

# Prediction of mortality risk and duration of hospitalization of COVID-19 patients with chronic comorbidities based on machine learning algorithms

DIGITAL HEALTH  
Volume 9: 1–12  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076231170493  
journals.sagepub.com/home/dhj



Parastoo Amiri<sup>1</sup> , Mahdieh Montazeri<sup>2</sup> , Fahimeh Ghasemian<sup>3</sup> ,  
Fateme Asadi<sup>4</sup>, Saeed Niksaz<sup>3</sup>, Farhad Sarafzadeh<sup>5</sup> and Reza Khajouei<sup>2</sup>

## Abstract

**Background:** The severity of coronavirus (COVID-19) in patients with chronic comorbidities is much higher than in other patients, which can lead to their death. Machine learning (ML) algorithms as a potential solution for rapid and early clinical evaluation of the severity of the disease can help in allocating and prioritizing resources to reduce mortality.

**Objective:** The objective of this study was to predict the mortality risk and length of stay (LoS) of patients with COVID-19 and history of chronic comorbidities using ML algorithms.

**Methods:** This retrospective study was conducted by reviewing the medical records of COVID-19 patients with a history of chronic comorbidities from March 2020 to January 2021 in Afzalipour Hospital in Kerman, Iran. The outcome of patients, hospitalization was recorded as discharge or death. The filtering technique used to score the features and well-known ML algorithms were applied to predict the risk of mortality and LoS of patients. Ensemble Learning methods is also used. To evaluate the performance of the models, different measures including F1, precision, recall, and accuracy were calculated. The TRIPOD guideline assessed transparent reporting.

**Results:** This study was performed on 1291 patients, including 900 alive and 391 dead patients. Shortness of breath (53.6%), fever (30.1%), and cough (25.3%) were the three most common symptoms in patients. Diabetes mellitus (DM) (31.3%), hypertension (HTN) (27.3%), and ischemic heart disease (IHD) (14.2%) were the three most common chronic comorbidities of patients. Twenty-six important factors were extracted from each patient's record. Gradient boosting model with 84.15% accuracy was the best model for predicting mortality risk and multilayer perceptron (MLP) with rectified linear unit function (MSE = 38.96) was the best model for predicting the LoS. The most common chronic comorbidities among these patients were DM (31.3%), HTN (27.3%), and IHD (14.2%). The most important factors in predicting the risk of mortality were hyperlipidemia, diabetes, asthma, and cancer, and in predicting LoS was shortness of breath.

**Conclusion:** The results of this study showed that the use of ML algorithms can be a good tool to predict the risk of mortality and LoS of patients with COVID-19 and chronic comorbidities based on physiological conditions, symptoms, and demographic information of patients. The Gradient boosting and MLP algorithms can quickly identify patients at risk of death or long-term hospitalization and notify physicians to do appropriate interventions.

<sup>1</sup>Student Research Committee, Kerman University of Medical Sciences, Kerman, Iran

<sup>2</sup>Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran

<sup>3</sup>Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

<sup>4</sup>Student Research Committee, School of Management and Medical Information, Kerman University of Medical Sciences, Kerman, Iran

<sup>5</sup>Infectious and Internal Medicine Department, Afzalipour Hospital, Kerman University of Medical Sciences, Kerman, Iran

## Corresponding author:

Reza Khajouei, Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Haft-bagh Highway, Kerman 7616911313, Iran.

Email: r.khajouei@kmu.ac.ir; r.khajouei@yahoo.com



## Keywords

COVID-19, chronic comorbidities, machine learning, mortality, hospital length of stay, prediction

Submission date: 28 November 2022; Acceptance date: 31 March 2023

## Introduction

In December 2019, coronavirus disease (COVID-19) emerged in China and spread rapidly around the world.<sup>1</sup> This virus is highly contagious in humans.<sup>2</sup> In January 2020, the World Health Organization announced the outbreak of COVID-19 at the Public Health Emergency of International Concern.<sup>3</sup> In October 2020, nearly 40 million people in more than 180 countries became infected and more than one million died.<sup>4</sup> Data from the early months of the outbreak showed that COVID-19 is more common in patients with chronic comorbidities such as cardiovascular disease, kidney disease, type 2 diabetes, hypertension, and cancer.<sup>5,6</sup> A retrospective study in China also showed that out of 138 patients with COVID-19, 64 (46.4%) had one or more chronic comorbidities.<sup>7</sup> Evidence shows that the outbreak of the COVID-19 virus has been a serious threat to patients with chronic comorbidities because these patients have a more severe form of respiratory problems than other COVID-19 patients, as it can increase their length of stay (LoS) in hospitals or even their death rate.<sup>8,9</sup>

Numerous studies have used machine learning (ML) algorithms to predict survival and calculate the LoS of patients with chronic comorbidities.<sup>10–12</sup> Survival can be defined as the time interval between the diagnosis of the disease and the death of the patient.<sup>13</sup> Factors such as chronic comorbidities and pandemics can reduce patient survival. The LoS is one of the criteria for measuring the utilization of a hospital<sup>14</sup> because the reduction of patients' LoS will lead to the optimal use of medical resources available in the hospital, such as hospital beds, staff, etc.<sup>15</sup>

So far, ML as a subset of artificial intelligence techniques has been successful in predicting the rapid recovery of many chronic comorbidities such as diabetes<sup>16,17</sup> and cancer<sup>18</sup> and reducing the LoS<sup>19</sup> and mortality from cardiovascular disease.<sup>20</sup> Accurate prediction of the mortality risk and reducing the LoS of patients reduce the pressure on healthcare systems and support medical decisions. Recently, several studies have used ML algorithms to predict the risk of mortality<sup>21–24</sup> and calculate the LoS<sup>12,25–27</sup> in patients with COVID-19. Wang et al.<sup>28</sup> used two ML models based on clinical and laboratory

features to predict the mortality risk of COVID-19 patients. In South Korea, a study demonstrated Lasso and linear support vector machine (SVM) had high sensitivities and specificities to predict the mortality risk of COVID-19 patients.<sup>29</sup> In addition, Jimenez-Solem et al.<sup>30</sup> used data from patients with COVID-19 in Denmark and the United Kingdom to develop mortality prediction models for COVID-19 patients. To predict the LoS, Ebinger et al.<sup>12</sup> trained three ML models with an accuracy of 0.765, which were based on the analysis of electronic health records of 966 COVID-19 patients in a large educational and medical center in the United States. In Saudi Arabia, a study predicted the LoS of COVID-19 patients in intensive care unit (ICU) with the highest accuracy (94.16%) using the Random Forest model.<sup>27</sup> A study in Iran showed that among the seven ML techniques, the SVM algorithm with an average accuracy of 99.5%, average specificity of 99.7%, and average sensitivity of 99.4% had the best performance on the laboratory data of 1225 COVID-19 patients.<sup>31</sup> A systematic study reported that variables such as age, gender, and chronic comorbidities including hypertension and diabetes played an important role in increasing the risk of death and LoS of patients with COVID-19.<sup>32</sup> Also, several studies have shown a higher risk of death, and LoS in COVID-19 patients with a history of chronic comorbidities such as hypertension, diabetes, and acute respiratory distress syndrome than others.<sup>33–40</sup> Although these studies have yielded interesting results, some of them used standard biostatistics methods for their calculations, leaving room for ML approaches.<sup>34–36</sup> In addition, some studies just included a specific type of inpatients, such as patients in ICU.<sup>37–39</sup> The objective of this study was to predict the risk of mortality and LoS of COVID-19 patients with a history of any chronic comorbidities by comparing the performance of selected ML algorithms and identifying the most important clinical variables. In general, we seek to answer the following questions concerning COVID-19 in patients with a history of any chronic comorbidities:

1. What are the best ML algorithms for predicting their risk of mortality?

2. What are the best ML algorithms for predicting their LoS?
3. What are the most important clinical variables for predicting their risk of mortality?
4. What are the most important clinical variables for predicting their LoS?

We present this article in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) reporting checklist (Supplemental Appendix 1).

## Methods

This retrospective study was conducted on all patients with COVID-19 who visited Afzalipour Hospital (the main treatment center for COVID-19 patients in Kerman, the largest province in Iran) from March 2020 to January 2021. Only patients whose real-time polymerase chain reaction test was positive for COVID-19 acute respiratory syndrome and had at least one chronic comorbidity were included in this study. Patients under 18 years of age were excluded from the study because they are admitted to a children's hospital. Pregnant women were also excluded from the study. These patients should be included in the scope of pregnancy exploration. The names of the variables were extracted from the patient medical records.

Based on the review of various studies<sup>32,33,41,42</sup> and the approval of two infectiologist, factors essential for predicting mortality and LoS of patients with COVID-19 with chronic comorbidities were identified. Data of all patients were extracted from the hospital information system and

based on their Electronic Health Records (EHRs). Since not all the data were recorded in patients' EHRs, the rest of the required data were collected by reviewing paper records. Demographic information, chronic comorbidities, patients' symptoms at admission, as well as their discharge status (alive/dead) and LoS were collected. The data were entered into an Excel sheet. Two output variables were considered. The first variable indicates the patient health status at discharge. The values for this variable were 0 for dead and 1 for alive status. Patients that were discharged in stable condition and without any symptoms were regarded as improved patients. Patient mortality was not considered after their discharge and out of the hospital. The second variable was for the LoS of patients in the hospital. Figure 1 shows the flowchart of the patient selection stage. Out of 1538 patients with COVID-19, 1291 were included in our analysis.

In this study, only clinical information recorded in the patient medical records was used and the identity information of patients was not used. Therefore, waiver of consent report.

## Data preparation

For the data preprocessing stage, first, missing data in Excel were identified and deleted. For data normalization, the Standard Scaler method was used. In this method,  $\mu$  represents the mean and  $s$  represents the standard deviation. As a result, the standard value of sample  $x$  is obtained from the following equation:

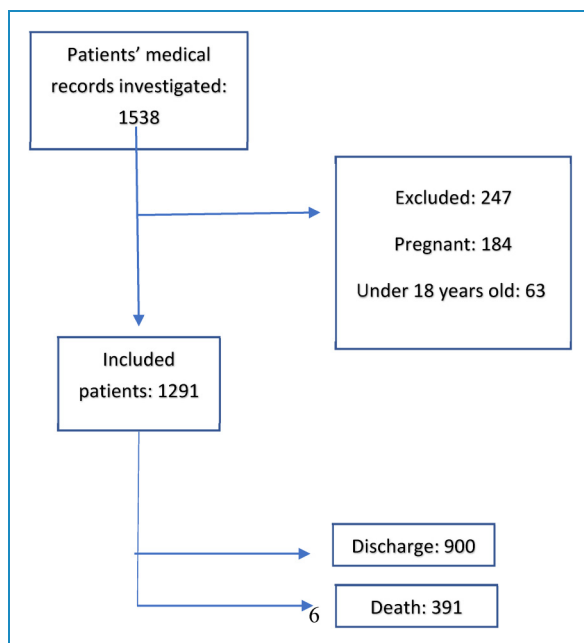
$$Z = (x - \mu)/s$$

Our dataset includes 26 features. The degree of importance and priority of these features were determined using filtering methods. To have an accurate and unbiased model, we made sure that our data set was normal. Also, the data samples in the training set were randomly selected and were completely separated from the test data.

## Predictive analytics algorithms

Various studies were reviewed to select the best ML algorithms to predict mortality and LoS of COVID-19 patients with a history of chronic comorbidities.<sup>25,39,43</sup> Since the performance of each ML algorithm depends on data structure and type of work,<sup>44</sup> the performances of several algorithms were evaluated using different criteria. In this study, Random Forest, Multilayer Perceptron, K-Nearest Neighbor (KNN), AdaBoost, Naïve Bayes, and SVM algorithms were used for the prediction of mortality. Multilayer perceptron (MLP), ElasticNet, support vector regression (SVR) and Lasso, and Ridge algorithms were also used for the prediction of LoS.

We configured the random forest algorithm<sup>45</sup> with 10, 50, and 100 trees in the forest. MLP was used to create a



**Figure 1.** Flowchart for selecting patients to participate in the study. ROC: receiver operating characteristic.

neural network model. The effective factors in predicting mortality were the inputs of neural network ( $n = 31$ ) and its output was the target variable (mortality and LoS). For KNNs 1, 3, 5, 10, 15, 30, and 50 neighbors were used. For random forest analysis, bagging with 100 iterations and base learner was used.

We also used ensemble learning methods. Ensemble learning is the process in which two or more ML model are combined to get better results and improve robustness over a single estimator. There are different approaches to ensemble. These methods decrease the variance of a base estimator and minimize the overfitting of data.<sup>46</sup>

In this study, we calculated the overall accuracy to compare ML algorithms. In addition, receiver operating characteristic (ROC) was generated for each production algorithm and their area under curve (AUC) and confusion matrix were calculated. ROC is often used to determine the strength of a model. In medicine, ROC is used to evaluate the precision of diagnostic tests. The precision of 0 to 0.5 indicates random classification and 0.5 to 1 indicates the overall recognition ability of the model.<sup>47</sup> We also ensured that there was no interference between the training and test datasets at any level.

### Statistical analysis and performance evaluation

The development of a model for predicting mortality and LoS for COVID-19 patients with chronic comorbidities was performed based on Python Scikit-learn package version 3.8. For performance evaluation, 70% of the data were considered for training and 30% for testing. The efficiency of patient mortality risk prediction models was evaluated by calculating AUC, ROC, precision, specificity, accuracy, F1 score, and recall. These criteria are defined and calculated using the confusion matrix components (Table 1).

Assume that the number of positive examples (number of alive patients) and negative (number of dead patients) are P and N, respectively, the following definitions can be given:

**Table 1.** Confusion matrix.

		Prediction value	
		Alive (-)	Dead (+)
Output	Alive (-)	TN	FP
	Dead (+)	FN	TP

FN: dead people incorrectly identified as alive; FP: alive people incorrectly identified as dead; TN: alive people correctly identified as dead; TP: dead people correctly diagnosed as dead.

FP = Alive people incorrectly identified as dead  
 TP = Dead people correctly diagnosed as dead  
 TN = Alive people correctly identified as dead  
 FN = Dead people incorrectly identified as alive

$$\text{True positive rate} = \frac{TP}{P}$$

$$\text{False positive rate} = \frac{FP}{N}$$

$$\text{Classification accuracy} = \frac{TP + TN}{P + N}$$

Accuracy refers to the number of alive and dead people who have been correctly diagnosed as alive or dead.<sup>48</sup> Precision refers to the number of people who have died and the model has correctly identified them as dead. Sensitivity refers to people who have died, and the model has correctly identified them as dead. Thus, more sensitivity indicates a more accurate diagnosis of the number of dead. Specificity refers to the proportion of people who are alive, and the model correctly identifies them as alive.<sup>49</sup> Mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) were used to compare the models for predicting the LoS of patients in the hospital.

### Ethical considerations

This study was approved by the ethics committee of Kerman University of Medical Sciences (code of ethics: IR.KMU.REC.1400.055). To protect the privacy and confidentiality of the patients, we concealed the unique identifying information of all the patients in the data collection process.

## Results

### Patients' characteristics

The dataset contained 1291 patient records including 900 (69.6%) alive and 391 (30.3%) dead patients (Figure 1). The mean ages of dead and alive patients were 66.2 and 53.9 years, respectively. Also, 54.6% of patients were male.

Shortness of breath (53.6%), fever (30.1%), and cough (25.3%) were the three most common symptoms in patients. Diabetes mellitus (31.3%), hypertension (27.3%), and ischemic heart disease (IHD) (14.2%) were the three most common chronic comorbidities of patients.

In total, based on the studies and experts' opinions, 26 characteristics were specified to predict the risk of mortality and LoS of COVID-19 patients with a history of chronic comorbidities (Table 2). The degree of importance of each feature in predicting mortality risk and LoS is shown in Table 2. The most effective chronic comorbidity for

**Table 2.** Risk factors of mortality and LoS of COVID-19 patients with chronic disease.

Factors		Alive (%)	Dead (%)	Degree of importance in predicting mortality	Degree of importance in predicting LoS	
Demographic information	Age	0-30	6.8	2.1	2	9
		31-40	12.6	5.2		
		41-50	15.3	8.6		
		51-60	19.8	14.6		
		61-70	25.2	24.2		
		>70	20.3	45.3		
	Sex	Male	53.1	55.9	6	17
		Female	46.9	44.1		
	Marital status	Single	13.6	5.4	20	23
		Married	86.4	89.1		
Unknown		0	5.5			
Chronic comorbidities	Hypertension	21.7	42.6	3	7	
	HLP	2.2	7.6	25	26	
	IHD	11.4	22.0	21	18	
	DM	32.4	32.5	19	3	
	COPD	15.8	4.3	5	4	
	ESRD	1.5	4.0	17	5	
	CKD	5.4	4.7	4	14	
	Asthma	4.1	1	18	1	
	TSH	2.8	4.0	24	20	
	Cancer	2.7	7.3	22	6	
Symptoms	Fever	24.3	43.8	23	25	
	Feeling fatigued	12.3	29	13	22	
	Sore throat	1.5	2.1	26	2	
	Cough	19.6	40.3	7	24	
	Chest pain	3.1	3.5	15	8	
	Shortness of breath	50.0	67.5	1	11	

(continued)

Table 2. Continued.

Factors	Alive (%)	Dead (%)	Degree of importance in predicting mortality	Degree of importance in predicting LoS
Dizziness	0.4	0.4	12	19
Confusion	0.4	13.3	8	10
Headache	4.3	5.5	14	15
Muscle pain	8.9	28.0	9	12
Vomiting	3.4	13.1	11	13
Diarrhea	0.0	11.0	16	16
Abdominal pain	0.1	4.7	10	21

CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; DM: diabetes mellitus; ESRD: end stage renal disease; HLP: hyperlipidemia; IHD: ischemic heart disease; LoS: length of stay; TSH: hypothyroidism.

Table 3. The performance of Naïve Bayes and five machine learning models in mortality prediction.

No.	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC area
1	Naïve Bayes	63.92	<b>86.11</b>	26.27	40.25	0.77
2	KNeighborsClassifier-K_15	72.54	77.27	57.62	66.02	0.78
3	SVM	<b>80.00</b>	77.68	<b>79.66</b>	<b>78.66</b>	<b>0.85</b>
4	MLP hidden 512*256 activation logistic	79.61	77.50	78.81	78.15	0.84

\*In each column, the best result is shown in bold.

MLP: multilayer perceptron; ROC: receiver operating characteristic; SVM: COVID.

predicting mortality is hypertension, and for predicting LOS is asthma. Shortness of breath and sore throat are the most important symptoms for predicting mortality and LoS, respectively.

Prediction of mortality requires the use of classification methods, whereas the increased LoS is a regression problem. There are several reasons that, compared to traditional methods, ML models can predict better. For instance, ML models can make predictions based on a much larger data set than traditional methods. On the other hand, ML is not biased by human emotions or subjective opinions. Furthermore, ML models can adapt to changes quickly. Finally, they can identify patterns that are too complex.

Because the number of null values in the feature is not above a certain threshold (5% of total data), we adopt RandomOverSampler for balancing data. It over-samples the minority class(es) by picking samples at random with replacement.

To predict the mortality risk of COVID-19 patients with a history of chronic comorbidities, simple Bayesian and five

ML algorithms were implemented. A comparison of their results in terms of specificity, sensitivity, accuracy, and ROC curve is shown in Table 3. The results of this table show that SVM is the best model for predicting mortality risk based on most evaluation metrics. The simple Bayesian model is the weakest model based on the F1-score metric.

In Table 4, you can see the prediction based on ensemble learning methods. As observed, Gradient boosting achieved the best results compared to other methods.

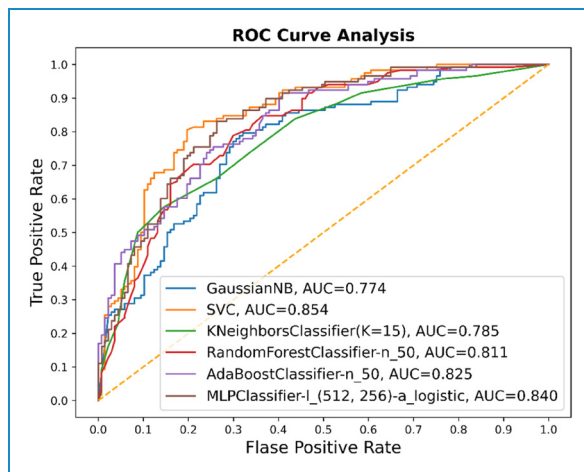
The performance of the selected models based on the ROC curve and confusion matrix is shown in Figures 2,3,4,5, respectively. The possible activation functions that can be used in Neural Networks are Sigmoid, Tanh(x), and rectified linear unit (ReLU) functions. According to the results of Table 3, the average accuracy was about 74.11%. This means that the prediction of about 74 out of every 100 data items given to the network is correct.

For the MLP method, instead of using a pretrained model, we built our model and trained it from scratch

**Table 4.** The performance of ensemble learning methods.

No.	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC area
4	Trees Random Forest (n = 100)	83.02	82.66	83.02	82.43	0.77
5	AdaBoost	82.26	81.89	82.26	81.97	0.78
6	Gradient Boosting(n = 50)	<b>84.15</b>	<b>83.86</b>	<b>84.15</b>	<b>83.64</b>	<b>0.79</b>
7	Hist Gradient Boosting(max_iter = 50)	83.40	83.05	83.40	82.86	0.78
10	Random Forest with Halving Grid Search	82.26	82.41	82.26	81.04	0.74

\*In each column, the best result is shown in bold. ROC: receiver operating characteristic.



**Figure 2.** Performance of mortality risk prediction models based on ROC curve. AUC: area under curve; ROC: receiver operating characteristic.

with our data. We adopted 100, 10, and 0.5 for epoch number, batch size, and dropout, respectively. We used the EarlyStopping function in the Keras library which monitors the accurate and loss values. If the loss is being monitored, training comes to halt when there is an increment observed in the loss values.

Model explanation is also a necessity in the perception of aggregated ML models. We adopt Local interpretable model-agnostic explanations (LIME) for this purpose to illustrate explanations for any single patient. For example, model explanation is depicted for patient #1 in Figure 6. The colors show the associations between features and the prediction. The colors blue and orange, depicting negative and positive associations, respectively. For instance, in Gradient boosting, diarrhea has positive association (orange), but chest tightness has negative association (blue). These values may differ for other patients.

The results of comparing the lowest error rate of ML algorithms in predicting the LoS of COVID-19 patients with comorbidities are shown in Table 5. According to

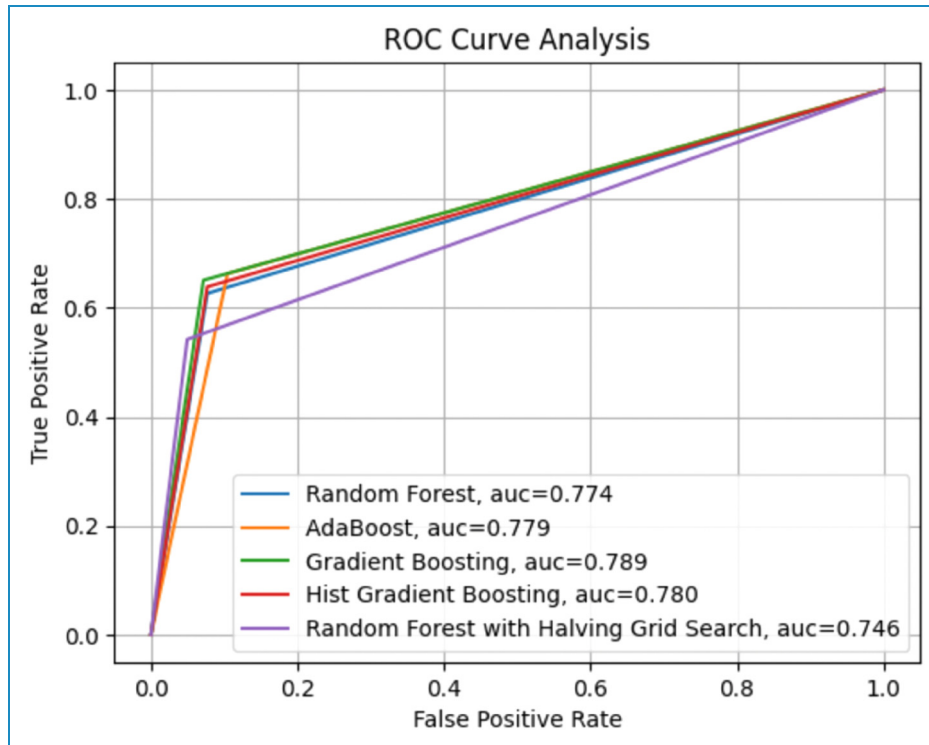
the results of this table, MLP (32\*1024\*32) with ReLU activity function is the best model for predicting the LoS of patients based on the considered metrics.

## Discussion

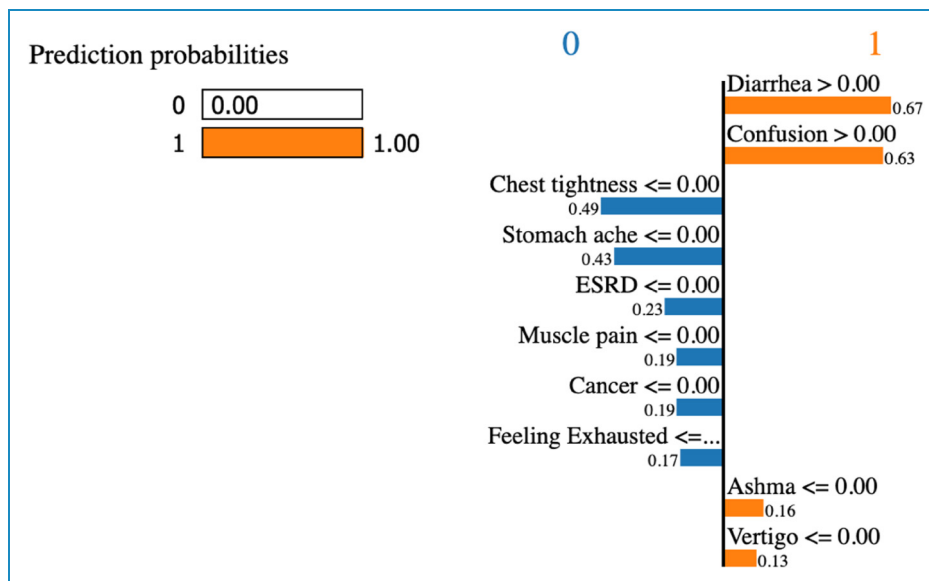
In the event of an outbreak, the prediction of mortality and LoS of patients is inevitable for resource management in healthcare facilities. In the present study, several ML algorithms were developed to predict the risk of mortality and LoS using demographic indicators, clinical symptoms, and chronic comorbidities of patients with COVID-19. This study was conducted on 900 alive and 391 dead patients. Based on the results, SVM and MLP algorithms with ReLU activation function had the best performance in predicting mortality risk and LoS of COVID-19 patients with chronic comorbidities, respectively.

Numerous studies have so far developed mortality prediction models for patients with COVID-19<sup>29,50,51</sup> but they did not focus specifically on patients with comorbidities. Among the six ML algorithms used in this study, SVM with an accuracy of 80% and ROC of 0.85 performed better than other models in predicting the risk of mortality in patients. These findings were consistent with the results of other studies. For example, in a study by Agieb et al.,<sup>52</sup> SVM was the most successful model in predicting mortality. Similarly, another study by Booth et al.<sup>53</sup> reported that SVM is the best model with 91% sensitivity and specificity.

Numerous studies have also developed models for predicting the LoS of patients with COVID-19<sup>38,54,55</sup> but they did not include chronic comorbidities patients or target a specific type of disease. In the present study, among the five ML algorithms used to predict the LoS of COVID-19 patients with chronic comorbidities, MLP with the ReLU activation function had the best performance (MAE = 0.434, RMSE = 0.624, and MSE = 0.389). These findings confirmed the results of other studies. For example, Bacchi et al.<sup>56</sup> reported that MLP is the best model for predicting the LoS of COVID-19 patients with the highest accuracy (MAE = 0.246, RMSE = 0.369, and



**Figure 5.** Performance of mortality risk prediction ensembling models based on ROC curve. AUC: area under curve; ROC: receiver operating characteristic.



**Figure 6.** LIME feature plot states the effect of each variable on the classification.

AUC = 0.864). In another similar study by Kulkarni et al.,<sup>57</sup> an MLP-based model predicted the LoS of COVID-19 patients with 90.87% accuracy.

On the other hand, several studies have investigated the role of chronic comorbidities in predicting outcomes of

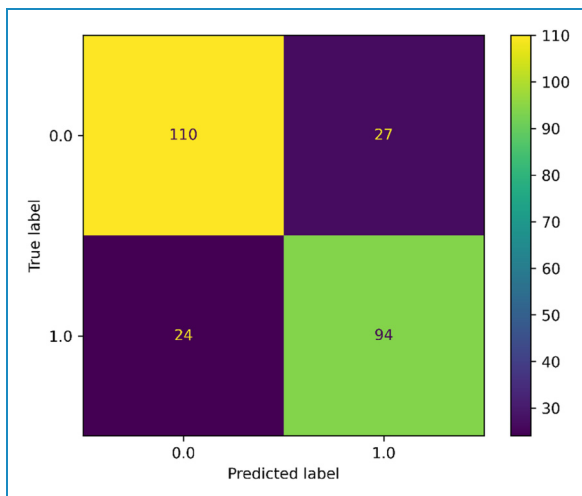
COVID-19.<sup>58,59</sup> Diabetes,<sup>60</sup> asthma,<sup>61</sup> cancer,<sup>62</sup> hypertension, and cardiovascular diseases<sup>63,64</sup> had a significant predictive role among chronic comorbidities. However, in the present study, Hyperlipidemia (HLP) was the most effective chronic comorbidity for predicting mortality and LoS. After



**Table 5.** The performance of various machine learning models for predicting the length of stay.

No.	Model	MSE	RMSE	MAE
1	Lasso	39.79	6.30	4.45
2	ElasticNet	39.92	6.31	4.45
3	Ridge	40.36	6.35	4.46
4	SVR	40.05	6.32	4.29
5	MLP (32*1024*32) activation ReLU	<b>38.96</b>	6.24	4.34

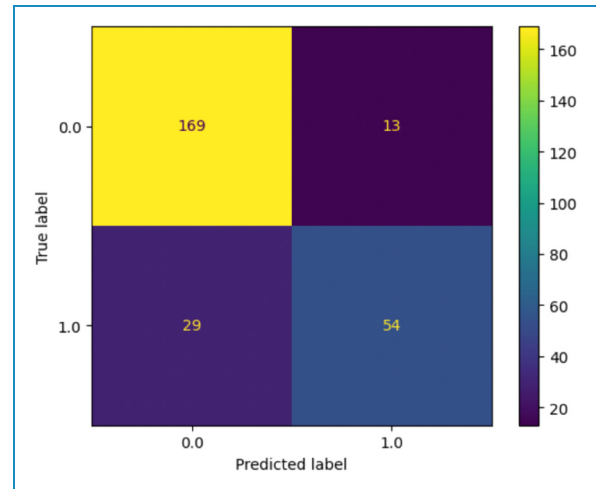
\*In each column, the best result is shown in bold. MAE: mean absolute error; MLP: multilayer perceptron; MSE: mean square error; ReLU: rectified linear unit; RMSE: root mean squared error; SVR: support vector regression.



**Figure 3.** Performance of the best mortality risk prediction model (SVM) based on the confusion matrix. SVM: support vector machines.

HLP, other chronic comorbidities such as diabetes, asthma, and cancer played a significant role in predicting patient mortality.

Clinical symptoms play a major role in the development of complications associated with COVID-19. In the present study, judging by the importance of the ranked features, shortness of breath was the most important symptom for predicting the mortality and LoS of COVID-19 patients with chronic comorbidities in the hospital. In addition to shortness of breath, symptoms such as sore throat, fever, diarrhea, and chest pain could effectively predict mortality, and symptoms such as fever, cough, fatigue, and abdominal pain were effective in predicting LoS. The best clinical symptoms in predicting mortality and longer LoS in other studies were fever, cough, shortness of breath, and diarrhea.<sup>55,64-68</sup>



**Figure 4.** Performance of the best mortality risk prediction ensembling model (Gradient Boosting) based on the confusion matrix.

This study showed that the patient's age effectively increases mortality and LoS. A study in the United Kingdom on 800 patients with COVID-19 and cancer showed that mortality in older patients is significantly higher.<sup>69</sup> In Iran, a study of 459 COVID-19 patients admitted to hospitals showed that the number of deaths increases with the age of patients.<sup>68</sup> In previous studies, age has been considered an independent and significant mortality index in diseases such as Middle East respiratory syndrome and Severe acute respiratory syndrome (SARS).<sup>70,71</sup> Finally, ML algorithms can be useful for physicians and administrators involved in treating patients with COVID-19 as well as COVID-19 patients with chronic comorbidities. The proposed algorithms can predict the mortality and LoS of patients with optimal accuracy, precision, sensitivity, specificity, and ROC. The results of these predictions can lead to the optimal use of hospital resources in treating patients with more critical conditions, help to provide better care and reduce medical errors resulting from fatigue and long working hours in hospital wards. Designing credible predictive models may improve the quality of care, increase patient survival, and reduce LoS. Therefore, predictive models for analyzing the risk of mortality and LoS can help identify high-risk patients and adopt the most effective care and treatment plans.

### Limitation

This study had two limitations. First, conducting this retrospective study in a single center may affect the quality of the data and the generalizability of the results. However, this hospital was the largest COVID-19 center in Kerman province and many patients from all over the province were hospitalized and treated there. Second, we did not include

important prognostic factors such as laboratory and radiological biomarkers.<sup>72–74</sup> However, according to the aim of the present study, it was sufficient to consider only the usual clinical features of patients at the time of admission.

## Conclusion

The results showed that the ML algorithms developed in this study can predict the risk of mortality and LoS in COVID-19 patients with chronic comorbidities with a mean accuracy of 74% and 85%, respectively, based on their physiological conditions, symptoms, and demographic information. Senility plays an important role in increasing mortality and LoS in these patients. Due to the high mortality rate of COVID-19 patients with chronic comorbidities, we recommend future studies monitor the course of the disease in patients with chronic comorbidities who have survived death from COVID-19.

**Availability of data and material:** Our data or material may be available from the corresponding author or first author upon reasonable request.

**Authors' contributions:** PA, FGh, and MM contributed to the study design; PA and FA collected the data; MM and SN analyzed the data; PA and MM drafted the manuscript; FA, RKh, and FS critically revised the manuscript for important intellectual content. All authors took part in the entire study and approved the final manuscript.


**Declaration of conflicting interests:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethics approval and consent to participate:** This article was extracted from an independent research project performed in the field of medical informatics at Kerman University of Medical Sciences without organizational support. This study was approved by the ethics committee of Kerman University of Medical Sciences (code of ethics: IR.KMU.REC.1400.055) and was performed according to the ethical guidelines of the Helsinki Declaration. Also, this study was supported by the Student Research Committee of Kerman University of Medical Sciences (code: 99000625). In addition, due to the retrospective nature of the study, the ethics committee of Kerman University of Medical Sciences waived the need for written informed consent.

**Funding:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iDs:** Parastoo Amiri  <https://orcid.org/0000-0002-5654-1987>

Mahdieh Montazeri  <https://orcid.org/0000-0001-6987-6199>

Fahimeh Ghasemian  <https://orcid.org/0000-0002-2176-7089>

**Supplemental material:** Supplemental material for this article is available online.

## References

1. Lu H, Stratton CW and Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *J Med Virol* 2020; 92: 401.
2. Liu Y, Gayle AA, Wilder-Smith A, et al. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine* 2020; 27(2): 13–21.
3. Bogoch II, Watts A, Thomas-Bachli A, et al. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med* 2020; 27: taaa008.
4. Dong E, Du H and Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; 20: 533–534.
5. Guan W-j, Liang W-h, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *European Respiratory Journal*. 2020; 55(5): 2000–2547.
6. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020; 323: 2052–2059.
7. 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: 1061–1069.
8. Jannat Z. COVID-19 and the elderly with chronic diseases: narrative review. *J Military Med* 2020; 22: 632–640.
9. Vekaria B, Overton C, Wiśniowski A, et al. Hospital length of stay for COVID-19 patients: data-driven methods for forward planning. *BMC Infect Dis* 2021; 21: 1–15.
10. Burke HB, Goodman PH, Rosen DB, et al. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997; 79: 857–862.
11. Lundin M, Lundin J, Burke H, et al. Artificial neural networks applied to survival prediction in breast cancer. *Oncology* 1999; 57: 281–286.
12. Ebinger J, Wells M, Ouyang D, et al. A machine learning algorithm predicts duration of hospitalization in COVID-19 patients. *Intell-Based Med* 2021; 5: 100035.
13. Somi M, Ahmadzadeh R, Farhang S, et al. Evaluation of treatment and survival rates in patients with esophageal cancer referred to Imam Khomeini Hospital, Tabriz, Iran. *Govareh* 2012; 17: 33–38.
14. Giambone GP, Smith MC, Wu X, et al. Variability in length of stay after uncomplicated pulmonary lobectomy: is length of stay a quality metric or a patient metric? *Eur J Cardiothorac Surg* 2016; 49: e65–e71.
15. Zolbanin HM, Davazdahemami B, Delen D, et al. Data analytics for the sustainable use of resources in hospitals: predicting the length of stay for patients with chronic diseases. *Information & Management*. 2020; 59(5): 103–282.
16. Dagliati A, Marini S, Sacchi L, et al. Machine learning methods to predict diabetes complications. *J Diabetes Sci Technol* 2018; 12: 295–302.
17. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic

- retinopathy in retinal fundus photographs. *JAMA* 2016; 316: 2402–2410.
18. Kehl KL, Elmarakeby H, Nishino M, et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol* 2019; 5: 1421–1429.
  19. Levin S, Barnes S, Toerper M, et al. Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. *BMJ Innovations*. 2021; 7(2): 100–110.
  20. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur Heart J* 2017; 38: 500–507.
  21. Doyle R. Prediction of COVID-19 mortality to support patient prognosis and triage and limits of current open-source data. medRxiv. 2021.
  22. Chowdhury ME, Rahman T, Khandakar A, et al. An early warning tool for predicting mortality risk of COVID-19 patients using machine learning. *Cogn Comput* 2021; 21: 1–6.
  23. An C, Lim H, Kim DW, et al. Machine learning prediction for mortality of patients diagnosed with COVID-19: A nationwide Korean cohort study. *Sci Rep* 2020; 10: 1–1.
  24. Banoei MM, Dinparastisaleh R, Zadeh AV, et al. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. *Crit Care* 2021; 25: 1–4.
  25. Mahboub B, Bataineh MTA, Alshraideh H, et al. Prediction of COVID-19 hospital length of stay and risk of death using artificial intelligence-based modeling. *Front Med (Lausanne)* 2021; 8: 592336.
  26. Garcia-Gutiérrez S, Esteban-Aizpiri C, Lafuente I, et al. Machine learning-based model for prediction of clinical deterioration in hospitalized patients by COVID 19. *Scientific reports*. 2022; 12(1): 70–97.
  27. Alabbad DA, Almuhaideb AM, Alsunaidi SJ, et al. Machine learning model for predicting the length of stay in the intensive care unit for COVID-19 patients in the eastern province of Saudi Arabia. *Inform Med Unlocked* 2022; 30: 100937.
  28. Wang K, Zuo P, Liu Y, et al. Clinical and laboratory predictors of in-hospital mortality in patients with coronavirus disease-2019: A cohort study in Wuhan, China. *Clin Infect Dis* 2020; 71: 2079–2088.
  29. An C, Lim H, Kim D-W, et al. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Sci Rep* 2020; 10: 1–11.
  30. Jimenez-Solem E, Petersen TS, Hansen C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients. *Sci Rep* 2021; 11: 1–12.
  31. Afrash M, Kazemi-Arpanahi H, Ranjbar P, et al. Predictive modeling of hospital length of stay in COVID-19 patients using machine learning algorithms. *J Med Chem Sci* 2021; 4: 525–537.
  32. Jutzeler CR, Bourguignon L, Weis CV, et al. Comorbidities, clinical signs and symptoms, laboratory findings, imaging features, treatment strategies, and outcomes in adult and pediatric patients with COVID-19: a systematic review and meta-analysis. *Travel Med Infect Dis* 2020; 37: 101825.
  33. Alimohamadi Y, Sepandi M, Dadgar A, et al. Hospital length of stay among COVID-19 patients: an application of competing risk analysis. *Journal of Biostatistics and Epidemiology*. 2021; 7(3): 224–234.
  34. Bertsimas D, Lukin G, Mingardi L, et al. COVID-19 mortality risk assessment: an international multi-center study. *PLoS One* 2020; 15: e0243262.
  35. Sousa G, Garces T, Cestari V, et al. Mortality and survival of COVID-19. *Epidemiology & Infection*. 2020; 25(2): 148.
  36. Butt AA, Kartha AB, Masoodi NA, et al. Hospital admission rates, length of stay, and in-hospital mortality for common acute care conditions in COVID-19 vs. pre-COVID-19 era. *Pub Health* 2020; 189: 6–11.
  37. Subudhi S, Verma A, Patel AB, et al. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *NPJ Digit Med* 2021; 4: 1–7.
  38. Dan T, Li Y, Zhu Z, et al. (eds.) Machine learning to predict ICU admission, ICU mortality and survivors' length of stay among COVID-19 patients: toward optimal allocation of ICU resources. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020. IEEE.
  39. Magunia H, Lederer S, Verbuecheln R, et al. Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Crit Care* 2021; 25: 1–14.
  40. Mahdavi M, Choubdar H, Zabeh E, et al. A machine learning based exploration of COVID-19 mortality risk. *PLoS One* 2021; 16: e0252384.
  41. Aktar S, Talukder A, Ahamad MM, et al. Machine learning approaches to identify patient comorbidities and symptoms that increased risk of mortality in COVID-19. *Diagnostics* 2021; 11: 1383.
  42. Alinejad H, Vazirinejad R, Sayadi A, et al. The relationship between COVID-19-induced death and chronic diseases. *J Client-Centered Nurs Care* 2021; 7: 167–174.
  43. Pourhomayoun M and Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* 2021; 20: 100178.
  44. Lau AY, Gabarron E, Fernandez-Luque L, et al. Social media in health—what are the safety concerns for health consumers? *Health Inform Manag J* 2012; 41: 30–35.
  45. Liu Y, Wang Y and Zhang J (eds.) New machine learning algorithm: random forest. International Conference on Information Computing and Applications, 2012. Springer.
  46. Zhang Z, Chen L, Xu P, et al. Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. *Laparosc Endosc Robot Surg* 2022; 5: 25–34.
  47. Mossman D and Somoza E. ROC curves, test accuracy, and the description of diagnostic tests. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 1991; 3(3): 330–333.
  48. Baratloo A, Hosseini M, Negida A, et al. Part 1: simple definition and calculation of accuracy, sensitivity and specificity. *Emergency* 2015; 6(4): 120–133.
  49. Liu J, Lan H, Fu Y, et al. Analyzing electricity consumption via data mining. *Wuhan Univ J Nat Sci* 2012; 17: 121–125.
  50. Shanbehzadeh M, Orooji A and Kazemi-Arpanahi H. Comparing of data mining techniques for predicting in-hospital mortality among patients with COVID-19. *J Biostatist Epidemiol* 2021; 7: 154–173.
  51. Chen X and Liu Z. Early prediction of mortality risk among severe COVID-19 patients using machine learning. 2020.

52. Agieb R. Machine learning models for the prediction the necessity of resorting to ICU of COVID-19 patients. *International Journal of Advanced Trends in Computer Science and Engineering*. 2020; 9(5): 6980–6984.
  53. Booth AL, Abels E and McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Mod Pathol* 2021; 34: 522–531.
  54. Orooji A, Shanbehzadeh M, Kazemi-Arpanahi H, et al. Predictive modeling of hospital Length of Stay in COVID-19 patients using Artificial Neural Networks. *Researchsquare* 2021; 10(9): 170–198.
  55. Zhang J, Wang X, Jia X, et al. Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. *Clin Microbiol Infect* 2020; 26: 767–772.
  56. Bacchi S, Gluck S, Tan Y, et al. Prediction of general medical admission length of stay with natural language processing and deep learning: a pilot study. *Intern Emerg Med* 2020; 15: 989–995.
  57. Kulkarni H, Thangam M and Amin AP. Artificial neural network-based prediction of prolonged length of stay and need for post-acute care in acute coronary syndrome patients undergoing percutaneous coronary intervention. *Eur J Clin Invest* 2021; 51: e13406.
  58. Wong CK, Wong JY, Tang EH, et al. Clinical presentations, laboratory and radiological findings, and treatments for 11,028 COVID-19 patients: a systematic review and meta-analysis. *Sci Rep* 2020; 10: 1–16.
  59. Zaki N, Alashwal H and Ibrahim S. Association of hypertension, diabetes, stroke, cancer, kidney disease, and high-cholesterol with COVID-19 disease severity and fatality: a systematic review. *Diabetes Metab Syndr: Clin Res Rev* 2020; 14: 1133–1142.
  60. Guo W, Li M, Dong Y, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diabetes Metab Res Rev* 2020; 36: e3319.
  61. Izquierdo JL, Almonacid C, González Y, et al. The impact of COVID-19 on patients with asthma. *European Respiratory Journal*. 2021; 57(3): 200–314.
  62. Lee LY, Cazier J-B, Starkey T, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol* 2020; 21: 1309–1316.
  63. Assaf D, Gutman Ya, Neuman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med* 2020; 15: 1435–1443.
  64. Zhou Y, He Y, Yang H, et al. Exploiting an early warning nomogram for predicting the risk of ICU admission in patients with COVID-19: a multi-center study in China. *Scand J Trauma Resusc Emerg Med* 2020; 28: 1–13.
  65. Wu S, Xue L, Legido-Quigley H, et al. Understanding factors influencing the length of hospital stay among non-severe COVID-19 patients: a retrospective cohort study in a Fangcang shelter hospital. *PLoS One* 2020; 15: e0240959.
  66. Alwafi H, Naser AY, Qanash S, et al. Predictors of length of hospital stay, mortality, and outcomes among hospitalised COVID-19 patients in Saudi Arabia: a cross-sectional study. *J Multidiscip Healthc* 2021; 14: 839.
  67. Guan W-j, Ni Z-y, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; 382: 1708–1720.
  68. Alamdari NM, Afaghi S, Rahimi FS, et al. Mortality risk factors among hospitalized COVID-19 patients in a major referral center in Iran. *Tohoku J Exp Med* 2020; 252: 73–84.
  69. Lee LY, Cazier J-B, Angelis V, et al. COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. *Lancet* 2020; 395: 1919–1926.
  70. Jia N, Feng D, Fang LQ, et al. Case fatality of SARS in mainland China and associated risk factors. *Trop Med Int Health* 2009; 14: 21–27.
  71. Hong K-H, Choi J-P, Hong S-H, et al. Predictors of mortality in Middle East respiratory syndrome (MERS). *Thorax* 2018; 73: 286–289.
  72. Yuan M, Yin W, Tao Z, et al. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PLoS One* 2020; 15: e0230548.
  73. Yue H, Yu Q, Liu C, et al. Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. *Annals of translational medicine*. 2020; 8(14): 859–910.
  74. Guan X, Zhang B, Fu M, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study. *Ann Med* 2021; 53: 257–266.
-