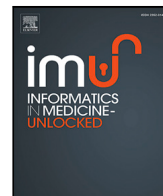




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh

Hafsa Binte Kibria*, Oishi Jyoti, Abdul Matin

Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh

ARTICLE INFO

Keywords:

Time series analysis
ARIMA model
COVID-19
Pandemic
Bangladesh

ABSTRACT

During the third wave of the coronavirus epidemic in Bangladesh, the death and infection rate due to this devastating virus has increased dramatically. The rapid spread of the virus is one of the reasons for this terrible condition. So, identifying the subsequent cases of coronavirus can be a great tool to reduce the mortality and infection rate. In this article, we used the autoregressive integrated moving average-ARIMA(8,1,7) model to estimate the expected daily number of COVID-19 cases in Bangladesh based on the data from April 20, 2021, to July 4, 2021. The ARIMA model showed the best results among the five executed models over Autoregressive Model (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Rolling Forest Origin. The findings of this article were used to anticipate a rise in daily cases for the next month in Bangladesh, which can help governments plan policies to prevent the spread of the virus. The forecasting outcome indicated that this new trend (named delta variant) in Bangladesh would continue increasing and might reach 18327 daily new cases within four weeks if strict rules and regulations are not applied to control the spread of COVID-19.

1. Introduction

The novel coronavirus originated from Wuhan, China, in 2019 has escalated all over the world so far. The variant of the coronavirus is highly infectious and has a global impact on more than 199,313,422 cases, claiming 4,245,582 lives till August 2, 2021 [1]. The symptoms shown by this virus can be mild, moderate, and intense, which builds severe respiratory distress syndrome [2]. Initially, no COVID-19 vaccination was found. In those circumstances, the only way to prevent the spread of the virus was to identify social distance, to identify positive cases using large-scale testing, and to isolate infected people [3]. Later on, several vaccines were invented to provide acquired immunity against the coronavirus. However, the mutation of COVID-19 prevents the vaccines from being fully effective. Maintaining the necessary measures and proper dose of vaccines, several nations have succeeded in controlling the disease, whereas Bangladesh has faced difficulties in managing the situation. The large population and inadequate supply of vaccines have made it very difficult to normalize the circumstances. Therefore, it is very important to prevent the spread of the virus. Because of the significant impact and easy spreading of the virus, the national government has enforced lockdown in most affected areas or all over the country [4]. In that case, it would be much more effective if the possible number of confirmed cases could be known in advance. Then the lockdown can be implemented in possible afflicted

areas in time. As a result, the spread will decrease as well as the number of confirmed cases.

As of August 02, 2021, the total number of infected people is over 1,309,910 in Bangladesh [1]. Bangladesh reports an average of 14,132 daily new cases, which are the highest of 98% peak as per August 02, 2021 [5]. These numbers are growing significantly and continue to have adverse effects on human lives, medical centers, and nations' currencies. Therefore, it is clear that predicting the future pandemic outbreak using existing data is significant for understanding the present circumstances and helping authorities take necessary steps and effective ideas for limiting new infections. Several research has been done to evaluate the spread and impact of the virus using machine learning models and various computational analysis [6].

The researchers have used four comparative models to predict coronavirus spread in Saudi Arabia in [7], and autoregressive integrated moving average-ARIMA (2,1,1) came out as the best model. Data availability was a concern because the research was conducted at an early stage. They have selected their ARIMA order using AIC (Akaike information criterion) approaches. Since it was the beginning period of COVID-19, their data was relatively linearly increasing. Other models were also used to predict the spread, and at last comparative analysis was done among them. Another study [8] used the ARIMA model for forecasting in the top five affected countries: India, Brazil, Spain, the

* Corresponding author.

E-mail addresses: hafsabintekibria@gmail.com (H.B. Kibria), oj.ruet@gmail.com (O. Jyoti), ammuaj.cseruet@gmail.com (A. Matin).

URL: <https://www.ece.ruet.ac.bd/AbdulMatin> (A. Matin).

US, and Russia. Considering there was no vaccine available at the time, their main focus was to prepare the government to take the necessary steps. Here also, ARIMA gave promising results.

In another study [9], ARIMA and nonlinear autoregressive (NAR) models have been used to forecast the COVID-19 in India. ARIMA(1,1,0) was selected based on BIC (Bayesian information criterion) values, and the NAR model consisted of ten neurons. In [10], adaptive neuro-fuzzy inference system (ANFIS) and long short-term memory networks (LSTM) were implemented to forecast COVID-19 pandemic growth in Bangladesh, and LSTM provided a satisfactory result. LSTM is preferable in forecasting the long term where ARIMA is for the short term, and in our work, we have only used the third wave COVID-19 dataset, which contains approximately three months records of the confirmed cases. That is why we used ARIMA for our prediction. Also, promising result has been found for forecasting COVID-19 cases in other different countries using ARIMA [8,11–13]. Different order of ARIMA model has been used for the forecasting in Italy, Spain, and France, with satisfactory results [14]. In [15], confirmed cases and mortality cases were anticipated using two data-driven approaches for some countries. Both ARIMA and LSTM produced favorable results.

Many kinds of research are being conducted in order to alleviate the pandemic's suffering. Many researchers choose ARIMA as their primary choice for time series analysis. Other algorithms, such as LSTM, NAR, VAR, and other machine learning algorithms, are also being applied. For the performance measurement, RMSE, MAE, and other performance matrices have been used. One of the problems was the availability of data and the correctness of the given data, as in many developing countries, all COVID-19 records were not updated. Even though researchers are trying their best to predict the possible cases or deaths associated with the COVID-19 pandemic in the near future, and many algorithms have provided promising outcomes.

ARIMA has shown promising results in various studies for COVID-19 predictions. In [8], the researchers used ARIMA to estimate the top five afflicted countries, and they got a good result for the first 18 days of July. The MAD and MAPE were found within the acceptable agreement. Also other works have been done using ARIMA [11–13] for COVID-19 forecasting. In [16] ARIMA model was used to forecast the new cases along with deaths of COVID-19 in Bangladesh. Researchers conducted the same model in [17] to forecast the new cases and deaths in COVID-19 using the European Centre for Disease Prevention and Control data. To identify the confirmed number of diagnoses, besides ARIMA, several models were assigned in [18]. Among these, ARIMA has been specified as the most suitable model.

In finance and the economy, ARIMA plays a significant role. It has been widely used in stock market predictions [19,20]. In [21], ARIMA was used to forecast the stock change of New York and Nigeria, and the findings show that it can compete favorably with existing techniques. Several researchers have used the ARIMA model for weather forecasting on time series data. ARIMA (2,0,2) and ARIMA (2,1,3) were used respectively for rainfall data and temperature data to predict the weather for fifteen years in Varanasi, India [22]. Also, ARIMA has been applied to other time-series data like stock prediction [21,23], detecting malaria [24,25] and it showed good results.

In this study, the ARIMA model, the Autoregressive Model (AR), Moving Average(MA), ARMA, and Rolling Forest Origin were utilized to forecast the number of new confirmed COVID-19 cases in Bangladesh on a daily basis. On data recorded from April 20, 2021, to July 4, 2021, the assigned models were implemented and verified based on data recorded between July 5, 2021, to July 28, 2021. The train and test ratio was 75:25. Because of its better prediction functionalities, the models are widely recognized and effective. We have also implemented ARIMA(8,1,7) to forecast the coming month and discovered that the confirmed cases might exceed 18327 by August 30, 2021. The rolling forecast has the lowest error of all the models, but it can only predict a single period in our proposed model. Because of this limitation, despite the lowest error, we recognized ARIMA as the best-performing model.

2. Materials and methodology

2.1. Data description

The data used in this work is a statistical report of COVID-19 cases of Bangladesh, which is available online [5]. They have the information of all the COVID-19 related data such as death, confirmed cases, and lab tests. For the time being, this web source is the most reliable source for everyday COVID-19 cases. We used confirmed case data in our research to forecast coronavirus spread in Bangladesh. Many academics have conducted their research using data collected from this internet source [10,26,27]. The range of the dataset is from April 20, 2021, to July 28, 2021. For training, the dataset was taken from April 20, 2021, to July 6, 2021, while the remaining days till July 28, 2021, were used for testing.

Fig. 1 represents the total number of daily confirmed COVID-19 cases reported in Bangladesh from April 20, 2021, to July 28, 2021. The number of the confirmed cases was limited to six thousand until June 24, 2021; later, it has significantly increased and surpassed that number. Then we observed a significant increase in the first week of July, which is referred to as the third wave in Bangladesh. From April 20, 2021, it took nearly 70 days to double and only 30 days to quadruple. Data was collected from [5], there was a pattern in the increasing number of confirmed cases of COVID-19. Though it increased from May 1, 2021, to July 15, 2021, then it dropped significantly for a few days displayed in Fig. 1. It could be for several reasons, but none of them have been confirmed. We collected data from a single source, and the variables were unchanged. So the possible reason can be a computational error, where the excluded number was added the other day. Or, because of the unpredictable nature of COVID-19, it might show this kind of sudden characteristics. The sudden cases drop in the test data also made it difficult to correctly predict as it was not a predefined characteristic, and it decreased the accuracy of the models. The rolling forecast model captured this characteristic, and that is why only this model was able to predict closely the actual values between July 21 to July 24.

We can see that the ARIMA (8,1,7) model outperformed the others except for the rolling forecast by comparing the results from Figs. 5, 6 and Table 2. The predicted confirmed cases from July 29, 2021 to August 30, 2021, are depicted in Fig. 9 and Table 4. The yellow line represents the forecasted curve of daily confirmed cases for the following month based on the ARIMA (8,1,7) model of COVID-19. Fig. 9 shows the current number of confirmed cases officially reported by the Bangladesh Ministry of Health from July 2, 2021, to July 5, 2021 (in blue).

2.2. Data preprocessing

In our dataset, the date was the input column, and the confirmed case was the output. First, the time column was arranged according to year-month-day. And we did not have any missing values in our dataset. For training, the dataset was divided into a 75:25 train test ratio. The split can be placed randomly for any other dataset since there is no dependence from one observation to another. In time-series data, samples are observed in fixed time intervals, and the test indices must be higher than the train. So during splitting, we took the first 75% for training and the remaining 25% for testing. In Fig. 3, the procedure of COVID-19 prediction has been displayed. After collecting the data, it was converted into time series and then checked for stationarity which will be discussed later. After preprocessing, it was split and then trained. Trained models used the testing data for forecasting. At last, we evaluated the performance of every model.

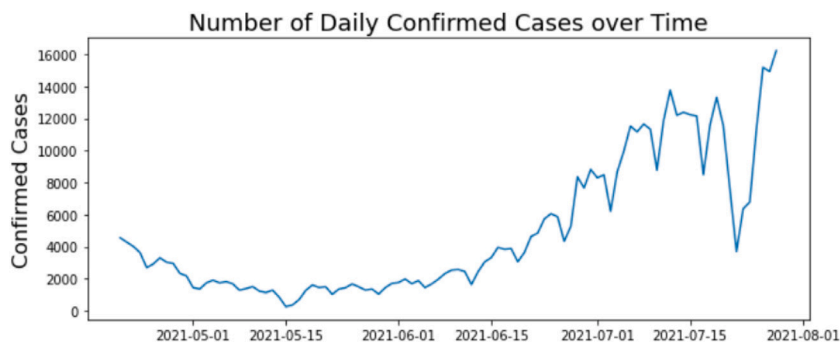


Fig. 1. Daily confirmed cases of COVID-19 in Bangladesh from April 20, 2021 to July 6, 2021.

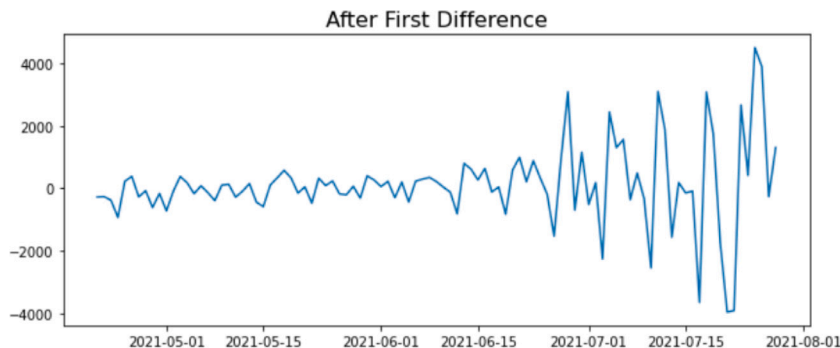


Fig. 2. Representation of dataset after first difference.

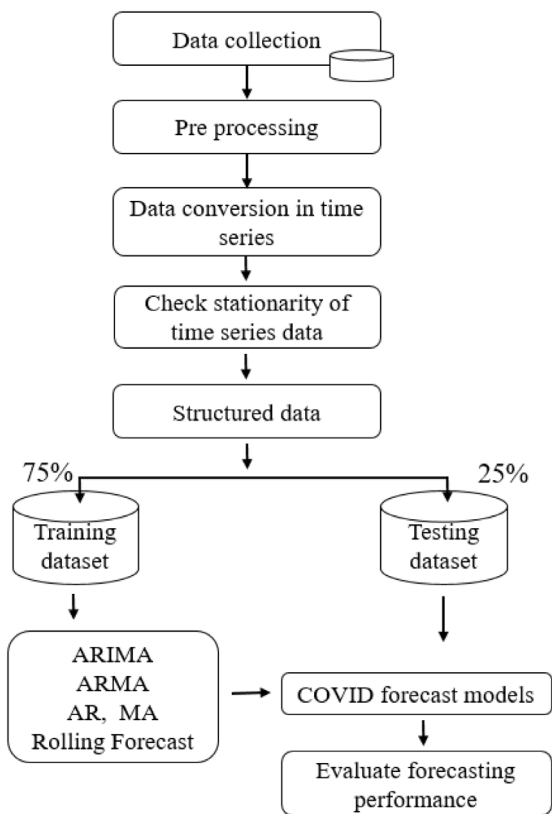


Fig. 3. Overall procedure of proposed model.

2.3. Description of models

This research has focused on using statistical models to predict the spread of the coronavirus in Bangladesh. The Autoregressive (AR) model is the simplest and frequently used model structure. It is called the autoregressive model, as the values are predicted based on past values. As its forecasts based on its values, that is why the model is referred to as auto, and the model is a regression model. Therefore, it is known as a regressive model. The current output m_t is expressed by previous values and coefficients (β) in the AR model in Eq. (1). Here β denotes the coefficient, time is t , p is the order of the parameters, and m_{t-p} is the previous value of p order.

$$m_t = \beta_0 + \beta_1 m_{t-1} + \beta_2 m_{t-2} + \dots + \beta_p m_{t-p} + \epsilon \tag{1}$$

In the MA model, the measurement of error from the previous period helps better estimate the next period. Here, the model always moves around average; that is why it is called the moving average model. It is always centered to its average.

$$f_t = \mu + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q} + \epsilon_t \tag{2}$$

In MA model, f_t is the predicted value, μ is the constant (the mean) in Eq. (2). The error is represented by ϵ_t , while the coefficient of error is represented by ϕ .

By merging AR and MA, there is a more improved structure known as the ARMA model, which is defined in (3).

$$m_t = \beta_0 + \beta_1 m_{t-1} + \beta_2 m_{t-2} + \dots + \beta_m m_{t-p} + \phi_1 \epsilon_{t-1} + \epsilon_{t-2} + \dots + \epsilon_{t-q} + \epsilon_t \tag{3}$$

A more advanced statistical model is Autoregressive Integrated Moving Average (ARIMA), which includes differences at least once. The ARIMA model formula is stated in Eq. (4).

$$z = \phi_1 z_{t-1} + \phi_2 z_{t-2} z_{t-2} + \dots + \phi_k z_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_k \epsilon_{t-q} + \epsilon_t \text{ where } z_t = a_{t+1} - a_t \tag{4}$$

Table 1
Order selection of ARIMA based on AIC approach.

q in MA	p in AR				
	0	7	8	9	14
0	1704	1682	1682	1676	1674
7	1691	1672	1671	1672	1671
9	1683	1677	1677	1678	1675
14	1688	1684	1686	1688	1680

AR and MA require stationarity, which is a statistical property of a time series. The mean and standard deviation need to be constant for the time series data. Also, there should be no seasonality — these are the criterion of a time series to achieve stationarity. As our model did not have the property, we took the first difference of our dataset. It was achieved by subtracting the previous observation from the current observation. The graph after the first-order difference represents in Fig. 2. But for ARIMA, the parameter d makes it stationary, so we do not need to make the difference by subtracting the observant. In our model, d is equal to 1.

ARIMA is used when the graph shows stationarity but has a linear moving trend. It forecasts the differences between time series rather than the time series itself. The ARIMA model is identified by arranging the three variables in the following order: p for the order of AR, d for the degree of differencing, and q for the order of MA.

2.4. Parameters estimation

The time-series data must be stationary before evaluating the parameters for all models. The stationarity of the data was checked with the augmented dickey fuller test(ADF), for which the null hypothesis H_0 indicates that the time series is not stationary. The ADF result suggested that time-series data was non-stationary ($p = .99$ in this case). After applying the first difference, the null hypothesis was rejected since the p -value was.004, which is less than .05.

Based on the Akaike information criterion (AIC), the parameters of the ARIMA model were selected, as shown in Eq. (5).

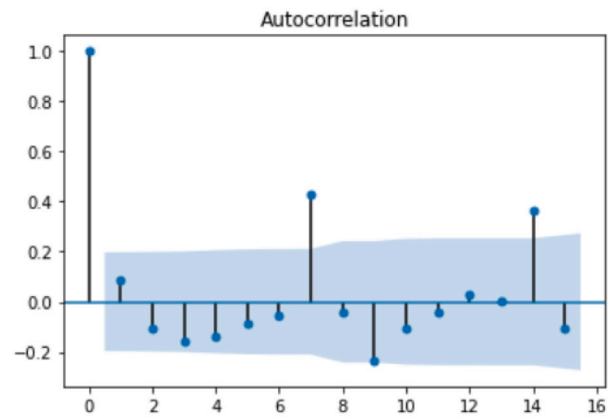
$$AIC = 2k - 2l \tag{5}$$

AIC assists in the selection of the best model. It is a prominent method in statistics for selecting any model. It is made up of two parts.

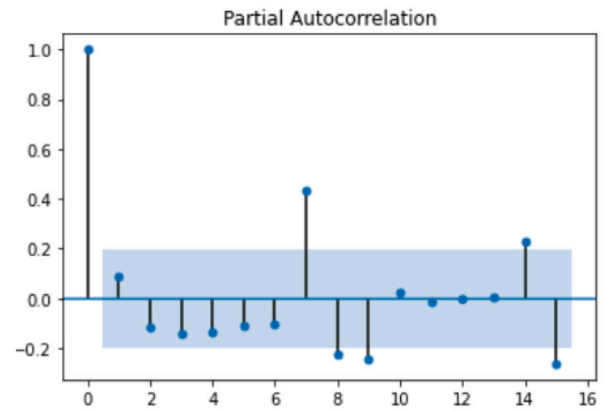
- Log-likelihood (l): It shows how strong the model fits the data. A more sophisticated model can fit the training data better but also introduces overfitting. Also, a complicated model may give incorrect results on testing data.
- Number of parameters in the model (k): the fewer the parameters, the simpler the model is. But a too simple model will introduce underfitting, while a too complicated model would introduce overfitting.

The AIC needs to be as small as possible, so we will select the model with the lowest AIC. A model with low AIC will have low k (number of parameters) that will keep the model as simple as possible. However, we also need to ensure that the model has a high log-likelihood (l), so it will select a model that fits the data well with relatively few parameters. As a result, it will choose a model that will strike a balance. The order we elected to calculate the AIC values is from the ACF and PACF graphs. For the AR order, the significant lags (0,7,8,9,14) have been chosen from the PACF graph, while the lags (0,7,8,9,14) that are significant in the ACF graph have been selected for the MA order.

Table 1 displays the AIC values for different p and q parameters. We evaluated numerous p and q parameters chosen from the ACF and PACF graphs of the first difference of the data. The ACF and PACF have been shown in Fig. 4. We have found that ARIMA(8,1,7) produces the lowest AIC values. Based on the ADF test, d was selected as 1 for the ARIMA model.



(a) Auto-correlation



(b) Partial auto-correlation

Fig. 4. ACF and PACF of the first difference of data.

2.5. Models performance measures

To evaluate the performance of each model given above, a relatively common accuracy measuring method was applied. These are the performance functions:

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (w_i - \tilde{w}_i)^2} \tag{6}$$

where w_i and \tilde{w}_i are real and predicted values, respectively.

- Mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |w_i - \tilde{w}_i| \tag{7}$$

- Coefficient of determination (R^2)

$$R^2 = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (w_i - \tilde{w}_i)^2}{\frac{1}{N} \sum_{i=1}^N (w_i - \bar{w})^2} \tag{8}$$

where \bar{w} is defined as $\bar{w} = \frac{1}{N} \sum_{i=1}^N w_i$

- Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\tilde{w}_i - w_i}{\tilde{w}_i} \right| \tag{9}$$

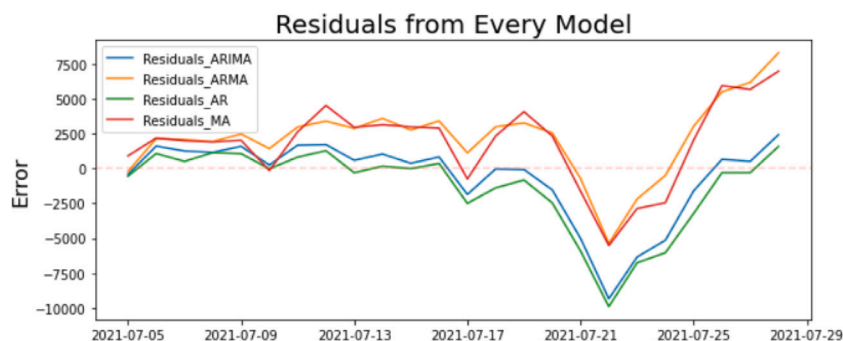


Fig. 5. Residuals of the ARIMA, ARMA, AR, and MA models.

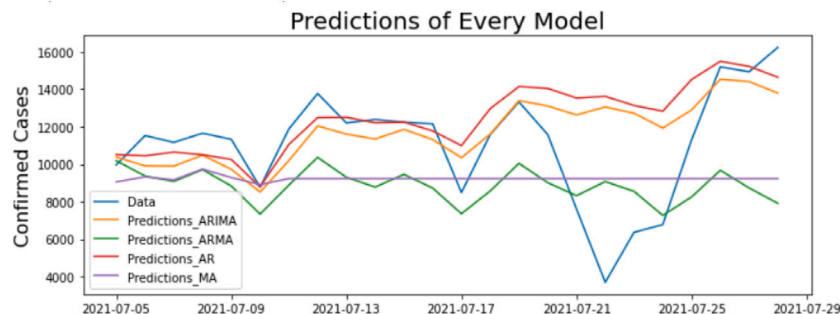


Fig. 6. Prediction result comparison for the daily confirmed cases in Bangladesh. ARIMA, ARMA, AR, and MA models were used for comparison.

Table 2

All proposed models' performance to forecast the total number of confirmed cases.

Model	RMSE	MAE	R ²	MAPE
ARIMA	2965	1965	-0.06	0.28
ARMA	3486	2968	-0.47	0.29
AR	3231	2023	-0.26	0.30
MA	3405	2963	-0.40	0.29
Rolling Forecast	2693	1920	0.12	0.24

In this work, we have used five statistical models to forecast the spread of COVID-19 in Bangladesh. AR, MA, ARMA, ARIMA, and Rolling forecast origin models are among them. Each models' performance has been measured using the above performance metrics. Table 2 shows that the rolling forecast model has the minimum RMSE, MAPE, and RMSRE values. As a result, the rolling forecast model surpasses all the models. But as in our proposed approach, this model predicts only for one day, observes the predictions, and then includes them in the training set to predict the upcoming day. For this reason, it cannot be employed when we need predictions one week or month in advance. If we apply ARIMA for one-week advance prediction, the performance is substantially better. As a result, we consider ARIMA the best model, with the ARMA model ranking second, followed by the AR and MA models. If we only need the next-day estimate; hence the rolling forecast model will be the best choice.

3. Results

As discussed, the expected number of confirmed cases was forecasted using five alternate models. The data was divided into two parts; one was used for training while the other was for testing the models. The data were divided into training and testing sets in the ratio of 75:25. After training, the models were used to forecast the number of confirmed cases using the test set. Fig. 5 shows the residuals of the four models, and Fig. 6 represents the prediction of the test data for the four models.

In Fig. 5, it is observed that the residual of ARIMA (8,1,7) is more close to zero than any other model. And in Fig. 6, all the models' predictions were near to the actual data until the time of July 21, 2021, when there was a drastic change. As the number of verified cases dropped dramatically, the models could not keep up with the actual data. For this decrease, the error of all the models' increases; otherwise, the models performed satisfactorily at all the other periods. Here, a thing to notice is that the prediction of the MA model fluctuates a little fast, then it is almost flat. The reason is that a simple moving average model predicts the constant mean of the value. A simple MA has the limitation of only being able to estimate the number of periods equal to the lag employed in MA order.

In our model, the order for MA was (0,7), which explains why the first seven days' prediction fluctuates. As a result, in Fig. 6, it fluctuates a little at first and then produces a constant mean of it. Also, in Table 3, we see the predictions of MA is constant after July 11, 2021. Table 3 shows the predicted confirmed cases for the test data from April 5, 2021, to April 28, 2021, along with the actual values. All the five models' predictions have been shown here, and compared to them, the rolling forecast model comes closest to the actual values, followed by ARIMA, ARMA, AR, and MA.

The residuals and prediction of the rolling forecast model are shown in Figs. 7 and 8. This rolling forecast model has used ARIMA(8,1,7) for prediction. As previously said, the rolling forecast only predicts the next day, and after observing the actual value, it learns and forecasts the next period. That is why this type of model is not always applicable.

4. Discussion

In the current world, there are still frightening increases in the number of cases of new coronavirus. The overall number of people affected is over 199 million, spreading across 222 countries, about 0.64% of Bangladesh. The primary goal of this study is to employ a commonly-used statistical analytic model, known as an ARIMA model, to observe and predict the epidemiology of COVID-19 in Bangladesh based on the data of daily confirmed cases publicly published by the Bangladesh

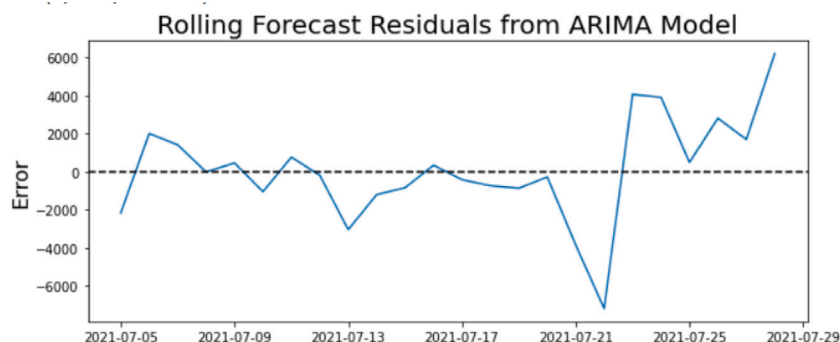


Fig. 7. Residuals of Rolling forecast model.

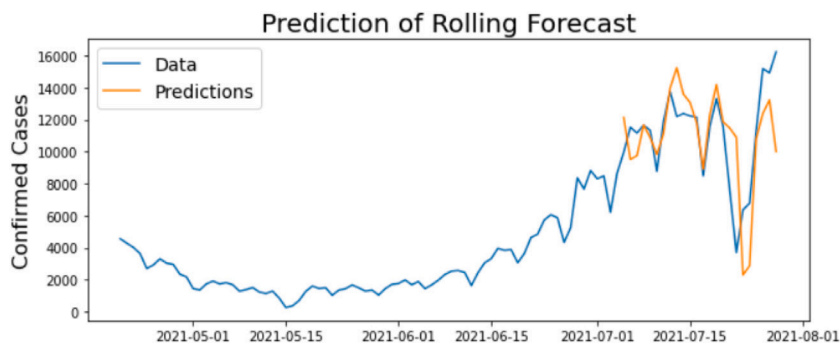


Fig. 8. Forecasting results for the daily number of confirmed cases in Bangladesh using rolling forecast origin.

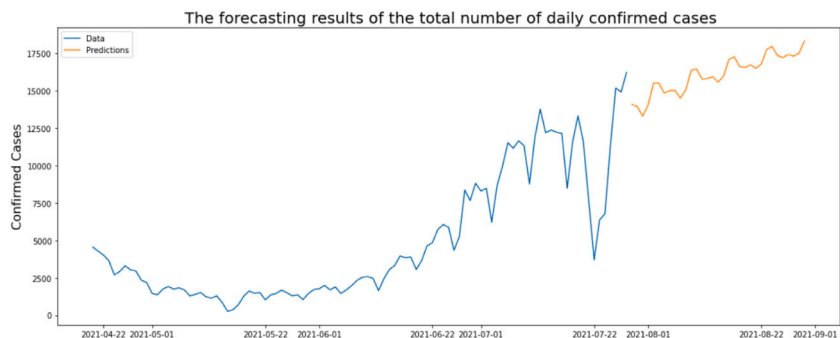


Fig. 9. Forecasting results of the total number of daily confirmed cases in Bangladesh from July 29, 2021 to August 30, 2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ministry of Health. We believe that theoretical studies based on statistical modeling are essential for understanding the pandemic features of the epidemic to forecast the COVID-19 pandemic’s possible trend. The results of such statistical models provide a complete picture of the present pandemic condition, allowing authorities to be active by developing plans and effective decisions to battle the pandemic and thereby limiting its impact on the economy, healthcare institutions, and society.

According to our findings, the ARIMA model’s prediction outperformed the other models described above. As such, after first-order difference processing, the daily number of confirmed COVID-19 cases in Bangladesh was forecasted using the ARIMA(8,1,7) model. The statistics suggest that the number of daily cases in Bangladesh might increase from 1492 to 17311 daily cases within one month. From this worrying increase in confirmed cases, it is crystal clear that Bangladeshi authorities must take immediate prevention and control strategies, particularly in the worst-affected cities such as Dhaka, Chittagong, Comilla, and Sylhet.

To combat the pandemic’s growing impact, different parties, including individuals and administration entities led by the Ministry of

Health, must work together intensively. The first and most important proactive strategy is to ensure that everyone is vaccinated. Bangladesh started the first dose of COVID-19 vaccines on 21/2/7, while the second dose began on April 21, 2021. The faster mass vaccination begins, the less likely it is that people will become infected. Using the time series prediction model, lockdown can be applied to a specific district where the coronavirus rate is expected to be very high. Furthermore, more awareness is required, and a strong push to encourage people to practice proper hygiene and sanitation. The health ministry will be able to take effective measures in the coming days using this future prediction of coronavirus spread.

Building a reliable model to predict COVID-19 daily cases has constraints that begin with the uniqueness of the virus. It does not seem to follow any trend. After the first wave, the second wave arrived faster, and the third wave came in a flash. This work was completed on July 28, 2021, and we did not use the entire dataset for prediction. The infected rate in Bangladesh was very low at the start of 2020. All death and confirmed case records have surpassed in the third wave, which starts in April 2021. That is why we only used that dataset to predict for the upcoming days.

Table 3
Comparison of the predicted and actual values of the ARIMA, ARMA, AR, and MA models.

Date	Actual	ARIMA	ARMA	AR	MA	Rolling prediction
2021-07-05	9964	10381	10180	10511	9053	12132
2021-07-06	11525	9908	9370	10446	9328	9521
2021-07-07	11162	9897	9082	10644	9161	9752
2021-07-08	11651	10482	9708	10505	9737	11665
2021-07-09	11324	9725	8836	10250	9287	10869
2021-07-10	8772	8514	7336	8790	8910	9826
2021-07-11	11874	10192	8883	11050	9232	11113
2021-07-12	13768	12041	10362	12483	9232	13962
2021-07-13	12198	11601	9302	12497	9232	15250
2021-07-14	12383	11340	8779	12206	9232	13591
2021-07-15	12236	11858	9456	12242	9232	13081
2021-07-16	12148	11313	8726	11779	9232	11809
2021-07-17	8489	10346	7357	10987	9232	8911
2021-07-18	11578	11602	8572	12961	9232	12319
2021-07-19	13321	13387	10042	14139	9232	14195
2021-07-20	11579	13106	9014	14040	9232	11860
2021-07-21	7614	12627	8321	13529	9232	11472
2021-07-22	3697	13052	9072	13618	9232	10901
2021-07-23	6364	12717	8552	13128	9232	2298
2021-07-24	6780	11927	7267	12826	9232	2867
2021-07-25	11291	12890	8250	14512	9232	10804
2021-07-26	15192	14527	9675	15485	9232	12382
2021-07-27	14925	14404	8733	15216	9232	13233
2021-07-28	16230	13793	7911	14637	9232	10014

Table 4
The forecasted outcome of the COVID-19 pandemic in Bangladesh for the next month using the ARIMA model.

Date	Confirmed Case
2021-07-29	14092
2021-07-30	13947
2021-07-31	13307
2021-08-01	14044
2021-08-02	15511
2021-08-03	15513
2021-08-04	14844
2021-08-05	15009
2021-08-06	15015
2021-08-07	14518
2021-08-08	15072
2021-08-09	16366
2021-08-10	16460
2021-08-11	15785
2021-08-12	15827
2021-08-13	15938
2021-08-14	15579
2021-08-15	15986
2021-08-16	17110
2021-08-17	17270
2021-08-18	16620
2021-08-19	16561
2021-08-20	16737
2021-08-21	16505
2021-08-22	16799
2021-08-23	17759
2021-08-24	17965
2021-08-25	17359
2021-08-26	17220
2021-08-27	17431
2021-08-28	17311
2021-08-29	17521
2021-08-30	18327

Although the rolling forecast model outperformed ARIMA, we consider ARIMA to be the best model in this case. Because the rolling forecast only predicts for a single day, then it trains itself and predicts the next period based on the accuracy of the previous prediction. As a result, despite having a low error rate, it is never a good choice. A forecast for a single day will not assist control makers in making

any decisions. Although the ARIMA model is often an excellent choice for such a forecast problem, it does have a few drawbacks, including the lack of automatic updates. When new data is added, a new run is required. In addition, the structure of the ARIMA model is linear, but there is non-linearity in our problem, which is why accuracy is likely to be reduced.

5. Conclusion

More research is being conducted to forecast the COVID-19, as the outcome of this disaster affects our daily lives. Many researchers have been using different types of prediction models. In this work, we used AR (8,0), MA (0,7), ARMA (8,7), ARIMA (8,1,7), and rolling forecast with ARIMA (8,1,7) models to forecast the spread of the COVID-19 outbreak in Bangladesh. The best forecasting model for predicting the daily confirmed cases' trend in Bangladesh was discovered to be ARIMA (8,1,7). The daily production number of confirmed cases was estimated for the next month using this model. By August 20, 2021, the estimated daily confirmed cases could reach 18327. These findings could motivate policymakers to take the required steps to combat the COVID-19 pandemic, such as implementing new movement restrictions and conducting mass vaccinations for COVID-19 cases across the country.

However, further research is needed to analyze which type of forecasting method or model is the most accurate for different situations around the world. Different countries have different patterns in COVID-19 characteristics, and based on their variants, appropriate algorithms need to be selected. There is a high possibility that the input data was not completely accurate for various reasons, such as an infected individual being asymptomatic, not being tested, or not being registered in the database. Despite this, the gradual learning strategy can overcome the inaccuracy of the incoming data. Furthermore, an unidentified suspect in the community believes that a few countries are submitting fraudulent data for political purposes. Many countries, including Bangladesh, imposed social distance and lockdown, which impacted the number of cases and casualties. By taking these aspects into account, the forecasts' outcomes may be affected. While we attempted to produce a promising result for the Bangladesh data, we still need to test on many other databases for more accurate findings. Nonetheless, the data show AI's promise and success in forecasting a pandemic.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to acknowledge the Department of Electrical and Computer Engineering (ECE), Rajshahi University of Engineering & Technology (RUET), to facilitate the work.

References

- [1] Bangladesh: WHO coronavirus disease (COVID-19) dashboard with vaccination data. 2021, [Online]. URL <https://covid19.who.int/region/searo/country/bd/>, [Accessed on 02.08.2021].
- [2] Surveillances V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* 2020;2(8):113–22.
- [3] Haghani M, Bliemer MC, Goerlandt F, Li J. The scientific literature on coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. *Saf Sci* 2020;129:104806.
- [4] Kumar N, Susan S. Covid-19 pandemic prediction using time series forecasting models. In: 2020 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE; 2020, p. 1–7.
- [5] Coronavirus cases: worldometer. 2021, [Online]. URL <https://www.worldometers.info/coronavirus/country/bangladesh/>, [Accessed on 02.08.2021].

- [6] Mahalle P, Kalamkar A, Dey N, Chaki J, Shinde G. Forecasting models for coronavirus (COVID-19): a survey of the state-of-the-art. 2020, TechRxiv.
- [7] Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infection Public Health* 2020;13(7):914–9.
- [8] Sahai AK, Rath N, Sood V, Singh MP. Arima modelling & forecasting of COVID-19 in top five affected countries. *Diabetes Metab Syndr: Clin Res Rev* 2020;14(5):1419–27.
- [9] Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *J Safety Sci Resilience* 2020;1(1):12–8.
- [10] Chowdhury AA, Hasan KT, Hoque KKS. Analysis and prediction of COVID-19 pandemic in Bangladesh by using ANFIS and LSTM network. *Cogn Comput* 2021;13(3):761–70.
- [11] Kufel T, et al. Arima-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilib Q J Econ Econ Policy* 2020;15(2):181–204.
- [12] Dehesh T, Mardani-Fard H, Dehesh P. Forecasting of covid-19 confirmed cases in different countries with arima models. 2020, MedRxiv.
- [13] Tandon H, Ranjan P, Chakraborty T, Suhag V. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. 2020, arXiv preprint arXiv:2004.07859.
- [14] Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Environ* 2020;729:138817.
- [15] Kumar M, Gupta S, Kumar K, Sachdeva M. Spreading of COVID-19 in India, Italy, Japan, Spain, UK, US: a prediction using ARIMA and LSTM model. *Digit Gov: Res Practice* 2020;1(4):1–9.
- [16] Kundu LR, Ferdous MZ, Islam US, Sultana M. Forecasting the spread of COVID-19 pandemic in Bangladesh using arima model. *Asian J Med Biol Res* 2021;7(1):21–32.
- [17] Bayyurt L, Bayyurt B. Forecasting of COVID-19 cases and deaths using ARIMA models. 2020, Medrxiv.
- [18] Barria-Sandoval C, Ferreira G, Benz-Parra K, López-Flores P. Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study. *Plos One* 2021;16(4):e0245414.
- [19] Banerjee D. Forecasting of Indian stock market using time-series ARIMA model. In: 2014 2nd international conference on business and information management (ICBIM). IEEE; 2014, p. 131–5.
- [20] Meher BK, Hawaldar IT, Spulbar CM, Birau FR. Forecasting stock market prices using mixed ARIMA model: A case study of Indian pharmaceutical companies. *Invest Manag Financial Innov* 2021;18(1):42–54.
- [21] Ariyo AA, Adewumi AO, Ayo CK. Stock price prediction using the ARIMA model. In: 2014 UKSim-AMSS 16th international conference on computer modelling and simulation. IEEE; 2014, p. 106–12.
- [22] Shivhare N, Rahul AK, Dwivedi SB, Dikshit PKS. ARIMA based daily weather forecasting tool: A case study for varanasi. *MAUSAM* 2019;70(1):133–40.
- [23] Devi BU, Sundar D, Alli P. An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *Int J Data Min Knowl Manag Process* 2013;3(1):65.
- [24] Anokye R, Acheampong E, Owusu I, Isaac Obeng E. Time series analysis of malaria in Kumasi: Using ARIMA models to forecast future incidence. *Cogent Soc Sci* 2018;4(1):1461544.
- [25] Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malar J* 2016;15(1):1–10.
- [26] Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. Covid-19 pandemic prediction for hungary; a hybrid machine learning approach. 2020, MedRxiv.
- [27] Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ* 2020;728:138762.