

Evaluations of Machine Learning Algorithms Using Simulation Study

Nasrin Khatun

Department of Statistics and Data Science, Jahangirnagar University, Dhaka, Bangladesh

Email: nasrin@juniv.edu

How to cite this paper: Khatun, N. (2025) Evaluations of Machine Learning Algorithms Using Simulation Study. *Open Journal of Statistics*, 15, 41-52.
<https://doi.org/10.4236/ojs.2025.151003>

Received: July 29, 2024

Accepted: February 15, 2025

Published: February 18, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

1st cases of COVID-19 were reported in March 2020 in Bangladesh and rapidly increased daily. So many steps were taken by the Bangladesh government to reduce the outbreak of COVID-19, such as masks, gatherings, local movements, international movements, etc. The data was collected from the World Health Organization. In this research, different variables have been used for analysis, for instance, new cases, new deaths, masks, schools, business, gatherings, domestic movement, international travel, new test, positive rate, test per case, new vaccination smoothed, new vaccine, total vaccination, and stringency index. Machine learning algorithms were used to predict and build the model, such as linear regression, K-nearest neighbours, decision trees, random forests, and support vector machines. Accuracy and Mean Square error (MSE) were used to test the model. A hyperparameter was also applied to find the optimum values of parameters. After computing the analysis, the result showed that the linear regression algorithm performs the best overall among the algorithms listed, with the highest testing accuracy and the lowest RMSE before and after hyper-tuning. The highest accuracy and lowest MSE were used for the best model, and for this data set, Linear regression got the highest accuracy, 0.98 and 0.97 and the lowest MSE, 4.79 and 4.04, respectively.

Keywords

Linear Regression, K-Nearest Neighbours, Decision Tree, Random Forest, Support Vector Machine, Hyper-Tuning

1. Introduction

According to the World Health Organization, the number of COVID-19 patients in Bangladesh is now recorded at about 2.05 million. This large number of people has been increasing for many reasons, called variables, which are included in my

study. The variables responsible for growing new cases are called independent variables, and the number of new cases is selected as a dependent variable. These large numbers of people faced many challenges, including people who had other diseases like diabetes and heart disease and pregnant women who also had an increased risk of death; they developed another disease due to COVID-19, like increased Allergy problems and breathing problems. Besides, those countries faced economic crises; hence, private company employees lost their jobs. As this problem involves public health, the main aim is to apply predictive modelling, including four different machine learning algorithms and identify the best one based on data. If this type of virus arises again with these responsible variables further, one can use this model for their study by referencing this study.

The 2019 new Coronavirus illness (COVID-19) was first reported in Wuhan, Hubei Province, China, at the end of December and quickly spread worldwide in weeks [1]. Using the official dataset of the Iranian Ministry of Health and Medical Education and considering the effect of control measures on the spread of COVID-19, this study tries to accurately forecast the number of new cases, fatalities, and recovered patients in Iran for 180 days. The best performing methodology for the particular case study is chosen after four distinct types of forecasting methodologies, time series, and machine learning algorithms, have been constructed. We take into account the four algorithms under the time series, including the Prophet, Long Short-Term Memory, Autoregressive, and Autoregressive Integrated Moving Average models. When we compared the various methodologies, we discovered that deep learning algorithms produce better outcomes than time series forecasting algorithms. More precisely, seasonal ANN and LSTM models are found to have the lowest value for error measures. According to our research, if precautions are taken seriously, there will be fewer new cases and fatalities overall, and there won't be any in September 2021 [1]. There are different types of machine learning algorithms explained including linear regression, decision tree, random forest, and support vector machine [2]-[4]. Foreseeing the global COVID-19 pandemic has been the subject of numerous research projects. Since COVID-19 spreads mostly through human interaction, it was essential to only take into account those nations with large populations and densities. For the top 10 most populous and densely inhabited nations, a COVID-19 outbreak prediction system was constructed using the methods described in this study. For our outbreak prediction algorithm, we took into consideration the top 10 nations with the highest population and densities. It was done using data from the following nations: Bangladesh, India, China, Pakistan, Germany, Nigeria, Ethiopia, Democratic Republic of the Congo, Philippines, and Indonesia. The proposed prediction models employ nine different machine learning algorithms, including ARIMA, ARMA, Support Vector Regression, Linear Regression, Bayesian Ridge Regression, Linear Regression (Polynomial), Random Forest, Holt Wailer Regression, and XGBoost Regression, to forecast the number of new cases that are likely to arise over the course of five consecutive days. As there were no discernible trends, it was impossible to obtain

the findings for each country using all 9 algorithms. We had very poor accuracy for the Philippines because there was a dramatic decrease of about 1400 cases to 0 and an increase of 4500 the next day in the case count. The COVID-19 count in the nation altered significantly as a result of the shift in government regulations as well [5]. A novel coronavirus outbreak is increasingly becoming more widely known day by day at an exponential rate. Flu-like symptoms were observed in people's health, along with coughing, dyspnea, weariness, and fever. Millions have died as a result of the epidemic, and some others were in danger in some circumstances [6]. The Ministry of India dataset and the Worldometer dataset from the months of February to July 2020 are utilized in this study to forecast the occurrence of COVID-19 instances. A variety of characteristics, including the quantity of affected individuals, confirmed cases, and fatalities, were projected at both the worldwide and national levels using the hybrid EAMA model. The number of ongoing cases and fatalities was forecast down to the state level, particularly in India [7]. Many practices have been implemented worldwide to stop the disease's transmission, including indoor confinement, social withdrawal, hand washing, travel restrictions, lockdowns, etc. Some of these measures, like lockdowns, are exceedingly severe, entail serious economic repercussions, and have an unparalleled impact on daily human activity. Several public health and social measures (PHSMs) have been implemented globally since early 2020 to lessen the COVID-19 pandemic brought on by SARS-CoV-2 [7]. The COVID-19 epidemic is a dilemma for low or middle-income countries because most citizens lack the savings or financial resources necessary to live without employment and afford necessities like rent and food. Given the uncertainty surrounding case statistics and the high risk of infection, predicting the number of infected people in a given area is challenging over the short term. The government is looking into new ways to organize medical equipment and set up hospitals to address this issue. This includes developing existing COVID-19 treatment centers and establishing brand-new ones. By taking these steps, the healthcare system will be better able to respond to the pandemic and give the populace essential medical care. It has become a top field of research globally due to the current emphasis on analyzing environmental data to predict future trends. The present and future states of the data can be projected or anticipated depending on the type of data prediction and analytical techniques used. Past or present data are analyzed using various modelling, statistics, data mining, artificial intelligence (AI), and machine learning approaches to predict future trends [7]. The dynamic method of infectious disease transmission takes place in a crowd. To accurately predict the future pattern of contagious diseases, frameworks can be constructed for this method to analyze and verify the transmission mechanism of contagious diseases. Hence, research into and evaluation of predictive models for infectious diseases has been a popular area in science to monitor or lessen the harm caused by contagious diseases [8]. The purpose of this article is to develop a simple average aggregated machine learning approach to forecast the quantity, magnitude, and duration of COVID-19 cases across India, as well as

their wind-up phase. On the basis of prediction precision, it was determined that the proposed strategy performed better than previously available practical models. Therefore, putting preventive measures in place can effectively control the spread of COVID-19, as well as lower and eventually eliminate the fatality rate in India and other countries [8]. Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment. However, some will become seriously ill and require medical attention, and some will die. The Institute of Epidemiology, Disease Control and Research (IEDCR) confirmed the first incidence in Bangladesh on March 8, 2020, and the number of infected cases rose quickly [9]. On March 8, 2020, the first case of the novel coronavirus illness (COVID-19) in Bangladesh was recorded. By July 2020, there had been over 175,000 confirmed cases. Several regression-based machine learning models were applied to infected case data in order to construct a cloud-based machine learning short-term forecasting model for Bangladesh that estimates the number of COVID-19-infected people over the course of the next seven days. By using the sample data from the previous 25 days that was recorded on our web application as training, our method can predict with accuracy the number of infected patients per day [9]. SARS-CoV2 is transmitted from person to person via personal contact, respiratory droplets, and touching contaminated surfaces; however, the most challenging aspect of COVID-19 transmission is that infected but asymptomatic patients are the primary source of the virus. Due to the population's low level of disease awareness, complicated and unreliable social-political elements accelerated the virus's rapid spread. With about 161.4 million inhabitants, Bangladesh has a high population density of around 1115 persons per square kilometer, with many residing in cramped cities and villages [9]. According to Arthur Samuel, in 1959, machine learning was the field of study that allowed computers to learn without being explicitly programmed. According to Tom Mitchell in 1997, A computer program is said to learn from experience E concerning some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E . Machine learning involves coding programs that automatically adjust their performance by their exposure to information in data. This learning is achieved via a parameterized model with tunable hyper-parameters that are automatically adjusted according to different performance criteria and make intelligent decisions. There are four types of machine learning supervised, unsupervised, semi-supervised, and re-enforcement learning. If output variables are labelled data and categorical, then classification problems are raised in supervised learning [10]. Data collected from the World Health Organization [11]. COVID-19 has sparked widespread worry about the epidemic's trajectory of growth and transmission. Much mathematical research examined the modelling of COVID-19's development and the impact of measures on the spread's containment [12].

The primary purpose of this study is to predict COVID-19 cases using different

machine-learning algorithms before and after tuning the parameters. Various machine learning algorithms worked together, including linear regression, K-nearest neighbours, decision trees, random forests, and support vector machine algorithms. To determine which algorithm performs best to predict new cases due to COVID-19. Also, try to find the best parameters using tuning systems to get the best model, accuracy, and MSE.

2. Methods and Materials

Regression prediction algorithms of Machine learning will be used to analyze the disease data. The secondary data was collected from WHO [11]. Using correlation and Spearman's correlation to check the association. The dataset consists of 15 variables such as new cases, new deaths, masks, schools, business, gatherings, domestic movement, international travel, new test, positive rate, test per case, new vaccination smoothed, new vaccine, total vaccination, and stringency index. New cases of this dataset are considered as dependent variables, and the remaining variables are considered independent variables. To investigate the trend and the relationship between deaths along with other independent factors, employed new deaths as an independent variables name. As a measure of disease severity, the inclusion of new deaths provides insight into the severity of the disease burden over the research period. Furthermore, examining the correlation between new fatalities and other elements (dependent variables name) aids in our comprehension of how these actions affect the course of the illness. In order to provide a more comprehensive understanding of the disease's dynamics and the efficacy of the interventions put in place, we sought to assess whether and how changes in mortality correspond with other independent factors by integrating new deaths. In analyzing the data, different machine learning algorithms were used, including linear regression, decision tree, random forest, K-nearest neighbors, and support vector machine. The dataset split into two types including training data and testing data. 75% of the whole data is considered as training dataset and 25% considered as testing dataset. Also, we used seed 100 for comparing the models. Cross validation is 5 used in hyperparameter tuning and used iterations number is 50. To select the best method, I used training accuracy, testing accuracy, accuracy score, and root mean sum square error (RMSE). The technique has the highest training accuracy, testing accuracy, and accuracy score, especially in greater than 90% and the lowest RMSE, so the method is the best-fitted model, otherwise rejected model. All of the included algorithms are used when response variable is quantitative. Here, new cases are quantitative variables. Linear regression is a useful tool for predicting a quantitative response. Through this algorithm, finding independent variables are responsible for changing COVID-19 new cases, and this algorithm is best or not for analyzing the data. The linear regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon.$$

Here the β 's are unknown parameters or coefficients [4]-[6]. K-Nearest Neighbors is a supervised machine-learning algorithm used for quantitative response variables.

In KNN, select the value of “K”; it may be 1, 3, 5, etc., and find Euclidean distance using the mean values of all variables. Using K, select the smallest number of values and then calculate the predicted mean values. This way, a new value is expected, the process is continued, and the model is prepared for prediction [2]-[4]. Decision trees and Random Forests are unique machine-learning algorithms for classification and regression problems. A decision tree is used for a single tree, and a random forest is used for multiple decision trees. The final tree is selected using the ID3 technique, Entropy, Information Gain, and Gini Index. Each tree comprises root nodes, internal nodes, and leaves, which predict the final result [10]. Support Vector Machine is another machine-learning algorithm used for classification and regression problems. The dataset is separated using a maximal margin classifier if the dataset is linearly separable; otherwise, the kernel function is applied to transform the separable function. There are different types of kernel functions, including linear, polynomial, RBF, gaussian, etc. [10]. Hyper-tuning is a technique used to select the optimal parameter for the dataset of algorithms chosen that fit the model. There are several types of tuning methods, such as Grid Search, Random search, Bayesian Optimization, Genetic Algorithm, and Gradient descent and in this research, the Grid search method is applied.

3. Results and Discussion

3.1. Descriptive Information

Information on several variables pertaining to a certain issue is included in the data. An overview of each variable is provided below.

Table 1 shows the number of newly reported cases. There are 2181.9 new instances on average, with 1542.5 being the median. A minimum of 0 and 16,230.0 new instances have been reported. There has been a variance of $7.221269e+06$ in the number of new cases reported. The number of newly reported deaths have the median and mean number of new fatalities, respectively, are 24 and 34.84. The total number of newly reported deaths ranges from 0 to 264. The variation in newly reported fatalities is $1.007076e+05$. This variable represents the number of fresh tests run. The median is 14,140, with 15,822 new tests being completed on average. The total number of newly executed tests ranges from 0 to 55,284. $9.434399e+07$ is the variance of the number of new tests performed. The test positive rate has the median and mean favorable rates, which are both 0.1186. The positivity rate ranges from zero (the smallest) to 0.3037 (the maximum). The positivity rate’s variance is $5.996847e-03$. The variable test per case shows how many tests are performed for each positive instance. The median and mean number of tests per positive case are both 8.2 and 14.24, respectively. The minimum and maximum tests per positive case are 0 and 94.10, respectively. The variance of the tests performed for each positive instance is $3.118024e+02$. This variable indicates the total number of vaccinations given. The median is zero, with 10,557,614 vaccines provided as the mean. The lowest and highest vaccine doses are 0 and 129,277,282, respectively. $6.857820e+14$ is the variance in the number of immuniza-

tions given.

Table 1. Summary of different independent variables.

Variable names	Mean	Median	Variance	Minimum	Maximum
New cases	2181.9	1542.5	7.221269e+06	0	16230.0
New deaths	34.84	24	1.007076e+05	0	264.00
New test	15822	14140	9.434399e+07	0	55284
Positive rate	0.1212	0.1186	5.996847e-03	0	0.3037
Test per case	14.24	8.2	3.118024e+02	0	94.10
Total vaccination	10557614	0	6.857820e+14	0	129277282

Table 2. Summary of different independent variables.

Variable names	Least severe	Moderate	Most severe	Not applicable	Total percentage	Total
Masks	0	0	92.1	7.9	100.0	989
Schools	0.9	0	58.6	40.4	100.0	989
Business	24.6	16.8	9.2	49.4	100.0	989
Gatherings	7.7	11.4	51.5	29.4	100.0	989
Domestic movement	15.0	8.3	7.6	69.2	100.0	989
International travel	26.3	2.7	14.1	56.9	100.0	989

Table 2 represents that the least severe means that COVID-19 has had a very small or no impact on that factor. Moderate means that COVID-19 may have been affected by this factor. Most Severe means that COVID-19 has significantly or negatively impacted that factor. Not applicable means that COVID-19 has not affected that factor or is irrelevant to the analysis. From the table, 92.1% of people were affected with COVID-19 who did not use a mask and 7.9% of people were not affected with COVID-19 who did not use a mask. 58.6% of people were affected COVID-19 who went to school, 0.9% of people were affected by COVID-19 who went to school, and 40.4% of people were affected by COVID-19 who did not go to school. 9.2% of people affected by COVID-19 who did business, 24.6% of people affected by COVID-19 who did business, 16.8% of people affected by COVID-19 who did business, and 49.4% of people not affected by COVID-19 who did not do business. 51.5% of people were affected with COVID-19 who went to public gatherings, 29.4% of people were affected with COVID-19 who did not go to public gatherings, 7.7% of people were affected with COVID-19 who went to public gatherings, and 11.4% of people were affected with COVID-19 who went to public gatherings. 7.6% of people affected by COVID-19 who had domestic movements, 69.2% of people were affected COVID-19 who did not have domestic movements, 15% of people affected COVID-19 who had international travel, and 8.3% of people were not affected by COVID-19 who had international travel.

14.1% of people were affected with COVID-19 who had international travel, 56.9% of people were affected with COVID-19 who had international travel, 26.3% of people were affected by COVID-19 who had international travel, and 2.7% of people were not affected with COVID-19 who had international travel.

3.2. Correlation Analysis

Table 3. Association of COVID-19 cases with vaccination, test & policy, and policy responses.

Variable name	Correlation with No. of daily cases	P-value
New deaths	0.8	<2.2e-16
Masks	0.49	<2.2e-16
Schools	0.58	<2.2e-16
Business	0.33	<2.2e-16
Gatherings	0.5	<2.2e-16
Domestic movement	0.23	2.059e-15
International travel	0.37	<2.2e-16
New test	0.76	<2.2e-16
Positive rate	0.76	<2.2e-16
Test per case	-0.41	<2.2e-16
New vaccination smoothed	0.01	0.7778
New vaccine	0.02	0.5943
Total vaccination	-.18	6.975e-10
Stringency index	0.37	<2.2e-16

Table 3 shows that new-vaccination-smoothed and the number of new vaccines is not significant because the p-value of the test is higher than 0.05 (0.7778 and 0.59), indicating that the observed correlation is statistically insignificant at any reasonable significance level. The other variables are significant because the p-value of the test is very small, indicating that the observed correlation is statistically significant at any reasonable significance level.

Table 4 provided that new-vaccination-smoothed and no. of new-vaccine are not significant because the p-value of the test is higher than 0.05 (0.75 and 0.4218), indicating that the observed correlation is statistically insignificant at any reasonable significance level. The other variables are significant because the p-value of the test is very small, indicating that the observed correlation is statistically significant at any reasonable significance level. Therefore, these two variables should be subtracted for good analysis.

3.3. Machine Learning Algorithms

The estimated model is

Table 4. Association of COVID-19 deaths with vaccination, test & policy, and Policy responses.

Variable name	Correlation with No. of daily death	P-value
New cases	0.8	<2.2e-16
Masks	0.4	<2.2e-16
Schools	0.69	<2.2e-16
Business	0.46	<2.2e-16
Gatherings	0.63	<2.2e-16
Domestic movement	0.34	<2.2e-16
International travel	0.47	<2.2e-16
New test	0.59	<2.2e-16
Positive rate	0.65	<2.2e-16
Test per case	-0.37	<2.2e-16
New vaccination smoothed	-0.14	0.75
New vaccine	-0.02	0.4218
Total vaccination	-.26	<2.2e-16
Stringency index	0.38	<2.2e-16

New Cases

$$\begin{aligned}
 &= -5.94972337 - 0.00546254 \times \text{new deaths} + 0.0015042 \times \text{new test} \\
 &\quad + 0.00158245 \times \text{positive rate} - 0.012561 \times \text{test per case} \\
 &\quad - 0.00003016 \times \text{new vaccination smoothed} - 0.000005718 \times \text{new vaccine} \\
 &\quad + 0.000000407 \times \text{total vaccination} + 0.079399 \times \text{stringency index}
 \end{aligned}$$

When all the independent variables are constant, the number of new cases will decrease by 5.95 units. When one unit increases new deaths, the new cases will decrease by 0.00543, holding all other independent variables. When one unit increases the new test, the new cases will increase by 0.0015042 units, holding all other independent variables constant. When one unit increases the favorable rate, then the new cases will increase by 0.00158245 units, holding all other independent variables constant. When one unit increases the test per case, then the new cases will decrease by 0.012561 units, holding all other independent variables constant. When one-unit increases, new vaccinations are smoothed, and the new cases will decrease by 0.00003016 units, holding all other independent variables constant. When one unit increases the new vaccine, then the new cases will decrease by 0.000005718 units, holding all other independent variables constant. When one unit increases total vaccination, then the new cases will increase by 0.000000407 units, holding all other independent variables constant. When one unit increases the stringency index, then the new cases will increase by 0.079399 units, holding all other independent variables constant.

According to **Table 5**, the linear regression algorithm performs the best overall

among the algorithms listed, with the highest testing accuracy (95.6%) and the lowest RMSE (4.794). Here, decision trees and random forests have high training accuracy, testing accuracy, and accuracy scores, but RMSE is comparatively far from linear regression. Hence, it has been said that Linear regression is the best-fitted model for predicting the COVID-19 dataset in Bangladesh.

Table 5. Accuracy of before hyper-tuning.

Algorithm	Training accuracy	Testing accuracy	Accuracy score	RMSE
Decision tree	0.998	0.958	0.986	345.937
Random forest	0.993	0.979	0.989	300.976
Linear regression	0.981	0.956	0.943	4.794
KNN	0.881	0.710	0.831	12.257
SVM	-0.098	-0.118	-0.108	3034.745

Table 6. Accuracy of after hyper-tuning.

Algorithm	Training accuracy	Testing accuracy	Accuracy score	RMSE
Decision tree	0.991	0.964	0.983	594.209
Random forest	0.995	0.979	0.990	417.862
Linear regression	0.973	0.975	0.975	4.404
KNN	0.999	0.745	0.940	9.526
SVM	0.250	-0.010	0.172	2500.091

After tuning, according to **Table 6**, the linear regression algorithm also performs the best overall among the algorithms listed, with the highest training accuracy (97.3%), testing accuracy (97.5%), accuracy score (97.5%), and the lowest RMSE (4.404). In decision trees and random forests, training accuracy, testing accuracy, and accuracy score are high, but RMSE is not low enough to say it fits the model. In KNN, training accuracy and accuracy scores are high, RMSE is low but not less than linear regression, and testing accuracy is not greater than 90%. We cannot say that KNN is a good-fit model. In SVM, all the parameter values are not good enough, so it is also not a good-fit model. So, overall, linear regression is the best-fitted model.

So, comparing **Table 5** and **Table 6**, we found that the best method is linear regression before hyper-tuning and after hyper-tuning both.

4. Conclusions

The patients' average positive attack rate is 12.12%, with a maximum rate of 30%. The average number of identified new cases is 2182, with the maximum number of new cases being 16,230. The average new death is 35, with the maximum death

is 264. The critical factor that most seriously causes COVID-19 to increase if patients don't use masks at an effective rate is the highest (92.1%). The second essential factor is that students going to school will be affected by COVID-19, which is 58.6%. The third important factor is Gathering, which is affected by COVID-19 at 51.5%. The least important factors are domestic movement and business, which are less affected by COVID-19, with 7.6% and 9.2%, respectively. Among the 15 selected variables, according to the correlation test, the two variables are identified as insignificant, new-vaccination-smoothed, and the number of new vaccines, and the other 13 variables are significant. The variables, including new test, positive rate, total vaccine, and stringency index, are positively associated with increased new cases, with the most increased factor being the stringency index. On the other hand, the variables, including new test, test per case, new vaccination smoothed, and new vaccine, were negatively associated with decreased new instances.

The linear regression model is the best-fitted model in both cases, such as before hyper-tuning and after-hyper-tuning, as training accuracy, testing accuracy, and accuracy scores are high compared to others, and RMSR is low among all algorithms. So, overall, the number of COVID-19 patients will increase due to some factors, which are analyzed with the linear regression model, which is the best method for prediction.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Abdollahi, J., Irani, A.J. and Nouri-Moghaddam, B. (2021) Modeling and Forecasting Spread of COVID-19 Epidemic in Iran until Based on Deep Learning.
- [2] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021) An Introduction to Statistical Learning with Applications in R. 2nd Edition, Springer.
- [3] Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Vol. 2, Springer, 1-758.
- [4] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning. Vol. 112, Springer, 18.
- [5] Khakharia, A., Shah, V., Jain, S., Shah, J., Tiwari, A., Daphal, P., *et al.* (2020) Outbreak Prediction of COVID-19 Patients for Dense and Populated Countries Using Machine Learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3655759>
- [6] Mohammad Masum, A.K., Khushbu, S.A., Keya, M., Abujar, S. and Hossain, S.A. (2020) COVID-19 in Bangladesh: A Deeper Outlook into the Forecast with Prediction of Upcoming per Day Cases Using Time Series. *Procedia Computer Science*, **178**, 291-300. <https://doi.org/10.1016/j.procs.2020.11.031>
- [7] Mohan, S.A.J., Abugabah, A.M.A., Kumar Singh, S., kashif Bashir, A., *et al.* (2021) An Approach to Forecast Impact of Covid-19 Using Supervised Machine Learning Model. *Software: Practice and Experience*, **52**, 824-840. <https://doi.org/10.1002/spe.2969>
- [8] Ogundokun, R.O. and Awotunde, J.B. (2020) Machine Learning Prediction for COVID-19 Pandemic in India.

- [9] Satu, M.S., Howlader, K.C., Mahmud, M., Kaiser, M.S., Shariful Islam, S.M., Quinn, J.M.W., *et al.* (2021) Short-Term Prediction of COVID-19 Cases Using Machine Learning Models. *Applied Sciences*, **11**, Article No. 4266.
<https://doi.org/10.3390/app11094266>
- [10] Hastie, T., Tibshirani, R. and Friedman, J. (2008) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edition, Springer.
- [11] WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/>
- [12] Xiang, Y., Jia, Y., Chen, L., Guo, L., Shu, B. and Long, E. (2021) COVID-19 Epidemic Prediction and the Impact of Public Health Interventions: A Review of COVID-19 Epidemic Models. *Infectious Disease Modelling*, **6**, 324-342.
<https://doi.org/10.1016/j.idm.2021.01.001>