



Research Article

The machine learning approach for predicting the number of intensive care, intubated patients and death: The COVID-19 pandemic in Turkey

Pınar CİHAN^{1,*}

¹Department of Computer Engineering, Tekirdağ Namık Kemal University, Tekirdağ, Turkey

ARTICLE INFO

Article history

Received: 10 January 2021

Accepted: 10 March 2021

Key words:

COVID-19; Machine learning,
Intensive care, Intubated; Death

ABSTRACT

The coronavirus infection outbreak started in Wuhan city, China, in December 2019 (COVID-19) and affected more than 200 countries in a year. The number of patients dying from and infected with COVID-19 is increasing at an alarming rate in almost all affected countries. One of the most important factors in the COVID-19 death and case rates is the care of intensive care patients. The management of COVID-19 patients who need acute and/or critical respiratory care has created a significant difficulty for healthcare systems worldwide. To prevent the further spread of COVID-19 around the world and to fight the disease, non-clinical computer-aided quick solutions such as artificial intelligence and machine learning are needed. Prediction techniques evaluate past situations and enable predictions about the future situation. In this study, using the dataset created from the data received from the World Health Organization and national database, the numbers of intensive care, intubated patients, and deaths from COVID-19 in Turkey were predicted by the random forest, bagging, support vector regression, classification and regression trees, and k-nearest neighbors machine learning regression methods. In this study, the random forest method has been the most successful algorithm for predicting the number of intensive care patients ($r = 0.8698$, RMSE = 188.5, MAE = 135.1, MAPE = 13%), the number of intubated patients ($r = 0.9846$, RMSE = 47.1, MAE = 39.7, MAPE = 9.2%), and the number of deaths ($r = 0.9994$, RMSE = 1.2, MAE = 0.9, MAPE = 3.5%). The results in this study, it has been shown that machine learning methods, which have been successfully applied in other epidemic diseases, will be successfully applied in the COVID-19 pandemic.

Cite this article as: Pınar C. The machine learning approach for predicting the number of intensive care, intubated patients and death: The COVID-19 pandemic in Turkey. Sigma J Eng Nat Sci 2022;40(1):85-94.

*Corresponding author.

*E-mail address: pkaya@nku.edu.tr

*This paper was recommended for publication in revised form by
Regional Editor Ferkan Yılmaz*



INTRODUCTION

In December 2019, a novel coronavirus causing severe respiratory disease was detected in China. This virus, which also causes the infection, was named by the World Health Organization (WHO) as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 is rapidly spreading among humans globally. People do not have immunity to this novel identified virus. Fever, fatigue, and dry cough are the most common symptoms. The virus causes more severe symptoms, such as high fever, severe cough, and shortness of breath, which often indicate pneumonia. Difficulty in breathing or shortness of breath are serious symptoms and patients require mechanical ventilation. The mortality rate is very high in patients at this stage [1].

COVID-19 spread to 114 countries within first three months and was declared a pandemic by the World Health Organization, causing 4 291 deaths [2]. To date (March 2021), 219 countries have been affected by the COVID-19 outbreak, and over 2.5 million people have died. The first case of the coronavirus in Turkey was observed on March 11, 2020. The total number of intensive care and intubated patients was regularly reported between March 27 and July 28, 2020. As of July 28, a total of 227 982 cases were detected, and the number of deaths recorded until then was 5 645, the total number of intensive care patients was 1 280, and the total number of intubated patients was 403 (Figure 1).

COVID-19 is an inflammatory disease caused by a new virus, and there is no clinically proven vaccine and drug yet. The disease has symptoms such as cold, cough, fever and respiratory distress and causes respiratory problems in severe cases. To prevent or minimize the spread of the virus, measures such as increasing hand washing, reducing touching the face, the use of masks, and social distance have been taken in many countries of the world. In Turkey, to reduce the spread of the coronavirus epidemic, curfews were imposed for people older than 65 years of age and younger

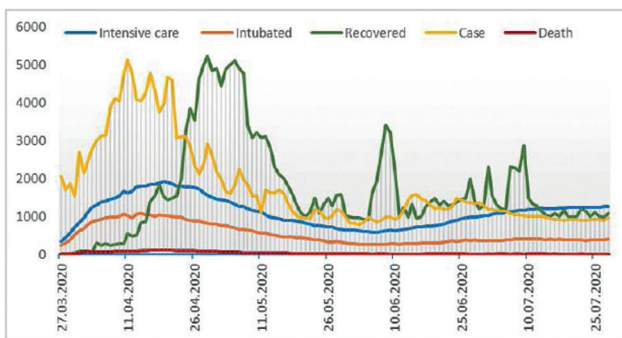


Figure 1. Distribution of the intensive care, intubated, recovered, case and death in COVID-19 pandemic for Turkey.

than 18 years of age, travel restrictions were brought, distance education was started in schools, weddings were banned, curfew restriction was introduced on some days, such as weekends or religious holidays when there is intensive interaction. These measures show the severity of the situation and that the virus is spread via mobility.

Machine learning (ML) is an artificial intelligence subset tool for transforms information into knowledge, provides the ability to automatically learn and improve from experience. ML methods are frequently used in different disciplines to solve classification and regression problems [3-6]. Predicting a numeric value is known as the regression method and many real-life problems can be solved by regression. Many regression methods can be used in solving real-life problems. The purpose is to find the best function for the used dataset. Hence, the method which minimizes the error between the true value and the predicted value is select.

Nowadays, artificial intelligence methods and clinical studies are used to reduce or stop the spread of the COVID-19 outbreak. Various models are widely used by the authorities and the media to predict the outbreak, determine the peak in advance, and predict the mortality and case rate [7]. Using machine learning methods, mathematical and statistical methods, predictions are made about the unknown with what has been learned from past data. In other words, prediction methods are modelled with computers and make inferences from the data, and thus predictions are made.

The treatment of patients, who have developed pneumonia due to coronavirus, in whom the lungs have remained insufficient in providing sufficient oxygen to the body and who need support by a breathing apparatus, is continued in intensive care units. If patients are experiencing respiratory failure, in other words, if they are unable to breathe adequately, they perform the breathing process with the artificial respiration connected to them. The procedure of placing a special tube through the mouth or nose into the respiratory tract of intubated patients is called intubation. The mortality rate in patients who are intubated and connected to a respiratory support device due to COVID-19 infection is above 50% [8].

Expanding the bed capacity, organizing the necessary health workforce, providing the necessary materials and equipment for both patients and staff are among the important elements of the COVID-19 outbreak. Insufficiencies in these elements can seriously affect the health systems of countries. Estimating information such as the number of intensive care patients, the number of intubated patients, and the number of deaths can contribute to the planning of the needed elements in the health system and the preparation of the infrastructure.

The motivation of this study is to compare the predictive performances of machine learning methods in estimating the number of daily deaths, the number of intensive care patients, and the number of intubated patients,

independent of time, from different variables. For this purpose, random forest (RF), bagging, support vector regression (SVR), classification and regression trees (CART), and k-nearest neighbors (KNN) machine learning regression methods were developed for prediction. Root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), correlation coefficient (r), and coefficient of determination (R^2) metrics were used to compare the ML models' prediction performances. As the case study, confirmed data announced by the Turkish Ministry of Health were used.

RELATED WORKS

Machine learning methods have been accepted as a computational technique with significant potential in outbreak prediction. Some of the ML methods that have successfully predicted the previous outbreaks are as follows. The random forest method was successful in predicting the swine flu outbreak [9, 10], neural networks in predicting the H1N1 flu, dengue fever and norovirus outbreaks genetic programming in predicting the norovirus outbreak [11], CART in predicting the dengue fever outbreak [12], and the Bayesian network was successful in predicting the dengue fever and Aedes outbreak [13]. ML prediction models are also needed to fight the COVID-19 outbreak.

The number of artificial intelligence techniques is increasing every day, due to the advantages they offer in controlling the COVID-19 outbreak. Li et al. [14], developed a prediction model that uses machine learning algorithms to predict cumulative confirmed cases, the number of new cases, and death cases in China between January 20 and March 1. Wang and Wong [15], aimed to detect COVID-19 cases from chest radiography images using the deep convolutional neural network method. Jiang et al. [16], tried to predict which patient affected by COVID-19 would develop acute respiratory distress syndrome, using the decision trees, random forests, and support vector machines methods. Machine learning models were trained with samples obtained from 53 patients. Yan et al. [17], used a machine learning algorithm to predict the risk of death

of an infected person. Data obtained from 29 patients were used in the study. Furthermore, many review studies examining artificial intelligence techniques for COVID-19 have been carried out [18-21]. Ceylan [22] forecasted the number of COVID-19 cases in Italy, Spain and France with the ARIMA model. The best models in forecasting the number of cases in the study were determined as ARIMA (0,2,1) for Italy, ARIMA (1,2,0) for Spain and ARIMA (0,2,1) for France. The mean absolute percentage error of these models in estimating the number of cases was 4.75%, 5.85% and 5.63%, respectively. Alzahrani et al. [23] estimated the number of COVID-19 cases in Saudi Arabia with the ARIMA model. It was determined that the ARIMA (2,1,1) model for Saudi Arabia estimates the closest to the real value, and the number of COVID-19 cases for the next month was estimated with this model.

It is observed that the studies conducted are generally on the prediction of the number of COVID-19 cases and deaths. Estimation studies of the number of intensive care patients and the number of intubated patients are quite limited. In addition to machine learning methods, the ARIMA model, which is a time series method, was frequently used in the studies. The ARIMA model does not contain independent variables, the forecast is based on time.

MATERIALS AND METHODS

Data preparation

The first COVID-19 case in Turkey was observed on March 11, 2020. The government started to announce the total number of intensive care and intubated patients regularly after March 27, 2020. However, this information has started not to be given after July 28, 2020. Therefore, in the study, total numbers of intensive care, total numbers of intubated patients and number of daily death were predicted using the data announced by the ministry between March 27 and July 28, 2020. The dataset was created with this announced information in order to develop machine learning models and make predictions. Features in the dataset, abbreviations of these features, input and output information of machine learning structures are given in Table 1.

Table 1. Parameters, abbreviations, input and output types

Parameter	Abbreviation	Prediction of intensive care	Prediction of intubated	Prediction of Death
Cumulative intensive care patients	IC	output	input	input
Cumulative intubated patients	I	input	output	input
Daily number of test	T	input	input	input
Recovered patients	R	input	input	input
Daily cases	C	input	input	input
Daily death	D	input	input	output

The COVID-19 data for 124 days were divided into two parts: 75% for training and 25% to test the success of the model. In other words, the training process of the machine learning model was carried out with the data for 93 days, and the independent test data for 31 days were used to evaluate the prediction performance of ML models. In the study, the RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE), correlation coefficient (r), and coefficient of determination (R^2) criteria were used to compare the model prediction performances.

Prediction models

In this study, the estimation success of different regression methods was compared. The methods used in the study are briefly mentioned below.

Random forest (RF)

RF is a supervised learning algorithm that can be used in both classification and regression problems. RF is an ensemble learning technique in which individually created classification and regression trees (CART) come together to form a decision forest. The results obtained during the formation of the decision forest are combined, and the final prediction is made. The RF algorithm is an improved version of the bagging algorithm. Unlike bagging, the randomness feature has been added to the RF algorithm. Each dataset is generated from the original data set with displacement. Afterward, trees are developed using random feature selection. The developed trees are not pruned. This strategy makes the accuracy of RF unique. At the same time, RF is very fast, resistant to overfitting and works with as many trees as desired [24]. In this study, the *train* function in the *caret* package was used.

Bagging

Bootstrap aggregating, also called bagging, is a machine learning-based ensemble learning technique that increases the accuracy of statistical classification methods. Basically, it is aimed to create more than one training sample by producing different combinations of training data. Assuming that the training data consist of N , training data with N samples are generated by random selection. In this case, some samples are not included in the training data, while some may appear more than once. Each decision tree is trained with training data containing different samples produced in this way, and the result is determined by majority voting. Bagging is usually applied to decision tree methods and can be used with any method [25]. Bagging was used with the CART method in the study. In this study, the *bagging* function in the *ipred* package was used.

Support vector regression (SVR)

Support Vector Regression (SVR) is a special version of Support Vector Machine (SVM) developed for numerical prediction problems. It is more successful than other traditional learning methods in solving nonlinear problems. The

purpose of SVM is to find the optimum separator plane that separates the data into two separate classes. There are two situations that can be encountered in SVM. One of them is that the data can be separated linearly, and the other one is having a structure that cannot be separated linearly. To distinguish linearly classifiable data from each other, it determines the largest margin among many possible linear functions. If the dataset cannot be separated completely or linearly with a certain error rate, the dataset can be moved to a higher dimensional space by performing transformations with the help of basic functions. Thus, it is ensured that it corresponds to a linear model in this new space. Different kernel functions (linear, polynomial, radial, sigmoid) can be used to categorize data in multidimensional space [26, 27]. In this study, the *svm* function in the *e1071* package was used.

Classification and regression trees (CART)

In this approach, if the target value in the training data is categorical, the classification tree is called the regression tree, if it is continuous. As can be understood from the name, a tree structure is used. The main purpose of classification and regression trees is to divide independent variables (inputs) into homogeneous subgroups according to the dependent variable (output). In the creation of subgroups, data are presented in a hierarchical order in the form of a branching tree. The dependent variables that have made the best separation are shown at the intermediate nodes in the tree shape. The critical value of the differential dependent variables is given in the branches of these nodes. Leaves show the values of the dependent variable. There are lines from the root node (first node) to the leaves (last node). Along these lines, the rules of separation are shown where the separation between classes is maximized and variation within each class is minimized [28]. It is a fast and flexible algorithm that does not require any special data preparation. In this study, the *rpart* function in the *rpart* package was used.

K-nearest neighbor (KNN)

The K-nearest neighbor (KNN) algorithm is a nonparametric machine learning method used for classification and regression. The basic logic in the KNN method is found using the nearest k neighbor distance measure of the sample to be classified. The average value of this k neighbor is the value of our test sample [29]. In this study, the *knn.reg* function in the *FNN* package was used.

Model performance evaluation criteria

Evaluation metrics compare the target and output value for each machine learning method and calculate an index score for the performance and accuracy of the methods developed. Thus, model results can be compared, and differentiation between models can be made. In this study, RMSE, MAE, MAPE, r , and R^2 metrics were used to evaluate

the prediction performance of machine learning regression models. These statistics compare the target and output values and calculate a score as an index for the performance and accuracy of the methods developed.

RMSE is the measure of the distance between the values predicted by the machine learning regression model and the actual values. It is frequently used in the literature to evaluate the prediction success of regression models. RMSE is the standard deviation of prediction errors, and its value is always positive. A low RMSE value is a desirable condition. It is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - P_t)^2} \quad (1)$$

MAE is the average of the difference between the values predicted by the machine learning regression model and the actual values. This value can range from zero to infinity. The model with a low mean absolute error exhibits a better prediction performance. It is defined as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - P_t| \quad (2)$$

MAPE is the average percentage of the difference between the values predicted by the machine learning regression model and the actual values. It is used to measure the predictive accuracy of the model. It is defined as follows:

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - P_t}{A_t} \right| \quad (3)$$

r is a criterion frequently used to evaluate the goodness of fit of machine learning prediction models. It defines the relationship between the actual and predicted values. It is defined as follows:

$$r = \frac{\sum_{t=1}^n (A_t - \bar{A})(P_t - \bar{P})}{\sqrt{\sum_{t=1}^n (A_t - \bar{A})^2} \sqrt{\sum_{t=1}^n (P_t - \bar{P})^2}} \quad (4)$$

Here, n is the number of data used for the test, A is the actual value, P is the output value predicted by the model.

R², summarizes the explanatory power of the regression model. It is '1' if the regression model has predicted perfectly and '0' if it has failed completely. In other words, no variance can be explained by a regression model. It is defined as follows:

$$R^2 = r \cdot r \quad (5)$$

EXPERIMENTAL RESULTS

The dataset used in the study contains the intensive care, intubated, test, recovered, confirmed cases, and daily death information announced by the Ministry of Health between 27 March 2020, and 28 July 2020. The reason for using data in this range in the study is that the numbers of intensive care and intubated patients were regularly announced between these dates. The statistical information of the dataset is presented in Table 2.

The relationship between the features in the dataset was examined, and the correlation matrix is presented in Figure 2. Accordingly, it is observed that the number of deaths is highly correlated with the number of intubated patients and intensive care patients (r = 0.96, P < 0.00001; r = 0.81, P < 0.00001, respectively). Since intubated and intensive care patients are in the critically ill group, a high mortality rate is expected. Thus, it is known that 50–97% of intubated patients die [8]. There is a significant correlation between the number of patients in the intensive care

Table 2. Variables and statistics of the dataset

Variable	Min.	Max.	Mean	Std.
IC	344	1922	1118	388
I	241	1087	541	264
T	7533	57829	37151	10850
R	26	5231	1717	1287
C	786	5138	1823	1130
D	14	127	45	34

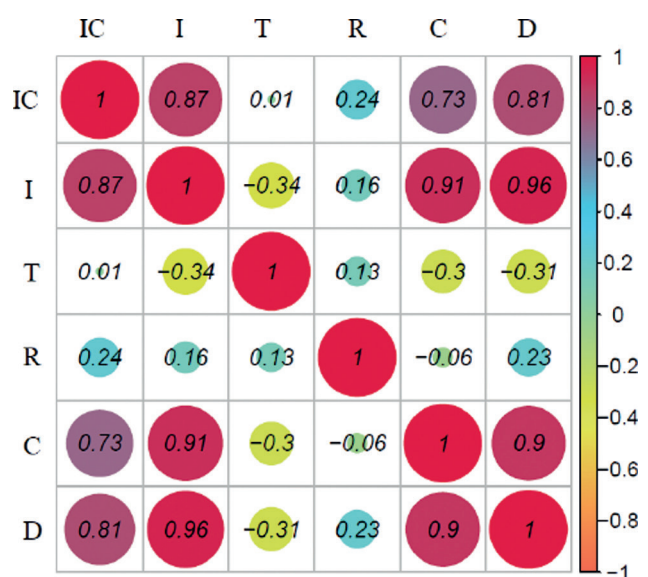


Figure 2. Correlation matrix.

unit and the number of intubated patients ($r = 0.87$, $P < 0.00001$). There is a significantly high correlation between the number of cases and the number of intensive care, intubated patients, and deaths ($r = 0.73$, $r = 0.91$, $r = 0.90$, $P < 0.00001$, respectively).

Prediction of the number of intensive care patients

In this section, the number of intensive care patients was predicted using RF, bagging with CART, SVR, CART, and KNN prediction methods, and the results obtained are presented in Table 3. The goodness-of-fit was evaluated using statistical values such as RMSE, MAE, and MAPE. The lower these statistical values are, the higher the prediction precision of the model is. When the prediction performances of the models are compared, it is observed that the method that predicts the number of intensive care patients with the least error is RF (RMSE = 188.5, MAE = 135.1, MAPE = 13%). At the same time, it is observed that the RF method makes more successful predictions than other methods with an error rate of 13%.

The comparative performances of machine learning methods in estimating the number of intensive care patients are given with the Taylor diagram presented in Figure 3A. The normalized standard deviation, normalized RMSE, and r results of the prediction models are shown

comparatively in the Taylor diagram. In the Taylor diagram, it is seen that the normalized standard error is the lowest, the correlation is the highest, and the closest model to the actual value is the RF. The RF method has the highest correlation in predicting the number of IC patients ($r = 0.88$). Due to its low error in predicting the number of IC patients (Table 3) and higher R-value, the actual vs predicted values with the RF model were illustrated in Figure 3B. Also scatter plot of the actual IC and predicted IC with the RF model are given in Table 6. It is observed that there is a high correlation between the RF model predicted values and actual values ($R^2 = 0.77$).

Prediction of the number of intubated patients

In this section, the number of intubated patients predicted with different machine learning methods and the comparative results obtained are presented in Table 4. When the prediction performances of the models are examined, it is observed that the RF model (RMSE = 47.1, MAE = 39.7) and bagging (RMSE = 46.6, MAE = 39.0) methods predict the number of intubated patients very close to each other and with the lowest error. The RF model predicted the number of intubated patients with an error rate of 9.2%, and the bagging method predicted the number of intubated patients with an error rate of 9.9%.

Table 3. Comparing the prediction results of the number of intensive care patients

Metric	RF	Bagging	SVR	CART	KNN
RMSE	188.5	198.9	255.6	269.0	198.7
MAE	135.1	150.1	195.6	192.5	163.1
MAPE (%)	13.4	13.7	18.4	16.9	16.7

Table 4. Comparing the prediction results of the number of intubated patients

Metric	RF	Bagging	SVR	CART	KNN
RMSE	47.1	48.5	62.5	57.8	100.6
MAE	39.7	41.0	56.8	47.4	71.2
MAPE (%)	9.2	10.4	14.3	11.2	17.4

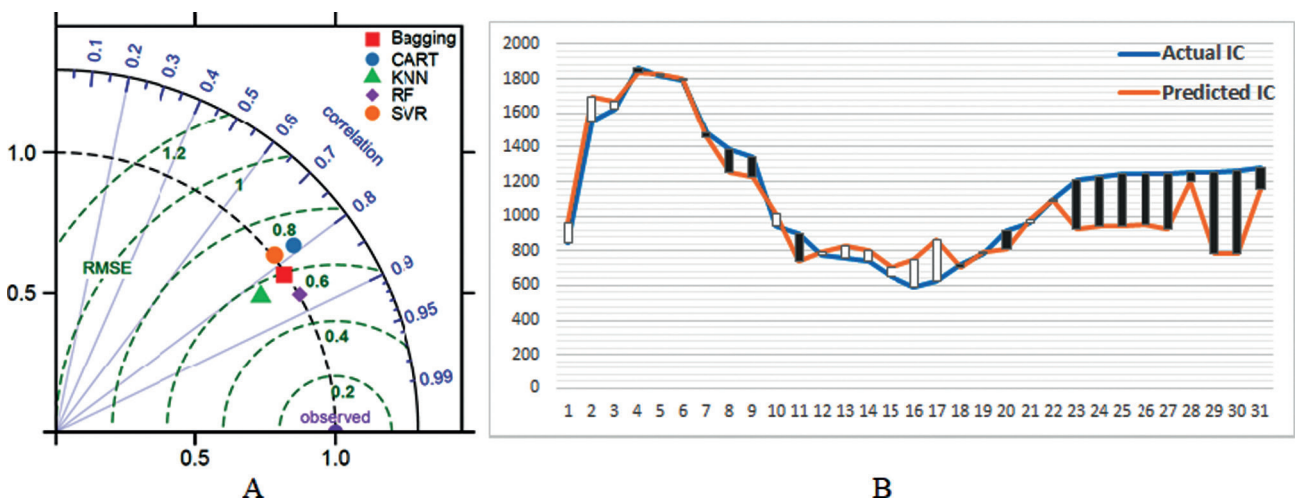


Figure 3. (A) Taylor diagram of prediction models, (B) Scatter plot of the actual IC and predicted IC with the RF method.

The prediction performances of the number of intubated patients by different machine learning models are compared in Figure 4A. It was seen that the RF model has the highest correlation in predicting the number of intubated patients and the bagging model follows it ($r = 0.9847$, $r = 0.9816$, respectively). In the Taylor diagram, the method with the lowest standard deviation was KNN, while the lowest RMSE and the highest correlation value were obtained from the RF model. There is a high correlation between the predicted values of the number of intubated patients of the RF model and the actual values ($R^2 = 0.97$). In Figure 4B, the values predicted with the RF method and the actual values are presented with a scatter plot. It is seen that the residuals between the actual values and the estimated values are low. This shows that the method makes good predictions.

Prediction of the number of death

In this section, the number of mortality was predicted using RF, bagging with CART, SVR, CART, and

KNN prediction methods, and the results are given in Table 5. When the prediction performances of the models are examined, it is observed that the method that predicts the mortality with the lowest error is RF (RMSE = 1.2, MAE = 0.9). It is revealed that the RF model is quite successful in predicting the number of mortality with an error rate of 2.9% compared to other methods.

Table 5. Comparing the prediction results of the number of death

Metric	RF	Bagging	SVR	CART	KNN
RMSE	1.2	3.9	2.3	7.4	9.5
MAE	0.9	3.2	2.0	5.5	7.3
MAPE (%)	2.9	12.0	7.6	15.5	22.1

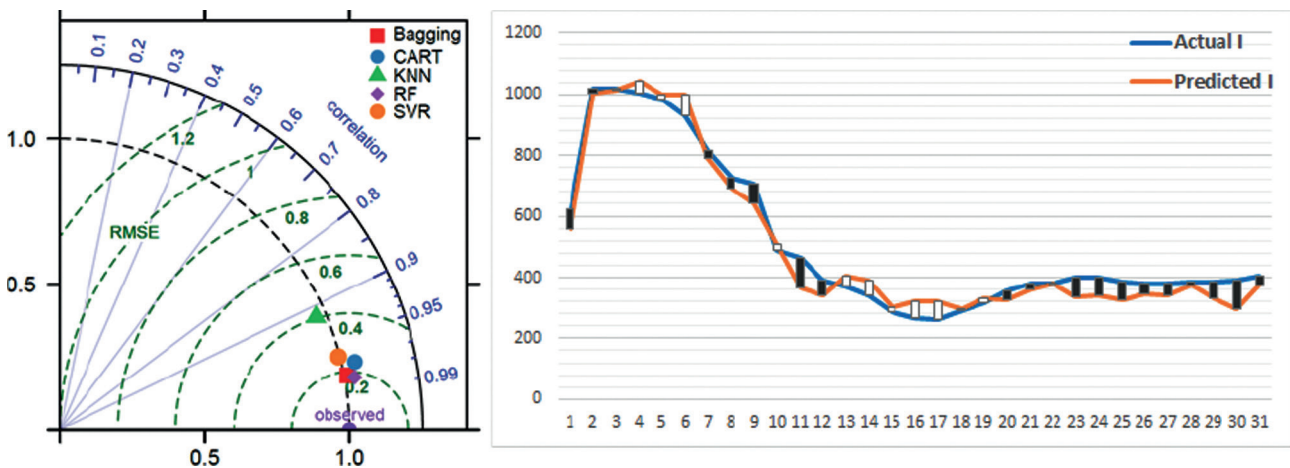


Figure 4. (A) Taylor diagram of prediction models, (B) Scatter plot of the actual I and predicted I with the RF method.

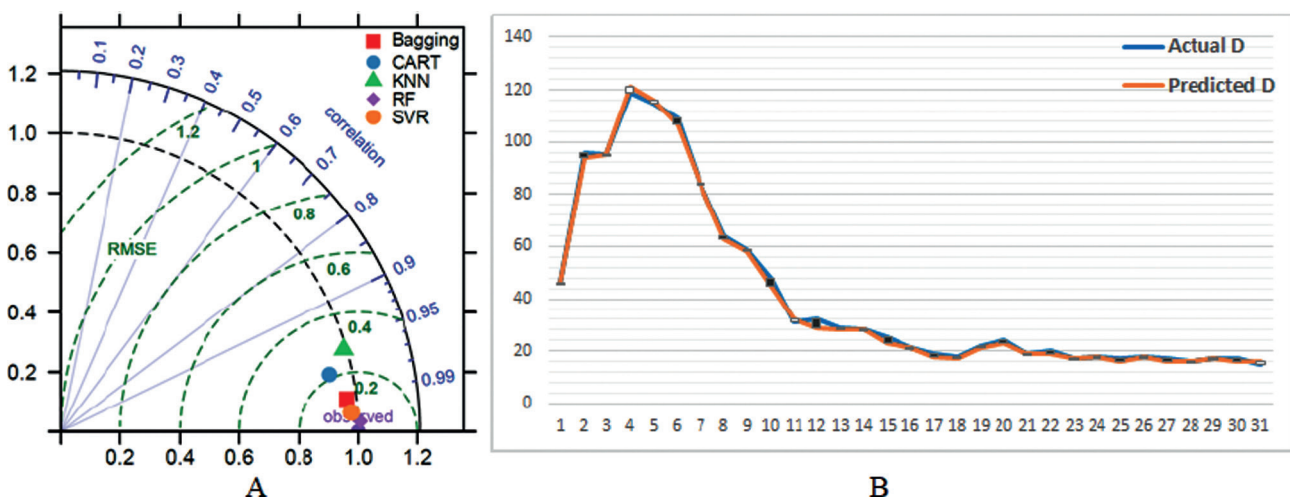


Figure 5. (A) Taylor diagram of prediction models, (B) Scatter plot of the actual D and predicted D with the RF method.

Table 6. Actual and predicted values

Actual IC	Predicted IC	Actual I	Predicted I	Actual D	Predicted D
847	962	622	560	46	46
1552	1692	1017	1004	96	94
1626	1668	1021	1014	95	95
1865	1832	1006	1042	119	121
1816	1826	982	1000	115	116
1790	1803	929	999	109	107
1480	1454	818	792	84	84
1384	1256	727	692	64	63
1338	1227	707	644	59	58
944	1011	490	510	48	45
903	740	463	364	31	32
775	790	388	339	32	29
756	828	371	404	29	28
739	805	338	389	28	28
648	701	287	300	25	23
591	746	264	320	21	21
625	863	261	319	19	18
722	704	291	295	18	17
781	796	318	331	22	21
914	809	356	325	24	23
963	975	382	361	19	19
1093	1091	378	378	20	19
1206	930	401	336	17	17
1226	947	402	341	18	18
1243	944	385	326	17	16
1246	952	379	344	18	18
1248	930	379	338	17	16
1252	1203	386	377	16	16
1249	781	387	332	17	17
1263	781	392	294	17	16
1280	1156	403	375	15	16

Models' number of death prediction performances are compared in Figure 5. Upon examining the Taylor diagram (Figure 5A), it is observed that the RF method has the highest correlation in predicting the mortality ($r = 0.9994$). It was determined that there is a high correlation between the death number prediction values with the RF model and the actual values ($R^2 = 0.9989$). When the death numbers predicted with the RF model and the actual death numbers are compared, it is seen that the estimates are very close, that is, the residuals approach zero (Figure 5B).

Actual values and predicted values with the Random Forest model for three scenarios are presented in Table 6.

CONCLUSION

The COVID-19 pandemic has spread rapidly, affecting the whole world. The sudden increase in the number of intensive care patients causes intensive care units to fill up and medical materials and personnel who take care of patients to become insufficient. The rapidly increasing number of patients is reflected in health systems worldwide as a significant capacity burden. Mortality rates can be reduced by planning and adjusting factors that will affect the health system, such as the early determination of the number of intensive care and intubated patients, allocating

medical resources on time, increasing the number of personnel, and increasing the capacity of intensive care units. In addition to the successful predictions of machine learning methods in other areas, their success in epidemics has also been proven. Machine learning methods are also used in the COVID-19 outbreak, and models have been developed to predict the outbreak, determine the peak in advance, and predict the death and case rates, and they are used by the authorities and the media [7].

In this study, different machine learning regression methods were used to predict the total number of intensive care patients (scenario 1), total intubated patients (scenario 2), and the number of daily death (scenario 3) caused by COVID-19. The prediction performances of machine learning models for three different prediction scenarios were compared and the most successful ML model was determined for each scenario.

According to the results obtained, it has been determined that the most successful method for predicting the number of intensive care patients was the Random Forest model ($r = 0.88$, RMSE = 188.5, MAE = 135.1, and MAPE = 13.4%). The Random Forest model was determined to be a successful method for estimating the number of intubated patients ($r = 0.9847$, RMSE = 47.1, MAE = 39.7, and MAPE = 9.2%). Finally, again the Random forest model has shown high performance in estimating the number of patients who died from COVID-19 ($r = 0.9994$, RMSE = 1.2, MAE = 0.9, MAPE = 2.9%).

The number of Intensive care patients, intubated patients, and death was estimated with different machine learning regression methods, and the comparative results showed that Random Forest was the best regression method, for all three scenarios (with mean absolute percentage error of 13.4%, 9.2%, 2.9%, respectively). In addition, it has been observed that Bagging was the best prediction model after the RF model in all scenarios. This shows that ensemble methods make better predictions and achieve better performance than any single model in this study. Ensemble learning methods are a collection of many single learning methods. They make reliable and more accurate predictions by combining decisions of individual methods.

In this study, it was found that the number of deaths was significantly correlated with the number of intubated patients and the number of intensive care patients ($r = 0.96$, $P < 0.00001$; $r = 0.81$, $P < 0.00001$, respectively). A significant correlation was determined between the number of intensive care patients and the number of intubated patients ($r = 0.87$, $P < 0.00001$). A significantly high correlation was found between the number of cases and the number of intensive care, intubated patients, deaths ($r = 0.73$, $r = 0.91$, $r = 0.90$, $P < 0.00001$, respectively).

FUNDING

Not applicable.

AVAILABILITY OF DATA AND MATERIAL

Data are freely available from WHO

CODE AVAILABILITY

R code, available upon request

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

- [1] Rahim F, Amin S, Noor M, Bahadur S, Gul H, Mahmood A, et al. Mortality of patients with Severe COVID-19 in the intensive care unit: an observational study from a Major COVID-19 receiving hospital. *Cureus* 2020;12:10906. [\[CrossRef\]](#)
- [2] Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 2020;76:71–76. [\[CrossRef\]](#)
- [3] Cihan P, Gökçe E, Atakişi O, Kirmizigül AH, Erdoğan HM. Prediction of immunoglobulin G in lambs with artificial intelligence methods. *Kafkas Univ Vet Fak Derg* 2021;27:21–27.
- [4] Cihan MT. Prediction of concrete compressive strength and slump by machine learning methods. *Adv Civil Eng* 2019;2019:3069046. [\[CrossRef\]](#)
- [5] Cihan P, Kalipsız O, Gökçe E. Yenidoğan kuzularda bilgisayar destekli tanı. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi* 2020;26:385–391. [\[CrossRef\]](#)
- [6] Cihan P, Gökçe E, Kalipsız O. A review of machine learning applications in veterinary field. *Kafkas Univ Vet Fak Derg* 2017;23:673–680.

- [7] Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020;395:10231. [CrossRef]
- [8] Auld S, Caridi-Scheible M, Blum JM, Robichaux CJ, Kraft CS, Jacob JT, et al. ICU and ventilator mortality among critically ill adults with COVID-19. *medRxiv* 2020. [CrossRef]
- [9] Liang R, LuY, Qu X, Su Q, Li C, Xia S, et al. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transbound Emerg Dis* 2020;67:935–946.
- [10] Tapak L, Hamidi O, Fathian M, Karami M. Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of health-care centers in Iran. *BMC Res Notes* 2019;12:353. [CrossRef]
- [11] Chenar SS, Deng Z. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res* 2018;128:20–37. [CrossRef]
- [12] Muurlink OT, Stephenson P, Islam MZ, Taylor-Robinson AW. Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach. *Infect Dis Model* 2018;3:322–330. [CrossRef]
- [13] Raja DB, Mallol R, Ting CY, Kamaludin F, Ahmad R, Ismail S, et al. Artificial intelligence model as predictor for dengue outbreaks. *Malaysian J Public Health Med* 2019;19:103–108. [CrossRef]
- [14] Li M, Zhang Z, Jiang S, Liu Q, Chen C, Zhang Y, et al. Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach. *medRxiv Preprint* 2020. doi: 10.1101/2020.03.18.20038117. [CrossRef]
- [15] Wang L, Wong A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *arXiv preprint arXiv:2003.09871*, 2020. [CrossRef]
- [16] Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Comput Mater Contin* 2020;63:537–551. [CrossRef]
- [17] Yan L, Zhang H-T, Xiao Y, Wang M, Sun C, Liang J, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan, *MedRxiv*, 2020. Preprint 2020. doi: 10.1101/2020.03.18.20038117. [CrossRef]
- [18] Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr* 2020;14:337–339. [CrossRef]
- [19] van der Schaar M, Alaa A. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Mach Learn* 2021;110:1–14. [CrossRef]
- [20] Nguyen TT. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. *arXiv:2008.07343*. Preprint.
- [21] Naudé W. Artificial Intelligence against COVID-19: An early review. *IZA Discussion Paper No. 13110*, 2020.
- [22] Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Envir* 2020;729:138817.
- [23] Alzahrani SI, Aljamaan IA, and Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infect Public Health* 2020;13:914–919.
- [24] Breiman L. Random forests. *Machine Learning* 2001;45:5–32.
- [25] Breiman L. Bagging predictors. *Machine Learning* 1996;24:123–140.
- [26] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing* 2004;14:199–222.
- [27] Vapnik V. *The Nature of Statistical Learning Theory*. Berlin: Springer Science & Business Media, 2013.
- [28] Loh WY. *Classification and regression trees*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2011;1:14–23.
- [29] Fix E. *Discriminatory analysis: nonparametric discrimination, consistency properties*. Ohio: USAF School of Aviation Medicine, 1951.