(1):27-34.
- Hu Y, Tao SY, Deng JM, Hou ZK, Liang JQ, Huang QG, et al. Prognostic Value of NRAS Gene for Survival of Colorectal Cancer Patients: A Systematic Review and Meta-Analysis. *Asian Pac J Cancer Prev* 2018;19(11):3001-8.
- Ok F, Erdogan O, Durmus E, Carkci S, Canik A. Predictive values of blood urea nitrogen/creatinine ratio and other routine blood parameters on disease severity and survival of COVID-19 patients. *J Med Virol* 2021;93(2):786-93.
- Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology* 2018;23(2):130-8.
- Gansevoort RT, Hilbrands LB. CKD is a key risk factor for COVID-19 mortality. *Nat Rev Nephrol* 2020;16(12):705-6.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Wadsworth & Brooks. Cole Statistics/Probability Series. 1984.
- Zhang H, Singer BH. Recursive partitioning and applications. Springer Science & Business Media; 2010.
- Zhao Y, Zhang Y. Comparison of decision tree methods for finding active objects. *Advances in Space Research* 2008;41(12):1955-9.
- Breiman L, Cutler A, Liaw A, Wiener M. Breiman and Cutler's Random Forests for Classification and Regression. UTC. 2018.
- Ripley BD. Classification and Regression Trees. UTC. 2019.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M, Siegert S, Doering M. pROC: display and analyze ROC curves. R package version. 2018;1.
- Cheng A, Hu L, Wang Y, Huang L, Zhao L, Zhang C, et al. Diagnostic performance of initial blood urea nitrogen combined with D-dimer levels for predicting in-hospital mortality in COVID-19 patients. *Int J Antimicrob Agents* 2020;56(3):106110.
- Ronco C, Reis T. Kidney involvement in COVID-19 and rationale for extracorporeal therapies. *Nat Rev Nephrol* 2020;16(6):308-10.
- Bahl A, Van Baalen MN, Ortiz L, Chen NW, Todd C, Milad M, et al. Early predictors of in-hospital mortality in patients with COVID-19 in a large American cohort. *Intern Emerg Med* 2020;15(8):1485-99.

20. Yanez ND, Weiss NS, Romand JA, Treggiari MM. COVID-19 mortality risk for older men and women. *BMC Public Health* 2020;20(1):1742.
21. Opal SM, Girard TD, Ely EW. The immunopathogenesis of sepsis in elderly patients. *Clin Infect Dis* 2005;41 Suppl 7:S504-12.
22. Liu Y, Du X, Chen J, Jin Y, Peng L, Wang HXH, et al. Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. *J Infect* 2020;81(1):e6-e12.
23. Gonzales JN, Lucas R, Verin AD. The Acute Respiratory Distress Syndrome: Mechanisms and Perspective Therapeutic Approaches. *Austin J Vasc Med* 2015;2(1):1009.
24. Azkur AK, Akdis M, Azkur D, Sokolowska M, van de Veen W, Brügggen MC, et al. Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. *Allergy* 2020;75(7):1564-81.
25. Huang I, Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. *J Intensive Care* 2020;8:36.
26. Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang YQ, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduct Target Ther* 2020;5(1):33.
27. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci* 2020;12(1):8.
28. Liao YC, Liang WG, Chen FW, Hsu JH, Yang JJ, Chang MS. IL-19 induces production of IL-6 and TNF-alpha and results in cell apoptosis through TNF-alpha. *J Immunol* 2002;169(8):4288-97.
29. Fischer K, Hoffmann P, Voelkl S, Meidenbauer N, Ammer J, Edinger M, et al. Inhibitory effect of tumor cell-derived lactic acid on human T cells. *Blood* 2007;109(9):3812-9.
30. Trabulus S, Karaca C, Balkan II, Dincer MT, Murt A, Ozcan SG, et al. Kidney function on admission predicts in-hospital mortality in COVID-19. *PLoS One* 2020;15(9):e0238680.
31. Lippi G, Wong J, Henry BM. Hypertension in patients with coronavirus disease 2019 (COVID-19): a pooled analysis. *Pol Arch Intern Med* 2020;130(4):304-9.
32. Algassim AA, Elghazaly AA, Alnahdi AS, Mohammed-Rahim OM, Alanazi AG, Aldhuwayhi NA, et al. Prognostic significance of hemoglobin level and autoimmune hemolytic anemia in SARS-CoV-2 infection. *Ann Hematol* 2021;100(1):37-43.
33. Weiss G, Ganz T, Goodnough LT. Anemia of inflammation. *Blood* 2019;133(1):40-50.