Forecasting COVID-19 Dynamics and Endpoint in Bangladesh: A Data-driven Approach
— Source link ↗

Al-Ekram Elahee Hridoy, Mohammad Naim, Nazim Uddin Emon, Imrul Hasan Tipo ...+3 more authors

Institutions: University of Chittagong, North South University, International Islamic University, Chittagong, University of Dhaka ...+2 more institutions

Related papers:

- COVID-19 Outbreak Situations in Bangladesh: An Empirical Analysis

- System dynamics of peak projection covid-19 in indonesia

- Trend Analysis and Forecasting of COVID-19 outbreak in India

- An Epidemic Model SIPHERD and its application for prediction of the spread of COVID-19 infection in India

- Covid-19 Pandemic- Pits and falls of major states of India.

Share this paper: 📘 🐦 in ✉

View more about this paper here: https://typeset.io/papers/forecasting-covid-19-dynamics-and-endpoint-in-bangladesh-a-83mjsw90uz

Forecasting COVID-19 Dynamics and Endpoint in Bangladesh: A Data-driven Approach

Al-Ekram Elahee Hridoy[1], Mohammad Naim[2], Nazim Uddin Emon[3], Imrul Hasan Tipo[4], ,

Safayet Alam[5] Abdullah Al Mamun[6], Mohammad Safiqul Islam[7*]

[1]Department of Geography and Environmental Studies, University of Chittagong, Chittagong – 4331, Bangladesh

[2]Department of Electrical and Computer Engineering, North South University, Dhaka – 1229, Bangladesh

[3]Department of Pharmacy, Faculty of Science and Engineering, International Islamic University Chittagong, Chittagong – 4318, Bangladesh

[4]Department of Biochemistry and Molecular Biology, University of Chittagong, Chittagong – 4331, Bangladesh

[5]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Dhaka, Dhaka – 1000, Bangladesh

[6]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong – 4349, Bangladesh

[7]Department of Pharmacy, Faculty of Science, Noakhali Science and Technology University, Noakhali-3814, Bangladesh

*Correspondence:

Dr. Mohammad Shafiqul Islam

Professor, Department of Pharmacy, Noakhali Science and Technology University, Email:

research_safiq@yahoo.com, Mobile: +8801727658650, ORCID ID: 0000-0003-4924-5319.

Highlights

1. According to the Logistic curve fitting analysis, the inflection point of the COVID-19 pandemic has recently passed, which was approximately between May 28, 2020, to May 30, 2020.

2. It is estimated that the total number of confirmed cases will be around 187-193 thousand at the end of the epidemic. We expect that the actual number will most likely to in between these two values, under the assumption that the current transmission is stable and improved stringent policies will be in place to contain the spread of COVID-19.

3. The estimated total death toll will be around 3600-4000 at the end of the epidemic.

4. The epidemic of COVID-19 in Bangladesh will be mostly under control by the 2nd of January 2021 if stringent measures are taken immediately.

Forecasting COVID-19 Dynamics and Endpoint in Bangladesh: A Data-driven Approach

Abstract

On December 31, 2019, the World Health Organization (WHO) was informed that atypical pneumonia-like cases have emerged in Wuhan City, Hubei province, China. WHO identified it as a novel coronavirus and declared a global pandemic on March 11th, 2020. At the time of writing this, the COVID-19 claimed more than 440 thousand lives worldwide and led to the global economy and social life into an abyss edge in the living memory. As of now, the confirmed cases in Bangladesh have surpassed 100 thousand and more than 1343 deaths putting startling concern on the policymakers and health professionals; thus, prediction models are necessary to forecast a possible number of cases in the future. To shed light on it, in this paper, we presented data-driven estimation methods, the Long Short-Term Memory (LSTM) networks, and Logistic Curve methods to predict the possible number of COVID-19 cases in Bangladesh for the upcoming months. The results using Logistic Curve suggests that Bangladesh has passed the inflection point on around 28-30 May 2020, a plausible end date to be on the 2nd of January 2021 and it is expected that the total number of infected people to be between 187 thousand to 193 thousand with the assumption that stringent policies are in place. The logistic curve also suggested that Bangladesh would reach peak COVID-19 cases at the end of August with more than 185 thousand total confirmed cases, and around 6000 thousand daily new cases may observe. Our findings recommend that the containment strategies should immediately implement to reduce transmission and epidemic rate of COVID-19 in upcoming days.

Keywords: COVID-19, Recurrent neural networks, Epidemic transmission, Bangladesh, Epidemic prediction, Logistic Curve

Introduction

Coronavirus disease 2019 (COVID-19), the ongoing worldwide pandemic, is a highly contagious disease that predominantly causes respiratory complications ranging from the cough to acute respiratory distress syndrome. At the time of writing this, it has claimed more than 4,46000 lives, spread in 213 countries around the world. Since the 8th December 2019, atypical pneumonia cases have identified in Wuhan City, Hubei province, China (Zhu et al., 2020), later this disease broke out within China and continued to spread all over the world, causing startling panic worldwide. Since then, it poses unprecedented challenges to global health and a challenging task for the scientific community to identify its behavior. In the following months, the World Health Organization (WHO) declared it a global pandemic due to its highly contagious behavior (Wang, Wang, Ye, & Liu, 2020). The virus severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), named by the International Committee on Taxonomy of Viruses (ICTV), that causes COVID-19 belongs to genus Beta coronavirus, and it is a positive-sense single-stranded RNA (+ssRNA) virus and thus can mutate frequently to cope up with changing environments (Kaneko, Nimmerjahn, & Ravetch, 2006). The mean incubation period of COVID-19 is estimated 6.5 days (2-14 days) while patients remain asymptomatic or experience a little symptom and thus spreading the virus silently (Lai, Shih, Ko, Tang, & Hsueh, 2020; Rothan & Byrareddy, 2020). As of now, there is no proven vaccine of COVID-19, that altogether contributed to its seriousness. Symptoms of COVID-19 vary from person to person. However, common symptoms include respiratory stress, fever, cough, shortness of breath, and breathing difficulties (Yang et al., 2020). It is estimated that the case-fatality rate of COVID-19 is 2.3%, but older people with comorbidities, the rate is higher (Novel, 2020). WHO (2020) recommended that preventive measures for COVID-19 includes maintaining social distancing, washing hands frequently, avoiding touching the mouth, nose, and face. The first three cases of COVID-19 was reported

in Bangladesh on 8th March 2020 origin from Italy. It spreads to the maximum of the districts of the country within a month. After the confirmation of the first three cases of COVID-19 on 8th March 2020, the government of Bangladesh declared a 10 days lockdown from 26th March 2020 to 4th April 2020. However, lockdown periods were later extended several times until 30th May 2020, and then zone-based lockdown strategies have been taken. In addition, All the educational institutions closed down from 17th March 2020 and still ongoing. The government of Bangladesh had suspended all national and international flights from 30th March 2020 and reopened international flights on a small scale from 16th June 2020.
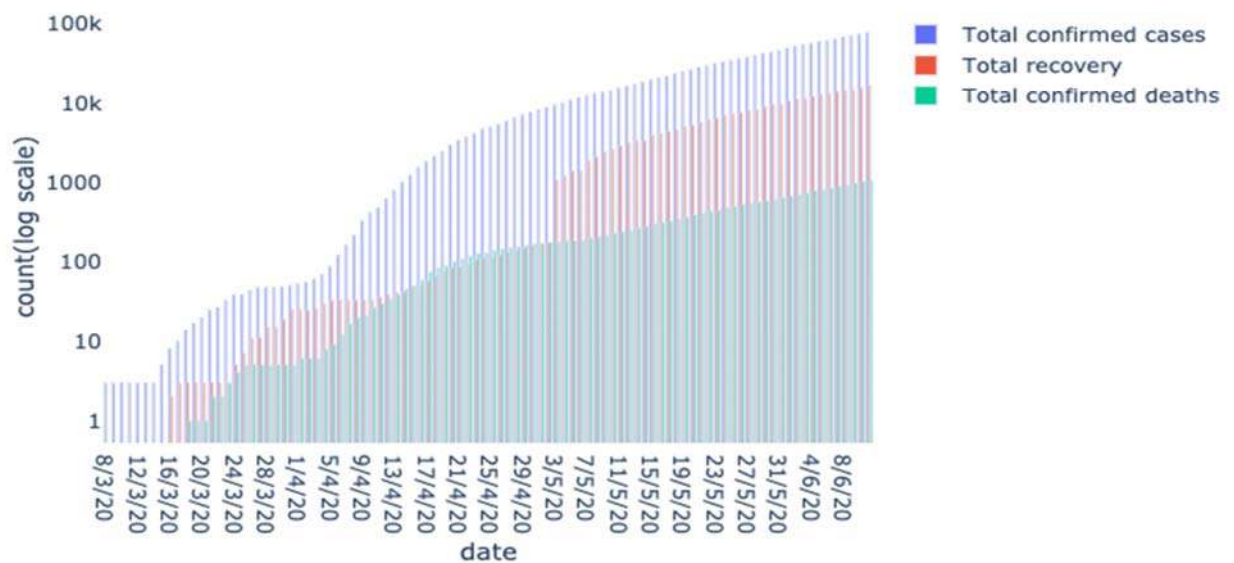

Figure 1 Total confirmed, recovered and death cases in Bangladesh up to June 8, 2020

At the time of writing this, the confirmed cases in Bangladesh have surpassed 100 thousand and more than 1343 death toll. The administration is struggling to accommodate COVID-19 patients. Therefore, some predictive models will be an excellent tool to predict future cases in upcoming days (Tobías, 2020; L. Wang et al., 2020). The COVID-19 data depends on non-

linear function; thus, using linear methods might fail to capture the COVID-19 dynamics (Chimmula & Zhang, 2020). Several statistical methods such as Auto-Regressive Moving Average (ARIMA), Moving Average (MA), Auto-Regressive (AR) methods have been used to understand the COVID-19 dynamics (Benvenuto, Giovanetti, Vassallo, Angeletti, & Ciccozzi, 2020; Dehesh, Mardani-Fard, & Dehesh, 2020) but these methods did not fit well COVID-19. To overcome such barriers, we adopted data-driven Deep Learning-based LSTM networks and Logistic Curve fitting methods to predict future likely cases and deaths. The proposed methods approximately fit real data; thus, policymakers can prepare to take the rush of COVID-19 in upcoming days. Besides, early prediction using mathematical models would help government officials of Bangladesh and also the health administration of Bangladesh to prepare beforehand and take the precautionary measures to compensate for the causalities of COVID-19 Pandemic. Mathematical models are often used to understand infectious disease dynamics. Several initial studies were performed for COVID-19 transmission. Reproduction Number (R_0) is often used in order to model infectious disease. An R_0 is, basically, how many people can get the infection by the infected person. An R_0 value 3 represent infected people will infect 3 persons around him/her. Liu, Gayle, Wilder-Smith, and Rocklöv (2020) estimated initial R_0 of COVID-19 in China was 4.2. The R_0 for the Diamond Princess, a cruise ship was estimated at 2.1 (Zhao et al., 2020). However, such an approach often leads to difficulties in predicting future cases. The Logistic Growth model and the Susceptible-Infectious-Recovered (SIR) model often use for epidemiological prediction and also have been employed on COVID-19 data. Roosa et al. (2020) used a Generalized Logistic Growth model to assess short-term forecasts of the total confirmed cases in China. Vattay (2020), using the Logistic Growth model, predicted the ultimate death numbers in Hubei, China and Italy. Zhou et al. (2020), using the Susceptible-Exposed-Infectious-Recovered (SEIR) and Logistic Growth model, forecasted the epidemic size of eight different countries. Rahman,

Ahmed, Hossain, Haque, and Hossain (2020) using SIR models projected final infection size in Bangladesh in assuming eight different scenarios. al Azad and Hussain (2020) using the SIR model estimated that the R_0 of COVID-19 in Bangladesh as of May is 2.88.

However, reviewing the literature, it is observed that a data-driven approach to the prediction of COVID-19 dynamics is limited in Bangladesh. Therefore, this study adopted a deep learning method LSTM and Logistic Curve model to predict likely future cases of COVID-19, its peak time, and the plausible end-date of the COVID-19 pandemic in Bangladesh. Our analysis may help policymakers and health professionals to take necessary measures toward COVID-19 in upcoming months.

The layout of this article is as follows: In section 2, similar studies have been presented. In section 3, methods and models have been explained. In section 4, results and discussion has been discussed, and in section 5, concluding remarks have been presented.

Related Works

Several works had been done globally with Machine learning approaches since the beginning of the pandemic COVID-19, some of them are mentioned below:

Chimmula and Zhang (2020), using the LSTM model, they predicted the possible ending point of the Pandemic COVID-19 outbreak in Canada would be around June 2020.

Kafieh et al. (2020) used a different machine learning approach to predict the future number of cases in Iran and found that M-LSTM was the most accurate model for their study.

Tomar and Gupta (2020) used the LSTM and curve-fitting model for prediction of the number of COVID-19 cases in India 30 days ahead and showed the effect of preventive measures like social isolation and lockdown on the spread of COVID-19.

Hu, Ge, Jin, and Xiong (2020) developed Modified Auto-Encoders (MAE) to forecast the number of accumulative and new confirmed cases of Covid-19 in China.

Fountain-Jones et al. (2019) used Machine learning models to predict disease risk in animal populations by pathogen exposure.

Methods & Models

Dataset

The COVID-19 data used in this research is collected from Johns Hopkins University's GitHub repository (https://github.com/CSSEGISandData/COVID-19). The dataset also includes the number of total confirmed cases, fatalities, and recovered patients by the end of each day. This study has been used the COVID-19 data of Bangladesh from 8th March 2020 to 13th June 2020 for LSTM models. The deep learning models are trained with 98 occurrences of total confirmed, recovered and death COVID-19 cases. The dataset is divided into 97% training set and 3% test set. For the Logistic Curve fitting model, data from 8th March 2020 to 18th June 2020 has been used.

Time series (TS)

A sequence of data points over a regular interval of time is called time-series (TS). In the regression predictive modeling approach, it is desired that the observation is in temporal structure. In other words, it will remain consistent over time. In TS terminology, the consistency is referred to as time series being stationary, which indicates it has constant mean

and variance in respect to time. These are important characteristics in TS; however, it can be easily violated by presences of trend, seasonality, and error in TS. A TS is said to have a trend when a certain pattern repeats over regular intervals of time. A non-stationary time series has trends, seasonal effects; as a result, its statistical properties change over time. Time series datasets of COVID-19 may have non-stationary patterns due to lockdown, social distancing, etc. Thus, it is important to know the nature of TS before applying forecasting methods on the given TS dataset. This study adopted the Augmented Dickey-Fuller (ADF) test (Cheung & Lai, 1995) to check the nature of TS. ADF test checks the stationarity of TS dataset. ADF is a unit root test that checks the impacts of trends in a given TS. The results are interpreted by the p-value from the test. If the p-value is between a threshold of (1-5) %, it suggests that we can reject the null hypothesis (i.e., it does not have a unit root, and it can be regarded as stationary series); otherwise, a p-value above the threshold suggests we fail to reject the null hypothesis (i.e., it does have a unit root, and it can be regarded as non-stationary series). The ADF test of our study suggests we fail to reject the null hypothesis; the data has a unit root, and data is non-stationary.

Long Short-Term Memory (LSTM) Network

Recurrent neural network (RNN) is a type of artificial neural network, which is used in temporal domains to learn sequential patterns. Recurrent LSTM networks can address the limitations of traditional time series forecasting techniques by adapting nonlinearities of given COVID-19 dataset and can result in a state of the art results on temporal data (Chimmula & Zhang, 2020). The Long Short-Term Memory model (LSTM) (M. Zhang, Geng, & Chen, 2020; Q. Zhang, Gao, Liu, & Zheng, 2020) is an advancement from the recurrent neural network. However, RNN suffers from vanishing gradient problems, meaning networks cannot learn from long data sequences. To overcome this barrier, the LSTM was

proposed by (Hochreiter & Schmidhuber, 1997). Each block of LSTM operates at different time steps and passes its output to the next block until the final LSTM block generates the sequential output (**Figure 2**). The basic structure of LSTM consists of four gates- input gate, forget gate, control gate and output gate.
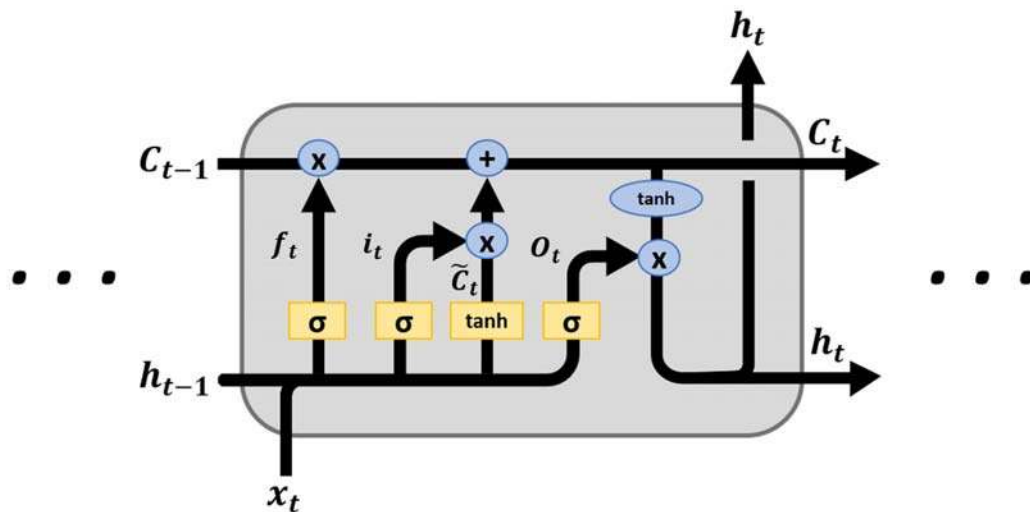

Figure 2 basic LSTM cell structure

In figure 2, i_t, f_t, C_t, O_t are the input gate, forget gate, control gate and output gate, respectively. The details of these four gates are enlightened below:

The input gate decides which information can be transferred from the previous cell to the current cell. The input gate is defined as

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{1}$$

The forget gate decides it should store the information from the input of previous memory or not. The forget gate is defined as

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{2}$$

The control gate controls the update of the cell and is defined as

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3}$$

Finally, the output gate updates the hidden layer (h_{t-1}) and also updates the output and is given by the following equations:

$$O_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o)$$

$$h_t = O_t * tanh(C_t)$$

(4)

In the above equations, x_t is the input and $W \ and \ b$ are the corresponding weight matrices and bias, respectively; C_{t-1} and C_t are the previous and current block memory, respectively; h_{t-1} and h_t are previous and current blocks the output, respectively. Moreover, $tanh$ is the hyperbolic tangent function that used to scale the values into range -1 to 1 and σ is the sigmoid activation function, which gives the output in between 0 to 1.

Logistic Curve Fitting Model

The Logistic Curve model is widely used to describe the growth of a population. An infectious disease outbreak like COVID-19 can be seen as the growth of the population of a pathogen agent; thus, a logistic model seems reasonable to model the progression of the agent (Bertolaccini & Spaggiari, 2020). The most generic expression of a logistic function is given below:

$$f(x, a, b, c) = \frac{c}{1 + e^{-(x-b)/a}}$$

(5)

Where x is the time and a, b, c are parameters; a refers to the infection speed, b is the day with the maximum infections occurred and c is the total number of recorded infected people at the infection's end.

Evaluation Metrics

Evaluating the model accuracy is a crucial part of a machine learning model, which gives us the idea of how well the model is predicting. There are many types of evaluation matrices and vary according to the model type. It is wise to evaluate a model with multiple metrics to

ensure the correct and optimal operation of the model. In our study, the following evaluation matrices have been considered:

Mean Squared Error (MSE)

MSE metric represents the difference between the true and predicted values extracted by squared the average difference over the dataset. The equation of MSE is given below:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}| \qquad (6)$$

Root Mean Squared Error (RMSE)

RMSE is nothing but the error rate by the square root of MSE and is defined by

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}|} \qquad (7)$$

R-squared (R^2)

The R^2 or the coefficient of determination metric represents the coefficient of how well the predicted values fit compared to the true values of the dataset. The value from 0 to 1 interpreted as percentages where higher value means the better model. The equation is given below:

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2} \qquad (8)$$

In the above equations, y_i is i^{th} indexed value, $\hat{y}$ and $\bar{y}$ are the predicted value and the mean value of y, respectively.

Mean Absolute Error Percentage (MAPE)

MAPE metric is used to measure the size of the error in percentage terms regarding the actual values and is defined as

$$MAPE = \frac{100}{N} \sum_{t=1}^{n} ||\frac{X_t - \hat{X}_t}{X_t}|| \qquad (9)$$

where X_t is the actual value and $\hat{X}_t$ is the corresponding estimated value for t^{th} sample from all n available samples.

Results and Discussion

In the following section, we present the experimental results for the validation and analysis of the proposed model introduced in this study. We hope that our approaches and predictive models will shed some light on the COVID-19 dynamics in Bangladesh and may help us to be aware of upcoming undesired circumstances and take necessary measures in advance and possible suppression of the COVID-19 Pandemic in Bangladesh.

Data-driven approach

This study has been used the COVID-19 data of Bangladesh from 8th March 2020 (When the first case of COVID-19 was registered in Bangladesh) to 13th June 2020. The deep learning models are trained with 98 occurrences of total confirmed, recovered and death COVID-19 cases. The dataset is divided into 97% training set (from 8th March 2020 to 10th June 2020) and 3% test set (from 11th June 2020 to 13th June 2020). For Logistic Curve fitting data from 8th March 2020 to 18th June 2020 has been considered. The performance evaluation of LSTM models is examined with mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE). Before selecting the best-performed LSTM model,

we considered three types of LSTM models (i.e. Vanilla, Stacked, Bidirectional) for univariate time series forecasting. We executed each model for total confirmed, recovered and deceased data, then the best-performed model is considered based on performance metrics. The performance results of different LSTM models are described in Table 1.

Table 1 performance results of different LSTM models

Total Confirmed				
LSTM Model(s)	**MSE**	**RMSE**	R^2	**MAPE**
Vanilla	3164020	1778.77		
Stacked	**352556.652**	**593.764**	**0.95**	**1.76%**
Bidirectional	4040000	2009.970		
Total Recovered				
Vanilla	842289	917.763		
Stacked	1240980	1113.990		
Bidirectional	**54610.71**	**233.689**	**0.72**	**1.18%**
Total Death				
Vanilla	5499.64	74.1595		
Stacked	1359.87	36.876		
Bidirectional	**406.60**	**20.164**	**0.75**	**1.25%**

From Table 1, we can observe that for total confirmed cases, the stacked LSTM model performed better than the vanilla and bidirectional LSTM model. The RMSE score of selected stacked LSTM model for total confirmed cases is 593.36, R^2 score 0.95 and accuracy 98.24%. For total recovered, the bidirectional LSTM model performed better than others; the RMSE is 233.68 and R^2 score 0.72 and accuracy 98.82%. For total death,

bidirectional LSTM performed better than others; the RMSE is 20.16, R^2 score 0.75, and accuracy 98.75%.
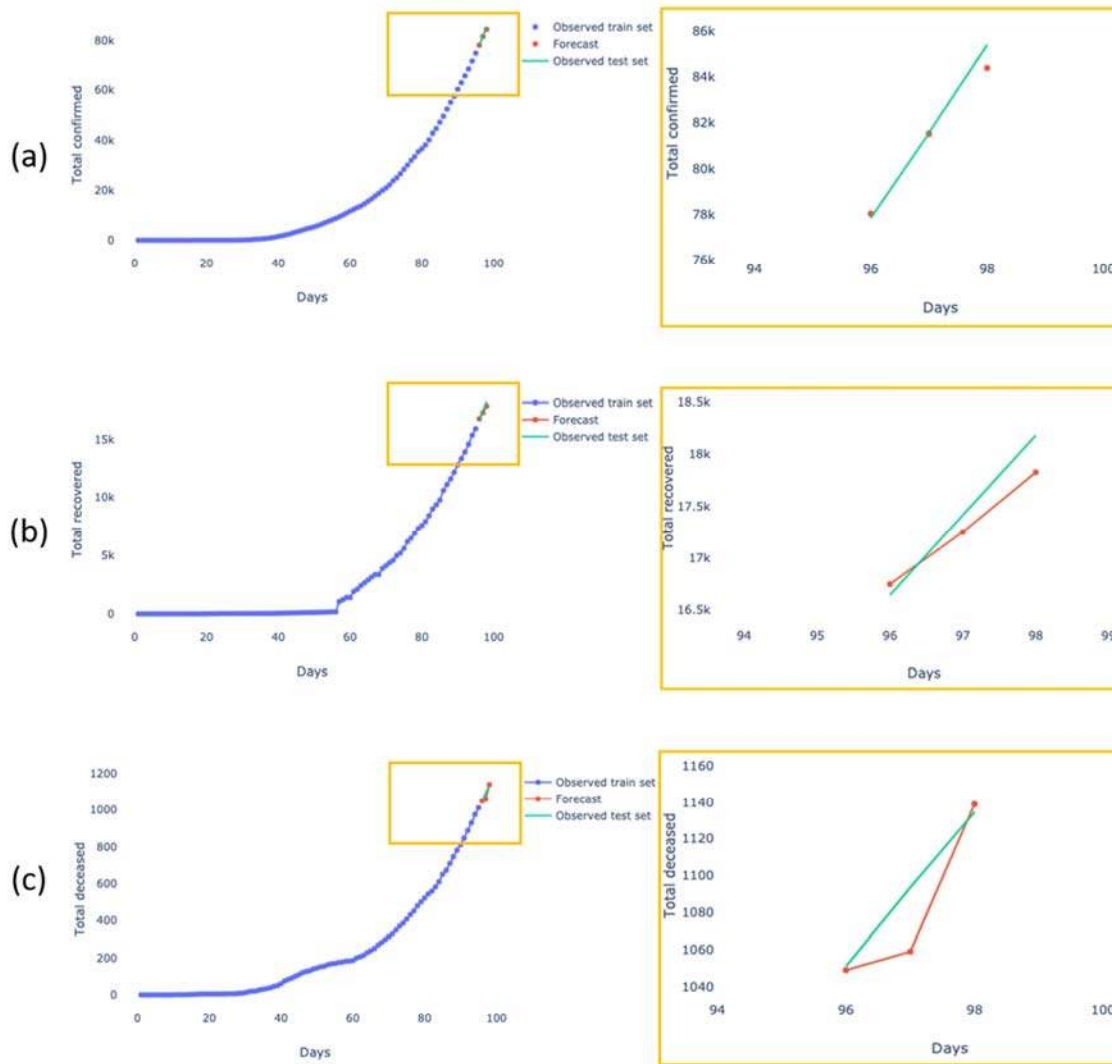

Figure 3 (a) Total number of confirmed cases prediction using LSTM (b) Total number of recovered cases prediction using LSTM (c) Total number of deceased cases prediction using LSTM.

LSTM results suggest that in the near future, the total numbers of confirmed cases will continue to grow exponentially. Thus, strict-lockdown and maintaining social distancing is necessary to reduce the transmission of COVID-19. The predicted data from LSTM model is very close to actual.

Logistic Curve fitting Model

Traditional infectious disease modeling uses differential equations based on the dynamics of
the diseases where too many factors are also associated; thus, it often leads to over-fitting. It
is observed that the Logistic Curve fits the COVID-19 data of Bangladesh very well. The
parameters of the Logistic Curve are obtained by using a non-linear optimization algorithm,
and results are shown in Table 2. The COVID-19 data of Bangladesh from 8th March 2020 to
18th June 2020 has been considered for Logistic Curve fitting.

Table 2 parameters obtained for total confirmed cases using the logistic model with standard
errors.

Parameter	a	b	c	R^2	Plausible end of the infection
Total confirmed cases	14.702 ± 0.187	15/06/2020	187169 ± 5610	0.99918	~= 02/01/2021
Total deceased cases	18.262 ± 0.365	16/06/2020	3692 ± 311	0.99810	

Total number of confirmed cases

As of 18th June 2020, the total confirmed cases of COVID-19 in Bangladesh are 102,292 and
ranked in 17th most infected countries in the world. The number of recorded cases has
increased dramatically between 40-50 days since the first confirmed cases, which represents
the sudden changes from where the number of infected cases started following an exponential
trend. It is observed that the Logistic curve fitting model (LG) approximates the total number

of confirmed cases very well. The fitting parameters are included in Table 2, where a is the infection speed, b is the day with maximum infections occurred and c refers to the total number of confirmed infected people at the infection's end and R^2 is the fitting goodness of the total confirmed cases.
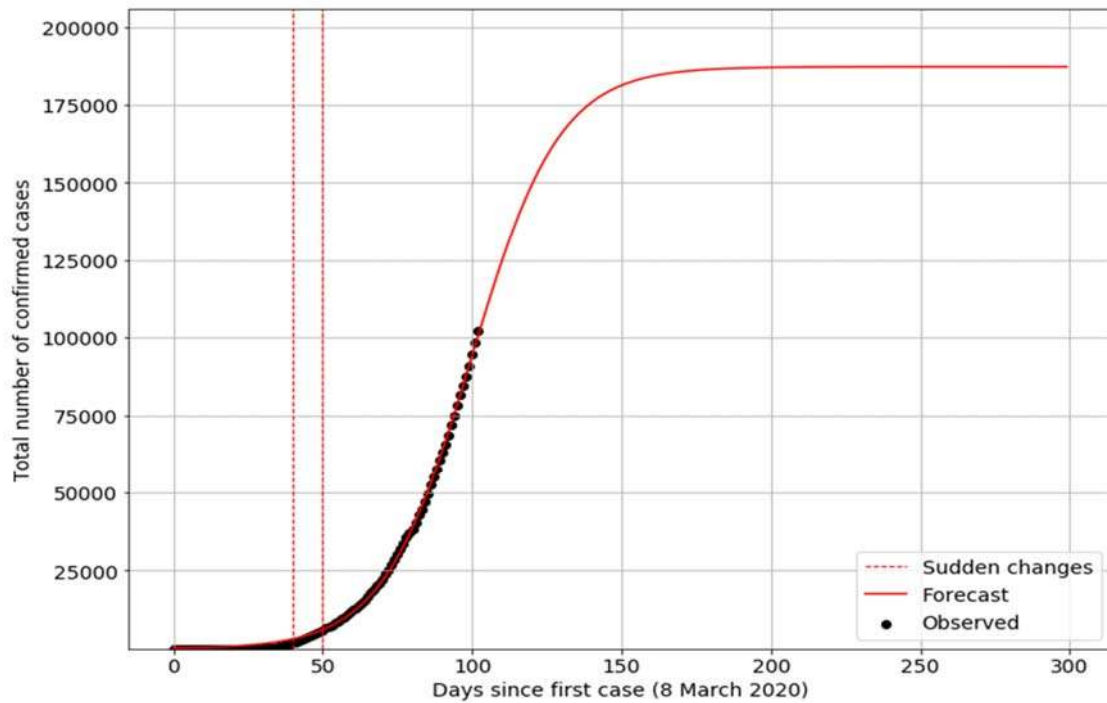

Figure 4 Logistic Curve fitting on the total number of confirmed cases

The logistic curve of Figure 4 suggests that Bangladesh is approaching its peak. Since the data is approximately fitting with the curve, this study suggests that assuming the people will maintain social distancing and the government of Bangladesh will keep various kinds of interventions, at the end of June 2020, there will be 135-140 thousand total confirmed cases. At the end of July 2020, there will be 178-180 thousand total confirmed cases, and at the end of August 2020, there will be more than 186 thousand confirmed cases, where the attainment of peak cases will be reached. Figure 4 also suggests that the curve will flatten at the

beginning of September 2020 if strict-lockdown in place. The plausible end of the infection is at the beginning of January 2021 with total confirmed cases in between 187 to 193 thousand.

Inflection point

Turning point or inflection point is the point where the exponential growth of transmission rate starts to decrease; more formally, its concavity changes from upwards to downwards. It is the midpoint of the spread of an infectious disease. In order to know the inflection point, we used Logistic curve-fitting on the new confirmed cases in Figure 5. Based on Figure 5, Bangladesh has recently passed the rising inflection point and approaching the peak of the infections.
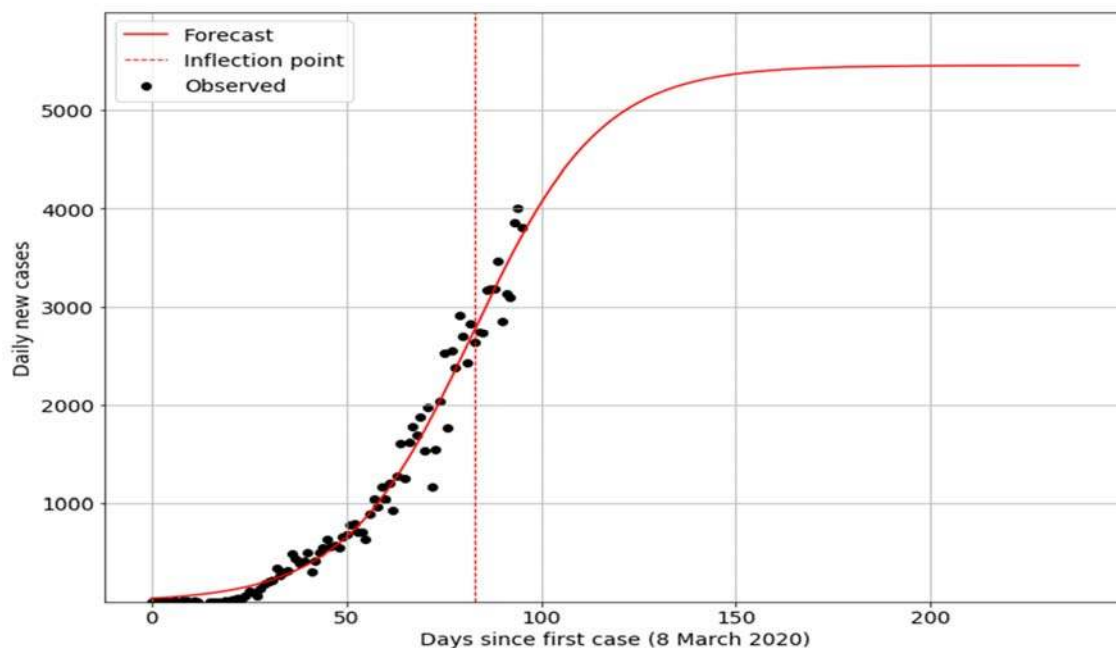

Figure 5 Logistic Curve fitting on the total number of new cases

Figure 5 also suggests that a rising inflection point was passed around May 28 – 30, 2020. It was the 83-85 days later since the first confirmed cases. Since Bangladesh has recently passed the inflection point in terms of the COVID-19, it indicates that COVID-19 cases are nearing its peak. Based on Figure 4 and 5, the Logistic Curve predicted that the highest peak

will be around at the end of August 2020 with more than 186 thousand confirmed cases, where daily new cases may surpass more than 6,000. However, it is worth noting that transmission of the COVID-19 depends on several exogenous factors such as lockdown, social-distancing; thus, this prediction might slightly vary to reality.

Total death toll

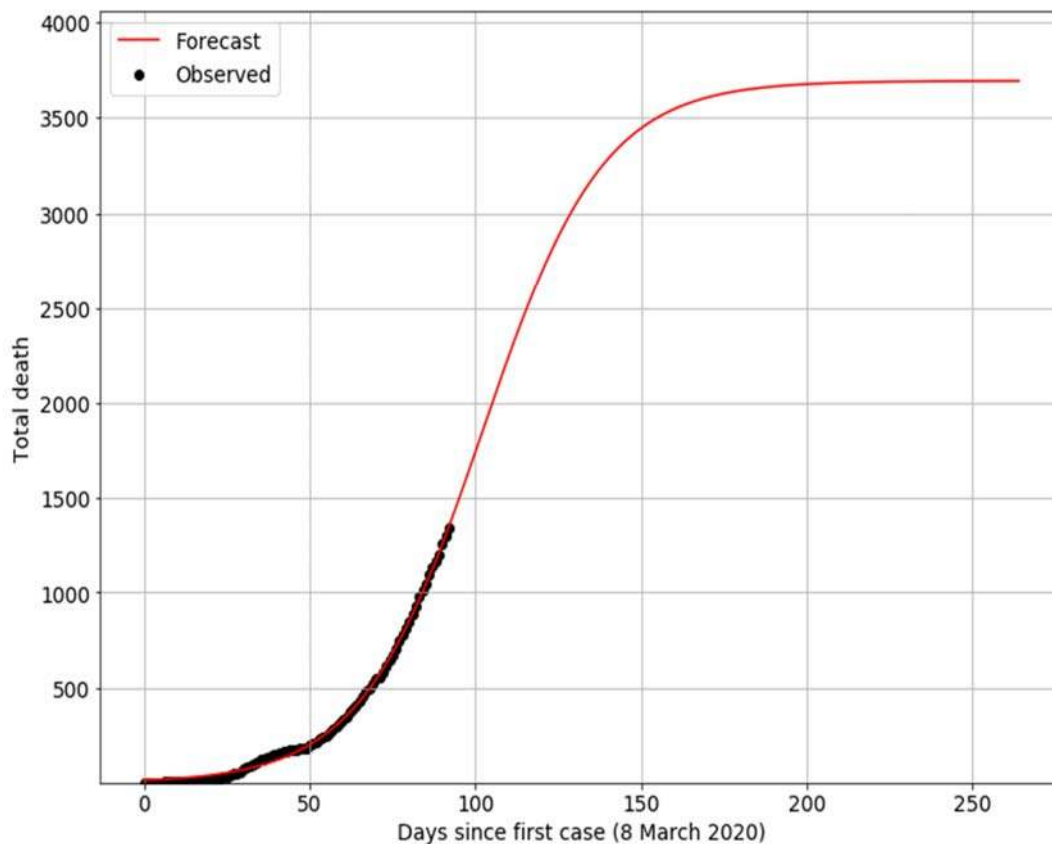Logistic model used to predict the total death toll in Bangladesh in upcoming days.

Figure 6 Logistic Curve fitting on the total number of death cases.

It is observed that the Logistic model approximates the total number of death tolls very well. The fitting parameters are included in Table 2, where a is the death speed, b is the day with maximum deaths occurred, and c refers to the total number of deaths at the infection's end and R^2 is the fitting goodness of the total deaths. The LG curve suggests that at the end of

June 2020, there will be around 200 death tolls, and at the end of July 2020, there will be more than 3150 death tolls. The maximum death toll will reach around the end of August 2020, with more than 3600 total death cases. Since the number of total deaths depends on the health care capacity of a country, we suspect this number will vary to reality.

Conclusion

In this study, a data-driven forecasting/estimation method has been adopted to estimate possible total confirmed and deceased cases in upcoming months. The rate of transmission in Bangladesh, still, is following exponential growth despite taking several intervention strategies by the government of Bangladesh. This is the first study, to best of our knowledge, to model COVID-19 transmission in Bangladesh using a deep learning approach. However, it worth noting that despite of having small-dataset, our model gave promising accuracy. Thus, these models will help policymakers to understand the future of the COVID-19 trajectory in Bangladesh and implementing interventions as well as estimate the impact of interventions. Our LSTM models can give valuable insights into the COVID-19 transmission in Bangladesh and possibly predict future outbreaks. In addition, we also adopted the Logistic curve fitting model to estimate confirmed cases and the death toll. However, it worth noting that the prediction and conclusion drawn from the results are under the preconditions that control measures for the COVID-19 in Bangladesh are stable and reliable, and the virus will not mutate as well in the future. The conclusion has drawn from fitting 103 days data where 65 days (25th March 2020 to 30th June 2020) of nation-wide lockdown was in place. The prediction will improve if new data are generated in the upcoming days.

Acknowledgment

Ethical Consideration

Not applicable

Funding

For this research, authors did not receive any grant from any agency.

Conflict of interest

Authors declare no conflict of interest.

Author's contribution

AEEH: conceptualization, design, manuscript writing, software. MN: design, data curation, manuscript writing. NUE: design, data analysis, final drafting. IHT: data curation, data analysis. MAAM: software, data analysis. SA: monitoring, draft revision. MSI: supervison, final drafting.

Consent to publication

All authors have read the final version of the manuscript and approved the manuscript.

References

al Azad, A. K., & Hussain, A. M. (2020). Modeling and Analysis of The Early-Growth Dynamics of COVID-19 Transmission.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, 105340.

Bertolaccini, L., & Spaggiari, L. (2020). The hearth of mathematical and statistical modelling during the Coronavirus pandemic: Oxford University Press.

Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics, 13*(3), 277-280.

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 109864.

Dehesh, T., Mardani-Fard, H., & Dehesh, P. (2020). Forecasting of covid-19 confirmed cases in different countries with arima models. *medRxiv*.

Fountain-Jones, N. M., Machado, G., Carver, S., Packer, C., Recamonde-Mendoza, M., & Craft, M. E. (2019). How to make more from exposure data? An integrated machine learning pipeline to predict pathogen exposure. *Journal of Animal Ecology, 88*(10), 1447-1461.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in china. *arXiv preprint arXiv:2002.07112*.

Kafieh, R., Arian, R., Saeedizadeh, N., Minaee, S., Yadav, S. K., Vaezi, A., . . . Javanmard, S. H. (2020). COVID-19 in Iran: A Deeper Look Into The Future. *medRxiv*.

Kaneko, Y., Nimmerjahn, F., & Ravetch, J. V. (2006). Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *science, 313*(5787), 670-673.

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *International journal of antimicrobial agents*, 105924.

Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine*.

Novel, C. P. E. R. E. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi, 41*(2), 145.

Rahman, M. M., Ahmed, A., Hossain, K. M., Haque, T., & Hossain, M. A. (2020). Impact of control strategies on COVID-19 pandemic and the SIR model based forecasting in Bangladesh. *medRxiv*.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., . . . Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling, 5*, 256-263.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity*, 102433.

Tobías, A. (2020). Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Science of the Total Environment*, 138539.

Tomar, A., & Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Science of The Total Environment*, 138762.

Vattay, G. (2020). Predicting the ultimate outcome of the COVID-19 outbreak in Italy. *arXiv preprint arXiv:2003.07912*.

Wang, L.-s., Wang, Y.-r., Ye, D.-w., & Liu, Q.-q. (2020). A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence. *International journal of antimicrobial agents*, 105948.

Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Cao, Y., . . . Gu, T. (2020). Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Science of the Total Environment*, 138394.

WHO. (2020). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/ad-ice-for-public.

Yang, C.-T., Chen, Y.-A., Chan, Y.-W., Lee, C.-L., Tsan, Y.-T., Chan, W.-C., & Liu, P.-Y. (2020). Influenza-like illness prediction using a long short-term memory deep learning model with multiple open data sources. *The Journal of Supercomputing*, 1-27.

Zhang, M., Geng, G., & Chen, J. (2020). Semi-Supervised Bidirectional Long Short-Term Memory and Conditional Random Fields Model for Named-Entity Recognition Using Embeddings from Language Models Representations. *Entropy, 22*(2), 252.

Zhang, Q., Gao, T., Liu, X., & Zheng, Y. (2020). Public Environment Emotion Prediction Model Using LSTM Network. *Sustainability, 12*(4), 1665.

Zhao, S., Cao, P., Gao, D., Zhuang, Z., Chong, M. K., Cai, Y., . . . He, D. (2020). Epidemic growth and reproduction number for the novel coronavirus disease (COVID-19) outbreak on the Diamond Princess cruise ship from January 20 to February 19, 2020: A preliminary data-driven analysis. *Available at SSRN 3543150*.

Zhou, X., Ma, X., Hong, N., Su, L., Ma, Y., He, J., . . . Zhu, W. (2020). Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model. *medRxiv*.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., . . . Lu, R. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*.