Original Article 10.30477/cjm.2025.481343.1092

Application of Machine Learning Models for Predicting COVID-19 Mortality: A Retrospective Cohort Study Using Random Forest Algorithm

Ghazal Sherkat[1], Mohamad Amin Pourhoseingholi[2], Seyyedeh Zahra Mostafavian[3]

1. Innovative Medical Research Center, Faculty of Medicine, Mashhad Medical Sciences, Islamic Azad University, Mashhad, Iran
2. Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Department of Community Medicine, Faculty of Medicine, Mashhad Medical Sciences, Islamic Azad University, Mashhad, Iran.

ABSTRACT

Background: The COVID-19 pandemic has posed immense challenges to healthcare systems worldwide, making it critical to identify predictors of patient mortality. Machine learning algorithms have shown great promise in predicting outcomes for COVID-19 patients based on clinical and laboratory data. This study aimed to evaluate the performance of three machine learning models—Random Forest, Support Vector Machine (SVM), and Logistic Regression—in predicting mortality among COVID-19 patients using clinical and laboratory data.

Methods: A retrospective cohort study was conducted on 2,500 COVID-19 patients admitted to three major hospitals in Tehran, Iran, between 2020 and 2021. Demographic, clinical, and laboratory data were collected. Machine learning models were trained to predict mortality, and their performance was evaluated using the area under the curve (AUC), sensitivity, specificity, and positive predictive value (PPV).

Results: The Random Forest model outperformed both SVM and Logistic Regression with an AUC of 83.3%, sensitivity of 63.0%, and specificity of 90.5%. Important predictors of mortality included age, gender, comorbidities (such as diabetes, ischemic heart disease, and cancer), ICU admission, and laboratory markers (e.g., ALT, white blood cell count, and creatinine levels).

Conclusion: The Random Forest algorithm demonstrated superior predictive performance compared to SVM and Logistic Regression in predicting mortality among COVID-19 patients. These findings suggest that machine learning models, particularly Random Forest, can be instrumental in identifying high-risk patients and supporting clinical decision-making.

Keywords: COVID-19 mortality; Machine learning; Random Forest

Recived Date: 02 October 2024
Revise Date: 13 January 2025
Accept Date: 26 February 2025
Publish Date: 08 March 2025
Editor: AA. Moodi Ghlibaf (Conflict of Interests: None).
Reviewers: A. Madadi Mahani (Conflict of Interests: None), S. Karkon Shayan(-Conflict of Interests: None), M. Ataee Karizmeh(Conflict of Interests: None).

***Correspondence to:** Seyyedeh Zahra Mostafavian, Department of Community Medicine, Faculty of Medicine, Mashhad Medical Sciences, Islamic Azad University, Mashhad, Iran.
Email: mhp85mhp@gmail.com

Cite as : Ghazal Sherkat, Mohamad Amin Pourhoseingholi, Seyyedeh Zahra Mostafavian. Application of Machine Learning Models for Predicting COVID-19 Mortality: A Retrospective Cohort Study Using Random Forest Algorithm. Canon Journal of Medicine. 2025; 4, 109-112.

INTRODUCTION

The COVID-19 pandemic has had devastating effects on the global community and healthcare systems, causing significant economic, social, and psychological challenges while negatively impacting public health. The rapid spread of the coronavirus has already claimed numerous lives. This pandemic was declared a global health emergency due to its alarming transmission rate. The virus is highly transmissible between individuals through respiratory droplets, direct contact with viral secretions, and oral, nasal, and ocular routes (1). Despite remarkable advances in combating diseases, infectious diseases continue to hold particular importance in epidemiology and public health.

Identifying various aspects of this disease in at-risk groups, based on the presence or absence of symptoms, can greatly assist in preventing infections and expediting recovery in affected patients. With stress overwhelming healthcare centers, it is essential for governments and healthcare institutions to identify and treat patients who are most likely to survive (2). Developing analytical models and algorithms that can effectively and with appropriate sensitivity and specificity, identify survival and mortality predictors is crucial for addressing the COVID-19 pandemic and reducing its burden (3).

The Random Forest method, introduced by Breiman in 2001, is a combined classification technique. Using a bagging approach, Random Forest constructs several distinct training sample sets. Each sample set builds a decision tree with randomly selected features. Random Forest offers several incomparable advantages over other models. Firstly, it has very high diagnostic efficiency, consistently ranking at the top in most models. Secondly, it disregards the assumption of data normality and can handle outlier data fields more effectively. It also allows for the analysis of high-dimensional data and overfitting, which may not appear in simpler linear models (4). Thirdly, it can measure the importance of each variable and effectively eliminate irrelevant variables. Due to these analytical functions, Random Forest has been employed in COVID-19 studies (4,5).

In a notable study, the Random Forest model was combined with the AdaBoost algorithm to predict the severity and outcome (recovery or death) of COVID-19 patients. The results identified a set of demographic and clinical variables, including age, gender, and disease duration, as significant factors influencing the outcome (6).

In this research, we aim to use data mining, Random Forest, and machine learning techniques to predict the survival of COVID-19 patients in Iran. While several studies in Iran have applied Random Forest machine learning to predict the final status of COVID-19 patients, most of these studies were limited by sample size and the number of variables analyzed (7,8). In our study, 2,500 samples and approximately 30 variables were examined, making it one of the most comprehensive studies of its kind. The purpose of this research is to predict the survival of COVID-19 patients using the Random Forest and machine learning methods during the 2020 pandemic.

METHODS

STUDY POPULATION AND DATA COLLECTION

This retrospective cohort study was designed to evaluate the predictive performance of machine learning algorithms in forecasting COVID-19 patient outcomes using a dataset of 2,500 COVID-19 patients. These patients were admitted to three major hospitals affiliated with Shahid Beheshti University of Medical Sciences in Tehran, Iran, between 2020 and 2021. Data were collected from electronic health records, including demographic data (age, gender), clinical data (symptoms, ICU admission, length of hospital stay), laboratory data (white blood cell count [WBC], alanine aminotransferase [ALT], creatinine levels), and comorbidity information (pre-existing conditions such as hypertension, diabetes, cardiovascular diseases, and chronic respiratory illnesses). The inclusion criteria for the study were patients who were admitted to the hospital and had complete records of demographic, clinical, and laboratory variables. Patients who were discharged within 24 hours or had missing critical outcome data were excluded from the analysis to maintain data integrity and model robustness. The study focuses on using the Random Forest algorithm for prediction, with logistic regression and support vector machines (SVM) included as baseline models for comparison.

DATA PREPROCESSING

Missing values were handled using multiple imputation techniques to avoid bias that might arise from incomplete records. For continuous variables such as age, laboratory markers (e.g., ALT, WBC), missing values were imputed using multiple imputation methods. Specifically, we used predictive mean matching (PMM) to replace missing continuous values, which is less sensitive to distributional assumptions and often preferred in clinical datasets with skewed data. For categorical variables, such as comorbidities and gender, we employed mode imputation, assuming the most common value among the available data points for each category (9).

The dataset was normalized to ensure that all continuous variables were on a comparable scale, which is particularly important for algorithms like logistic regression and SVM. Additionally, one-hot encoding was applied to categorical variables (e.g., gender, presence of comorbidities) to convert them into numerical form suitable for analysis.

MODEL DEVELOPMENT AND VALIDITY

The primary machine learning model used in this study was the Random Forest algorithm, chosen for its ability to handle complex, high-dimensional data and its robustness to overfitting. The Random Forest model works by building an ensemble of decision trees, where each tree is trained on a random subset of the data and features. The Random Forest model was trained using key features, including age, gender, ICU admission status, laboratory markers (e.g., WBC, ALT), and comorbidities (e.g., hypertension, diabetes). For comparison, two traditional models—logistic regression and support vector machines (SVM)—were also trained on the same dataset. Logistic regression was chosen due to its widespread use in clinical outcome prediction, while SVM was included for its ability to perform well in high-dimensional spaces (10).

To assess model reliability, we employed both internal and external validation. For internal validation, we applied 10-fold cross-validation, where the dataset was divided into 10 equal parts, using nine subsets for training and one for testing iteratively across all subsets. This approach minimizes overfitting and offers a robust measure of model performance. For external validation, we simulated real-world scenarios by intentionally removing a portion of the data (20%) from the complete dataset, introducing missingness across demographic and clinical variables to mimic incomplete records. The models were retrained and evaluated on these modified datasets, focusing on maintaining stable performance metrics (AUC, sensitivity, specificity) despite data incompleteness. This dual-validation approach provides confidence in model generalizability and resilience to variations in data quality.

RESULTS

DEMOGRAPHICS AND CLINICAL CHARACTERISTICS

The study analyzed a total of 2,544 COVID-19 patients, with 22.6% in the mortality cohort and 77.4% in the survived cohort. The average age of patients in the mortality cohort was 69.8 years (± 15.1), significantly higher than the 57.8 years (± 16.8) observed in the survived cohort ($P < 0.001$). Male patients were predominant in both groups; however, a higher proportion of males (61.3%) was noted in the mortality cohort compared to 55.7% in the sur-

Table 1. Demographics and Clinical Characteristics of COVID-19 Patients

Variables	Mortality Cohort (22.6%)	Survived Cohort (77.4%)	P-Value
Age (years)	69.8 (15.1)	57.8 (16.8)	<0.001
Male (%)	61.3	55.7	0.018
Diabetes Mellitus (%)	37.8	29.2	<0.001
Ischemic Heart Disease (%)	23.8	13.8	<0.001
Cancer (%)	8.3	4.4	<0.001
Congestive Heart Failure (%)	3.3	1.9	0.03
Chronic Obstructive Pulmonary Disease (%)	1.4	1.6	0.46
Asthma (%)	2.3	3.2	0.15
Fever (%)	45.7	48.3	0.01
Weakness (%)	43.6	35.9	0.001
Cough (%)	44.1	53.6	<0.001
ICU Admission (%)	49.8	14.0	<0.001
Intubation (%)	37.0	1.5	<0.001
Platelet Count (x10^3/µL)	201.2 (96)	212.7 (91.9)	0.009
White Blood Cell Count (x10^3/µL)	10.57 (8.46)	8.08 (6.77)	<0.001
ALT (IU/L)	60.8 (8.02)	43.9 (6.01)	0.001
Creatinine (mg/dL)	108.0 (15.1)	150.0 (24.3)	0.001

vived group (P = 0.018). Comorbidities were more prevalent in the mortality cohort. Diabetes mellitus was present in 37.8% of the mortality cohort compared to 29.2% in the survived cohort (P < 0.001). Similarly, ischemic heart disease (IHD) and cancer were more common in the mortality group, with 23.8% and 8.3%, respectively, compared to 13.8% and 4.4% in the survived cohort (P < 0.001 for both). A notable proportion of patients in the mortality cohort required intensive care unit (ICU) admission (49.8%) and intubation (37.0%), significantly higher than 14.0% and 1.5%, respectively, in the survived cohort (P < 0.001 for both). Laboratory findings showed elevated ALT and creatinine levels in the mortality cohort, with ALT at 60.8 IU/L and creatinine at 108.0 mg/dL, both significantly higher than in the survived cohort (P < 0.001 for both) (**Table 1**).

LOGISTIC REGRESSION ANALYSIS FOR PREDICTIVE VARIABLES

Logistic regression analysis identified several variables as significant predictors of mortality. Age was a critical predictor, with an odds ratio (OR) of 1.05 (95% CI: 1.04–1.055, P < 0.001). Male sex was also a significant predictor (OR: 1.26, 95% CI: 1.04–1.52, P = 0.02). Other key predictors included diabetes mellitus (OR: 1.47, 95% CI: 1.21–1.79, P < 0.001), ischemic heart disease (OR: 1.95, 95% CI: 1.55–2.46, P < 0.001), and cancer (OR: 1.97, 95% CI: 1.36–2.83, P < 0.001). ICU admission (OR: 6.09, 95% CI: 4.95–7.49, P < 0.001) and intubation (OR: 39.23, 95% CI: 26.19–58.75, P < 0.001) were strongly associated with mortality. Among laboratory parameters, elevated ALT (OR: 1.002, 95% CI: 1.001–1.003, P = 0.001) and creatinine (OR: 1.002, 95% CI: 1.001–1.003, P < 0.001) were significant predictors (**Table 2**).

VALIDATION OF MACHINE LEARNING MODELS

Table 2. Logistic Regression Analysis for Predictive Variables

Variables	OR	Lower 95% CI	Upper 95% CI	P-Value
Age*	1.05	1.04	1.055	<0.001
Male*	1.26	1.04	1.52	0.02
Diabetes Mellitus*	1.47	1.21	1.79	<0.001
Ischemic Heart Disease*	1.95	1.55	2.46	<0.001
Cancer*	1.97	1.36	2.83	<0.001
Congestive Heart Failure*	1.78	1.01	3.12	0.04
Chronic Obstructive Pulmonary Disease	0.88	0.40	1.92	0.75
Fever*	0.79	0.66	0.96	0.02
Weakness*	1.38	1.14	1.66	0.001
Muscle Pain*	0.46	0.37	0.58	<0.001
Cough*	0.68	0.57	0.82	<0.001
ICU Admission*	6.09	4.95	7.49	<0.001
Intubation*	39.23	26.19	58.75	<0.001
ALT*	1.002	1.001	1.003	0.001
Creatinine*	1.002	1.001	1.003	<0.001

The performance of three machine learning models—Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR)—was evaluated in predicting COVID-19 mortality. Random Forest exhibited the highest area under the curve (AUC) at 83.3%, followed by SVM (80.0%) and Logistic Regression (79.1%). In terms of sensitivity, Random Forest again outperformed the other models with 63.0%, compared to 37.6% for SVM and 41.1% for Logistic Regression. Specificity was highest in SVM (93.4%), while Random Forest and Logistic Regression showed comparable specificity (90.5% and 90.3%, respectively). Positive predictive values were similar across the models, with Random Forest at 57.2%, SVM at 62.9%, and Logistic Regression at 57.4% (**Table 3**).

Table 3. Validation of Machine Learning Models for Predicting Mortality

Model	AUC (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)
Random Forest (RF)	83.3	63.0	90.5	57.2
Support Vector Machine (SVM)	80.0	37.6	93.4	62.9
Logistic Regression (LR)	79.1	41.1	90.3	57.4

IMPORTANCE OF VARIABLES IN PREDICTING MORTALITY

The importance of various clinical and laboratory variables in predicting mortality was assessed using the Gini index. The most important predictors were elevated ALT levels (Gini index: 41.51), white blood cell (WBC) count (23.35), and platelet count (21.02). Muscle pain, weakness, and dyspnea were also identified as significant clinical predictors of mortality, with Gini indices of 9.15, 8.63, and 7.64, respectively. Among comorbidities, ischemic heart disease (Gini index: 7.82) and congestive heart failure (Gini index: 5.96) were notably influential in predicting mortality **(Table 4)**.

Table 4. Importance of Variables in Predicting Mortality Using Gini Index

Variables	Gini Index
ALT (IU/L)	41.50693
White Blood Cell Count (x10^3/µL)	23.35322
Platelet Count (x10^3/µL)	21.01522
Creatinine (mg/dL)	20.13921
Vomiting	7.434970
Fever	7.364482
Cough	6.673888
Headache	6.422400
Anosmia	6.113344
Sore Throat	5.988557
Congestive Heart Failure	5.9589897
Chronic Obstructive Pulmonary Disease	5.7143452
Diabetes Mellitus	4.3607490

DISCUSSION

In this study, 22.6% of patients admitted for COVID-19 had succumbed to the disease at the time of admission, while the remaining patients were discharged following recovery. The analysis demonstrates that all three machine learning models—Logistic Regression, Support Vector Machine (SVM), and Random Forest—exhibited considerable accuracy in predicting COVID-19 mortality based on input variables. Notably, despite marginal differences in overall performance, the Random Forest model achieved superior predictive accuracy, particularly when evaluated using sensitivity (the model's ability to correctly identify actual deaths).

The assessment of variable importance revealed that clinical symptoms such as muscle pain, weakness, fatigue, shortness of breath, nausea, and fever, along with underlying conditions including cardiovascular disease, chronic respiratory disease, and diabetes, were significant predictors of mortality. Furthermore, elevated levels of alanine aminotransferase (ALT), white blood cells, platelets, and creatinine were also associated with an increased risk of death in COVID-19 patients.

Our findings align with those reported by Molaei et al. (2022), who investigated predictors of COVID-19 mortality in Iran using machine learning. Their analysis, based on a dataset of 1,500 patients (1,386 survivors and 144 deceased), highlighted shortness of breath, ICU admission, and oxygen therapy as the top predictors of death. While their study identified smoking, ALT levels, and platelet counts as among the least predictive variables, our results indicate that ALT levels and platelet counts were more critical in predicting mortality. This discrepancy may be attributed to the differences in the number and selection of variables examined between the two studies (7).

Similarly, Talebi et al. (2022) employed a decision tree-based Random Forest model in a study involving 1,078 COVID-19 patients. They achieved 90% accuracy in predicting mortality using only CT lung scan data, without incorporating clinical or demographic biomarkers. Despite this methodological difference, our findings corroborate the effectiveness of the Random Forest model, even when only clinical biomarkers are used (8).

The role of laboratory biomarkers in predicting mortality was further emphasized in a study by Moslehi et al. (2022), which analyzed data from 2,740 COVID-19 patients in Hamadan, 25% of whom died. They identified biomarkers such as BUN, PTT, and ESR as significant predictors. While our study also highlights the importance of laboratory biomarkers, the number of variables we could include was limited by incomplete patient records. This may

explain the variation in biomarker significance between the two studies (11).

Elevated ALT was the most significant predictor in our model (according to Gini index), suggesting that liver dysfunction is closely associated with COVID-19 mortality. This is consistent with findings from Cai et al. (2020), which observed that elevated liver enzymes, including ALT, were strongly linked to poorer outcomes in COVID-19 patients (12). Also, Herold et al. (2020) demonstrated that increased WBC count, particularly in the form of neutrophilia, is often associated with severe inflammation and cytokine storm, which are known to worsen COVID-19 prognosis that is in consistent to our finding for WBC as the second most important predictor of mortality (13). Platelet Count emerged as another important factor. Thrombocytopenia (low platelet count) has been frequently reported in severe COVID-19 cases, as shown in studies by Lippi et al. (2020) (14). Also, elevated creatinine underscores the impact of renal function on COVID-19 outcomes. Hirsch et al. (2020) have highlighted that renal impairment, as indicated by elevated creatinine, significantly increases the risk of mortality, likely due to the critical nature of kidney function in maintaining overall homeostasis and managing the body's response to infection (15).

Gupta et al. (2021) applied Random Forest, SVM, decision trees, and neural networks to predict COVID-19 outcomes in India. Their study, like ours, found that Random Forest outperformed the other models, further validating the robustness of this method for predicting COVID-19 mortality across different populations (16).

Watson et al. (2020) combined a Nonlinear Bayesian Hybrid Model with a Random Forest algorithm to predict COVID-19 infections and deaths in the United States. Their integrated model achieved a mean error of 0.40 for infections and 0.32 for deaths, demonstrating the utility of Random Forest for mortality prediction over a two-week evaluation period. The consistency between their findings and ours underscores the effectiveness of Random Forest in estimating pandemic mortality (17).

In another study by Wang et al. (2020) in Wuhan, China, clinical data from 119 COVID-19 patients revealed that elevated myoglobin and LDH levels were strongly associated with increased mortality risk. Our results similarly emphasize the predictive value of laboratory parameters, although differences in sample size and population diversity may account for the variations in reported predictive biomarkers (18).

Finally, Mohapatra et al. (2020) employed various machine learning models, including Random Forest, to predict COVID-19 mortality and future positive cases based on data from China and India. Their findings, consistent with ours, support the precision and reliability of Random Forest models in predicting mortality across different datasets and geographic regions (19).

This study has several limitations that may impact the generalizability and interpretation of the results. First, the data were collected retrospectively from electronic health records (EHR), which can introduce issues related to incomplete or inconsistent documentation. Although we applied multiple imputation techniques to handle missing values, any residual data inaccuracies could affect model performance. Second, the lack of longitudinal data limits our ability to evaluate the model's effectiveness in predicting mortality outcomes over an extended period. This is particularly relevant for evolving diseases like COVID-19, where patient outcomes can vary significantly based on treatment advancements and emerging variants. Our models may thus be less accurate when applied to later phases of the pandemic or in predicting long-term outcomes. Another limitation is that we could not include some potentially influential biomarkers (e.g., interleukin-6 or D-dimer levels) due to incomplete availability across all patients and finally, while we employed simulated missing data for external validation, real-world applications often encounter complex patterns of missing data that may not be fully represented by our simulations. Thus, while the models demonstrated robustness to controlled missingness, real-world performance could differ in settings with more complex missing data structures.

CONCLUSION

In conclusion, this study demonstrates that the Random Forest algorithm offers higher accuracy and sensitivity in predicting the survival of COVID-19 patients compared to other machine learning approaches. Key predictors of mortality include cardiovascular disease, diabetes, respiratory conditions, and specific clinical and laboratory markers. These results suggest that Random Forest models can be instrumental in identifying high-risk patients and improving clinical decision-making during pandemics.

The Random Forest model's predictive strength and interpretability have significant practical implications in clinical settings, particularly in triaging and managing COVID-19 patients. With its ability to identify key predictors of mortality such as age, comorbidities, and certain laboratory markers, the model provides clinicians with a reliable tool to assess patient risk upon admission. In high-demand healthcare environments where resources like ICU beds and ventilators are limited, using the Random Forest model to stratify patients by mortality risk could inform early intervention decisions and optimize resource allocation. Overall, the Random Forest model offers a promising approach to patient stratification in clinical settings, supporting rapid, data-driven decisions. Further research on its real-time application and periodic updates of model training to reflect evolving COVID-19 profiles would strengthen its clinical relevance and adaptability.

CONFLICT OF INTERESTS

There are no conflicts of interest in terms of the present manuscript.

ABBREVIATIONS

SVM, Support Vector Machine; WBC, white blood cell count; ALT, alanine aminotransferase; PMM, predictive mean matching; ICU, intensive care unit.

REFERENCES

1. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. Front Public Health. 2020;8:357.
2. Kenny G, Mallon PW. COVID-19 clinical presentation and therapeutic

considerations. Biochem Biophys Res Commun. 2021;538:125-31.
3. Fogarty H, Townsend L, Ni Cheallaigh C, Bergin C, Martin-Loeches I, Browne P, et al. COVID-19 coagulopathy in Caucasian patients. Br J Haematol. 2020;189(6):1044-9.
4. Albahri AS, Hamid RA, Alwan JK, Al-Qays Z, Zaidan A, Zaidan B, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. J Med Syst. 2020;44:1-11.
5. Rasheed J, Jamil A, Hameed AA, Al-Turjman F, Rasheed A. COVID-19 in the age of artificial intelligence: a comprehensive review Interdiscip Sci. 2021;13:153-75.
6. Cobb JS, Seale MA. Examining the effect of social distancing on the compound growth rate of COVID-19 at the county level (United States) using statistical analyses and a random forest machine learning model. Public Health. 2020;185:27-9.
7. Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting COVID-19 mortality. BMC Med Inform Decis Mak. 2022;22(1):2.
8. Talebi A, Borumandnia N, Jafari R, Pourhoseingholi MA, Jafari NJ, Ashtari S, et al. Predicting the COVID-19 patients' status using chest CT scan findings: A risk assessment model based on Decision tree. Adv Exp Med Biol. 2023;1412:237-50.
9. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. BMJ. 2020;7;369.
10. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis. 2020;34:101623.
11. Moslehi S, Rabiei N, Soltanian AR, Mamani M. Application of machine learning models based on decision trees in classifying the factors affecting mortality of COVID-19 patients in Hamadan, Iran. BMC Med Inform Decis Mak. 2022;22(1):192.
12. Cai Q, Huang D, Yu H, Zhu Z, Xia Z, Su Y, et al. J Hepatol. COVID-19: Abnormal liver function tests. 2020 Sep;73(3):566-74.
13. Herold T, Jurinovic V, Arnreich C, Lipworth BJ, Hellmuth JC, von Bergwelt-Baildon M, et al. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. J Allergy Clin Immunol. 2020 Jul;146(1):128-36.
14. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta. 2020 Jul;506:145-8.
15. Hirsch JS, Ng JH, Ross DW, Sharma P, Shah HH, Barnett RL, et al. Acute kidney injury in patients hospitalized with COVID-19. Kidney Int. 2020 Jul;98(1):209-18.
16. Gupta VK, Gupta A, Kumar D, Sardana A. Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. Big Data Min Anal. 2021;4(2):116-23.
17. Watson GL, Xiong D, Zhang L, Zoller JA, Shamshoian J, Sundin P, et al. Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model. PLoS Comput Biol. 2021;17(3):e1008837.
18. Wang J, Yu H, Hua Q, Jing S, Liu Z, Peng X, et al. A descriptive study of random forest algorithm for predicting COVID-19 patients outcome. PeerJ. 2020;8:e9945.
19. Majhi R, Thangeda R, Sugasi RP, Kumar N. Analysis and prediction of COVID-19 trajectory: A machine learning approach. J Public Aff. 2021;21(4):e2537.

Author Contribution: All authors have contributed in this study.
Funding statement: This research has received no funding support.