

Analysis and Prediction of COVID-19 Pandemic in Bangladesh by using Long short-term memory network (LSTM) and Adaptive neuro fuzzy inference system(ANFIS)

Anjir Ahmed Chowdhury

American International University-Bangladesh

Khandaker Tabin Hasan (✉ tabin@aiub.edu)

American International University-Bangladesh <https://orcid.org/0000-0002-9849-5720>

Khadija Kubra Shahjalal Hoque

American International University-Bangladesh

Research Article

Keywords: LSTM, ANFIS, Covid-19, Machine Learning, Epidemic Prediction Model

Posted Date: October 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-55617/v5>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Cognitive Computation on August 10th, 2020. See the published version at <https://doi.org/10.1007/s12559-020-09786-6>.

Analysis and Prediction of COVID-19 Pandemic in Bangladesh by using Long short-term memory network (LSTM) and Adaptive neuro-fuzzy inference system (ANFIS)

Anjir Ahmed Chowdhury · Khandaker Tabin Hasan · Khadija Kubra Shahjalal Hoque

Received: October 21, 2020/ Accepted: date

Abstract

Objectives: The dangerously contagious virus named "COVID-19" has struck the world strong and has locked down billions of people in their homes for stopping the further spread. All the researchers and scientists in various fields are working constantly to come up with a vaccine and prevention method to aid the world from this challenging situation. However, a reliable prediction of the epidemic may help to control this contiguous disease until the cure is available. The machine learning techniques are one of the frontiers in predicting the future trend and behavior of this outbreak. Our research is focused on finding a suitable machine learning algorithm that can predict the COVID-19 daily new cases with higher accuracy.

Methods: In this research, we have 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 results of the experiments and it can be forenamed that LSTM has shown more satisfactory results.

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 means square error

(RMSE), 6.55 and Correlation Coefficient 0.75.

Conclusion: This study is expected to shed light on COVID-19 prediction models for researchers working with machine learning techniques and help avoid proven failures especially for small imprecise datasets.

Keywords LSTM, ANFIS, COVID-19, Machine Learning, Epidemic Prediction Model

1 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 the city of 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 declared 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 as well.

Anjir Ahmed Chowdhury
American International University-Bangladesh
E-mail: anzira431@gmail.com

Khandaker Tabin Hasan (✉)
American International University-Bangladesh
E-mail: tabin@aiub.edu

Khadija Kubra Shahjalal Hoque
American International University-Bangladesh
E-mail: khadijakubra197@gmail.com

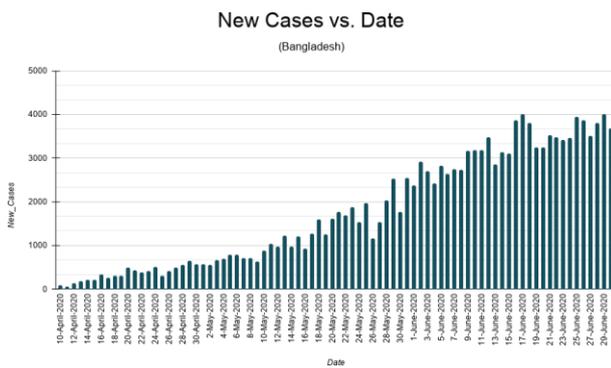


Fig. 1: Bangladesh new cases data from April 10-June 30, 2020 [4].

Furthermore, to break the chain of this contagious virus, many countries [3] took preventive measures, which include travel bans, home-office, country lockdown, and most importantly social distancing. However, Bangladesh, a lower-middle-income economy, and also one of the most populated countries in the world 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. From Figure 1, we can see 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 effective measures beforehand. The objective of this study is to develop a prediction model and to 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 be helpful in the efficiency of forecasting. At the very beginning of an epidemic, it is necessary to take proper and effective measures for preventing it from crossing the geological boundaries [5] which in result can save many lives with less impact. Therefore, forecasting and the proper study of the pattern of disease spread can be a strategic plan in the planning of control strategy. The predictive models can be utilized 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 using 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 and can determine the non-linearity in the time-series and can 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) that has used the WHO (World Health Organization) official data for predicting the confirmed cases for the upcoming 10 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 is time-series data that compiles the number of confirmed cases where the cases are increasing endlessly over time until it arrives at a certain concentrated peak curve. From there the LSTM captures the pattern of the dynamic growth of graphs with a minimum RMSE compared to the 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 6, we gave the conclusion of the work.

2 Methods

We wanted to conduct our experiment in two different types of neural network systems. Therefore, we chose one algorithm from the fuzzy inference system and another one from the recurrent neural network. We chose ANFIS 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 systems in a single framework. Furthermore, ANFIS is considered as 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 our 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 named, “Forgot gate“ and “Output gate” which make LSTM more efficient than other recurrent neural networks. LSTM also has a feedback connection that allows it to not only process single data points but also 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.

2.1 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 non-linearity. [15, 16]

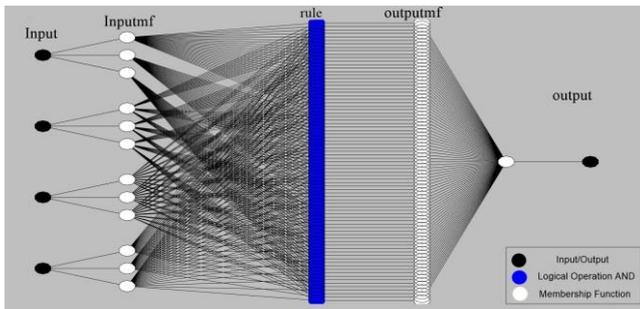


Fig. 2: Architecture of the developed ANFIS.

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 outputs of the first layer the second layer is created which carries the preceding values of the Member Functions (MF). 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 rules. 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 output 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 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).

2.2 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 that RNN is satisfying at processing data and manifests promising results in the time-series prediction through hoarding enormous historical information in its internal state [19]. However, it has an inconvenience of evanescence and the gradient detonate problems which leads 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.

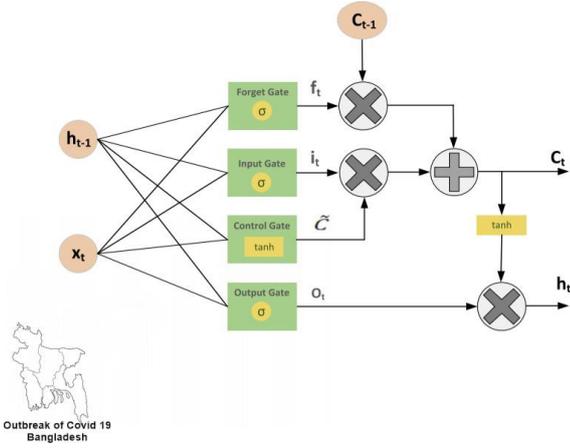


Fig. 3: The basic structure of the LSTM [20].

Some equations describe the gates of LSTM. Before start to explain the equation we need to know some variables which are used in these equations. The weight matrices are W and the activation function is σ 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 input of the current timestamp. Now, The Input gate (i_t) is described as,

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

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

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

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).

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

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

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

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

$$h_t = o_t * \tanh(C_t) \quad (4)$$

2.3 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 the Tensorflow 2 and Keras as the API. For the supportive package Jupyter notebook, pip3, NumPy, Matplotlib, and spacy 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 800Mhz. It also had the NVIDIA gtx 1060 6 GB graphics processing unit.

3 Dataset

3.1 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 it 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 very important 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.

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

Table 1: Three proposed scenarios for time-series prediction of COVID-19 in Bangladesh.

3.2 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 $x(t-1)$. From Table 2 we can see 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 find out the number of inputs which is optimal for our test. We used the LSTM algorithm with fixed parameters which are also mentioned in Table 2 to find out which the 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.

No. of inputs	Units	Batch Size	RMSE	MAPE	Corr Coeff.
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: No. of inputs selection test results.

We can also understand from Table 2 that there are multiple ways to take the prior 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 way is best for our experiment. That is why in Table 1 we proposed three ways (Scenarios). There might be other combinations like we can take the first two consecutive then one day gap then next two consecutive days or similar like this. We did not include this type of way because any other combination is made by these three combinations. These three ways are the base of other combinations.

3.3 The Correlation Matrix

Before implementation, we examine the correlation between each input and output. Correlation is a very common term in the statistical area. It is a single number

that gives acknowledgment of the degree of relationship between the inputs and output. The correlation matrix summarizes the dataset and visualizes the pattern in the dataset and gives feedback on whether the dataset contains any unwanted data or not. Although the output column of each scenario is the same, the input columns are different in every scenario. Every scenario is different in terms of input. So that 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,

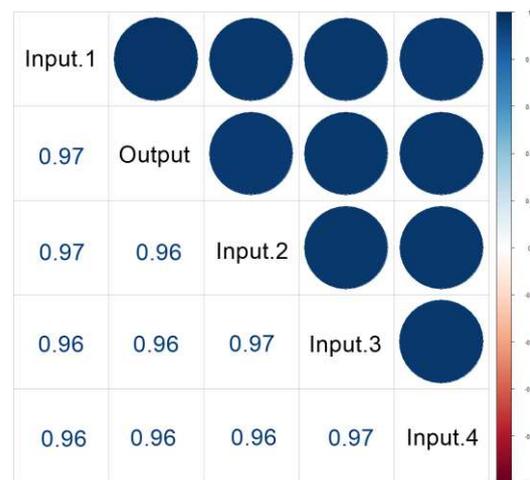


Fig. 4: The co-relation Matrix of Scenario 1

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,



Fig. 5: The co-relation Matrix of Scenario 2.

The correlation matrix of this scenario is also shown to have better correlation as in the previous scenario. Lastly, for the third scenario, we took the consecutive previous four days to predict the output and the correlation matrix of this scenario is,

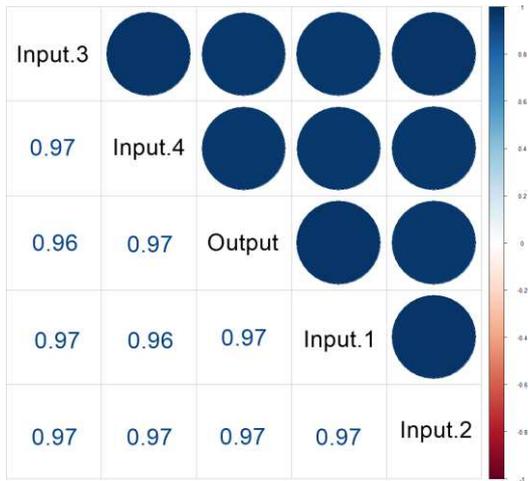


Fig. 6: The co-relation Matrix of Scenario 3.

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

3.4 Evaluation

The appraisals were conducted by finding the correlation coefficient (R), Mean Absolute Percentage Error (MAPE), and the Root Mean Square Error (RSME).

The performances of the models are evaluated by using these factors. below shows the formulas used to calculate the factors.

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

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

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

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

4 Results

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

4.1 Adaptive Neuro-Fuzzy Inference System(ANFIS)

We did not set the parameters 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 find out which optimum method is suitable for our experiment and we found that the error of the hybrid method is much higher than the backpropagation method. Hence, why we chose backpropagation. There are various ANFIS parameters, for example, no. 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 which are 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 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 No of MF is 3. As shown in Table 3.

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

Table 3: ANFIS training results.

From the first observation, from the above table, 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*.

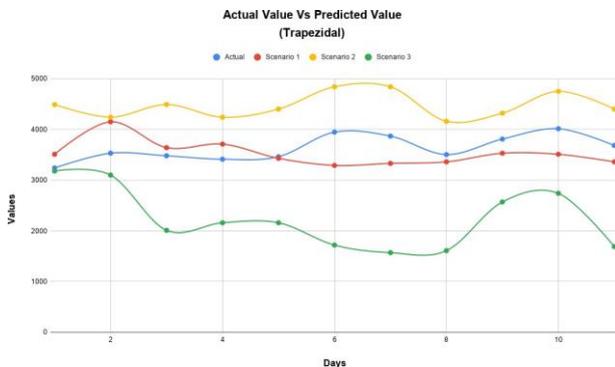


Fig. 7: Plot Diagrams for the prediction of Daily (Trapezoidal).

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.

4.2 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 below shows the dropout values and the epoch value.

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

Table 4: LSTM fixed training parameters.

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

Table 5: LSTM training results.

From the above table, it can be clearly said that *Scenario 2* with the batch size 16 has the best RMSE value and also 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 taken as the best case.

4.3 Validation

Table 6 and Figure 8 represent the best prediction results of the ANFIS and LSTM model for the period of 19-30 June 2020. The LSTM has presented more convincing values for RMSE and correlation coefficient for forecasting the outbreak.

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

Table 6: Best training results from both algorithms.

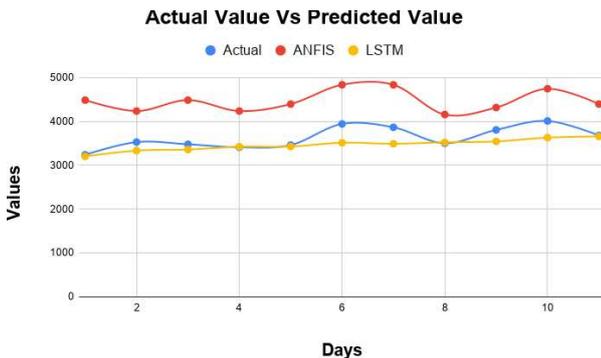


Fig. 8: Plot diagram for the best prediction results of ANFIS and LSTM Model.

5 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 that are rising high and arriving at a certain 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 compared to the other two scenarios. Additionally, LSTM was able to capture 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 would be 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 but most of the time testing various configurations and evaluating the outcome gives better results. In our case also, we test various configurations and find out that one configuration 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.

6 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. This might occur due to a lack of test time processing. There is another threat of an outlier to occur, even though this dataset of Bangladesh did not have an outlier. However, there is no guarantee if this will not happen as if it occurred in China on 12 Feb 2020 [23]. Before that day on 11 Feb the daily new cases were 2015, and later that day on 13 Feb the daily new cases were 5090. In comparison to these two days, 12 Feb is an outlier data for the model. Therefore, we can exclude this type of day while training.

7 Conclusion

A lot of research is going on predicting the COVID-19 as our daily livelihood depends on the result of this crisis. Various researches are done using various 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 all over the world. The input data is not surely correct either as there are various reasons like an infected person is asymptomatic or did not get tested or maybe 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. 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.

Conflict of Interest Statement: 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.

Authors and Contributors: This work was carried out in close collaboration between all co-authors. first defined the research theme and contributed an early design of the system. further implemented and refined the system development. wrote the paper. All authors have contributed to, seen and approved the final manuscript.

Compliance with Ethical Standards: 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.

References

- Punn NS, Sonbhadra SK, Agarwal S. COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. medRxiv. 2020;p. 2020.04.08.20057679. Available from: <http://medrxiv.org/content/early/2020/04/11/2020.04.08.20057679.abstract>.
- World Health Organization. Coronavirus (COVID-19) events as they happen; 2020. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>.
- Anwar S, Nasrullah M, Hosen MJ. COVID-19 and Bangladesh: Challenges and How to Address Them. *Frontiers in Public Health*. 2020 apr;8:154.
- Worldometer. Bangladesh Coronavirus: 190,057 Cases and 2,424 Deaths - Worldometer; 2020. Available from: <https://www.worldometers.info/coronavirus/country/bangladesh/>.
- 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. Available from: <https://doi.org/10.1007/s42979-020-00209-9>.
- Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*. 2020;12(3):165–174.
- Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*. 2020 jul;34(7):959–972. Available from: <http://link.springer.com/10.1007/s00477-020-01827-8>.
- Elmousalami HH, Hassanien AE. Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations. arXiv. 2020 mar; Available from: <http://arxiv.org/abs/2003.07778>.
- Jang JSR. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man and Cybernetics*. 1993;23(3):665–685.
- Al-Qaness MAA, Ewees AA, Fan H, Aziz MAE. Optimization method for forecasting confirmed cases of COVID-19 in China. *Applied Sciences*. 2020;9(3).
- Yudistira N. COVID-19 growth prediction using multivariate long short term memory. arXiv. 2020 may;14(8):1–8. Available from: <http://arxiv.org/abs/2005.04809>.
- Tahmasebi P, Hezarkhani A. Comparison of Optimized Neural Network with Fuzzy Logic for Ore Grade Estimation. *Australian Journal of Basic and Applied Sciences*. 2010 5;4.
- Tahmasebi P, Hezarkhani A. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers and Geosciences*. 2012 5;42:18–27. Available from: <https://pubmed.ncbi.nlm.nih.gov/24268588/>.
- Abraham A. *Adaptation of Fuzzy Inference System Using Neural Learning*. Springer, Berlin, Heidelberg; 2005.
- Chang FJ, Chang YT. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources*. 2006;29(1):1–10.
- Polat K, Güneş 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.
- Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. *SSRN Electronic Journal*. 2020;.
- Jiang W, Schotten HD. Deep Learning for Fading Channel Prediction. *IEEE Open Journal of the Communications Society*. 2020;1(April):320–332.
- Connor JT, Martin RD, Atlas LE. Recurrent Neural Networks and Robust Time Series Prediction. *IEEE Transactions on Neural Networks*. 1994;5(2):240–254.
- Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Science of the Total Environment*. 2020;728:138762. Available from: <https://doi.org/10.1016/j.scitotenv.2020.138762>.
- Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997;9(8):1735–1780.
- Brownlee J. How to Use Dropout with LSTM Networks for Time Series Forecasting; 2020. Available from: <https://machinelearningmastery.com/use-dropout-lstm-networks-time-series-forecasting/>.
- Worldometer. China Coronavirus: 85,591 Cases and 4,634 Deaths - Worldometer; 2020. Available from:

<https://www.worldometers.info/coronavirus/country/china>.

Figures

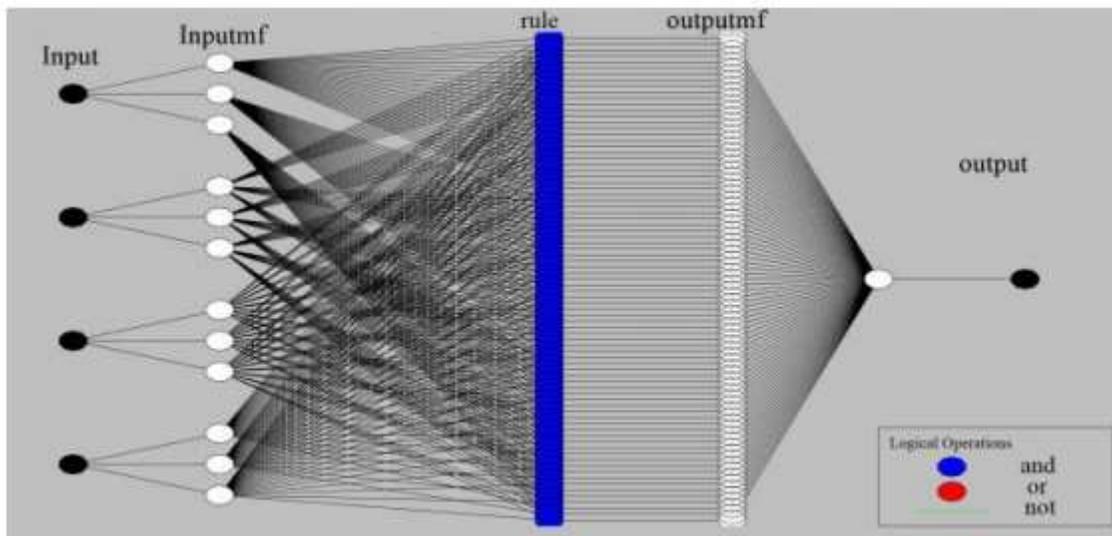


Figure 1

Architecture of the developed ANFIS.

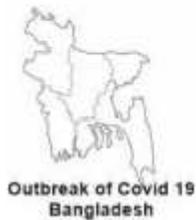
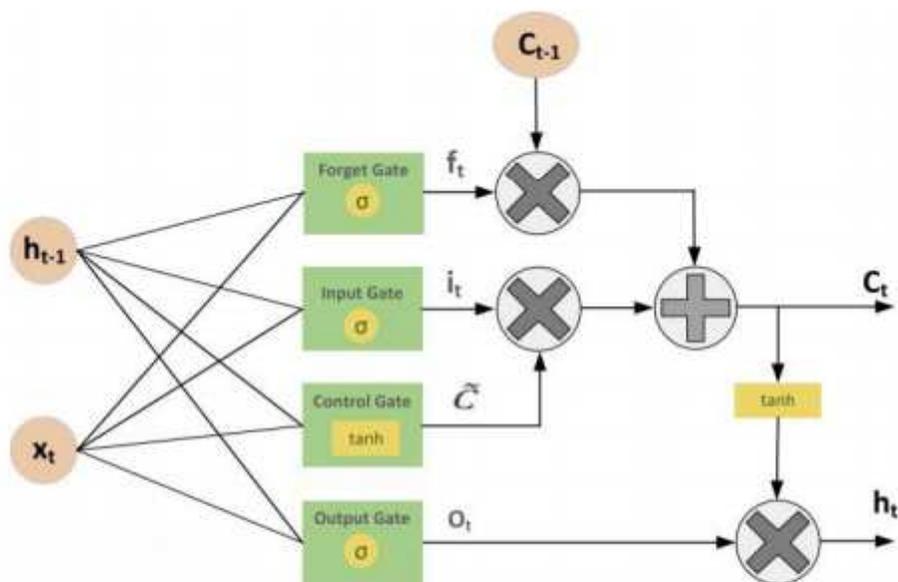


Figure 2

The basic structure of LSTM [16].

New Cases vs. Date

(Bangladesh)

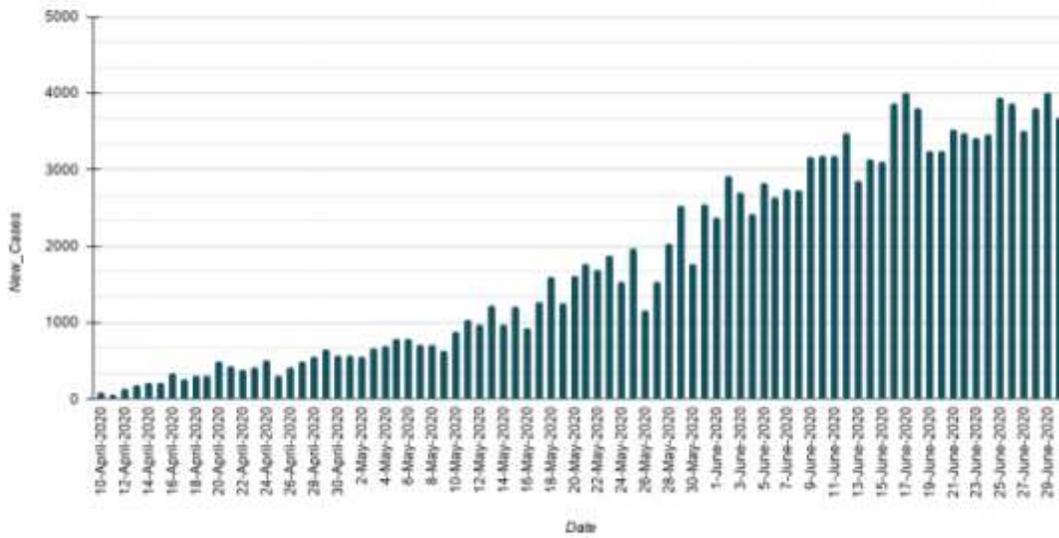


Figure 3

Bangladesh new cases data from April 10 - June 30, 2020.

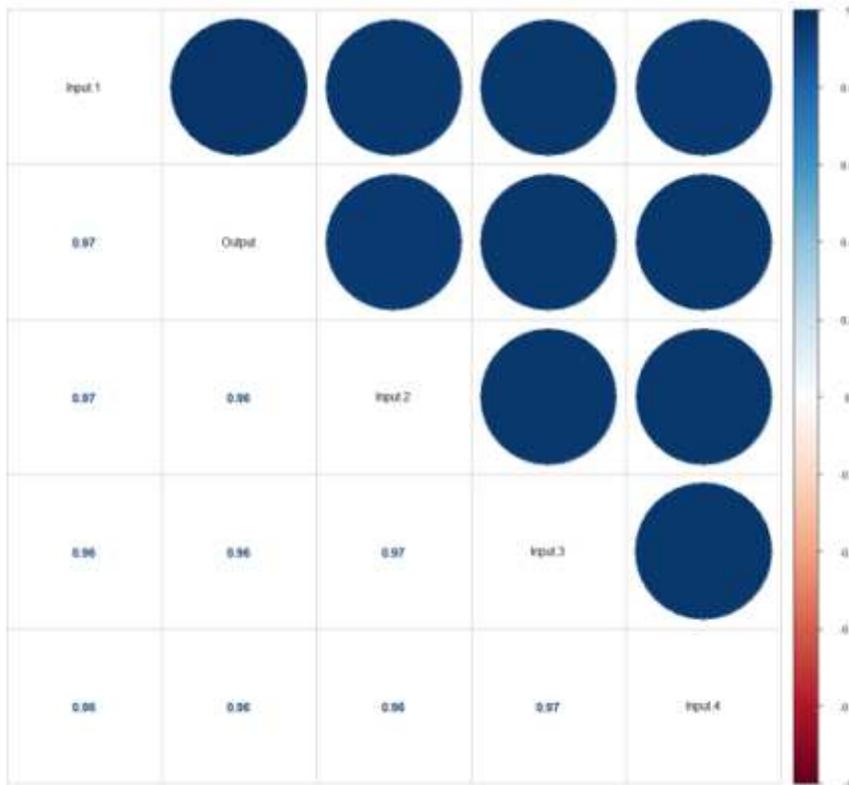


Figure 4

The co-relation Matrix of Scenario 1

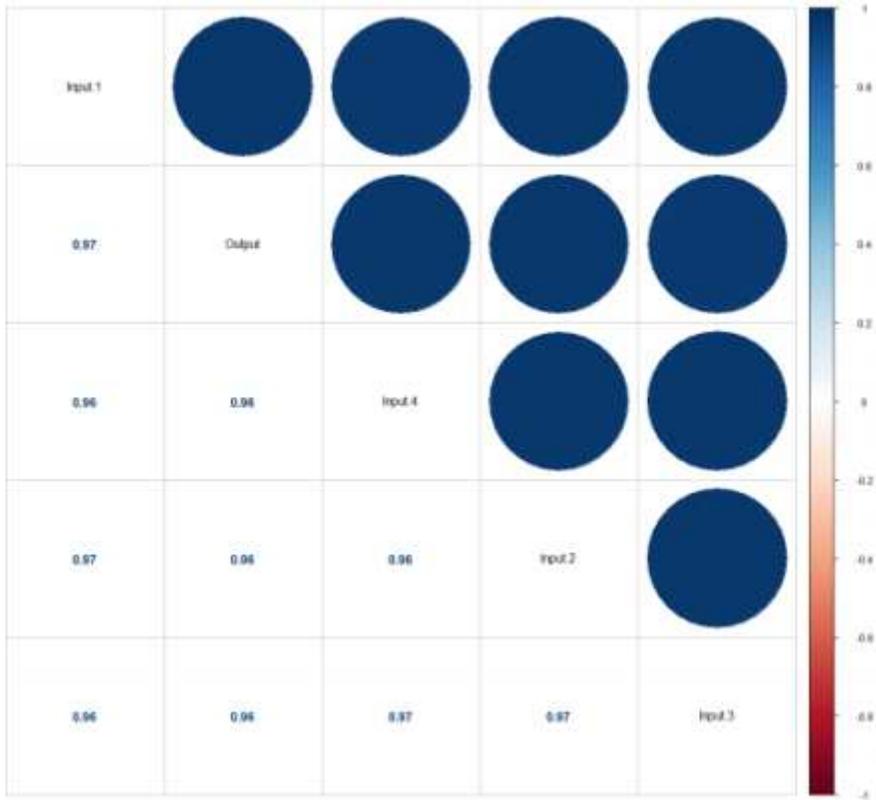


Figure 5

The co-relation Matrix of Scenario 2.

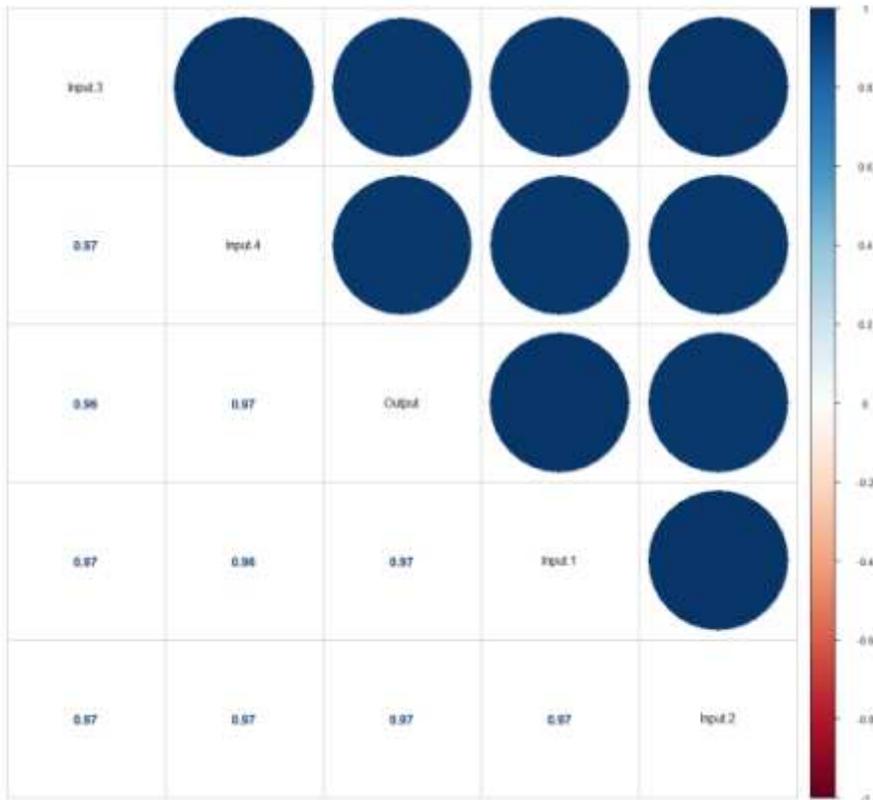


Figure 6

The co-relation Matrix of Scenario 3

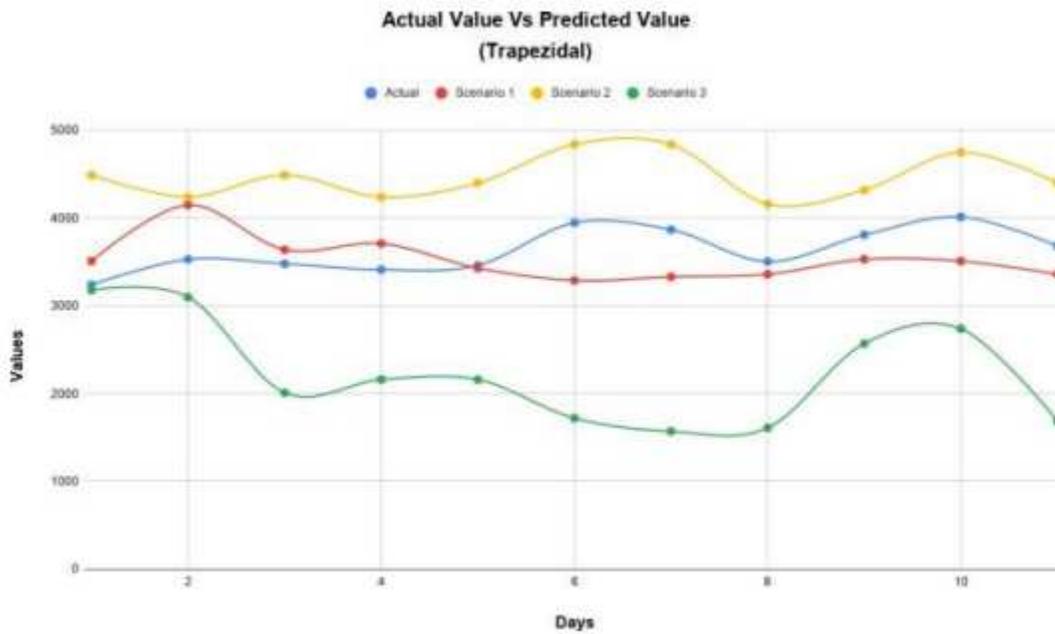


Figure 7

Plot Diagrams for the prediction of Daily (Trapezoidal).

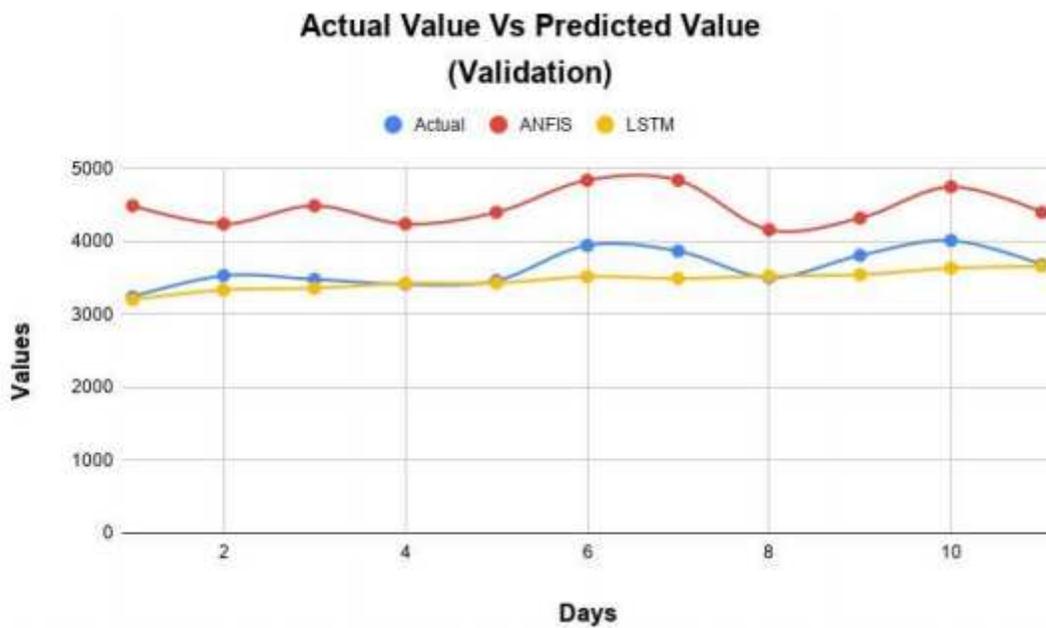


Figure 8

Plot diagram for the best prediction scenario of ANFIS and LSTM Model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [20200807covid19aiub.csv](#)