# Analysis and Prediction of COVID-19 Pandemic in Bangladesh by Using ANFIS and LSTM Network

Anjir Ahmed Chowdhury[1] · Khandaker Tabin Hasan[1] · Khadija Kubra Shahjalal Hoque[1]

## Abstract

The dangerously contagious virus named "COVID-19" has struck the world strong and has locked down billions of people in their homes to stop the further spread. All the researchers and scientists in various fields are continually developing a vaccine and prevention methods to aid the world from this challenging situation. However, a reliable prediction of the epidemic may help control this contiguous disease until the cure is available. The machine learning techniques are one of the frontiers in predicting this outbreak's future trend and behavior. Our research is focused on finding a suitable machine learning algorithm that can predict the COVID-19 daily new cases with higher accuracy. This research has used the adaptive neuro-fuzzy inference system (ANFIS) and the long short-term memory (LSTM) to foresee the newly infected cases in Bangladesh. We have compared both the experiments' results, and it can be forenamed that LSTM has shown more satisfactory results. Upon study and testing on several models, we have shown that LSTM works better on a scenario-based model for Bangladesh with mean absolute percentage error (MAPE)—4.51, root-mean-square error (RMSE)—6.55, and correlation coefficient—0.75. This study is expected to shed light on COVID-19 prediction models for researchers working with machine learning techniques and avoid proven failures, especially for small imprecise datasets.

**Keywords** LSTM · ANFIS · COVID-19 · Machine learning · Epidemic prediction model

## Introduction

Recently, the novel virus named "coronavirus" or we all know it as "COVID-19" is a subject that is under intense study. SARS-CoV-2 is a large family [1] of viruses that can cause fatal health issues in human beings. Researchers and scientists are running against time to find a vaccine and further prevention for this virus. The current outbreak of the virus was reported in late December of 2019 [1], in Wuhan, China. This virus is so contagious that it has spread all around the world in a short period. On January 30, 2020, WHO (World Health Organization) Director-General Dr.

Tedros Adhanom Ghebreyesus declared this outbreak as a Public Health Emergency [2] of International Concern. Later on, March 11, WHO announced this unseen, dangerous, infectious virus as pandemic [2], when more cases and deaths were reported. The world has been facing many challenges due to these pandemics periodically over the centuries, and the effects of these pandemics have an enormous impact on the world. Moreover, the world does not only crash economically [3] but also the overall strengths and morals of the people are impacted.

Furthermore, to break the chain of this contagious virus, many countries [3] took preventive measures, including travel bans, home-office, country lockdown, and most importantly, social distancing. However, Bangladesh, a lower-middle-income economy and one of the most populated countries globally, is facing many challenges. Therefore, it is essential to know the accurate prediction and the natural progression of this virus for a country like Bangladesh. Figure 1 shows the trend of the new cases of Bangladesh. The field of statistics, ML (machine learning) and AI (artificial intelligence), has provided us with various techniques and models to predict the next stage of this virus and take

✉ Khandaker Tabin Hasan
  tabin@aiub.edu

  Anjir Ahmed Chowdhury
  anzira431@gmail.com

  Khadija Kubra Shahjalal Hoque
  khadijakubra197@gmail.com

1 Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka, Bangladesh
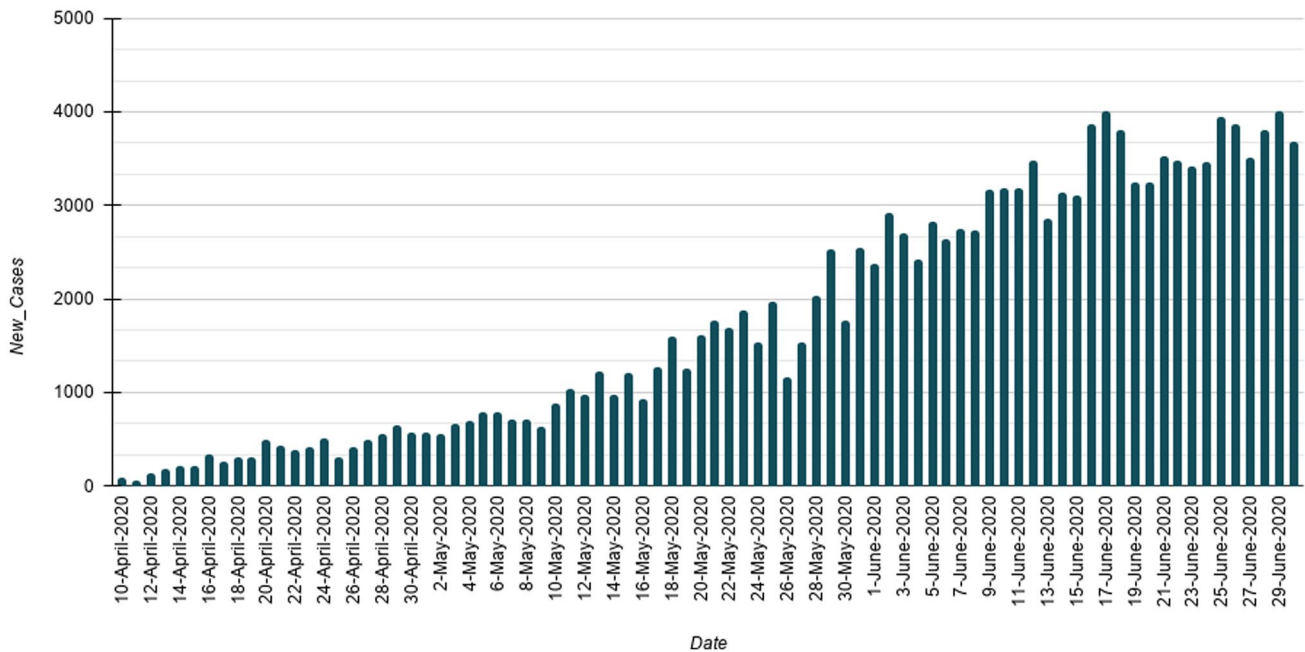
# New Cases vs. Date

## (Bangladesh)



**Fig. 1** Bangladesh new cases data from April 10–June 30, 2020 [4]

effective measures beforehand. This study aims to develop a prediction model and facilitate the decision-making process, where data are analyzed to get the perspective. Due to the data insufficiency in such a short period, which is not sufficient to train the AI models. Nevertheless, an effectively trainable AI model for time-series data is required, which can help forecast efficiency. At the very beginning of an epidemic, it is necessary to take proper and practical measures to prevent it from crossing the geological boundaries [5], which can save many lives with less impact. Therefore, forecasting and the proper study of the disease spread pattern can be a strategic plan in the planning of control strategy. Researchers can utilize predictive models to generate the prognosis, which depends on the underlying forecasting algorithm. In the case of COVID-19, the scenario is changing intensely with each passing day. In this situation, deep learning algorithms can provide a better mirror of the eye and upsurge [5]. Yang et al. have used an AI model taking the previous Severe Acute Respiratory Syndrome (SARS) dataset, which also shows the assurance for the future expectations of the plague [6]. They also introduced the dynamic SEIR (Susceptible-Exposed-Infectious-Recovered) model for foreseeing the COVID-19 contagion apex and sizes. More research works such as Sujatha and Chatterjee [7] have proposed a model that can be useful for perceiving the COVID-19 spread. They have used the linear regression,

Multilayer perceptron, and vector autoregression model on the Kaggle data for COVID-19 to visualize the hygienic example of the illness and measure of COVID-19 cases in India. Haytham H. Elmousalami et al. [8] also has utilized the time-series models and numerical detailing to represent the correlation of the day level to determine the COVID-19 influenced cases. For the time series prediction and forecasting problems, the adaptive neuro-fuzzy inference system (ANFIS) [9] is universally applied as it has demonstrated a good performance in many extant applications. Furthermore, the adaptive neuro-fuzzy inference system (ANFIS) is flexible. It can determine the nonlinearity in the time series and combine both the Artificial Neural Network (ANN) and the fuzzy logic systems. Al-qaness et al. has proposed the FPASSA-ANFIS model, which is the enhanced model of ANFIS and flower pollination algorithm (FPA) using the salp swarm algorithm(SSA). They have used the WHO (World Health Organization) official data for predicting the confirmed cases for the upcoming ten days, namely in China and USA [10]. More researches like the paper [11] has suggested more effective models handle the non-linearity and the complexity of the COVID-19 time-series data. Moreover, long short-term memory (LSTM) is a model that holds the correlation of the time series dynamics. The COVID-19 data are time series data that compiles the number of confirmed cases where the cases increase endlessly over time until it

arrives at a specific concentrated peak curve. From there, the LSTM captures the pattern of the dynamic growth of graphs with a minimum RMSE compared to RNN [11]. This paper has been organized as follows. In Section 2, the methods of LSTM and ANFIS are explained in detail. In Section 3, we have discussed how the dataset was prepared. Then, in Section 4, we have presented our results. Later, in Sections 5 and 6, we have discussed the findings and limitations of the results and possible threats to validity, respectively. Lastly, in Section 7, the conclusion of the work is given.

## Methods

The experiment was conducted in two different types of neural network systems. Therefore, one algorithm was chosen from the fuzzy inference system and another from the recurrent neural network. ANFIS was selected from the fuzzy inference system due for multiple reasons. This algorithm is a combination of ANN and Takagi–Sugeno fuzzy inference systems. Thus, it takes all the advantages of both approaches in a single framework. Furthermore, ANFIS is considered a universal estimator. Therefore, there is always an opportunity to find ANFIS more efficient and optimal if we can find the best parameter by genetic algorithm [12, 13]. This system also can process non-linear functions by using the sets of fuzzy "IF–THEN" rules [14]. That is why ANFIS is a perfect candidate for this experiment. We chose another algorithm from recurrent neural networks. There are so many networks like simple RNN, Gated Recurrent Unit (GRU), LSTM, etc. Each network has its advantages and disadvantages. For example, simple RNN does not have any gates, whereas GRU introduces gates to decide whether it will pass the previous input to the next cell or not. GRU also has a memory unit. The reason why we chose LSTM is that it has both advantages of RNN and GRU. Besides, it has two more extra gates called "Forgot gate" and "Output

gate" making LSTM more efficient than other recurrent neural networks. LSTM also has a feedback connection that allows it to process single data points and sequential data points. Hence, LSTM is another best candidate for our case as here we used time-series data, which is sequential. If we chose LSTM over RNN or GRU, then we can utilize so many extra features that give us the control to mix up inputs and weights. LSTM not only gives us the most control ability and flexibility but also gives better results.

### Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS architecture is a modified part of the artificial neural networks (ANNs). With the assimilation of the Takagi–Sugeno fuzzy which is the modification of the fuzzy logic [15, 16] system, and it prospers a high performance in both computing and learning techniques which are dealing with nonlinearity [15, 16] (See Fig. 2).

There are five main layers in the ANFIS model [17]. The layer starts with the input layer, which takes in the parameters and then constructs them into the model. This layer is also the input layer of the fuzzy system. Later, using the first layer's outputs, the second layer is created, which carries the member functions' (MF) preceding values. Using the nodes on the second layer, which steadfast the degree of activity on fuzzy rules are concluded. Moreover, the third layer of the ANFIS model expands the degree of the activity of any regulations. The second last layer normalizes the functions, and the nodes facilitate the production of the outputs and finally send them to the final layer, which is the output layer. Furthermore, the accuracy of the ANFIS model is determined using the number and type of MFs, the optimum method, and the output of the MF type [17]. The input parameters are set as the independent variables on each scenario, and the outcome was the number of cases. Among the eight MFs type, we chose the best three MFs types, namely triangular, trapezoidal, and Gaussian, was used to train the ANFIS

**Fig. 2** Architecture of the developed ANFIS
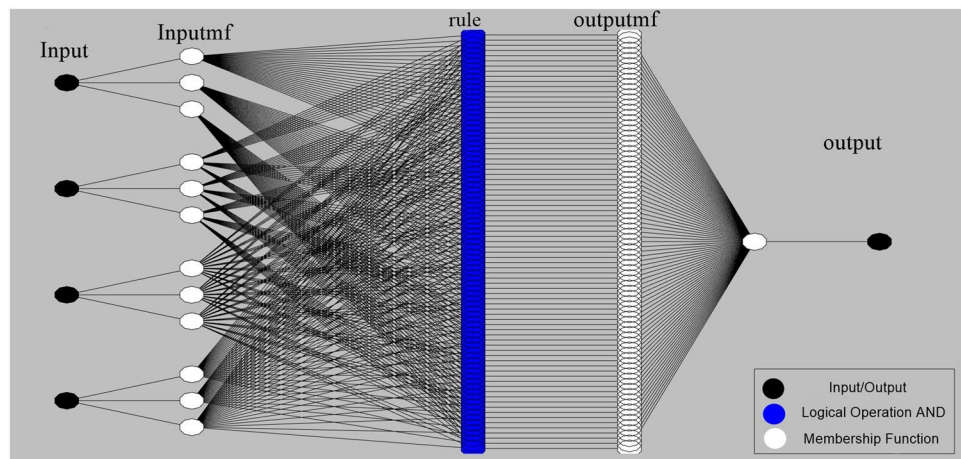
model so that the best MFs can be determined. Furthermore, for reducing the errors, the output membership function type was selected as linear. The optimum backpropagation (BP) method and the "0" value of error tolerance were done on training the fuzzy inference system (FIS).

## Long Short-Term Memory (LSTM)

From the various deep learning methods, we can say the recurrent neural network(RNN) has been convinced to be the most robust for prediction. As it can automatically excerpt the necessary features from the training samples, delivering the activation from the previous time step as the load for the present time step and the network self-connections [18]. As Connor et al. mentioned, RNN is satisfying at processing data and manifests promising results in the time-series prediction by hoarding enormous historical information in its internal state [19]. However,

it has an inconvenience of evanescence and the gradient detonate problems, leading to the abundant training time or the training does not work at all [20]. Therefore, to overcome this drawback, Hochreiter and Schmidhuber have designed the long short-term memory RNN structure in the year 1997 by dealing with the long-term dependency with the help of the multiplicative gates. These gates will manage the information flow, and the memory cells in the recurrent hidden layer[21]. The architecture of LSTM subsists of four gates, namely, input gate, forget gate, control gate, and the output gate [20]. Figure 3 shows the basic structure of LSTM.

Some equations describe the gates of LSTM. Before explaining the equation, we need to know some variables used in these equations. The weight matrices are W, and the activation function is $\sigma$, which is taken as the sigmoid. $h_{t-1}$ represents the output of the previous LSTM block, and b represents the bias for the respective gates. Lastly, $x_t$ is the
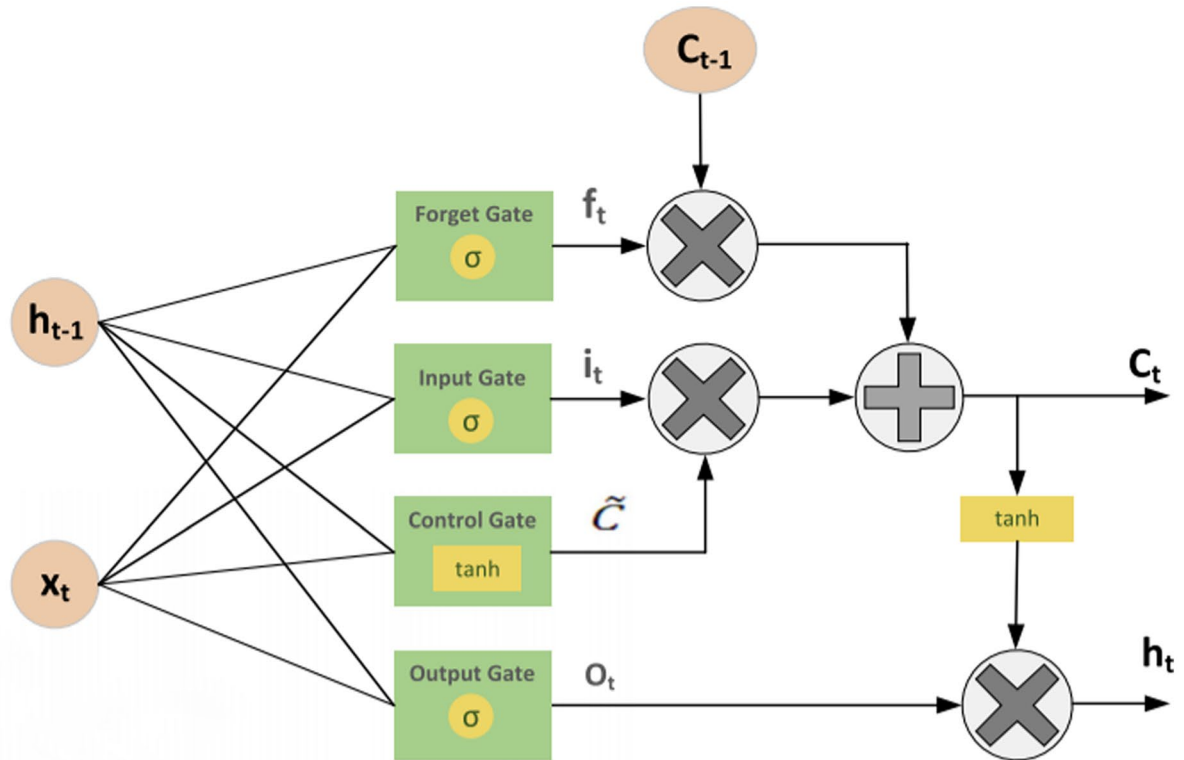


**Fig. 3** The basic structure of the LSTM [20]

input of the current timestamp. Now, The Input gate ($i_t$) is described as,

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

This equation chose information that can be passed to the cell. The data from the input side of a previous memory, which is to be ignored, is decided by the forget gate ($f_t$) by the following equation,

$$f_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \tag{2}$$

Control gate ($C_t$) controls the update of the cell by the following formula where tanh is used to normalize the values into range -1 to 1 and $\tilde{C}$ is the candidate for cell state at timestamp (t).

$$\tilde{C} = tanh(W_c * [h_{t-1}, x_t] + b_c)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3}$$

The output layer ($o_t$) updates both hidden layer $h_{t-1}$ and output as is given by,

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t) \tag{4}$$

## Hardware and Software Setup

The algorithms were not implemented from scratch. For the ANFIS algorithm, we used the fuzzy logic toolbox of Matlab 2016 software, and for the LSTM algorithm, we used python 3.7. We also used Tensorflow 2 and Keras as the API. For the supportive package, Jupyter notebook, pip3, NumPy, Matplotlib, and spicy were installed. All of these python packages performed their action on a Linux operating system. The hardware had the Intel core i5 4th generation processor with the clock speed of 3.20 GHz. It has 16 GB of DDR3 Ram, and the bus speed of the ram is around 1600Mhz. It also had the NVIDIA GTX 1060 6 GB graphics processing unit.

## Dataset

### Data Collection

The dataset is the statistical report of COVID-19 cases of Bangladesh, which is available online [4]. Right now, this online source is the most authentic source where daily covid cases are recorded. They collected data from other authentic sources. Many researchers are doing their research by collecting the data from this online source [17, 20]. The dataset range is from 10 April 2020 to 30 June 2020, for training the dataset is taken from 10 April 2020

to 19 June 2020, and the remaining days till 30 June (2020) is taken for testing the model. It is essential to train the algorithm at the initial step for developing a firm model [17]. Moreover, a forecasting model is recommended to be classified into two categories, i.e., input(s) (as independent variable(s)) and output(s) (as dependent variable(s)) [17]. The time-series data of Bangladesh was taken into three scenarios. The first and second scenario is designed for the desired output x(t) days where we have taken the last four consecutive odd and even days, respectively. Lastly, for the third scenario, the last four consecutive days are taken as the input for output x(t). The ANFIS model and the LSTM both were implemented on this dataset for achieving reliable prediction.

## Data Preparation

After collecting the data, we applied a strategy named teacher forcing. The teacher forcing method means if we expect an output y(t), then its input will be the prior time step x(t-1). Table 1 shows that we have to take four previous time steps as input because it is optimal for our case. We did a quick small test to determine the number of optimal inputs for our test. We used the LSTM algorithm with fixed parameters, which are also mentioned in Table 1 to determine which number of inputs gives us a better result. For this test, we considered prior to consecutive three steps, four steps, and five steps as inputs.

We can also understand from Table 1 that there are multiple methods to take the previous days. Consecutive is not the only way to take prior days. We also can consider prior odd or even days. However, we did not know which method is best for our experiment. That is why in Table 2, we proposed three ways (Scenarios). There might be other combinations like taking the first two consecutive days, one day gap, then the next two consecutive days, or different possible combinations. We did not include those possible combinations as they are made from the proposed three combinations.

## The Correlation Matrix

Before implementation, we examine the correlation between each input and output. Correlation is a ubiquitous term in the statistical area. It is a single number that gives acknowledgment

**Table 1** No. of inputs selection test results

| No. of inputs | Units | Batch Size | RMSE | MAPE | Corr Coff. |
|---|---|---|---|---|---|
| 3 | 5,10,15,20 | 16 | 113.73 | 9.87 | 0.19 |
| 4 | 5,10,15,20 | 16 | 60.15 | 6.71 | 0.51 |
| 5 | 5,10,15,20 | 16 | 301.22 | 0.07 | -0.27 |

**Table 2** Three proposed scenarios for time-series prediction of COVID-19 in Bangladesh

| Scenarios | Input | Output |
|---|---|---|
| Scenario 1 | $x_{t-1}, x_{t-3}, x_{t-5}$ and $x_{t-7}$ | $x_t$ |
| Scenario 2 | $x_{t-2}, x_{t-4}, x_{t-6}$ and $x_{t-8}$ | $x_t$ |
| Scenario 3 | $x_{t-1}, x_{t-2}, x_{t-3}$ and $x_{t-4}$ | $x_t$ |

of the degree of relationship between the inputs and output. The correlation matrix summarizes the dataset and visualizes the pattern in the dataset, and provides feedback on whether the dataset contains any unwanted data or not. Although each scenario's output column is the same, the input columns are different in every scenario. Every scenario is different in terms of input. So, we have to check the correlation of each scenario. As we mentioned earlier, that Table 1 has three different scenarios, which are used for training and testing. The first scenario was constructed with the previous four odd days as the input of each output. So the correlation matrix of the scenario is shown in Fig. 4,

From the matrix, we can summarize that the correlation between each column lies from 0.96 to 0.97. There are no anomalies found in this scenario. Then the second scenario is the vice versa of scenario 1 where instead of taking odd days, we took even days, and the correlation matrix of this scenario is shown in Fig. 5,

This scenario's correlation matrix is also shown to have a better correlation as in the previous scenario. For the third scenario, we took the consecutive last four days to predict the output and the correlation matrix of this scenario (Fig. 6). The reason behind choosing interval and alternative days is that most Machine Learning approaches can not predict more than 10 days ahead of input data on the small linear dataset.
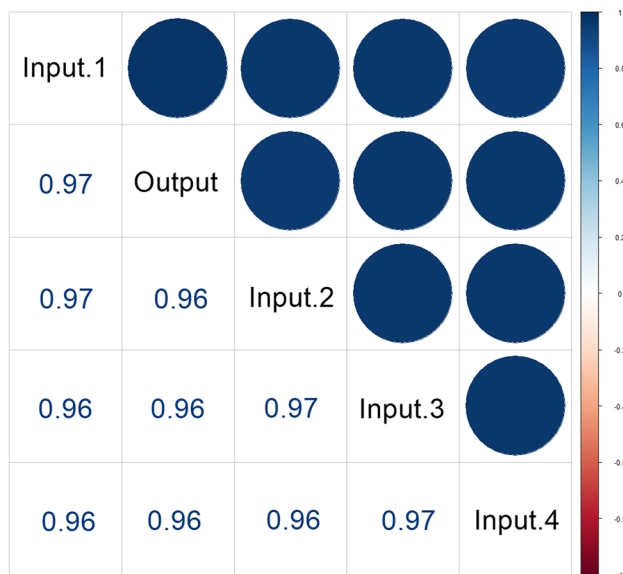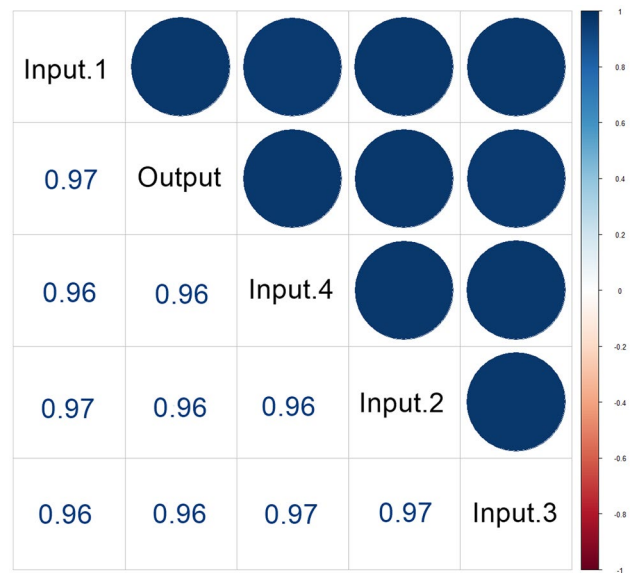


**Fig. 5** The Correlation Matrix of Scenario 2

Checking all the correlation matrices of each scenario, we can summarize that each scenario shows a high correlation between Inputs and output and does not contain any anomalies.

## Evaluation

The appraisals were conducted by finding the correlation coefficient (R), mean absolute percentage error (MAPE), and the root-mean-square rrror (RSME). The performances of the models are evaluated by using these factors. Below are the formulas used to calculate the factors.
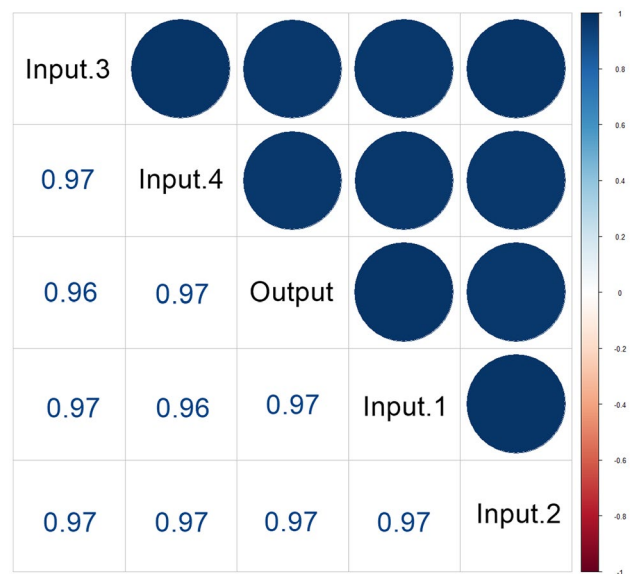


**Fig. 4** The Correlation Matrix of Scenario 1



**Fig. 6** The Correlation Matrix of Scenario 3

**Table 3** ANFIS training results

|  | MF Type | RMSE | MAPE | Corr Coff. |
|---|---|---|---|---|
|  | Triangular | 297.89 | 54.25 | 0.27 |
| Scenario 1 | Trapezoidal | 97.08 | 9.39 | -0.35 |
|  | Gaussian | 261.71 | 30.08 | 0.32 |
|  | Triangular | 946.14 | 62.54 | 0.50 |
| Scenario 2 | Trapezoidal | 216.48 | 23.30 | 0.66 |
|  | Gaussian | 234.58 | 36.729 | 0.42 |
|  | Triangular | 1065.8 | 93.28 | -0.22 |
| Scenario 3 | Trapezoidal | 600.61 | 38.076 | -0.28 |
|  | Gaussian | 835.48 | 70.97 | -0.24 |

$$R = \frac{N \sum(xy) - \sum(x) * \sum(y)}{\sqrt{[n \sum(x^2) - \sum(x)^2][n \sum(y^2) - \sum(y)^2}} \tag{5}$$

$$MAPE = \frac{1}{N} \sum_{t=1}^{n} \left| \frac{x_t - y_t}{x_t} \right| \tag{6}$$

$$RMSE = \frac{1}{N} \sum_{t=1}^{n} \sqrt{(x_t - y_t)^2} \tag{7}$$

Where $N$ represents the number of data points and $x$ and $y$ are the actual and predicted value respectively.

## Results

The ANFIS model's performance and the LSTM are evaluated using both the training and testing data, and the best-case scenario is defined.

### Adaptive Neuro-Fuzzy Inference System (ANFIS)

The parameters were not set arbitrarily. There are two optimum methods. One is backpropagation, and another is hybrid. In our case, the hybrid optimum method is not suitable for our dataset. We have performed a small test to determine which optimum method is ideal for our experiment. We found that the hybrid method's error is much higher than the backpropagation method; hence, we chose backpropagation. There are various ANFIS parameters, for example, number of MFs, MF types, or optimum methods, which can be modified to test the dataset. We played with those parameters and came up with optimal settings mentioned in Table 3. As mentioned earlier, we picked three MF types that are the best among the nine MF types. We also set the number of input MFs to 3, which is also optimal, as the other values we
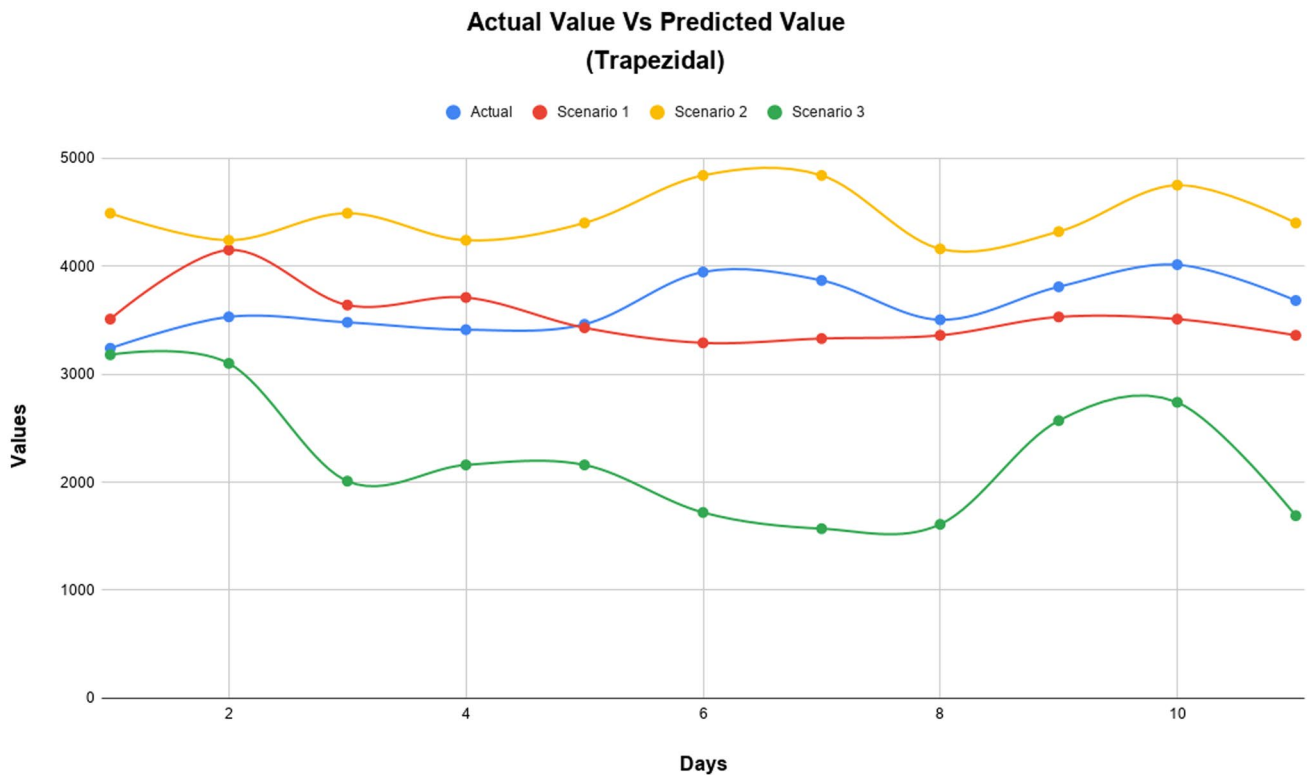


**Fig. 7** Plot diagrams for the prediction of daily (Trapezoidal)

have tried did not give promising results for the experiment. Firstly, the ANFIS model was carried out on the dataset using the three MF types, namely, triangular, trapezoidal, and Gaussian, where the back propagation (BP) was used as the optimum method. All outputs of MF are linear, and the number of MF is 3, as shown in Table 3.

From the first observation, from Table 3, we can see that the best membership function for the ANFIS model is trapezoidal. However, *Scenario 3* seems to have all the correlation coefficients negative, meaning the inputs have a negative relationship with the output. It can be said that using the trapezoidal MF for *Scenario 2*, we get a positively high correlation and lower RMSE value compared to the other two MF types in *Scenario 2*.

As we can see from the graph that the actual cases and *Scenario 2* predicted cases curve has a similar wave, though there is still a difference in the value (See Fig. 7).

## Long Short-Term Memory Network(LSTM)

The LSTM was also applied in the three scenarios for prediction. The dropout regularization method was the input in this recurrent connection to the LSTM units, which reduces the overfitting and improves the model performance by updating the weights while training a network [22]. The batch size decreases from 32, 16, and 8, respectively. The four-layer units were increased by 5 gradually such that the first layer had 5, the next layer 10, and so on till the fourth layer. Table 4 shows the dropout values and the epoch value.

From the Table 5, it can be clearly said that *Scenario 2* with the batch size 16 has the best RMSE value and the best correlation coefficient between input and output. However, *Scenario 1* with batch size 16 has a higher correlation coefficient, but the RMSE value is pretty high; therefore, it can not be considered the best case.

## Validation

Table 6 and Fig. 8 represent the ANFIS and LSTM model's best prediction results from 19th to 30th June 2020. The LSTM has presented more convincing values for RMSE and correlation coefficient for forecasting the outbreak.

**Table 4** LSTM fixed training parameters

| Parameter | Fixed Values |
| --- | --- |
| 1st Layer Dropout | 0.1 |
| 2nd Layer Dropout | 0.2 |
| 3rd Layer Dropout | 0.25 |
| 4th Layer Dropout | 0.3 |
| Epoch | 200 |
| Optimizer | Adam |
| Activation | Relu |

**Table 5** LSTM training results

| | Batch Size | RMSE | MAPE | Corr Coff. |
| --- | --- | --- | --- | --- |
| | 32 | 8.38 | 5.16 | 0.27 |
| Scenario 1 | 32 | 129.55 | 11.81 | 0.58 |
| | 32 | 29.07 | 4.39 | 0.68 |
| | 16 | 93.89 | 11.01 | 0.77 |
| Scenario 1 | 16 | 94.70 | 10.20 | 0.60 |
| | 16 | 27.56 | 5.00 | 0.66 |
| | 8 | 92.48 | 9.21 | 0.73 |
| Scenario 1 | 8 | 69.96 | 7.91 | 0.62 |
| | 8 | 85.09 | 7.64 | 0.69 |
| | 32 | 65.49 | 5.80 | 0.56 |
| Scenario 2 | 32 | 17.16 | 5.56 | 0.71 |
| | 32 | 73.79 | 5.28 | 0.65 |
| | 16 | 51.91 | 6.52 | 0.54 |
| Scenario 2 | 16 | 6.55 | 4.51 | 0.75 |
| | 16 | 45.27 | 7.99 | 0.67 |
| | 8 | 142.32 | 15.18 | 0.59 |
| Scenario 2 | 8 | 30.54 | 7.57 | 0.69 |
| | 8 | 11.94 | 4.83 | 0.64 |
| | 32 | 0.329 | 5.59 | 0.33 |
| Scenario 3 | 32 | 11.76 | 5.09 | 0.42 |
| | 32 | 82.64 | 5.82 | 0.62 |
| | 16 | 60.98 | 6.85 | 0.57 |
| Scenario 3 | 16 | 96.93 | 10.27 | 0.45 |
| | 16 | 15.87 | 4.64 | 0.54 |
| | 8 | 117.60 | 10.81 | 0.20 |
| Scenario 3 | 8 | 37.96 | 6.56 | 0.30 |
| | 8 | 153.86 | 10.97 | 0.32 |

## Discussion

We have used the ANFIS and LSTM based prediction model to forecast the COVID-19 pandemic growth in Bangladesh. These time-series data are the compiled number of confirmed cases are rising high and arriving at a specific coincided peak curve. These two machine learning approaches are used only for predicting and making no assumption in this outbreak. From our observation, we can say that LSTM has provided us with more satisfactory results compared to ANFIS for predicting the COVID-19 pandemic. Furthermore, we have taken three scenarios for our experiment, from which the *Scenario 2* has been showing better results than the other two

**Table 6** Best training results from both algorithms

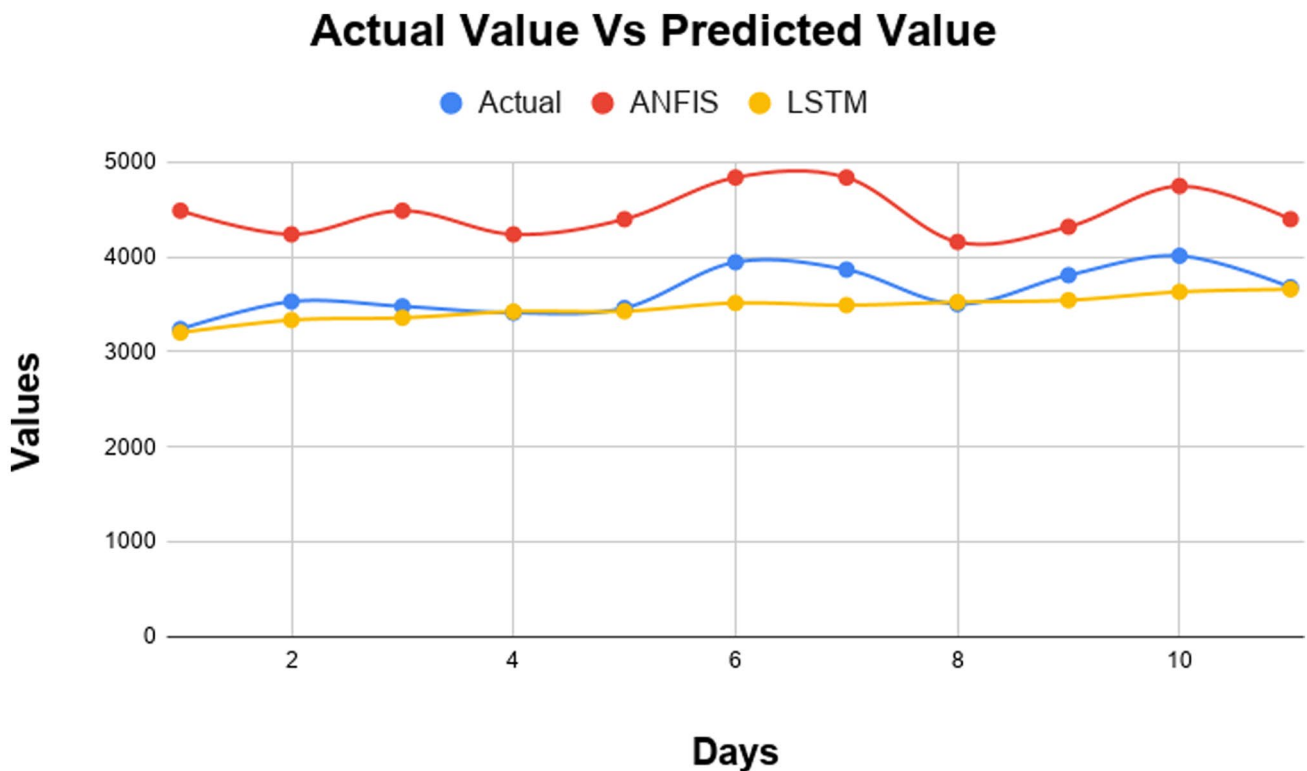| scenarios | Model | Param Info | RMSE | MAPE | Corr.Coff |
| --- | --- | --- | --- | --- | --- |
| Scenario 2 | ANFIS | Trapezoidal | 216.48 | 23.3 | 0.66 |
| Scenario 2 | LSTM | Batch size 16 | 6.55 | 4.51 | 0.75 |

## Actual Value Vs Predicted Value



**Fig. 8** Plot diagram for the best prediction results of ANFIS and LSTM Model

scenarios. Additionally, LSTM captured the pattern of this hyped-up growth of the graph with a better correlation value and RMSE. However, both the models have shown to have a decent potential in predicting time-series data. Moreover, we cannot forget the convoluted anomaly of this outbreak. Further research is still very fundamental to endorse the results and improve the quality of prediction. In some cases choosing the right activation function can give us a better result. However, most of the time, testing various configurations and evaluating the outcome provides better results. We also tested multiple configurations and found out that one of the configurations of LSTM gives the best result from all other configurations. From our observation at the end of the experiment, it is worth mentioning that ANFIS consumes more time and hardware resources for training than LSTM.

### Threats to Validity

There are few mentionable threats for the experiment; that is, there might be some problems with the dataset as not all the published cases are not of that specific day only. It might occur due to a lack of test time processing. There is another threat of an outlier to appear, even though this dataset of Bangladesh did not have outliers. However, there is no guarantee if this will not happen as if it occurred in China on 12 Feb 2020 [23]. Before 11 February, the daily new cases were 2015, and later, on 13 February, the daily new cases were 5090. In comparison to these two days, 12 February is an outlier data for the model. Therefore, we can exclude this type of days while training.

### Conclusion

Much research is ongoing to predict the COVID-19 as our daily livelihood depends on the result of this crisis. Various studies are done using different prediction models. We have selected the ANFIS model and the LSTM for predicting the COVID-19 cases in Bangladesh. However, more work is required for verifying which forecasting method or model is competent in all cases for all the different populations worldwide. The input data is not correct either as there are various reasons like an infected person is asymptotic or did not get tested or did not get listed in the database. Nevertheless, the gradual learning approach can overcome these imprecise input data. Also, there is an unknown suspect in the community that few countries are reporting false data for political reasons. Many countries, including Bangladesh,

implemented social distancing and lockdown, resulting in an impact on the cases and casualties. By considering these factors, the results of the predictions can be a little inhibited. While we tried to provide a promising result for the Bangladesh data, yet we need to further test for more optimized and far more accurate results on various other databases. Nonetheless, the given results are displaying promise and success for AI in predicting a pandemic.

**Author Contributions** This work was carried out in close collaboration between all co-authors. First, we defined the research theme and contributed an early design of the system. We further implemented and refined the system development and wrote the paper. All authors have contributed to, seen, and approved the final manuscript.

## Declarations

**Ethical Standard** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

**Conflicts of Interest** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Punn NS, Sonbhadra SK, Agarwal S. COVID-19 Epidemic analysis using machine learning and deep learning algorithms. medRxiv 2020. page 2020.04.08.20057679. Available from: http://medrxiv.org/content/early/2020/04/11/2020.04.08.20057679.abstract.

2. World Health Organization. Coronavirus (COVID-19) events as they happen, 2020. Available from: https://www.who.int/emergencies/diseases/novelcoronavirus-2019/events-as-they-happen.

3. Anwar S, Nasrullah M, Hosen MJ. COVID-19 and Bangladesh: challenges and how to address them. Front Public Health. 2020;8:154.

4. Worldometer. Bangladesh Coronavirus: 190,057 Cases and 2,424 Deaths - Worldometer, 2020. Available from: https://www.worldometers.info/coronavirus/ country/bangladesh/.

5. Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Computer Science. 2020;1(4):1–15. https://doi.org/10.1007/s42979-020-00209-9.

6. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liu P, Cao X, Gao Z, Mai Z, Liang J, Liu X, Li S, Li Y, Ye F, Guan W, Yang Y, Li F, Luo S, Xie Y, Liu B, Wang Z, Zhang S, Wang Y, Zhong N, He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165–74.

7. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Env Res Risk A. 2020;34(7):959–72. Available from: http://link.springer.com/10.1007/s00477-020-01827-8.

8. Elmousalami HH, Hassanien AE. Day level forecasting for coronavirus disease (COVID-19) Spread: analysis, modeling and recommendations. arXiv, 2020. Available from: http://arxiv.org/abs/2003.07778.

9. Jang JSR. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern. 1993;23(3):665–85.

10. Al-Qaness MA, Ewees AA, Fan H, Aziz MAE. Optimization method for forecasting confirmed cases of COVID-19 in China. Appl Sci. 2020;9(3).

11. Yudistira N. COVID-19 growth prediction using multivariate long short term memory. arXiv. 2020;14(8):1–8. Available from: http://arxiv.org/abs/2005.04809.

12. Tahmasebi P, Hezarkhani A. Comparison of optimized neural network with fuzzy logic for ore grade estimation. Aust J Basic Appl Sci. 2010;4:5.

13. Tahmasebi P, Hezarkhani A. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. Comput Geosci. 2012;42(5):18–27. Available from https://doi.org/10.1016/j.cageo.2012.02.004

14. Abraham A. Adaptation of fuzzy inference system using neural learning. Springer, Berlin, Heidelberg, 2005.

15. Chang FJ, Chang YT. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv Water Resour. 2006;29(1):1–10.

16. Polat K, Gunes S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing: A Review Journal. 2007;17(4):702–710.

17. Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 pandemic prediction for Hungary. SSRN Electronic Journal: A Hybrid Machine Learning Approach; 2020.

18. Jiang W, Schotten HD. Deep learning for fading channel prediction. IEEE Open Journal of the Communications Society. 2020;1(April):320–32.

19. Connor JT, Martin RD, Atlas LE. Recurrent neural networks and robust time series prediction. IEEE Trans Neural Netw. 1994;5(2):240–54.

20. Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci Total Environ. 2020;728:138762. Available from: https://doi.org/10.1016/j.scitotenv.2020.138762.

21. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

22. Brownlee J. How to use dropout with LSTM networks for time series forecasting. 2020. Available from: https://machinelearningmastery.com/use-dropout-lstm-networks-time-seriesforecasting/.

23. Worldometer. China coronavirus: 85,591 cases and 4,634 deaths - worldometer, 2020. Available from: https://www.worldometers.info/coronavirus/country/china.