The prediction of coronavirus disease 2019 outbreak on Bangladesh perspective using machine learning: a comparative study

Maqsudur Rahman[1], Md. Tofael Ahmed[1], Shafayet Nur[2], Abu Zafor Muhammad Touhidul Islam[3]

[1]Department of Information and Communication Technology, Faculty of Engineering, Comilla University, Comilla, Bangladesh
[2]Department of Computer Science and Engineering, Faculty of Science and Engineering, Port City International University, Chattogram, Bangladesh
[3]Department of Electrical and Electronics Engineering, Faculty of Engineering, Rajshahi University, Rajshahi, Bangladesh

Article Info

ABSTRACT

Coronavirus disease 2019 (COVID-19) has made a huge pandemic situation in many countries of the world including Bangladesh. If the increase rate of this threat can be forecasted, immediate measures can be taken. This study is an effort to forecast the threat of present pandemic situation using machine learning (ML) forecasting models. Forecasting was done in three categories in the next 30 days range. In our study, multiple linear regression performed best among the other algorithms in all categories with R2 score of 99% for first two categories and 94% for the third category. Ridge regression performed great for the first two categories with R2 scores of 99% each but performed poorly for the third category with R2 score of 43%. Lasso regression performed reasonably well with R2 scores of 97%, 99% and 75% for the three categories. We also used Facebook Prophet to predict 30 days beyond our train data which gave us healthy R2 scores of 92% and 83% for the first two categories but performed poorly for the third category with R2 score of 34%. Also, all the models' performances were evaluated with a 40-day prediction interval in which multiple linear regression outperformed other algorithms.

Corresponding Author:

Maqsudur Rahman
Department of Information and Communication Technology, Faculty of Engineering, Comilla University
Comilla, Bangladesh
Email: mrrajon@cou.ac.bd

1. INTRODUCTION

The coronavirus 2 is severe acute respiratory syndrome which is the main cause of coronavirus disease 2019 (COVID-19), an ongoing pandemic in many countries of the world [1]. Wuhan, Hubei, China is the place where this disease was first recorded in December 2019. About 36 million new cases and more than 1 million fatalities have been reported worldwide till date [2]. Many researchers and scientists are working hard to find a vaccine to prevent this virus. Because of this pandemic, the moral of people have crashed. Many countries are facing economic challenges. However, Bangladesh is a country of lower middle economy and also one of the most populated countries of the world is hit by this devastating pandemic. The world is taking precautions to lower the threat of this pandemic. So, it has become very essential also for Bangladesh to forecast the threat of the current situation which will help in taking precautions which can save many lives. Machine learning (ML) based predictive models can be a promising solution in this case. ML is proven to be a prominent field of study and also ML forecasting models can be utilized to analyze data to forecast future

scenario with great accuracy and performance. Various regression and neural network models are used to forecast threats of different diseases.

Lapuerta *et al.* [3] predicted the risk of coronary artery disease using neural network and Voyant *et al.* [4] presented methods for solar radiation forecasting using machine learning. Moreover, prediction equations for several cardiovascular disease endpoints were presented by Anderson *et al.* [5]. Furthermore, to predict breast cancer risks and to diagnosis breast cancer Bharat *et al.* [6] used ML models. Farhana *et al.* [7] presented an intrusion detection system using deep learning approach. These models can also be used to forecast the threats of COVID-19 as well like Yadav *et al.* [8] analyzed the risk factors of COVID-19 using machine learning approach and Rustam *et al.* [9] forecasted worldwide COVID-19 situation through the supervised machine learning approach and got the accuracy up to mark. Zeroual *et al.* [10] also used deep learning approach to forecast COVID-19 time series data. Besides many more research is done in this field recently to find out the possible friendly pattern. The following are some detailed discussions on some of the worldwide works on COVID-19 threat prediction.

Rustam *et al.* [9] forecasted COVID-19 threats in three categories, number of new positive cases, the number of deaths and the number of recoveries of upcoming 10 days. They used supervised ML models linear regression, lasso regression, support vector machine (SVM) and exponential smoothing to fulfill their work. In their study, exponential smoothing performed best among other algorithms and got R2 scores of 98%, 98% and 99% for the three categories. In this research, prediction was done for upcoming 10 days only but in our research, we predicted for upcoming 30 days.

Furthermore, Shastri *et al.* provided a comparative study of COVID-19 outbreak in India and USA [11]. They proposed a deep learning based predictive system and forecasted COVID-19 cases for 30 days ahead. Prediction was done in this research only in two categories for two countries such as confirmed cases and death cases. In our research, we contributed one more category which is the number of daily new recovered patients.

There were also some other previous works related to COVID-19 forecasting worldwide, like Wang *et al.* [12] used patient information based algorithm to forecast real-time prediction of mortality caused by COVID-19 and in Alazab *et al.* [13] analyzed the incidence of COVID-19 distribution across the world by presenting artificial-intelligence technique based on a deep convolutional neural network to detect COVID-19 patients from the real-world datasets. In [14], Elmousalami and Hassanien utilized time series models and numerical detailing to represent the correlation of the day level to determine COVID-19 cases. Furthermore, a flower pollination algorithm by using the salp swarm algorithm-adaptive neuro-fuzzy inference system (FPASSA-ANFIS) model and flower pollination algorithm (FPA) was used by Al-Qanees *et al.* [15] for predicting the confirmed cases for the upcoming 10 days in Chia and USA using World Health Organization (WHO) official dataset. Yudistira [16], suggested more effective models to handle the nonlinearity and the complexity of the COVID-19 time-series data. Table 1 contains a list of some previous works on the field of predicting worldwide COVID-19 risk factors and their followed methods. The following are some works related to COVID-19 prediction in Bangladesh.

Table 1. Previous worldwide works related to COVID-19 prediction

Article	Methods	Results
[7]	Supervised machine learning Models	Exponential Smoothing performed best with R2 scores 98% for new cases and new deaths and 99% for recovery rate
[9]	Recurrent neural network (RNN) based variants LSTM such as Stacked LSTM, Bi-directional LSTM and convolutional LSTM	97% (Convolutional LSTM for confirmed cases of India-USA)
[10]	RNN, long short-term memory (LSTM), Bidirectional LSTM, gated recurrent units (GRUs) and VAE algorithms	VAE outperformed all other algorithms
[11]	Prophet algorithm (PA), 0ARIMA, LSTM	99% accuracy (Prophet algorithm for confirmed cases in Australia and Jordan)
[12]	Comparison of day level forecasting models	Single Exponential Smoothing Performed more accurately
[13]	An improved ANFIS using an enhanced FPA by using the SSA	R2 score of 96% (FPASSA-ANFIS method)
[14]	LSTM	RMSE score of 1238.66 (LSTM)

Chowdhury *et al.* [17] foresaw the newly infected cases in Bangladesh. They used the adaptive neuro fuzzy inference system (ANFIS) and the long short-term memory (LSTM) in three scenarios where they took input of last four consecutive odd days for the first, even days for the second and last four consecutive days for the third scenario in evaluated the models. In their study, LSTM showed good root mean square error (RMSE) value and correlation coefficient in forecasting the outbreak. In this research

prediction was done only for newly infected cases but, in our research, we performed prediction for newly infected cases as well as daily new fatality and daily new recovery.

The study in [18] by Mahmud aims to predict and analyze the daily cases data of Bangladesh using the Facebook Prophet model for which they collected data from 9th march to 21st July and prediction was done up to 19th September with a R2 value of 89%. In our research we got 3% more accuracy for Facebook Prophet in daily new infected cases. Data-driven estimation methods were presented by Hridoy *et al.* in [19], for predicting the possible number of COVID-19 cases in Bangladesh. They used data from 8th March 2020 to 13th June 2020 for LSTM models and 8th March 2020 to 18th June 2020 for logistic curve fitting model. They used 3 LSTM models such as vanilla, stacked, bidirectional and got good R2 scores for total confirmed cases, total recovered cases, and total deaths prediction. They predicted in three categories but their data was limited.

In [20], Rahman *et al.* projected final infected number of COVID-19 patients in Bangladesh using standardized infection ratios (SIR) mathematical model with data collected from IEDCR Bangladesh. They obtained data from March 8 to April 20 and got reasonable results. This research lacks more data. This research also lacks the prediction in more than 1 category. Table 2 contains the list of previous works on the field of COVID-19 in Bangladesh and their followed methods.

Table 2. Previous works related to COVID-19 in Bangladesh

Article	Methods	Results
[17]	ANFIS and LSTM	0.75 correlation coefficients (LSTM)
[18]	Facebook Prophet	R2 score of 89%
[19]	LSTM networks and Logistic Curve methods	R2 score of 95% (Stacked LSTM)
[20]	SIR mathematical model	R2 score of 87%

From the discussed works of COVID-19 prediction in Bangladesh, [17] and [18] only predicted the newly infected cases and [20] predicted final or total infected cases which motivated us to predict in more categories like daily new confirmed cases, daily new fatality, and daily new recovery. Hridoy *et al.* [19] predicted total confirmed cases, total deaths and total recovery with limited data and got some reasonable R2 scores as their data was limited. In our study we added more data for training purpose and also predicted for upcoming 30 days which helps in making more suitable decisions. The detailed methodology is discussed in section 2.

2.   RESEARCH METHOD

As some of the previous works for predicting COVID-19 in Bangladesh lacks more categories in prediction like daily new fatality and daily new recovery and some of the works lack of updated and more data, we propose a machine learning based approach for predicting COVID-19 Bangladesh cases in three categories with updated data from 17th March to 30th September with great accuracy. In this paper, the predictions are done in three categories such as daily newly infected cases, daily new fatality, and daily new recovered patients in the next 30 days range. For that purpose, firstly we obtained the dataset and performed preprocessing and feature selection before training the data. For training, we used multiple linear regression, ridge regression, lasso regression and Facebook prophet and implemented using python programming language. After training, we evaluated the performance of the training model in terms of R^2 score, R^2 adjusted score, mean square error (MSE), mean absolute error (MAE), and RMSE. The proposed methodology shown in Figure 1.

2.2. Dataset

The most important phase of any research is the collection of data. We used two datasets in our study. We collected the datasets from GitHub repository maintained by our world in data [21] and the official website of WHO [22]. Both the datasets contains daily COVID-19 updates of more than 200 countries including Bangladesh. As the datasets contained data of countries other than Bangladesh, we extracted the data of Bangladesh from both datasets and combined them. We also confirmed the authenticity of the data with daily updates given by the Ministry of Health of Bangladesh [23].

2.3. Preprocessing and feature selection

As the dataset contained a lot of null values, we replaced all the null values with 0. For example, the number of total deaths on the 17th of March 2020 in the dataset was represented as null which means there were no deaths that day. So, null values were replaced by 0. Also, the dataset was standardized in some

situations before training. Some of the features of the dataset were necessary and some were not. The unnecessary features were removed and necessary features were used for training the models. The features which were necessary in a particular model was used in that particular model.
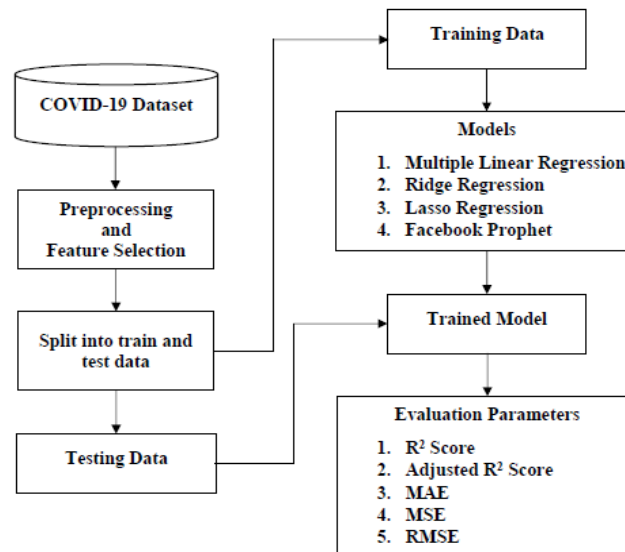

Figure 1. Proposed methodology

2.4. Splitting the dataset into train and test set

The dataset contains COVID-19 daily updates of Bangladesh from 17th March to 30th September. After the initial preprocessing and feature selection steps, we divided the dataset into two subsets, train, and test set. The training dataset contains data of COVID-19 from 17th March to 31st August (167 days) and the testing dataset contains data of 1st September to 30th September (30 days). After that the training dataset was used in training the models and testing dataset was used to test the models.

2.5. Machine learning models
2.5.1. Multiple linear regression model

Multiple linear regression, which is also known by the name multiple regression, is an extension of linear regression, which takes multiple independent variables as input (input variables) and finds the relation between the independent variables and a dependent variable (output) [24]. The relation between the independent variables and the dependent variable can be defined by (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_p x_p + \epsilon \tag{1}$$

Here, β_0 is the intercept and $\beta_1, \beta_2, \dots \beta_p$ are the coefficients, p is the number of independent variables, $x_1, x_2, \dots x_p$ are the independent variables, ϵ is the error term and y is the dependent variable that is to be predicted.

2.5.2. Ridge regression

Ridge regression is regularization technique, which reduces model complexity using a penalty score and computational expenses [25].

$$Ridge\ Regression = Loss + \alpha\ \|w\|^2 \tag{2}$$

Here, loss is measured from the difference between predicted value and actual value. Penalty is calculated in terms of product of α which is a constant value and $\|w\|^2$ where w is simply a vector of the coefficients. Penalty reduces loss and scales down coefficient magnitude.

$$\|w\|^2 = w_1^2 + w_2^2 + \cdots + w_p^2 \tag{3}$$

Here $w_1, w_2, \dots w_p$ are coefficients.

2.5.3. Lasso regression

Like ridge regression, Lasso regression is also a regularization technique which reduces model complexity using a penalty score [26]. It is defined in (4).

$$Lasso\ Regression = Loss + \alpha \|w\| \tag{4}$$

Same as ridge regression, loss is calculated from the difference between predicted value and actual value also. This loss is reduced using a penalty score. Penalty is calculated in terms of product of α which is a constant value and $\|w\|$ where w is simply a vector of the coefficients. Penalty reduces loss and scales down coefficient magnitude. Calculation of $\|w\|$ is shown in (5).

$$\|w\| = w_1 + w_2 + \dots + w_p \tag{5}$$

Here $w_1, w_2, \dots w_p$ are coefficients.

2.5.4. Facebook Prophet

Facebook Prophet was developed by Facebook's core data science team which is largely used in forecasting time series based on daily, weekly, monthly, or yearly trend [27]. This algorithm is open source and is used for producing many reliable forecasting outputs across Facebook. Forecast is done using a logistic growth trend model with a specified carrying capacity using Facebook Prophet. We implemented these algorithms using scikit-learn and Fbprophet library of python.

2.6. Evaluation parameters

We evaluated the accuracy of our model in terms of R-Squared (R2) Score [28], R-squared adjusted score (R^2_{adj}) [28], MSE [29], MAE [30] and RMSE. The equations (6)-(10) are used to calculate these evaluation parameters.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

$$R^2_{adj} = 1 - \frac{(1-R^2)(N-1)}{N-P-1} \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{10}$$

3. RESULTS AND DISCUSSION

3.1. Daily new infected cases forecasting

Figure 2 shows the daily new infected cases in Bangladesh from 17th March to 31st August (167 days). In our study, multiple linear regression and ridge regression performed best among the four models in predicting daily new cases. Table 3 contains the performance evaluation of all the models. Table 3 shows that multiple linear regression and ridge regression both performed equivalently and obtained R2 score of 0.99. They also got the same amount of MSE, MAE and RMSE error which are lowest among all algorithms. Facebook Prophet performed lowest among other algorithms in predicting daily new infected cases and got 0.92 R2 score with highest MSE, MAE and RMSE error scores. Figures 3(a) and 3(b) shows the performance of multiple linear regression, ridge regression, Lasso regression and Facebook Prophet in predicting new cases in the form of graph. The prediction graph of linear regression in Figure 3 is trending downwards at the beginning which means that new cases are decreasing but at the end of the graph we can see an upward trend which indicated that the daily new cases can increase in the upcoming days.
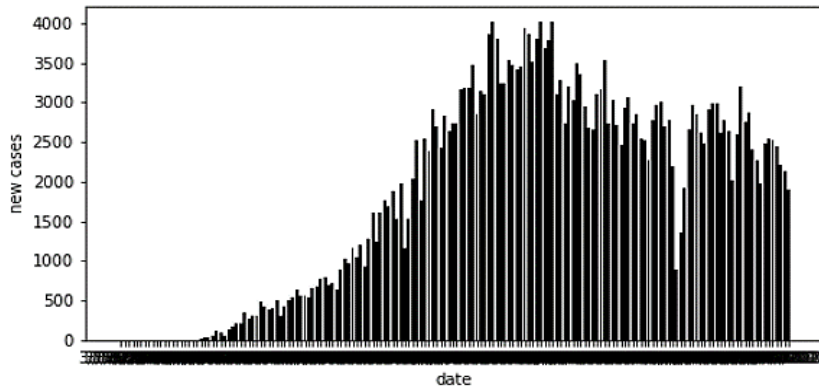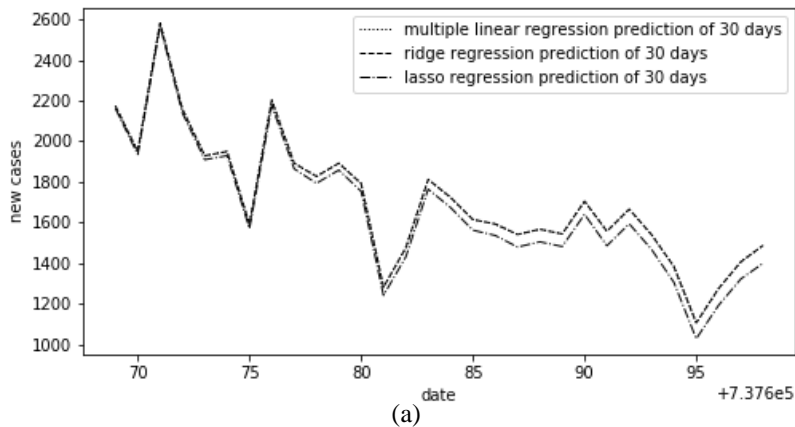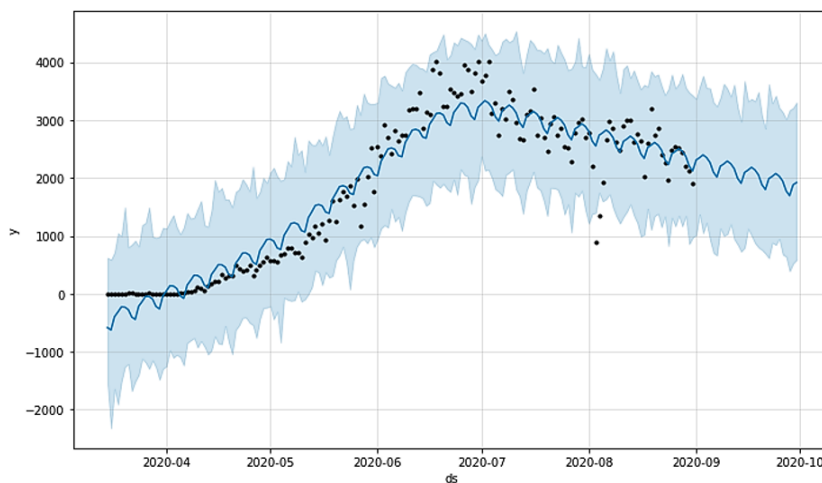
Figure 2. Daily new infected cases of 17th March to 31st August

Table 3. Performance of the models in new cases forecasting

MODELS	R2	R2 Adjusted	MSE	MAE	RMSE
MLR	0.99	0.99	0.004	0.05	0.06
RIDGE	0.99	0.99	0.004	0.05	0.06
LASSO	0.97	0.94	2884.15	48.27	53.70
PROPHET	0.92	0.92	130469.30	272.32	361.20

(a)

(b)

Figure 3. Prediction of new cases by (a) multiple linear regression, ridge regression, lasso regression and (b) Facebook Prophet

3.2. Daily new fatality cases forecasting

Figure 4 shows the daily new fatality in Bangladesh from 17th March to 31st August (167 days). In our study, multiple linear regression, ridge regression and Lasso regression performed best and same among the four models in predicting daily new fatality. Table 4 contains the performance evaluation of all the models.
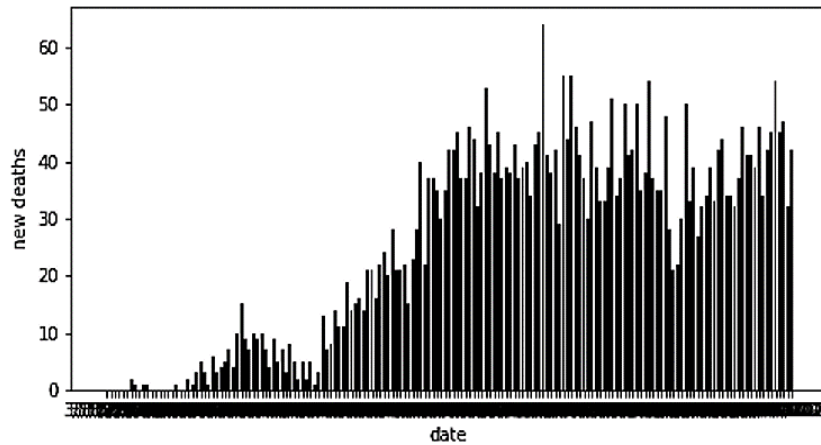
Figure 4. Daily new fatality of 17th March to 31st August

Table 4 shows that multiple linear regression, ridge regression and Lasso regression performed equivalently and obtained R2 score of 0.99. Multiple linear regression and ridge regression both got the lowest and same amount of MSE, MAE and RMSE error among all algorithms. Facebook Prophet performed lowest among all algorithms in predicting daily new fatality and got 0.83 R2 score with highest MSE, MAE and RMSE error scores.

Table 4. Performance of the models in new fatality forecasting

Models	R^2	R^2_{adj}	MSE	MAE	RMSE
MLR	0.99	0.99	0.003	0.05	0.06
RIDGE	0.99	0.99	0.003	0.05	0.06
LASSO	0.99	0.99	0.007	0.06	0.085
PROPHET	0.83	0.83	50.64	5.43	7.11

Figure 5(a) and 5(b) shows the performance of multiple linear regression, ridge regression, Lasso regression and Facebook prophet in predicting new fatality in the form of graph. As the three models got the same accuracy, we can see that the line plots overlapped one another. If we look at the daily new fatality graph of multiple linear regression in Figure 6, it indicated that the number of deaths will not fluctuate but increase of new confirmed cases will definitely increase the daily new deaths.

3.3. Daily new recovery forecasting

Figure 6 shows the daily new recovery in Bangladesh from 17th March to 31st August (167 days). In our study, multiple linear regression performed best among the four models in predicting daily new recovery. Table 5 contains the performance evaluation of all the models.

Table 5 shows that multiple linear regression performed best in prediction daily new recovery by obtaining R2 score of 0.99. Multiple linear regression also got the lowest amount of MSE, MAE and RMSE error among all algorithms. Facebook prophet performed very poorly by obtaining 0.34 R2 score with huge MSE, MAE and RMSE error scores.

Figure 7(a) and 7(b) shows the performance of multiple linear regression, ridge regression lasso regression and Facebook Prophet in predicting daily new recovery in the form of graph. In case of daily new recovery graph of multiple linear regression shown in Figure 8, there is a little bit fluctuation at end of the curve. More hospital facilities will definitely increase the number of daily new recovery.
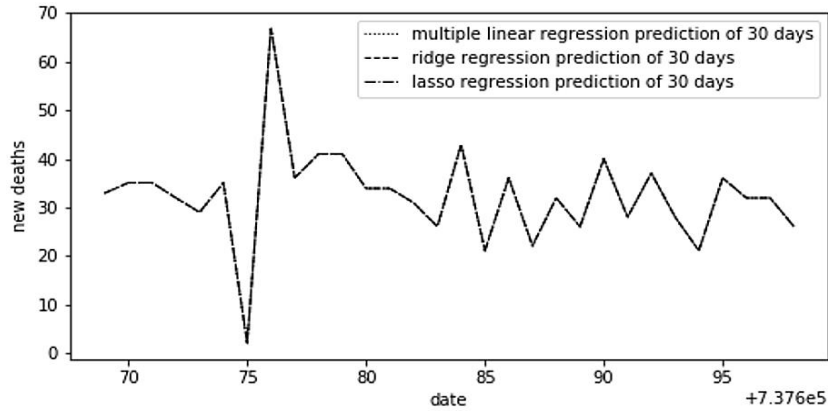
(a)
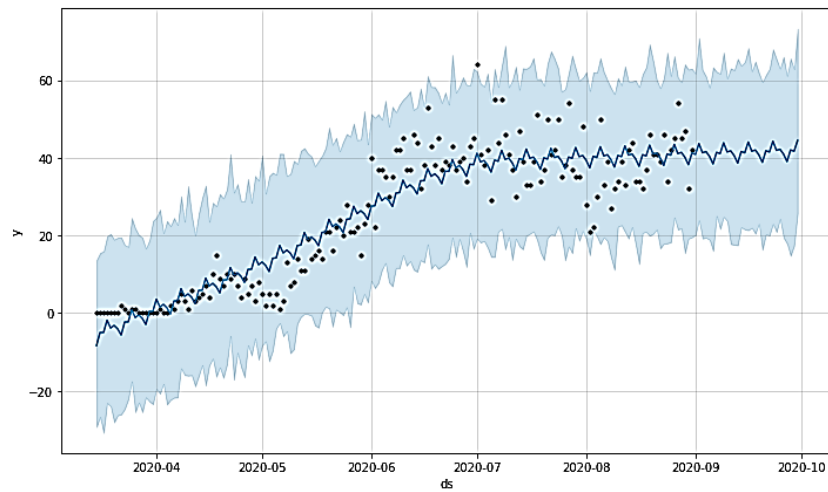
(b)

Figure 5. Prediction of new fatality by (a) multiple linear regression, ridge regression, lasso regression and (b) Facebook Prophet
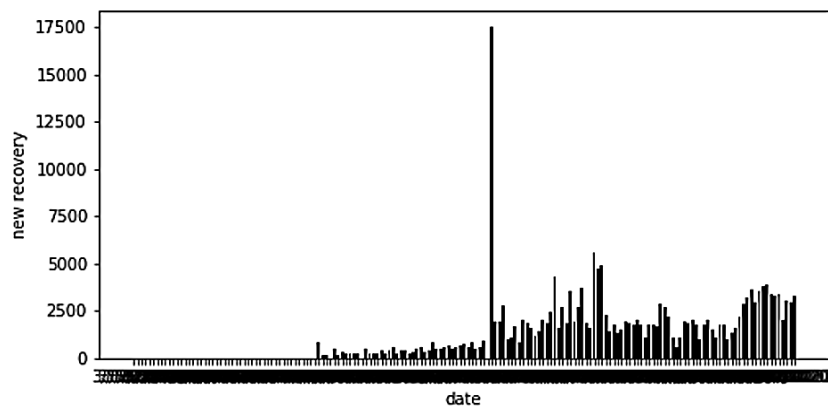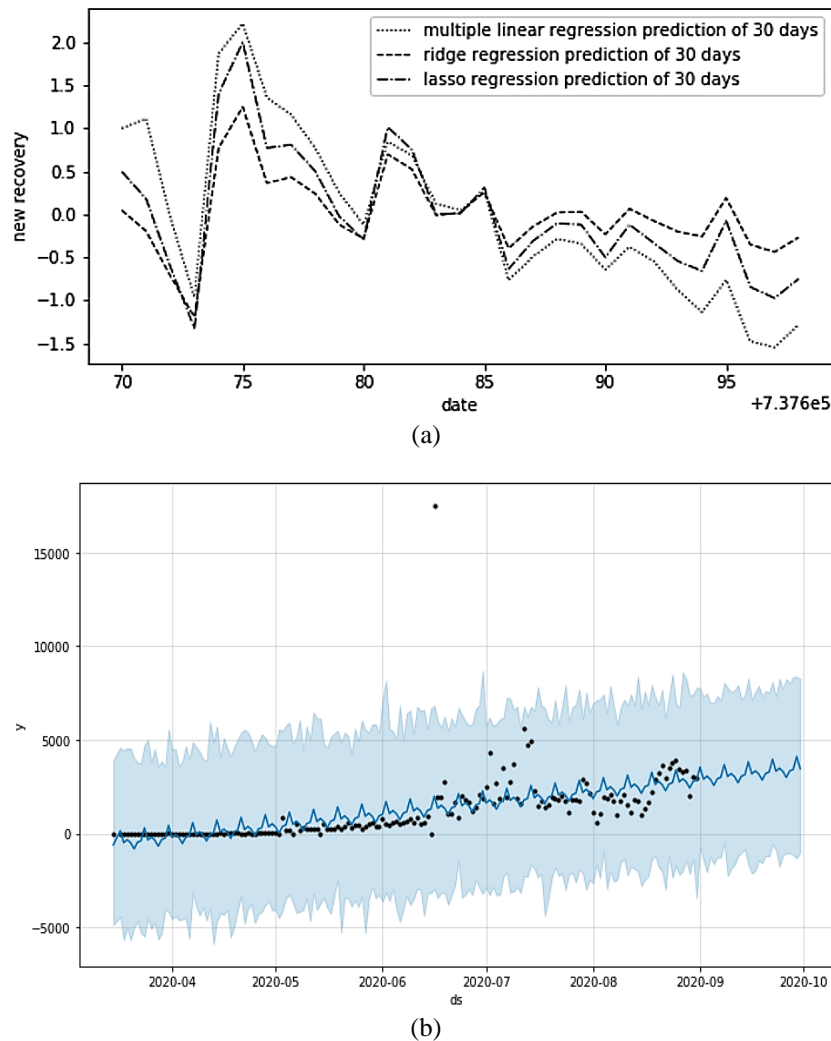

Figure 6. Daily new recovery cases of 17th March to 31st August

Table 5. Performance of the models in new recovery forecasting

MODELS	R2	R2 Adjusted	MSE	MAE	RMSE
MLR	0.94	0.94	0.05	0.16	0.23
RIDGE	0.43	0.42	0.56	0.61	0.74
LASSO	0.75	0.74	0.24	0.38	0.49
PROPHET	0.34	0.33	2026000.12	652.41	1423.37

(a)

(b)

Figure 7. Prediction of new recovery by (a) multiple linear regression, ridge regression, lasso regression and (b) Facebook Prophet

3.4. Performance evaluation of all the models with 40 days prediction interval

In the previous sections, four ML forecasting models were used to forecast COVID-19 threat in three categories such as daily new infected cases, daily new fatality, and daily new recovered patients in the next 30 days range. Among the four models, multiple linear regression performed better among other models. So, the model was used for further analysis where we trained the model in a 40 days prediction interval. At first, we trained all the models with from 17th March to 3rd May (48) days and predicted upcoming 30 days in which multiple linear regression performed best in first two categories and reasonably in the third category due to lack of data. The performance of multiple linear regression was followed by ridge regression, lasso regression and lastly Facebook Prophet which performed less among all algorithms. In the second interval, data of 17th March to 12th June (88 days) were used to train the models. Data of another 40 more days was added in this interval and prediction was made for upcoming 30 days in which the outcomes were same as the first interval. In the third prediction interval, data of 17th March to 22nd July (128 days) were used to train the models in which 40 more additional days were added to predict upcoming 30 days. In this interval, multiple linear regression outperformed other algorithms in all three categories. For the fourth interval, data of 17th March to 31st July (168 days) were used to train the models in which 40 more additional days were added to predict upcoming 30 days. The results of this interval was discussed in the previous sections where we showed that multiple linear regression outperformed all other algorithms.

Table 6 shows the performance details of all the models in forecasting new confirmed cases in the four intervals. Figure 8(a), 8(b) and 8(c) show the performance of multiple linear regression in all the categories (daily new cases, daily new fatality, and daily new recovery) in four intervals. Due to the lack of

highly correlated features, multiple linear regression showed reasonable in the third category but no other algorithms among the four except multiple linear regression could perform better in the third category.

Table 6. Performance of the models in new cases forecasting in four intervals
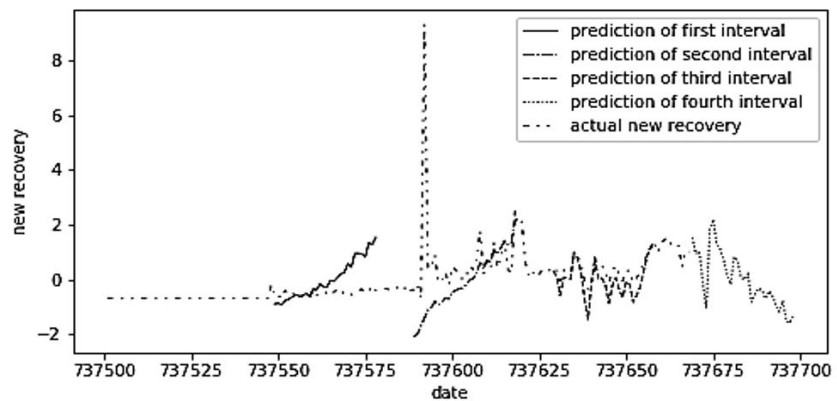
Dataset Size (Number of days)	MLR	RIDGE	LASSO	PROPHET
48	Best	Best	Best	Well
88	Best	Best	Better	Better
128	Best	Best	Better	Well
168	Best	Best	Best	Well
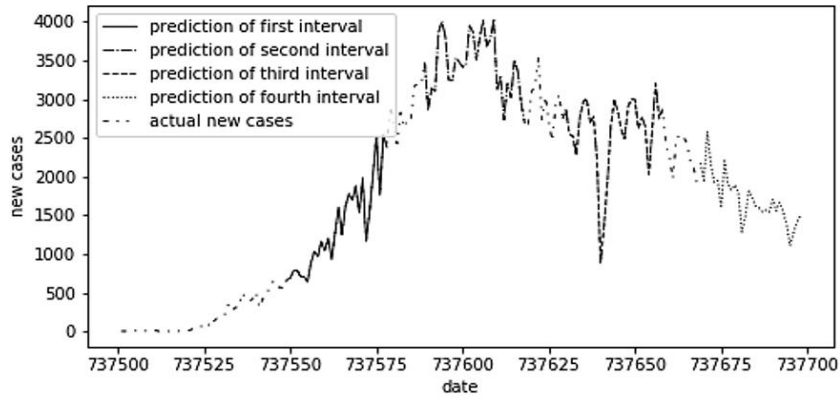
(a)
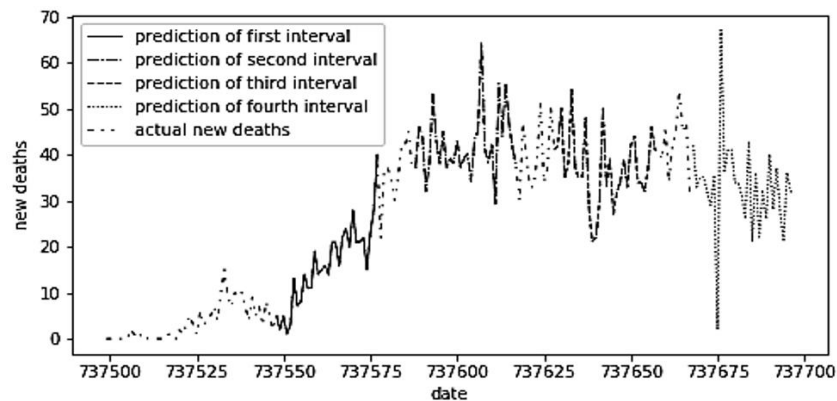
(b)

(c)

Figure 8. Performance of multiple linear regression with all intervals in prediction (a) daily new cases, (b) daily new fatality, and (c) daily new recovery

4. CONCLUSION

With the advancement of technology, the forecasting of any kind of threat is now in our grasp. COVID-19 is a huge threat for the whole world as well as for Bangladesh. In our research, we focused on assessing the current COVID-19 situation and make future predictions on the basis of those assessment. We focused on three categories as daily new infected cases, daily new fatality, and daily new recovery. In our research, we used four machine learning based predictive models such as multiple linear regression, ridge regression, lasso regression and Facebook Prophet. In our study, multiple linear regression performed the best with R^2 score of 99% on daily new infected cases and daily new fatality and 94% on daily new recovery. The forecast can be a great help to the authorities to understand the current pandemic situation and to take steps in relaxing the current COVID-19 crisis. In the future, we want to enhance this study more and work on real time live forecasting on COVID-19.

REFERENCES

[1]   WHO, "Naming the coronavirus disease (COVID-19) and the virus that causes it," *World Health Organization*, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it (accessed Dec. 27, 2020).

[2]   WHO, "ArcGIS Dashboards." https://covid19.who.int/ (accessed Oct. 01, 2020).

[3]   P. Lapuerta, S. P. Azen, and L. Labree, "Use of neural networks in predicting the risk of coronary artery disease," *Computers and Biomedical Research*, vol. 28, no. 1, pp. 38–52, Feb. 1995, doi: 10.1006/cbmr.1995.1004.

[4]   C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, May 2017, doi: 10.1016/j.renene.2016.12.095.

[5]   K. M. Anderson, P. M. Odell, P. W. F. Wilson, and W. B. Kannel, "Cardiovascular disease risk profiles," *American Heart Journal*, vol. 121, no. 1, pp. 293–298, Jan. 1991, doi: 10.1016/0002-8703(91)90861-B.

[6]   A. Bharat, N. Pooja, and R. A. Reddy, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739696.

[7]   K. Farhana, M. Rahman, and M. T. Ahmed, "An intrusion detection system for packet and flow based networks using deep neural network approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 5514–5525, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5514-5525.

[8]   M. Yadav, M. Perumal, and M. Srinivas, "Analysis on novel coronavirus (COVID-19) using machine learning methods," *Chaos, Solitons & Fractals*, vol. 139, Oct. 2020, doi: 10.1016/j.chaos.2020.110050.

[9]   F. Rustam *et al.*, "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[10]   A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-Series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, Nov. 2020, doi: 10.1016/j.chaos.2020.110121.

[11]   S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons & Fractals*, vol. 140, Nov. 2020, doi: 10.1016/j.chaos.2020.110227.

[12]   L. Wang *et al.*, "Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm," *Science of The Total Environment*, vol. 727, Jul. 2020, doi: 10.1016/j.scitotenv.2020.138394.

[13]   M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "COVID-19 prediction and detection using deep learning," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 12, pp. 168–181, 2020.

[14]   H. H. Elmousalami and A. E. Hassanien, "Day level forecasting for coronavirus disease (COVID-19) spread: analysis, modeling and recommendations," *arXiv preprint arXiv:2003.07778*, Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.07778.

[15]   M. A. A. Al-qaness, A. A. Ewees, H. Fan, and M. Abd El Aziz, "Optimization method for forecasting confirmed cases of COVID-19 in China," *Journal of Clinical Medicine*, vol. 9, no. 3, Mar. 2020, doi: 10.3390/jcm9030674.

[16]   N. Yudistira, "COVID-19 growth prediction using multivariate long short term memory," *arXiv preprint arXiv:2005.04809*, May 2020, [Online]. Available: http://arxiv.org/abs/2005.04809.

[17]   A. A. Chowdhury, K. T. Hasan, and K. K. S. Hoque, "Analysis and prediction of COVID-19 pandemic in Bangladeshby using long short-term memory network (LSTM) andAdaptive neuro fuzzy inference system (ANFIS)." Research Square Platform LLC, Oct. 2020.

[18]   S. Mahmud, "Bangladesh COVID-19 daily cases time series analysis using Facebook Prophet model," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3660368.

[19]   A.-E. E. Hridoy *et al.*, "Forecasting COVID-19 dynamics and endpoint in Bangladesh: a data-driven approach." Cold Spring Harbor Laboratory, Jun. 2020, doi: 10.1101/2020.06.26.20140905.

[20]   M. M. Rahman, A. Ahmed, K. M. Hossain, T. Haque, and A. Hussain, "Impact of control strategies on COVID-19 pandemic and the SIR model based forecasting in Bangladesh." Cold Spring Harbor Laboratory, Apr. 2020, doi: 10.1101/2020.04.19.20071415.

[21]   GitHub, "Our world in data," 2022. https://github.com/owid (accessed Oct. 01, 2020).

[22]   WHO, "Official COVID-19 data repository of World Health Organization." https://covid19.who.int/table (accessed Oct. 01, 2020).

[23]   "Official corona updates of Bangladesh." https://corona.gov.bd/ (accessed Oct. 01, 2020).

[24]   G. K. Uyanık and N. Güler, "A study on multiple linear regression analysis," *Procedia - Social and Behavioral Sciences*, vol. 106, pp. 234–240, Dec. 2013, doi: 10.1016/j.sbspro.2013.12.027.

[25]   D. W. Marquardt and R. D. Snee, "Ridge regression in practice," *The American Statistician*, vol. 29, no. 1, Feb. 1975, doi: 10.2307/2683673.

[26]   V. Roth, "The generalized LASSO," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 16–28, Jan. 2004, doi: 10.1109/TNN.2003.809398.

[27]   "Prophet." Accessed: Oct. 01, 2020. [Online]. Available: https://facebook.github.io/prophet/.

[28]   J. Miles, "R squared, adjusted R squared," in *Wiley StatsRef: Statistics Reference Online*, Wiley, 2014.

[29] H. Marmolin, "Subjective MSE measures," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 3, pp. 486–489, 1986, doi: 10.1109/TSMC.1986.4308985.

[30] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.

BIOGRAPHIES OF AUTHORS

Maqsudur Rahman Lecturer at Department of Information and Communication Technology, Comilla University, (Bangladesh). Teaching area includes Computer Programming, Computer Networks, data Communication and Mobile Communication. Research interests lie at Digital Communication, IoT and Network Security. He can be contacted at email: mrrajon@cou.ac.bd.

Md. Tofael Ahmed Associate Professor of ICT, Comilla University (Bangladesh). Completed PhD from India. Research area includes Cyberbullying Detection, Data Mining, Text Analytic, Big Data, Online Social Network analysis. He can be contacted at email: tofael@cou.ac.bd.

Shafayet Nur Teaching Assistant at Department of Computer Science and Engineering, Port City International University (Bangladesh). Research interest includes Machine Learning, Deep Learning Approaches, Natural Language Processing, performance Analysis. Teaching area includes Data Structure and Algorithms, Pattern Recognition. He can be contacted at email:shafayetnur.abir@gmail.com.

Abu Zafor Muhammad Touhidul Islam Professor of Electrical and Electronics Engineering, Rajshahi University, Bangladesh. Completed PhD from Japan. Research interest includes Big Data, Online Social Network analysis. He can be contacted at email: touhid.eee@ru.ac.bd.